UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A283 825

REPORT DOCUMENTA

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RES.....
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION / 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION ONR/ONT , ARL, VA 22217 / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

DTIC ELECTE AUG 2 9 1994 S G D

11 TITLE (Include Security Classification)
Eigenvalue Tests and Distributions for Small Sample Order Determination for Complex Wishart Matrices (U)

12 PERSONAL AUTHOR(S)
Curtis Irvin Caldwell

13a TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 1984 TO 1994	1994 August 13	941

16 SUPPLEMENTARY NOTATION

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
mathematical statistics, probability, hypothesis tests, complex variables, principal components analysis, complex Wishart distribution, eigen-structure signal processing,

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This thesis looks at tests to determine how many signal sources exist in the medium when constrained to using only a few samples. It applies classical hypothesis testing assuming complex multivariate Gaussian random variables. The critical issue is the derivation of probability density functions of appropriate test statistics.

This thesis includes a comprehensive development of the tools of statistics of complex variables for engineers and physicists. This includes complex matrix derivatives, changes of complex variables, and properties of the characteristic function of a complex multivariate random variable.

Probability density functions are derived for: the set of eigenvalues satisfying the generalized eigenvalue problem of two complex Wishart matrices, the matrix complex Normal distribution, a joint distribution needed to derive the density for the sphericity test statistic, ratio of averages of disjoint sums of sequential eigenvalues of a complex Wishart matrix, and several tests based on the ratio of an arbitrary eigenvalue to the

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Curtis Irvin Caldwell	22b TELEPHONE (Include Area Code) 814/234-8786 / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

DTIC QUALITY INSPECTED 1

18.(Continued) characteristic functions of multivariate complex variables, changes of multivariate complex variables, complex matrix derivatives, exterior products, wedge products, complex linear algebra, Lebesgue-Radon-Nikodym theorem, zonal polynomials of complex matrix argument, eigenvalues, eigenvectors, sphericity test, beamforming, acoustics, array signal processing.

19 (Continued) maximum, minimum, average, or sum of all the eigenvalues for a special case of the complex Wishart matrix.

This thesis also contains a few minor results regarding zonal polynomials of complex matrix argument, and a tutorial on the Lebesgue-Radon-Nikodym theorem use for estimation.

The Pennsylvania State University

The Graduate School

The Graduate Program in Acoustics

EIGENVALUE TESTS AND DISTRIBUTIONS FOR SMALL SAMPLE ORDER DETERMINATION FOR COMPLEX WISHART MATRICES

A Thesis in

Acoustics

by

Curtis Irvin Caldwell

(c) 1993 Curtis Irvin Caldwell

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 1994

Accesion For
NTIS CRA&I
DTIC TAB
Unannounced
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

94-27558
94 8 26 108

We approve the thesis of Curtis Irvin Caldwell

Date of Signature

Leon H. Sibul — June 13, 1994

Leon H. Sibul
Senior Scientist and Professor of Acoustics
Thesis Adviser
Chair of Committee

Carter L Ackerman — June 13, 1994

Carter L. Ackerman
Associate Professor of Engineering Research

Sabih Hayek — June 13, 1994

Sabih I. Hayek
Distinguished Professor
of Engineering Mechanics

Diana McCammon — June 13, 1994

Diana F. McCammon
Senior Research Associate
Associate Professor of Acoustics

Jiri Tichy — June 13, 1994

Jiri Tichy
United Technologies Professor of Acoustics
Chair of the Graduate Program in Acoustics

ABSTRACT

This is a theoretical thesis. The goal is to determine how many signal sources exist in the medium when constrained to using only a few samples. The need to make decisions based on only a few samples is motivated by the slow sound propagation speed and the time urgency to make decisions. This research treats the problem from the point of view of classical hypothesis testing assuming complex multivariate Gaussian random variables. This is the small sample complex principal components analysis problem. The critical issue is the derivation of probability density functions of appropriate test statistics. The goal has been partially achieved.

The probability density functions for several important distributions have been derived. In particular, these include the distribution for the set of eigenvalues satisfying the generalized eigenvalue problem of two complex Wishart matrices, the matrix complex Gaussian distribution, a joint distribution needed to derive the density for the sphericity test statistic, the density function for the ratio of averages of disjoint sums of sequential eigenvalues of a complex Wishart matrix, and several tests based on the ratio of an arbitrary eigenvalue to the maximum, minimum, average, or

sum of all the eigenvalues for a special case of the complex Wishart matrix. This thesis includes a derivation completely in the context of complex variables of the density function of the complex Wishart distribution and the distribution of its eigenvalues. It also includes a few minor results regarding zonal polynomials of complex matrix argument.

A comprehensive development of the tools of statistics of complex variables for engineers and physicists is provided. This includes a study of complex matrix derivatives, changes of complex variables, and properties of the characteristic function of a complex multivariate random variable. A derivation of the complex Hotelling's T^2 test statistic and distribution useful for tests on means is given. A tutorial on Kiefer and Wolfowitz' application of the Lebesgue-Radon-Nikodym theorem for the estimation approach is provided.

CONTENTS

LIST OF FIGURES xix

LIST OF TABLES xx

ACKNOWLEDGMENTS xxi

1 INTRODUCTION 1

1.1 Characterization of Thesis 1

1.1.1 Focus of Thesis 1

1.1.2 Discipline Home of Thesis 2

1.1.3 What This Thesis is Not 3

1.2 Order Estimation 4

1.3 Eigensolutions 5

1.4 Major Assumptions and Rationale for Approach 9

1.5 Organization of Thesis 13

2 MATHEMATICAL STATEMENT OF PROBLEM 17

2.1 Introduction 17

2.2 Specific Test Statistics 21

3 OTHER APPLICATIONS 29

3.1 Acoustic Emission Analysis 31

3.2 Acoustical Oceanography 36

4 OTHER APPROACHES **39**

4.1 General Discussion 39

4.2 Generalized Maximum Likelihood Estimators 41

4.2.1 Introduction 41

4.2.2 Lebesgue-Radon-Nikodym Theorem 43

4.2.3 Kiefer and Wolfowitz Development of Maximum Likelihood Estimators 50

4.2.4 Uniqueness of the Maximum Likelihood Estimator . . . 62

4.3 Specific Techniques 65

4.3.1 Akaike Information Criterion (AIC) 67

4.3.2 Bayesian Information Criterion (BIC) 71

4.3.3 Kashyap Information Criterion (KIC) 71

4.3.4 Hannan-Quinn (HQ) 72

4.3.5 Efficient Detection Criterion (EDC) 72

4.3.6 White Noise Tests (T_1, $TAIC$, $TBIC$, THQ) 75

4.3.7 Minimum Description Length (MDL) 76

4.3.8 Wang's Sphericity Test 78

4.3.9 Finite Markov Chain Maximum Entropy Order Estimator (FMCME) 79

4.3.10 Coherent MDL 82

4.3.11 Maximum Likelihood (ML) 83

4.3.12 Maximum Entropy (ME) 83

4.3.13 Criterion Autoregressive Transfer Function 84

4.3.14 Final Prediction Error Criterion (FPE) 85

4.3.15 Weak Parameter Criterion (WPC) 85

4.3.16 Singular Value Plot Criterion (R_t) 86

4.3.17 C_p Criterion 86

4.3.18 Bayesian Quickest Decision 87

4.3.19 Quickest Detection Sample Size 87

4.3.20 Likelihood Ratio Test 89

4.3.21 Guttman Lower Bound Criterion 89

4.3.22 Other Significance Tests 89

4.4 Comparisons and Evaluations 90

5 PREVIOUS WORK **94**

5.1 Array Processing 96

5.2 Statistics 103

5.2.1 Eigenvalue Distributions and Testing 103

5.2.2 Complex Statistics Other than Eigenvalue Testing . . . 118

5.3 Zonal Polynomials, Hypergeometric Functions, Group Representation Theory 124

6 STATISTICAL TESTS **132**

6.1 Tests Based on Two Independent Sets of Samples 134

6.1.1 F-Statistic from Ratio of Independent Sums of Sample Eigenvalues 134

6.1.2 Density of Eigenvalues of Sample Signal Plus Noise Covariance Matrix with Respect to Independent Sample Noise-Only Covariance Matrix 138

6.2 Tests Based on One Set of Samples 157

6.2.1 Tests that Require Specifying the Population Covariance Matrix 157

6.2.2 F# in MUSIC 165

6.2.3 Test that Requires Only Partial Knowledge of the Covariance Matrix 172

6.3 Tests Motivated by Krishnaiah 188

6.3.1 Joint Density of Ratio of Adjacent Sample Eigenvalues 190

6.3.2 Joint Density of the Ratio of an Arbitrary Sample Eigenvalue to the Smallest Sample Eigenvalue 198

6.3.3 Joint Density of Ratio of Sample Eigenvalues to Largest Sample Eigenvalue 204

6.3.4 Joint Density of Ratio of Arbitrary Sample Eigenvalue to Trace of Sample Covariance Matrix 209

7 SUMMARY AND CONCLUSIONS **217**

7.1 Results Related Directly to Order Determination 217

7.2 Complex Statistics Tools for Engineers and Physicists 223

7.3 Other Results 227

7.4 Linear Algebra Results Verified for the Complex Case 231

7.5 Other Simple Results 231

7.6 Proofs for Results Stated by Others 233

7.7 More Results 233

7.8 Comparisons. 234

8 FURTHER RESEARCH **236**

8.1 Extending This Research 238

8.1.1 Some References to Consider 238

8.1.2 Connecting Gross and Richards' Work to Stein and Weiss' Work 239

8.1.3 Computation of Zonal Polynomials 241

8.1.4 Calculus of Zonal Polynomials 244

8.1.5 Distributional Theory and Tools 245

8.1.6 Noncentral Distributions and Power Functions 246

8.2 Bridging Theory and Application 246

8.2.1 Important Authors to Consult 246

8.2.2 Small Sample Test Theory 248

8.2.3 Approximation Theory 249

8.2.4 Burnside's Theorem and Characteristic Functions . . . 249

8.2.5 Sturm Separation Theorem and Parallel Processing . . 250

8.2.6 Empirical Characteristic Functions 252

8.3 Acoustics and Signal Processing 252

8.3.1 Processor Structure 252

8.3.2 Time Variation of Noise Field 252

8.3.3 Patterned Arrays 253

8.3.4 Multipath Detection 254

8.3.5 Analogy of Temporal and Spatial Domain Signal Processing 255

8.3.6 State Space Processing 256

8.3.7 Application to Intensity Measurement 256

REFERENCES **258**

A MATHEMATICAL BACKGROUND **282**

A.1 Organization of Appendices 282

A.2 Mathematical Background References 284

A.2.1 Linear Algebra 284

A.2.2 Multivariate Statistics 286

A.2.3 Abstractions 288

A.3 A Real Complex Treatment 289

A.4 The Rest of the Story 293

B MATRIX DIFFERENTIAL OPERATORS 295

B.1 Complex Derivatives 295

B.1.1 Computation and Cauchy-Riemann Equations 296

B.1.2 Derivative of the Conjugate of a Variable 297

B.1.3 Derivative of the Magnitude 298

B.1.4 Derivative of the Magnitude Squared 299

B.1.5 Derivative with Respect to a Vector 300

B.2 Derivative with Respect to a Matrix 306

B.2.1 Linear Transformation of a Vector 311

B.2.2 Derivative of a Matrix with Respect to Itself 312

B.2.3 ~Athans and Schweppe Theorems 314

B.2.4 Derivative of Determinants 320

B.3 Differential Operator $D(Z) = \left(\frac{\partial}{\partial X} + i\frac{\partial}{\partial Y}\right)$ 332

B.3.1 Vector Case 333

B.3.2 Matrix Case 336

B.3.3 Differential Functions of Determinants 338

B.4 Complex Characteristic Functions 341

B.4.1 Definition of Characteristic Function of a Complex Random Matrix 345

B.4.2 Basic Properties of Characteristic Functions 346

B.4.3 Just a Moment 354

B.4.4 Uncharacteristic Functions for a Moment 367

C COMPLEX CHANGE OF VARIABLES 377

C.1 Introduction to Changing Variables for the Complex Case . . 377

C.1.1 Univariate Real Change of Variables 381

C.1.2 Traditional Multivariate Change of Variables 382

C.2 Exterior (Wedge) Products 385

C.2.1 Example of Wedge Products in Rectangular to Polar Coordinate Change of Variables 388

C.3 Jacobians for Complex Change of Variables 390

C.4 Jacobians Requiring Exterior Product Approach 417

D DISTRIBUTIONS, PART I 470

D.1 Complex Normal Distribution Introduction 470

D.1.1 Definition of a Vector Complex Random Variable . . . 471

D.1.2 Proof that Σ is Isomorphic to $2S$ 474

D.1.3 Density of Vector Complex Normal Distribution 478

D.1.4 Characteristic Function for the Standardized Vector Complex Normal Distribution 479

D.2 Matrix Complex Normal Distribution 485

D.2.1 Definition of the Matrix Complex Normal Distribution 485

D.2.2 Properties of the Matrix Complex Normal Distribution. 489

D.2.3 Specialization to the Vector Complex Normal Distribution 502

D.2.4 Matrix Complex Normal Density Function 506

D.2.5 Specialization to the Vector Complex Normal Density Function 508

D.3 Complex Wishart Distribution 509

D.3.1 Introduction 509

D.3.2 Properties of the Complex Wishart Distribution 512

D.4 Distribution of Hotelling's T^2 for Complex Variables 531

E DISTRIBUTIONS, PART II **539**

E.1 Complex Wishart Density 540

E.1.1 Arnold's Proof by Induction 541

E.1.2 Goodman's Construction 547

E.1.3 Srivastava's Derivation 552

E.2 General Theorem on Density of Eigenvalues 562

E.3 Joint Density of Eigenvalues of Complex Standard Wishart . . 565

E.4 Joint Density of Eigenvalues of Complex Wishart 566

E.5 Distribution of $F(2n, 2n)$ 574

E.5.1 Integration 577

E.5.2 Evaluation of Residue 580

E.5.3 Simplification 583

E.6 Ordered Versus Unordered Eigenvalues 586

F RESULTS FOR SIGNAL PROCESSING 590

F.1 Trace Distributions 590

F.2 Characteristic Function of the Complex Wishart Distribution . 601

F.3 Functions of a Wishart Matrix 606

F.4 Tague and Styan Properties 632

F.5 Tague Example: Signal-to-Noise Ratio 654

G ZONAL POLYNOMIAL COMMENTARY 659

G.1 History of Development 659

G.2 Gross and Richards' Development 661

H SOME GROUP THEORY 681

H.1 Basic Group Theory 681

H.2 Group Representation Theory 683

H.2.1 Group Representation Definition 683

H.2.2 Homomorphism Familiar Examples 684

H.2.3 Homomorphism Theorems 685

H.2.4 Transformation Groups 688

H.3 Topology and Basic Measure Theory 692

H.3.1 Topology 692

H.3.2 Measure Theory 695

H.4 Topological Group Theory 700

H.5 Matrix Groups 706

H.6 Group Invariance Property of the Vector Complex Normal Distribution 711

H.6.1 Construction of a Group 712

H.6.2 Action Set Definition 714

H.6.3 Invariance Demonstration 715

I ELEMENTARY HILBERT SPACE THEORY 717

J COMPLEX VECTORS 719

J.1 Abstract Field 719

J.2 Abstract Vector Space 720

J.3 Complex Vector Space 721

J.3.1 Verifying if a Field is Defined 724

J.4 Construction of Vector Space in $\mathbf{R}^{2n}$ Isomorphic to $\mathbf{C}^n$ 733

J.5 A More General Vector Space 735

K COMPLEX MATRIX ALGEBRA 737

K.1 Basic Definitions 737

K.2 Trace Identities 741

K.3 Inverse 753

K.3.1 Partitioned Matrix Inverse 753

K.3.2 Complex Matrix Inversion Lemmas 756

K.4 Determinants 763

K.4.1 Basic Properties 763

K.4.2 Partitioned Matrix Determinants 770

K.4.3 Other Determinants 773

L GRAM-SCHMIDT **778**

L.1 Algorithm Using $< x, y >= x^H y$ 778

L.1.1 Inner Product Gram-Schmidt Algorithm 779

L.1.2 Inner Product Gram-Schmidt Algorithm Proof 779

L.2 Algorithm Using $(x, y) = x^T y$. 781

L.2.1 Bilinear Gram-Schmidt Algorithm 781

L.2.2 Proof 782

M COMPLEX MATRIX DECOMPOSITIONS AND EIGENVALUES **784**

M.1 Decomposition to a Product of a Triangular Matrix 785

M.2 Similarity Transformation 797

M.3 Transformation to a Triangular Matrix with the Same Eigenvalues 799

M.4 Functions of Eigenvalues 804

M.5 Eigenvalue Decomposition 805

M.6 Hermitian Definiteness 810

M.7 Square Root Decomposition 813

M.8 Unitary Transformations 817

M.9 Cholesky or Bartlett Decomposition 825

M.10 Eigenvalues of Simply Modified Matrices 828

M.11 Singular Value Decomposition 833

N TRIGONOMETRY OF COMPLEX MATRICES 835

N.1 Matrix Exponential and Logarithm Properties 835

N.2 Matrix COS and SIN Functions 844

N.3 Relating Trace and Determinant 854

O USEFUL IDENTITIES 857

O.1 Sums 857

O.2 Combinatorics 858

O.3 Classical Distribution Properties 859

O.4 Functions of a Hermitian Positive Definite Matrix 860

O.5 Properties of Unitarily Invariant Functions 863

O.6 Some Special Definitions 865

O.7 Generalized Nested Operator 866

P INTEGRALS 868

P.1 Easy Chain Rule Bookkeeping Method 869

P.2 Mundane Integrals 870

P.2.1 Exponential Integral Definition 870

P.2.2 Integrals of Rational Functions 871

P.2.3 Integrals Related to the Gamma Function 875

P.2.4 Ratio of Exponential to Algebraic Term 880

P.2.5 Product of Rational Term and Exponential 882

P.2.6 Generalized Even and Odd Functions 886

P.2.7 Exponentials 888

P.3 Integrals with Hypergeometric Function of Matrix Arguments and Zonal Polynomials 894

P.4 Integrals with Complex Multivariate Gamma Function 902

Q NOTATION **908**

Q.1 Names of Variables 908

Q.2 Special Notation 909

Q.3 Selected Abbreviations 913

Q.4 Reference Author Names 913

Q.5 Taxonomy of Logic 914

LIST OF FIGURES

2.1 Array Geometry 18

4.1 Graphic Representation of the Lebesgue-Radon-Nikodym Theorem 45

4.2 Maximum Likelihood Estimate (MLE) Convergent Sequences of L 52

4.3 Modified Maximum Likelihood Estimator (MMLE) Convergent Sequences of L 54

4.4 Neighborhood Maximum Likelihood Estimator (NMLE) Convergent Sequences of L 56

8.1 Sturm's Eigenvalue Lattice 251

C.1 Partitioning of Domain and Range for Multivalued Transformation 385

D.1 Integration Contour to Get Characteristic Function 481

E.1 Integration Contour to Get Cumulative Distribution Function 578

LIST OF TABLES

C.1 Structure in Complex Variables 380

C.2 Variables on a Shaped Sample Space 384

C.3 Elementary Operator Matrices 458

E.1 Change of Variables 579

E.2 Line Integral Evaluation 580

E.3 Expansion of $\Pr[F \leq f]$ Summation 584

E.4 Pascal's Triangle 587

H.1 Notation for Types of Orthogonal Groups 707

H.2 Special Orthogonal Groups 708

M.1 $CT = D^2C$ Decomposition Pseudo-Code 786

M.2 $CT = CD^2$ Decomposition Pseudo-Code 788

ACKNOWLEDGMENTS

Im Namen des Vaters, und des Sohnes, und des Heilegen Geistes.

Whatever your work is, put your heart into it as if it were for the Lord and not for men. {COL 3:23}

If they are capable of acquiring enough knowledge to be able to investigate the world, how have they been so slow to find its Master?

Avoid getting into debt, except the debt of mutual love. {ROM 13:8}

I gratefully acknowledge the following people for their direct intellectual or financial contributions: Carter Lee Ackerman, Michael G. Akritas, Steven F. Arnold, Gutti Jogesh Babu, Anthony J. Baratta, Arujana Peter Chaiyasena, Curtis K. Church, Clint W. Coakley, Robert J. Ferlez, Sabih I. Hayek, Steven George Krantz, Debasis Kundu, Paul M. Lapsa, Adrian Ocneanu, Champak D. Panchal, Calyampudi Radhakrishna Rao, Leon H. Sibul, Alan Douglas Stuart, John A. Tague, Clarence Eugene Wayne, and William J. Wilson.

A few people helped birth and develop my professional love for underwater acoustics. They are: Edward Abrahams III, Diana Flory McCammon, Albert Theodore Mollegen, Jr., James Alvin Sagerholm, Paul Allman Siple, Bernard K. Swanson, David L. Westfall, and Clarence Lee Williams. The following have provided encouragement and counsel: Alister C. Anderson, Donald A.

Cario, Yap Siong Chua, James Whitfield Coaker, Bernard Levinson, Robert M. G. Libby, William Hines Linder, Constantine (Charlie) Varelas, and Donald White. Many others have contributed to my broader education or support. The list is much to large to enumerate. To those, also, thank you.

Christopher W. Casey of TCI Software Research, Inc. provided extraordinary support in programming this thesis format for The Pennsylvania State University. Ralph T. Muehleisen and Rodney Richard Korte gave many hours in nurturing computers into productive servants.

Families endure this kind of effort with varying degrees of accommodation. My wife, Susan Marion Belcher Caldwell, and my son, Joshua Benjamin Lee Caldwell have been with me throughout this period. My older brother, Harold Earnest Caldwell, is the best teacher I have ever had. My parents, Elmer Irvin Caldwell and May Alice Wing Caldwell, have provided the values and significant financial support to see this research come to its conclusion.

Finally, with pleasure, I acknowledge the financial support provided by the United States of America Office of Naval Research, the Office of Naval Technology, The Pennsylvania State University Applied Research Laboratory, the Department of Nuclear Engineering, and The Graduate Program in Acoustics.

Chapter 1

INTRODUCTION

1.1 Characterization of Thesis

1.1.1 Focus of Thesis

The focus of this thesis is the development of tools and construction of methods for determining the number of point sources present in a measured acoustic field. There are several good approaches to this problem. The approach examined in this thesis is that of Principal Components Analysis for the small sample case of signals and noise arising from the matrix complex normal probability distribution. This distributional assumption is a typical starting point for problems in array processing. The forms of test statistics applicable to this problem have been known by many people for a long time. The hard part of the problem is obtaining the sampling distributions of those statistics. The distributions for test statistics have been developed in this thesis for some of the simple (and, hence, unrealistic) cases. Although there remains much work to be done, this thesis does develop significant tools required for the further study of this problem and it partially develops the derivations of the ultimately desired distributions.

1.1.2 Discipline Home of Thesis

A major criticism levied against this thesis is the notion that it is not a thesis in acoustics. It is true that most of the work produced in the course of this study does not have the flavor most acousticians would recognize, yet it was originally (and still is) solidly motivated by a problem in acoustics. Because the terminal goal of this research has not been reached, it is not yet possible to demonstrate its application via experiment or simulation to acoustics. However, because of the research accomplishments of this thesis, the day when that might be possible is now closer (in event time measure).

The bulk content of this thesis is multivariate statistics of complex variables. Statisticians generally would not claim this work because of the extensive use of complex variables. The most difficult contributions of this thesis are grounded in topological group representation theory, yet mathematicians would not generally claim this work because it is too applied. Nevertheless, the key observation in this thesis (the justification of Gross and Richards' splitting theorem for zonal polynomials of two complex Hermitian matrix arguments, and its application to the derivation of the joint probability density of eigenvalues of an Hermitian Wishart matrix) requires such a treatment to establish it. It is appropriate to remark here that the most widely useful results of this thesis, which include the systematic redaction of the linear algebra, differential and integral calculus, and statistics of complex variables, is accessible to most

engineering juniors.

Although signal processing most often finds its academic home in electrical engineering, electrical engineers most often deal with applications which make use of large sample sizes. I am interested in the small sample size case. Further, the use of the exterior product in developing Jacobians for changes of complex variables is uncommon among electrical engineers. Signal processing is most properly classified as an information science, and is quite independent of the use of electrons to implement its ideas. Another difference is that the speed of acoustic signals is significantly slower than for the case of electromagnetically propagated signals. So, this thesis must reside in an interdisciplinary home. With this major impediment set aside, let us continue with the description of the background and content of the subject matter.

1.1.3 What This Thesis is Not

This thesis is not about:

- devising new signal processing structures
- faster or more robust algorithms
- inventing new statistical tests
- comparing old tests
- finding asymptotic distributions of test statistics

- examining Cramer-Rao bounds for estimators
- assessing estimator consistency
- simulating results

1.2 Order Estimation

This thesis concentrates on the problem of determining the number of significant sources present at an array in a noisy environment. This is known as system order determination or system identification in other contexts. More correctly, the question being investigated is the number of arrival paths containing signals that can be distinguished from noise. Often, the question is asked for a fixed frequency.

Several studies in signal processing assume that system order is given or can be obtained. One important work is the introduction of the MUltiple SIgnal Classification (MUSIC) algorithm by Schmidt [238]. He requires knowledge of the number of eigenvalues of the received data matrix that are associated with noise. Another study that requires knowledge of the number of received signals is the thesis on maximum likelihood estimation by Mirkin [183] (p. 37). In Tague's study [263] (p. 140) of stochastic operators and their matrix representations applied to estimator-correlator processors, he examined the relationship between system identification and receiver performance. He

showed that perfect identification is not required in order to improve processor gain, and that even poor identification conducted at a low signal-to-noise ratio (SNR) results in some improvement. The approach of this thesis relies on the testing of eigenvalues.

1.3 Eigensolutions

The number of arrival paths is related to the array element output covariance matrix eigenvalues and is independent of array geometry. I assume in this thesis that the array is unstructured. *The eigenvectors of this covariance matrix is a function of array geometry and the directions of arrival.*

Morrison [186] has a wonderful discussion on the geometric interpretation of eigenvalues and eigenvectors in his discussion on principal components. The eigenvectors define a coordinate system. The eigenvector associated with the largest eigenvalue defines that linear combination of data that produces the maximum variance in the data. The eigenvector associated with the second largest eigenvalue defines that linear combination of data that produces maximum variance subject to the restriction that the second eigenvector is orthogonal to the first eigenvector. Successive axes are defined in the same way. The sample eigenvalues are the variance estimates of the linear combinations of the data defined by the associated eigenvectors. When the eigenvectors are normalized to unit length, they can be thought of as direction cosines which

specify the rotation from the original response axes of the data to the axes given by the set of eigenvectors.

To understand the effect of eigenvalue separation on the accuracy of directions of arrival computed from associated eigenvectors, consider the eigenvectors as being axes of an n-dimensional ellipsoid. Think of the square root of the eigenvalues (the singular values) as being the lengths of the semi-axes. Now, visualize an ellipsoidal shell conforming to this geometry. The sharpness of curvature of the ellipsoidal shell can be thought of as a measure of the stability of the direction-of-arrival estimate or bearing accuracy.

If all the eigenvalues are equal, you have a ball! Hence, a test for equality of eigenvalues is often called a *sphericity test.* There are an infinite number of possible 3-dimensional orthogonal coordinate systems that you can fit to a 3-dimensional sphere. Assuming that the origin of all coordinate systems is at the center of this sphere, the first choice is an arbitrary point on the sphere, like the North Pole. The number of choices is uncountable. This fixes the first coordinate (eigenvector). The second coordinate is constrained to be orthogonal to the first, which places the second choice for a point on the sphere's equator. Even here the number of choices is uncountable. In general, for an n-dimensional sphere, the number of axes for which there are uncountable choices of orientations is n-1.

There will be as many non-zero eigenvalues as there are sources when

there are more sensors than sources and there is no noise. If there is noise then all the eigenvalues will be nonzero. The sensor outputs are random variables and hence the eigenvalues and eigenvectors of the sample covariance matrix are random variables. When the signal-to-noise ratio is large, the large eigenvalues are associated with the signal plus noise and the small eigenvalues are associated with the noise. When the signal-to-noise ratio is small, the determination of the exact number of sources is not as easy. The primary question of this thesis is as follows.

Given two eigenvalues (or groups of eigenvalues) from a noisy process,
is the difference between them due to mere chance,
or is it more likely due to some underlying real cause?

The sensitivity of the accuracy of eigenvectors as a function of (a) eigenvalue separation, (b) underlying distribution determined by the α-mixing of two Gaussian distributions, and (c) covariance estimation method (conventional sample covariance estimation, rank correlation, weighted M-estimate) was the subject of a simulation study by Moghaddamjoo [184]. The concept of α-mixing refers to the convex sum of two or more probability distributions. For the simple two-distribution case, one of the distributions can be called a contaminating distribution. Conventional estimation was best when there was no α-mixing. The rank-correlation (robust) method was best when the contamination factor was 0.1. The weighted M-estimate never was best. These results

were observed at all signal-to-noise ratios. As expected from the geometrical interpretation, when the signal-related eigenvalues were not well separated from each other or from noise, then the estimates of the related eigenvectors were very different from their true values. As long as good eigenvalue separation existed, then the space spanned by the estimated signal eigenvectors was almost the same as the true signal space. When a signal related eigenvalue was close to the noise eigenvalues, there was significant mixing between its corresponding eigenvector and noise related eigenvectors. The only remedy was to increase the overall array signal-to-noise ratio by increasing the number of sensors and filtering the noise as much as possible.

The problem reduces to looking at the sample eigenvalues to test if the corresponding population eigenvalues are the same or significantly different. More generally, the hypothesis I would like to test is $H : c_1^T \Lambda^2 c_1 = c_2^T \Lambda^2 c_2$ versus the alternative $A : c_1^T \Lambda^2 c_1 > c_2^T \Lambda^2 c_2$ where c_1 and c_2 are column vectors of real numbers that specifies linear combinations of eigenvalues contained in the diagonal matrix Λ^2. This is equivalent to a test proposed by Krishnaiah and Lee [153] without providing an expression for the distribution involved.

1.4 Major Assumptions and Rationale for Approach

Since an eigenvalue is the square of its related singular value, we can test the square of the sample singular values to determine the appropriate rank of an approximating covariance matrix for an eigensystem processor [212]. This rank is known as the system order. Once known, beams can be formed to maximize the signal-to-noise ratio in the desired look-directions by cancelling out the interfering point sources using methods described in Monzingo and Miller [185]. The mathematics for optimal processing has been worked out when the system order is known. Progress in the development of statistical estimation techniques that apply to this problem is still being made. For example, see the fascinating thesis by Kundu [158]. The hypothesis testing approach has received little attention.

The order estimation problem can be approached from a strategy of *estimation* or a strategy of *hypothesis testing.* If you choose an estimation strategy, you must know how good your answer is. A confidence level $(1 - \alpha)$ must be chosen to form a confidence interval. If you choose hypothesis testing, the size α of the test must be chosen to construct the critical value against which

the test statistic is compared. In both strategies the choice of α is subjective, whether the choice is made directly or indirectly, such as via cost and utility functions. Regardless of your strategy, you can construct a better an estimator or hypothesis test if you know more about the distributions involved. To even assume that data is drawn from an exponential family distribution is subjective, even when the hypothesis of such an event is not rejected by testing. Explicitly identified subjectivity is not necessarily bad. It enables us to build tractable models and efficiently achieve reasonable results. The charge of "subjectivity" lodged against hypothesis testing by proponents of estimation is an invalid defense of estimation and an invalid claim of advantage of estimation over testing. Estimation and testing both require a choice of α for the results to be meaningful and thus are based on the same underlying theory. Both are worthy candidates for investigation and development. The advantage of estimation over hypothesis testing is that less work is usually involved in obtaining an answer. The usefulness of the answer, however, can only be assessed by assuming a value for α and applying distributional theory.

One characteristic of *acoustic signal processing* that distinguishes it from *processing electromagnetic signals* is the comparatively slow *propagation speed* of acoustic signals. In radar, if you need more independent samples to satisfy applicability of the central limit theorem, you increase your pulse repetition rate. In acoustics, the speed at which data is propagated is slow compared to

the speed of light. This means that decisions must be based on a restricted number of independent samples available per unit time. This drives interest to the small sample case. The desirability of working with a *small sample size* distinguishes this problem in acoustics from one in electrical engineering which is usually satisfied by the large sample case.

The desire to work with *complex random variables* distinguishes the work in this thesis from work that might usually be found in statistics. Bandpass acoustic data is naturally represented with complex numbers. The primary interest in using complex variables in the development of theory is the natural and convenient representation of the time-dependency of physical variables by using the form $exp(i\omega t)$. By applying the Hilbert transform to the array element data, the resulting data stream can be represented as complex numbers. Application to actual data allows us to efficiently do phase comparisons and computations. A very nice discussion in the sonar context is in Ziomek's 1985 book [299] (pp. 176-189). Let our real data stream be the variable $x(t)$ and let the Hilbert transform of $x(t)$ be $y(t)$. The usual notation for the Hilbert transform of $x(t)$ is $\hat{x}(t)$. You can think of the Hilbert transform as being a quadrature filter having $x(t)$ as its input. Then our complex data stream is formed by $z(t) = x(t) + iy(t)$.

A common assumption for purposes of mathematical simplicity when first developing theory for an application in signal processing is that the process

is *stationary.* Application to array processing leads to consideration of complex *multivariate* distributions. The assumption of *Gaussian white noise* is a traditional starting point in signal processing studies because it simplifies the mathematics involved and it is not a bad model for a wide range of situations.

Wooding [293] is often cited as the beginning point for the work with complex normal random variables because he connected it with application to the envelope of a random noise signal. He considered the form of the covariance matrix and density function of the random variable $z_n(t) = x_n(t)+iy_n(t)$ where x_n and y_n are independent normal random variables. Thus, for the complex scalar $z_n(t)$, the real and imaginary parts, $x_n(t)$ and $y_n(t)$, are uncorrelated. He showed that the covariance matrix for the real and imaginary parts of two such complex normal random variables, z_m and z_n, satisfied the following conditions: $E\{y_m y_n\} = E\{x_m x_n\}$ and $E\{x_m y_n\} = -E\{x_n y_m\}$. He derived the density function and the characteristic function of the vector complex normal distribution for the zero mean case. Goodman [92] (p. 173), a pioneer in the study of complex Gaussian statistics, remarked that many stationary non-Gaussian processes become nearly Gaussian when "passed through" sufficiently narrowband filters. Bendat and Piersol [39] provide a cautionary remark that physical phenomena and measured data ultimately are limited by nonlinear restraints in the positive and negative direction, so no random data can be truly Gaussian. Therefore the Gaussian distribution is not appropriate

for looking at extreme events, which are events located in the tails of the distribution. This is precisely where our interest lies for the detection problem, and I will conveniently ignore their wise cautionary remark under the rubric that one should understand what is easy before trying to understand what is hard. Attention is focused on the *complex multivariate normal* and related distributions. The *complex Wishart* distribution is the natural distribution for examining the variability of a sample spectral density matrix.

This thesis focuses on hypothesis testing strategies.

The problem is examined in the context of a *complex variable small sample principal components analysis problem.*

1.5 Organization of Thesis

This thesis is organized as follows. The chapters contain the materials which I judged are mathematically accessible to most engineers and are most directly related to the hypothesis testing question. The appendices contain the supporting mathematical background or results which I judged not commonly accessible to most engineers.

Chapter 2 provides a mathematical statement of the problem as one of a small sample complex principal components test. Chapter 3 reviews other applications that can benefit from eigenvalue tests. Chapter 4 identifies ap-

proaches to the order estimation problem different than the one taken in this thesis. It also includes an exposition of Kiefer and Wolfowitz [140] generalizations of maximum likelihood estimators. This discussion provides an abstract setting within which the process of model order identification and estimation can be viewed as part of the same problem of selecting one or a family of probability measures from among candidates. Chapter 5 reviews previous work on order determination by hypothesis testing. Chapter 6 specifies some statistical tests of interest. Chapter 7 contains the summary and conclusions. Chapter 8 contains recommendations for further research.

The first appendix highlights the mathematical background necessary for this thesis. It identifies good preparatory references and gives examples that illustrate the need for the special care and attention to details. It also outlines the major structure of the three groups of appendices. The last appendix identifies notation conventions and defines special symbols and functions. It is located at the very end to make it easy to use.

The appendices are perhaps the most valuable part of this thesis. They lay the groundwork to support many other efforts. The experienced reader will have used many of these results, having found them in isolated literature, or will have independently developed the results. I know of no systematic thorough presentation of these results explicitly for the complex case. Perhaps the closest to achieving this is the fine text by Stewart [259]. Consequently, I

have taken the liberty of developing results related to my general theme even when they do not follow the very narrow line of reasoning expected in a law court to explore the stated thesis topic. This development was a labor of love initially patterned by Chapter 17 of the wonderful text by Arnold [31]. It expanded to include work derived in great measure by Muirhead [187] and Anderson [26]. These appendices are not in natural pedagogical order, but rather are grouped by my anticipation of which material would be useful to different kinds of readers.

Appendices A through F are accessible to most engineers and are directly related to this thesis. If this thesis is ever read, I expect that this group of appendices to be of the most use to other people. Those who insist on practical results can find some in the wonderful work by Tague [264], which is presented here with some steps that were omitted in his journal article due to lack of space. Appendices G through J are at a more abstract level. The most challenging contributions made in this thesis are given in equation G.10, material related to equation G.16, and theorem 98, all contained in appendix G. All of appendices G through J are necessary for complete understanding of this thesis. Much of it is not new knowledge, but is included to allow engineers to get access to the necessary mathematical background quickly. Appendices K through P form a repository of results that are mundane, useful (for the most part), and are not generally available elsewhere. Other than Appendices

H.1 through H.5 and I, the appendices are results which I have recast from real variables into the complex variables case, or are results I have not seen elsewhere even for the real variables case (yet). The most interesting results in this group of appendices are in appendix L, and the easy results that were fun to produce are in appendix N. The most important of this group of appendices is appendix M, and the most difficult to produce was appendix P.

Chapter 2

MATHEMATICAL STATEMENT OF PROBLEM

2.1 Introduction

In this chapter, I provide a mathematical statement of the problem and test statistics known to apply to the problem. In a later chapter, you will observe I have also included a few other statistics applicable to the order identification problem.

The basic mathematical problem can be stated as follows. Assume that we have m arbitrarily oriented sensors and p sources. In particular, I am not restricting this to a study requiring a linear array. We know m and we want to find p. The value of m is selected with the intention that the assumption $m > p$ is valid. Assume the Gaussian white noise is isotropic and independent of the signals and that the signals are mutually independent. We want the difference between a signal at various sensors to depend only on the time difference due to propagation between the source and the sensors. Therefore, accept the linearized equations for small amplitude acoustics and assume that the sensors are located in the acoustic (but not necessarily geometric) far field of the sources. I do not require an assumption of plane wave propagation

across the array. Those are issues related to beamformer assumptions which are not within the scope of this thesis. The geometry is illustrated in figure 2.1.

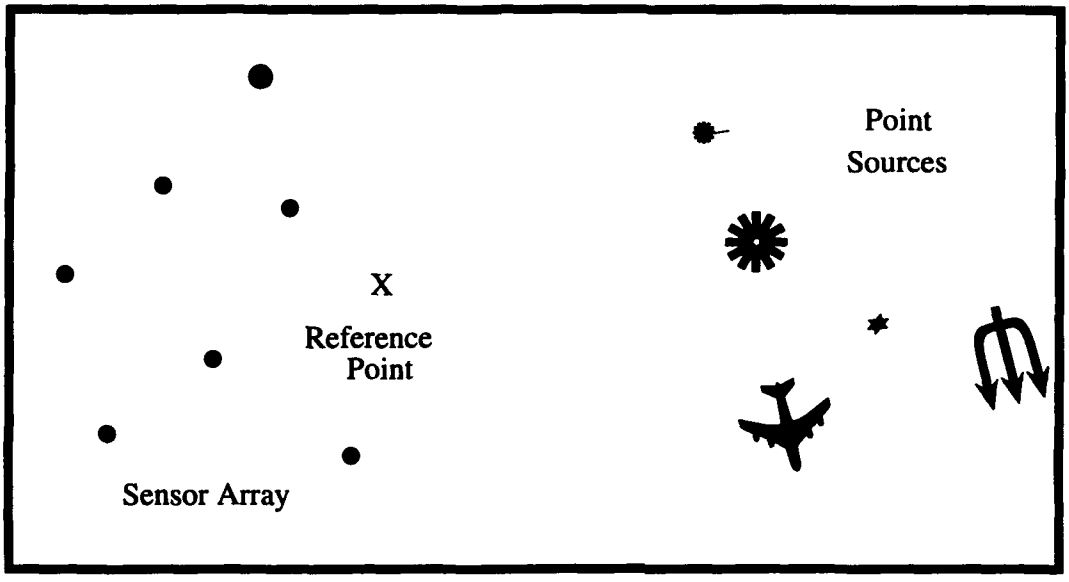


Figure 2.1. Array Geometry

If the array is sampled n times, where $n > m$, then we get the following matrix. Each $X_i(k)$ is a complex random variable.

$$X = \begin{pmatrix} X_1(1) & X_2(1) & \cdots & X_m(1) \\ X_1(2) & X_2(2) & \cdots & X_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ X_1(n) & X_2(n) & \cdots & X_m(n) \end{pmatrix}_{n \times m} \tag{2.1}$$

Regardless of the origin of the elements of matrix X, we can determine

the rank of X by determining the number of nonzero singular values of X. Alternately, we can determine the rank of X by examining the number of nonzero eigenvalues of either XX^H or X^HX. Independence of the samples is not required for the singular values to identify the rank of X. When I finally derive distributional results I will require that the samples be independent to simplify the mathematics. This will allow the assertion that the covariance matrix is a complex Wishart matrix. However, a future development should deal with X without the sample independence constraint, perhaps via studying the singular values.

In absence of noise, the rank of this matrix is the number of sources. I want to find a matrix A of lowest rank that is a best approximation of X in some sense. Then ν= rank(A) is the answer. The random variable ν is our approximation to p. I want to find out what is ν. Suppose that the data in matrix X includes noise. If some matrix Y consists of only the noise data, then we can examine the rank of $Z = X - Y$. We may examine the rank of X or Z directly by looking at their singular values obtained from a Singular Value Decomposition (SVD), or by looking at their eigenvalues obtained from an Eigenvalue Decomposition (EVD) of X^HX or Z^HZ. Eaton and Perlman [73] showed that X^HX is of full column rank m with probability 1. Okamoto [197] showed that the eigenvalues of such a matrix are all distinct. Let the SVD of X be given by $X_{n,m} = \sum_{i=1}^{m} l_i P_i Q_i^H$, and let the singular values of

X be ordered according to $l_1 > l_2 > \cdots > l_m$. Let L be the rectangular matrix containing the diagonal matrix of the singular values $\{l_i\}$ in its upper left corner. The norm

$$\min_{\rho(A)=\nu\leq m} \| X - A \|^2 = l_{\nu+1}^2 + \cdots + l_m^2$$

is attained when $A = \sum_{i=1}^{\nu} l_i P_i Q_i^H$. The $\{P_i\}$ are the left singular vectors, and they satisfy $P^H P = I_n$. The $\{Q_i\}$ are the right singular vectors, and they satisfy $Q^H Q = I_m$. The $\{l_i^2\}_{i=1}^m$ are the non-zero eigenvalues of both $X^H X$ and $X X^H$. Let $B = X^H X$. Define $B_j = \sum_{i=j+1}^m l_i^2 Q_i Q_i^H$. The matrix B_j is an approximation of the matrix B formed with the smallest $(m-j)$ eigenvalues and corresponding eigenvectors of matrix B. We will see these again in a moment.

I essentially want to perform a test for sphericity on the smallest $m - \nu$ eigenvalues. We seek to determine if they are the same for practical purposes, or if at least one of them is significantly different from the others. Proceed in a sequential manner with different values of j. We want to find out what's ν. The order in which you test is your test strategy. The order you choose depends on your confidence in which direction of testing, from small to large eigenvalues or large to small eigenvalues, will result in a successful identification of the rank of the non-noise contribution to X with the least amount of computational work.

Suppose you have no signals. With probability 1, no two sample eigen-

values will be the same even though there is only one underlying population eigenvalue. The smallest sample eigenvalue will underestimate the common population eigenvalue, and the largest sample eigenvalue from this noise-only matrix will overestimate the population eigenvalue. This means that if you want to estimate the smallest eigenvalue, you should use an average of the sample eigenvalues you have classified as belonging to the same population eigenvalue rather than using the smallest sample eigenvalue by itself. Doing the latter would bias your estimate. The testing situation may be different because the distribution of the sample eigenvalues accounts for this problem (and in fact, causes the problem). When testing a new sample eigenvalue for inclusion in a set associated with underlying equal population eigenvalues, you should include in your test as many sample eigenvalues as you have already classified as being the same population eigenvalue.

2.2 Specific Test Statistics

If you have an array with many sensors and an environment of only a few sources, then consider sequential tests of sphericity beginning with the full matrix. The usual test for sphericity uses the maximum likelihood ratio test statistic developed by Anderson [24]. Anderson determined the large sample

(asymptotic) distribution of T_1.

$$T_1 = n \ln \left[\frac{\left(\frac{1}{m-\nu} \sum_{i=\nu+1}^{m} l_i^2 \right)^{m-\nu}}{\prod_{i=\nu+1}^{m} l_i^2} \right] \tag{2.2}$$

A form for which a density function might be easier to derive is T_2. This is essentially equations (14) and (23) of Wax, Shan, and Kailath [279].

$$T_2 = \frac{\prod_{i=\nu+1}^{m} l_i^2}{\left[\frac{1}{m-\nu} \sum_{i=\nu+1}^{m} l_i^2 \right]^{m-\nu}} \tag{2.3}$$

For the special case of $m - \nu = 2$, the density function of T_2 is given as equation 6.17, and the cumulative distribution function is given as equation 6.18. This is the same as the statistic u that Muirhead [187] uses in the case of real variables. We will see that equation 6.15 is very similar to equation 2.3.

Another statistic to consider is T_3 or its inverse. This is suggested by C. R. Rao [212] (equations 3.10, 3.11, and 17.1).

$$T_3 = \frac{l_1^2 + \cdots l_m^2}{l_1^2 + \cdots l_\nu^2} = \frac{\mathrm{tr}(B)}{l_1^2 + \cdots l_\nu^2} \tag{2.4}$$

The density of T_3 can be obtained by theorem 8. The statistic $\frac{1}{T_3}$ has the interpretation as being the fraction of the total variance explained by those eigenvalues attributed to being influenced by the signals. Alternatively, you could test that the last few eigenvalues explain only a small fraction of the data as in T_4.

$$T_4 = \frac{\mathrm{tr}(B)}{l_{\nu+1}^2 + \cdots + l_m^2} \tag{2.5}$$

The density of T_4 can be obtained by theorem 8. As a point of convenience, note that $\mathrm{tr}(B) = \mathrm{tr}(X^H X) = \mathrm{tr}(X X^H)$. Another concept that is useful is to test if the largest p eigenvalues are significantly different than the smallest $m - \nu$ eigenvalues as in T_5.

$$T_5 = \frac{l_1^2 + \cdots + l_p^2}{l_{\nu+1}^2 + \cdots + l_m^2} \tag{2.6}$$

The density of T_5 can be obtained by theorem 8.

In a real ocean environment with multipath propagation, you may want to distinguish the direct (refracted) path from other paths using the assumption that the signal-to-noise ratio along the direct path is greater than by other paths. This is a bit simplistic, and a more intelligent model could be made. Then you might want several partitions of $\{l_i^2\}_{i=1}^m$ to test on. To really confuse the issue, you could go back to the sample covariance matrix and perform tests on selected entries in that matrix to compare elements to each other or to known constants.

Let the population eigenvalues be denoted by $\Lambda^2 = \mathrm{diag}(\lambda_1^2, \cdots, \lambda_m^2)$. Let a column vector of real constants $c \in R^m$ be used to construct linear combinations of the population and sample eigenvalues. Let e_j be a column vector in R^m with all zeros except for a 1 in the j^{th} position. We will construct our choices of various vectors c using sums of selected e_j. Construct a general test

statistic T_6.

$$T_6 = \frac{\frac{1}{n}\left(\frac{nc_1^H L^2 c_1}{c_1^H \Lambda^2 c_1}\right)}{\frac{1}{n}\left(\frac{nc_2^H L^2 c_2}{c_2^H \Lambda^2 c_2}\right)} \tag{2.7}$$

Let the distribution of this test statistic be dependent on n. Denote the distribution function by f_{06} when $c_1^H \Lambda^2 c_1 = c_2^H \Lambda^2 c_2$. The density of T_6 can be obtained by theorem 8.

Now, suppose that we want to test if there are exactly ν sources. Assume that it is already established that the last $m - \nu$ eigenvalues are identical. If $\lambda_\nu^2 = \lambda_{\nu+1}^2$, then $\lambda_\nu^2 = \frac{1}{m-\nu} \sum_{j=\nu+1}^{m} \lambda_j^2$. This leads to $c_1 = e_\nu$ and $c_2 = \sum_{j=\nu+1}^{m} \frac{1}{\sqrt{m-\nu}} e_j$. For this selection of c_1 and c_2 we have met the goal of $c_1^H \Lambda^2 c_1 = c_2^H \Lambda^2 c_2$. If $c_3 = \sum_{j=\nu+1}^{m} e_j$ we get the relationship $c_1^H \Lambda^2 c_1 = \frac{1}{m-\nu} c_3^H \Lambda^2 c_3 = c_2^H \Lambda^2 c_2$ when $\lambda_\nu^2 = \lambda_{\nu+1}^2$. When this is true, the test statistic becomes T_7.

$$T_7 = \frac{(m-\nu) c_1^H L^2 c_1}{c_3^H L^2 c_3} \tag{2.8}$$

The density of T_7 can be obtained by theorem 8. The null hypothesis is H_0 : $(m-\nu) c_1^H \Lambda^2 c_1 \leq c_3^H \Lambda^2 c_3$. Written out in terms of the individual population eigenvalues, this is $H_o : \lambda_\nu^2 \leq \lambda_{\nu+1}^2 = \cdots = \lambda_m^2$. The alternate hypothesis is given by $H_a : \lambda_\nu^2 > \lambda_{\nu+1}^2 = \cdots = \lambda_m^2$. If $T_7 \leq f_{07(1-\alpha)}$, then conclude to not reject H_0; otherwise reject H_0 and choose H_a.

Another desirable question is to ask if there are no more than ν sources. Suppose you have concluded that there are exactly ν sources. Then the best estimator of X is given by $X_{[\nu]} = \sum_{i=1}^{\nu} l_i P_i Q_i^H$. What is left over should be due

to noise alone and should therefore be spherical. Let $Y = \sum_{i=\nu+1}^{m} l_i P_i Q_i^H$. The statistic for testing the sphericity of $Y^H Y$ is given by T_8.

$$T_8 = n \ln \left[\frac{\left(\frac{1}{m-\nu} \sum_{i=\nu+1}^{m} l_i^2 \right)^{m-\nu}}{\prod_{i=\nu+1}^{m} l_i^2} \right] \tag{2.9}$$

Let $f_{08}(m-\nu, n)$ be the distribution of the test statistic T_8 when $\lambda_{\nu+1}^2 = \cdots \lambda_m^2$. The null hypothesis is $H_0 : \lambda_{\nu+1}^2 = \cdots \lambda_m^2$. The alternate hypothesis is H_a : one or more of the λ_i^2 are different from the rest, or equivalently, not all the λ_i^2 are equal. If $T_8 \leq f_{08(1-\alpha)}(m-\nu, n)$ then do not reject H_0. Otherwise, reject H_0 and conclude H_a.

Suppose that the noise covariance matrix R is known. Then we want to find the rank of the matrix $W = B - R$. There is a problem with a direct approach when all the eigenvalues of W are zero in that such distribution density functions become undefined. However, this is precisely what we want to look at. Alternatively, let the eigenvalue decomposition of W be $W = QL^2Q^H$. Then, let the eigenvalue decomposition of R be given by VD^2V^H. We can test if the last $m-\nu$ eigenvalues of B equal the last $m-\nu$ eigenvalues of R. Define the test statistic T_9.

$$T_9 = \frac{c_1^H L^2 c_1}{c_1^H D^2 c_1} \tag{2.10}$$

The density of T_9 can be obtained by theorem 8. Let $f_{09}(n)$ be the distribution of T_9 when $c_1^H \Lambda^2 c_1 = c_1^H D c_1$ is true. Let $D^2 = \text{diag}(d_1^2, \cdots, d_m^2)$. Then the

null hypothesis is $H_0 : \lambda_\nu^2 = d_\nu^2$ and the alternate hypothesis is $H_a : \lambda_\nu^2 \neq d_\nu^2$. If $f_{09(1-\frac{\alpha}{2})}(n) \leq T_9 \leq f_{09(\frac{\alpha}{2})}(n)$ then do not reject H_0, otherwise conclude that H_0 is rejected and therefore chose H_a. When H_0 is true, B is not of rank ν, and there are not ν significant sources. When H_0 is false, we reject H_0, and by default choose H_a, concluding that there are ν significant sources.

We can not use the sphericity test on W because all the tested eigenvalues are zero under the null hypothesis, and the density function of the test distribution possibly will not exist. We can test that there are no more than ν sources by comparing the sums of eigenvalues of B and the sums of eigenvalues of R. Assume that there are no more than $\nu+1$ sources. In practice, this should not be a problem for the proposed test. The null hypothesis is given by $H_o : \lambda_{\nu+1}^2 + \cdots + \lambda_m = d_{\nu+1}^2 + \cdots + d_m^2$, and the alternate hypothesis by H_a : equality does not hold. For this problem, let $c_1 = \sum_{i=\nu+1}^{m} e_i$ and compute test statistic T_9. If $f_{09(1-\alpha)}(n) \leq T_9$ then do not reject H_0, otherwise reject H_0 and conclude H_a. We are not looking at λ_ν^2. If the last $m - \nu$ eigenvalues of B equal those of R, then the rank of W is less than $\nu + 1$. Therefore, the rank of W is no more than ν.

Suppose now that you do not know R, but you have an estimate of R which we will call S. Then we want to find the rank of the matrix $V = B - S$. Let the eigenvalue decomposition of V be $V = QL^2Q^H$. You would like to know if the last $m - \nu + 1$ eigenvalues of V are small enough to be considered zero.

Define the test statistic T_{10}.

$$T_{10} = \sum_{i=\nu}^{m} l_i^2 \tag{2.11}$$

Let $f_{10}(n)$ be the distribution of T_{10} when $c_1^H \Lambda^2 c_1 = 0$ is true. Then the null hypothesis is $H_0 : \lambda_i^2 = 0$ and the alternate hypothesis is $H_a : \lambda_i^2 \neq 0$. If $T_{10} \leq f_{10(1-\alpha)}(n)$ then do not reject H_0, otherwise conclude that H_0 is rejected and therefore chose H_a. When H_0 is true, B is not of rank ν, and there are not ν significant sources. When H_0 is false, we reject H_0, and by default choose H_a, concluding that there are at least ν significant sources.

A sequential test for rank that begins with the largest eigenvalue may be practical in systems with a large number of sensors and a few expected sources. The idea is to test the ratio of the largest ν eigenvalues to the sum of all the eigenvalues, as given by statistic T_{11}.

$$T_{11} = \frac{l_1^2 + \cdots + l_\nu^2}{l_1^2 + \cdots + l_m^2} = \frac{l_1^2 + \cdots + l_\nu^2}{\mathrm{tr}(B)} \tag{2.12}$$

The density of T_{11} can be obtained by theorem 8. A sphericity test could similarly be constructed for the eigenvalues yet to be estimated, such as T_{12}.

$$T_{12} = n \ln \frac{\left[\frac{1}{m-\nu}\left(tr(B) - \sum_{i=1}^{\nu} l_i^2\right)\right]^{m-\nu}}{\left(\frac{\det(B)}{\prod_{j=1}^{\nu} l_j^2}\right)} \tag{2.13}$$

Of course, it is also possible to consider the difference between adjacent sample eigenvalues $l_i^2 - l_{i+1}^2$ or the ratio $l_i^2/(\beta l_{i+1}^2)$ where β is some real constant

of proportionality of interest. If you expect λ_{i+1}^2 to be a variance associated with noise and λ_i^2 to be signal plus noise variance, then you may be interested in $\beta = 10^{0.1}$ to correspond to a 1 dB $(S+N)/N$ ratio, or $\beta = 10^{0.5}$ for a 3 dB ratio. Similar statistics between any two sample eigenvalues of interest may also be appropriate, such as $l_i^2 - l_j^2$ or $l_i^2/(\beta l_j^2)$. Values of j that may be of special interest are 1 and p. You may want to use the average of the sample eigenvalues $\alpha = \frac{1}{p}\sum_{i=1}^{p} l_i^2$ instead of l_j^2. All test statistics (except of the form of T_1) are heuristically based, drawing from background in regression analysis.

In this chapter, the problem has been cast as a problem in complex principal components analysis. A number of testing situations with their commonly known test statistics have been presented. The goal of this thesis is to develop the density functions for these statistics We sill see some more easily derived density functions for closely related tests presented during our search for the densities of the stated tests. This is a problem in dimensionality reduction.

Chapter 3

OTHER APPLICATIONS

The purpose of this chapter is to briefly review applications of system identification in which the eigenvalue based hypothesis testing approach might be used. In addition, applications are suggested for acoustic emission analysis and acoustical oceanography. Those are areas I am interested in and have not seen evidence of order determination applied. Even though the abstract problem has its own beauty, this chapter shows that it also has usefulness. Application to acoustic emission analysis will be discussed in section 3.1 and acoustical oceanography will be discussed in section 3.2. This section briefly identifies a variety of other applications.

Goodman was an early pioneer in the distribution theory and application of complex random variables. He reported that geophysicists treat simultaneous measurements at several positions in the ocean of the height of gravity waves generated by the wind as multivariate complex normal records [92].

Krishnaiah and Waikar [147] reported that the distributions of the intermediate roots can be used for reduction of dimensionality in pattern recognition problems and principal component analysis. In nuclear physics, the distributions of any few consecutive ordered roots are useful for finding the distributions of the spacings between the energy levels of certain complicated systems [53][174]. Krishnaiah and Waikar referenced Wigner with regard to

applications in physics [284][285][287]. Krishnaiah and Shuurmann [151] applied methods similar to those developed in the present research to vertical and horizontal accelerometer data to examine the vibration at different locations of the cargo deck on a C-5A aircraft. They also referenced Cooper and Cooper's work [60] in non-supervised signal detection and pattern recognition. Horel [111] wrote a very nice article on the theory and practice of complex principal component analysis. He said it has been shown to be a useful method for identifying traveling and standing waves in geophysical data sets. The frequency domain principal component (FDPC) analysis is the most general of the available methods of studying propagating phenomena. Complex principal component (CPC) analysis in the time domain is considered an attractive alternative to FDPC analysis. CPC analysis is essentially FDPC analysis averaged over all frequency bands.

Krishnaiah [150] references Liggett's work [166] in passive sonar and Priestly's work [209] in system identification. Kelly et al. [134] applied concepts of statistics of complex variables in an active sonar acoustic imaging problem where noise $n(t)$ was distributed according to $\mathbf{C}N(0,\Sigma)$ where $\Sigma = \frac{N_0}{2}I$. Tague [264] used concepts developed during this thesis research for evaluating the signal-to-noise ratio of a beamformer output. The complex matrix normal distribution, whose form is verified in this thesis, is the natural setting for beginning analysis of two-dimensional spatial data such as found in rectangular

sonar arrays.

The solution to the problem of this thesis is also the solution to some applications involving remote sensing, such as data compaction. It will allow automation of a wide range of analyses now requiring application area specialists. The list of areas to which these methods apply is growing as people discover how to work with complex random variables. For some other applications, see references [165] and [128]. Other references to the statistics of complex variables include [30][70][132][133][127][62][178][179].

3.1 Acoustic Emission Analysis

Acoustic emission testing is the detection, location, and analysis of acoustic emissions from materials under static or dynamic stress. The term "acoustic emission" (AE) refers to the class of phenomena whereby transient elastic waves are generated by the rapid release of energy from localized sources within a material, or the transient elastic waves so emitted. Other (less preferred) terms used for the same phenomena are "stress wave emission" (SWE) and "microseismic activity". Standard definitions for terms relating to acoustic emission are given in reference [32].

Short [243] noted that the first major systematic approach to acoustic emission of materials under stress was by Kaiser. Kaiser concluded that the number of emissions increased with the applied stress, and that after unloading

there was no acoustic emission upon reloading until the previous maximum load was exceeded. This is known as the Kaiser effect, and is observed both in metals and composites at low loads. If a composite is not held at a load in the elastic region until all emissions have stopped and is unloaded, emissions then occur at a load lower than the previous maximum load.

Acoustic emissions are detected using one or more transducers, usually piezoelectric transducers, to obtain an electric signal proportional to the mechanical vibration at the location of the transducer. An array of transducers is required to locate the source of an emission by comparing the arrival times of acoustic transients at each transducer. A multichannel analyzer is used to cross-correlate the signals in the time domain.

An acoustic emission may be identified by its signature in the time and frequency domain. Within the time domain, the important parameters are the amplitude rise time and emission duration. Emissions are also classified by their frequency spectrum. Together, emissions are characterized by their time-dependent frequency distribution. This is a function of the type of material, geometry, structures coupled to it, the applied stress, and the mechanism producing the emission.

AE testing is still in its infancy. Theoretical work lags far behind its use in practical applications. A very basic open question is why growing cracks in some materials emit many AEs while in other materials growing cracks emit

hardly any AEs [168]. A procedure for AE testing for fiberglass reinforced plastic tanks is contained in [33].

Applications require recognition that sound propagation is dispersive. When the structure is liquid loaded, the analysis must also recognize that there is coupling between propagation modes. A very short time after emission, most (over 93%) of the energy is in bending waves, which means that a surface mounted transducer will be effective as a sensor [77].

The use of triangulation which works nicely against a point source is not optimum against a source that is spatially extended or against multiple sources [195]. Triangulation search for emission sites is time consuming and makes poor use of the data. Alternatives include the use of surface mounted arrays. This overcomes the problem of detecting and locating multiple sources, and mapping of sources that are not small enough to be considered point sources. This approach was examined by Simaan et al. [244]. The authors assumed a constant speed of sound and thus treated only longitudinal waves. However, by processing signals at a selected frequency, these concepts can be applied to bending (transverse) waves. By doing this at several frequencies, an added benefit is that the time-dependent frequency signature at the source location can be reconstructed which aids the classification of the type of emission.

AE data is very noisy. Deciding how many sources exist is typically determined by the judgment of the engineer in post-processing of the data. Re-

moval of the analysis engineer from the immediate test environment restricts the ability to take timely action for follow-up testing, or examination of the test environment for explanations of the signal that may be due to something possibly obvious to an on-site observer. Thus, *in situ* automated detection and emission site determination not only increases the efficiency of the testing, it allows observation of causes that otherwise would escape notice or explanation.

Because AE data is noisy, the covariance matrix will contain all nonzero eigenvalues. It is critical to determine which of the eigenvalues are associated with AE signals and which are associated with noise. When the ratio between adjacent eigenvalues is large, making this judgment by merely examining the eigenvalues without other processing is appropriate. Almost always, the large eigenvalues will be associated with an AE of interest, and the small ones will be associated with noise. When the ratio between adjacent eigenvalues is not large, then it is more difficult to make the judgment without a more formal approach.

In traditional AE testing, a tank or vessel is subjected to artificially induced forces to place the material under enough stress to produce emissions at existing flaws. For example, a tank might be pressurized well above its normal operating pressure. Enough emissions are produced, and the monitoring period is long enough, that the signal-to-noise ratio is large enough to produce a detectable and usable signal.

There are circumstances where this might be undesirable. For example, traditional methods of applying the stress might be very expensive, time consuming, or hazardous. This might be the case for testing the hull of a ship. The cost of testing might be significantly less by not requiring the ship to enter dry dock. By decreasing the required signal-to-noise ratio, it might be possible to use the normal operating forces of the industrial process, or the forces of nature, to provide the stress-inducing force needed to produce AE events under usually safe conditions.

If the industrial process is critical and possibly hazardous if corrective action is not taken shortly after the onset of a failure, continuous monitoring might be desirable. This means that monitoring must be done under normal operating conditions, which might not usually induce stresses large enough to generate enough high level emissions to be detectable by present means. An alternative monitoring technology is to embed or coat the object with optical fibers or very thin wires. When a crack occurs in the material, the fiber or wire breaks, detecting the existence of the first crack. The problem with this method is that only the first crack along the filament is detectable. Subsequent cracks along that filament are not detectable. A field monitoring technique, such as EM detection, is required. Monitoring a nuclear reactor vessel might be such an application.

Another motivation for wanting to make detection of AE events at a lower

signal-to-noise ratio is to increase the area under effective testing. This could speed testing of very large structures and thus decrease costs. This might be the case for large natural gas tanks or pipelines.

3.2 Acoustical Oceanography

The problem addressed in this thesis is the same as the problem of determining the number of different arrival angles at a vertical line array. In a propagation loss experiment, for a fixed frequency, each arrival angle can be associated with a different propagation mode. By determining the vertical directions of arrival of a test signal during propagation loss experiments, it is possible to determine more precisely the energy distribution of sound among the propagating modes. Such examination is useful *in situ* to determine the adequacy of the hypothesized propagation loss model used in planning the experiment, judging if the propagation conditions are acceptable for continuation of the experiment as planned, and planning the source placement for additional samples if any are needed to meet the experimental goals. When transients are used as sources, it is necessary to determine the number of received modes and directions of arrival at a given frequency based on only a few samples. Verification of mode presence early in an experiment and accounting for actual environmental conditions allows for adjusting sensor depths to construct a mode filter for use for the remainder of the experiment. This will increase the signal-to-noise ratio,

allowing better data capture and analysis.

The general statistical techniques developed in the process of research for this thesis can be applied to the multivariate analysis of ambient noise when environmental parameters are also recorded. Krishnaiah [150] notes that the problems of testing the hypotheses on complex multivariate populations play an important role in drawing inference on the multiple stationary Gaussian time series since certain suitably defined sample spectral density matrices of these time series are approximately distributed as complex Wishart matrices. Jobst and Adams [122] studied the statistics of ambient sea noise using two deep arrays in the North Atlantic separated in depth and by several miles. They reported that the statistical tests showed that most observations of narrow-band noise were consistent with the hypothesis that the in-phase and quadrature components of ambient noise are zero-mean Gaussian processes with equal power. Noise power is locally homogeneous over the array aperture, and stationary for periods up to 22 minutes at 75 Hz. As a function of frequency, narrow-band ambient noise measurements are consistent with the hypothesis of constant power in adjacent bands up to 0.22 Hz wide. When analyses were extended to 0.8 Hz bands the noise power was no longer constant.

Matsumoto [172] (p. 358) assumed isotropic Gaussian noise in reporting on characteristics of Sea MARC II phase data. McDaniel [173] considered the

underneath surface of the Arctic ice canopy to have a zero mean Gaussian height distribution with an rms roughness of 1-2 meters for the purpose of modeling high frequency forward scattering.

It is cautioned that the distributions that noise sources obey do vary according to their cause. For example, wind-driven sea surface noise has a different distribution that noise due to long range shipping. Further, these will be differently distributed than noise from snapping shrimp on a shallow ocean floor, porpoise and whale whistles and clicking in the ocean volume, or oil industry generated noise on the ocean floor.

Chapter 4

OTHER APPROACHES

4.1 General Discussion

The purpose of this chapter is two-fold. First, it presents a setting in which order estimation and parameter estimation are subsumed into one approach. In the abstract setting, the problem reduces to finding that probability measure, from all the candidate probability measures, which "best" explains the data. The second purpose is to present a very brief catalog of methods for order determination other than that being examined in this thesis.

There are other approaches to model order identification. Methods traditional to statisticians can be found in texts for statisticians on linear models. This is a question often asked when building regression models. Some techniques used for order determination for regression models include the maximum correlation squared, the C_p, forward step wise variable inclusion, backward step wise variable exclusion, and other criteria. Söderström [250] considered the use of Wilks' likelihood ratio statistic and the F-test for comparing two competing models. Prasad and Chandna [208] hint at use of canonical correlation between array subsets, where their application is bearing measurement. Methods traditionally used for model order determination in the context of linear regression analysis can be found in Neter and Wasserman [190]. Most

of the comments made in this section are taken from one or more references.

Kundu [158] discusses application of Cross Validation. The Cross Validation approach was studied by Lachenbruch (1975), Stone (1974 and 1977), Dawid (1974), and C. R. Rao (1988). Cremona and Brandon [63] refer to a Singular Value Plot criterion R_t. Others have looked at Jackknife procedures and Bootstrap procedures. Bouvet [41] considered a Bayesian approach.

Recent developments popular in the electrical engineering model order determination concentrate on techniques based on information-theoretic criteria. These techniques are usually referenced in the literature by their initials rather than their long title. An ancestor of these methods can be seen in the 1954 book by Savage (p.235 ff)[232]. He considers the evaluation of information given two neighboring values of the parameter of an estimation problem. He uses the concept of differential information which he says is even older than Fisher's information.

The recent motivation for the information-theoretic approach is based on the work by Akaike. His work is traceable to 1968, and he continued publishing at least as late as 1979. A listing of 22 of his publications ([2] through [23]) gleaned from other papers referencing his work appears in the bibliography. It was his innovative 1974 paper [16] discussing his method known as AIC (Akaike Information Criteria) that is primarily responsible for the tremendous subsequent world-wide activity in the information-theoretic approach.

Akaike explains (Section V, p.719 of [16]) that IC stands for information criterion and A is added so that similar statistics, BIC, DIC etc., may follow. Rissanen [224][225] and Schwartz [239] developed the MDL (Minimum Description Length) method in 1978 and 1982. In 1986, Zhao, Krishnaiah, and Bai [297][298] derived a statistically consistent estimator generalization of AIC which is called EDC (Efficient Detection Criterion) or GIC (General Information Criterion). In 1989, C. R. Rao and Y. Wu [219] proposed two discriminant criteria that are strongly consistent. Other methods are CAT (Criterion Autoregressive Transfer), by Parzen in 1974 [203], and FPE (Final Prediction Error). An *ad hoc* method is NEE (Noise Error Estimation).

4.2 Generalized Maximum Likelihood Estimators

4.2.1 Introduction

If you *choose the best probability measure to fit your random sample*, then *you have determined the order* of your system. Thus, we seek the measure that has a covariance matrix of the right rank and also the proper parameter values if the distribution family considered is parameterized. Note that this is stronger than just determining the order of a system.

More abstractly, families of distributions with covariance matrices of differ-

ent ranks, taken together, merely form a larger family of measures from which to choose. This can also be extended to sets of different kinds of distributions, such as considering simultaneously the normal and Poisson distributions. In fact, in a parameterized family of distributions, for each fixed parameter, you have an entirely different distribution. Except for computational convenience, there is no reason to explicitly consider parameters when finding a maximum likelihood estimator. A maximum likelihood estimator is merely the selection of that measure, from among all measures you are allowed to look at, that best fits the data from your random sample. Thus, you can even consider an unparameterized class of measures. You might properly argue that in establishing sequences, the imposed indexing becomes a parameter even though the index does not appear as part of a functional expression of the distribution.

The following discussion decodes remarks by Kiefer and Wolfowitz (p. 892-893) [140] on several ways of generalizing maximum likelihood estimators. The first set of generalizations treat the issue when the supremum of the likelihood estimators is not contained in the allowable set. The second set of generalizations repeat the first, but with the additional quality of using the Radon-Nikodym derivative as a generalized probability density function. Taken together, these approaches extend the classes of functions for which *a* maximum likelihood estimator can be obtained. Application of these concepts to the order determination problem was suggested by C. R. Rao [215].

The close reader will observe that the application of these ideas, where the allowable set of underlying covariance matrices are of different ranks and from different complex Wishart distributions, may be problematical. At issue is that all the measures under consideration must be defined on the same σ-algebra. In the idealized case, you end up with problems wanting to consider measures with different σ-algebras. For example, if you consider a singular bivariate distribution in $\mathbf{R}^2$, the Lebesgue measure $\lambda(\mathbf{R}^2)$ of a line is zero. Either you decide that the offending set is allowable, albeit of measure zero, or you decide that such a set is not in the σ-algebra. Under the first interpretation, the following theory applies. Under the second interpretation, the following theory does not apply. The physical world is much nicer because we never have the case of a truly deficient covariance matrix. The problem becomes one of testing for significant differences. This is, therefore, one case where the abstraction of an idea actually produces an approach that is very practical.

4.2.2 Lebesgue-Radon-Nikodym Theorem

In this section we present a statement of the subject theorem and define the terms which will be used in the study of the likelihood estimators of Kiefer and Wolfowitz. This material is from Rudin [230]. We begin with a few definitions. Let μ be a positive σ-finite measure on a σ-algebra $\mathcal{M}$ in a set X, and let λ be a complex measure on $\mathcal{M}$. Then

Definition 1 *$\lambda \ll \mu$ means $\lambda(E) = 0$ for all sets $E \in \mathcal{M}$ for which $\mu(E) = 0$.*

Definition 2 *If there is an $A \in \mathcal{M}$ such that $\lambda(E) = \lambda(A \cap E)$ for every $E \in \mathcal{M}$, then we say λ is* concentrated on *A.*

Definition 3 *Let λ_1, λ_2 be measures on $\mathcal{M}$. Let $A, B \in \mathcal{M}$ such that $A \cap B = \Phi$ (the empty set), where λ_1 is concentrated on A, and λ_2 is concentrated on B. Then λ_1 and λ_2 are* mutually singular, *and we write this condition as $\lambda_1 \perp \lambda_2$.*

Theorem 1 *The theorem of Lebesgue-Radon-Nikodym. Let μ be a positive σ-finite measure on a σ-algebra $\mathcal{M}$ in a set X, and let λ be a complex measure on $\mathcal{M}$. Then*

(a) There is then a unique pair of complex measures λ_a and λ_s on $\mathcal{M}$ such that

$$\lambda = \lambda_a + \lambda_s \qquad (\lambda \text{ is partitioned})$$

$$\lambda_a \ll \mu \qquad (\lambda_a \text{ is absolutely continuous with respect to } \mu)$$

$$\lambda_s \perp \mu \qquad (\lambda_s, \mu \text{ are mutually singular})$$

(b) There is a unique $h \in L^1(\mu)$ such that

$$\lambda_a(E) = \int_E h \, d\mu$$

for every set $E \in \mathcal{M}$.

Some remarks are in order regarding what is important about the above theorem.

1. $(\lambda_a, \lambda_s)_\mu$ is called the *Lebesgue decomposition of* λ *relative to* μ.

2. *Existence* of the decomposition is the significant part of (a).

3. Part (b) is known as the *Radon-Nikodym theorem.*

4. The function h is called the *Radon-Nikodym derivative* of λ_a with respect to μ.

The theorem, remarks, and definitions make much more sense after looking at figure 4.1.

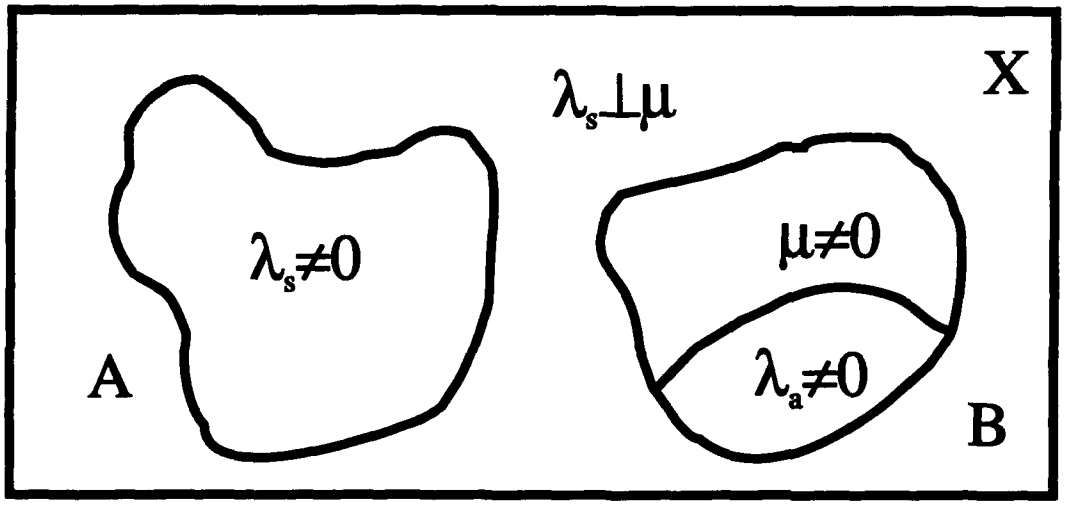


Figure 4.1. Graphic Representation of the Lebesgue-Radon-Nikodym Theorem

In this figure, the complete region inside the frame represents the set X. We have defined two *measurable sets* in the *same* σ-algebra $\mathcal{M}$. We will refer to *measures* which are defined on this *common* σ-algebra. Set A, in the left half of the figure, is the set of elements of X on which the measure $\lambda_s \neq 0$. We can

say that set A is the *support* of measure λ_s, or we say that λ_s is *concentrated* on A. Everywhere outside of A we know that $\lambda_s = 0$. At those points in X where $\lambda_s = 0$, we say that λ_s is singular. Similarly, set B is the set of elements of X on which measure $\mu \neq 0$. Thus, B is the set on which μ is concentrated. In this particular example, the sets A and B are disjoint. In those regions where both $\lambda_s = 0$ and $\mu = 0$, we say that λ_s and μ are mutually singular. We denote this by $\lambda_s \perp \mu$.

The notation for mutual singularity, $\lambda_s \perp \mu$, is suggestive of orthogonality. Mutual singularity is a mathematically stronger concept than orthogonality. All mutually singular functions are mutually orthogonal, but mutually orthogonal functions are not necessarily mutually singular. Functions that are mutually orthogonal may individually attain non-zero values on the common set over which the pair of functions are orthogonal.

Mutual singularity is a property of functions that are measures defined on a common sigma-algebra. It is useful to think in terms of these functions as having mutually exclusive support.

Orthogonality is a property of a pair of functions, a common domain, and a relation defined on those functions over the entire domain. Orthogonality does not require the pair of functions to be measures. Orthogonality is a concept usually dealt with when discussing inner product spaces. However, the inner product is a stronger concept than what orthogonality requires of its relational

operator.

Suppose we have another measure λ_a that is concentrated on some subset (possibly all) of B. Then everywhere λ_a is nonzero we know that μ is also nonzero. An alternate way of saying the same thing is that everywhere μ is zero, we require λ_a to also be zero. When this is true for every measurable set E belonging to $\mathcal{M}$, we say that μ *dominates* λ_a. We denote this by $\lambda_a \ll \mu$.

The Lebesgue-Radon-Nikodym theorem says that when you are given a positive σ-finite measure μ on σ-algebra $\mathcal{M}$ in a set X, and also given any complex measure λ also defined on $\mathcal{M}$, then this measure λ has a *unique* decomposition $\lambda = \lambda_a + \lambda_s$ satisfying the conditions that $\lambda_a \ll \mu$ and $\lambda_s \perp \mu$. Another way of saying this is that for any given pair of measures (λ, μ) that are defined on the same σ-algebra $\mathcal{M}$, then there exists some subset A of X on which $\lambda \neq 0$ when $\mu = 0$, and some subset B of X on which $\lambda \neq 0$ when $\mu \neq 0$. This is a partitioning of the regions of X on which $\lambda \neq 0$ where the partition is determined by the region of X where $\mu \neq 0$. In fact, the set A can be the B-complement of X, $A = X \backslash B = X_B^c$. When viewed in this way, it is obvious that this decomposition of λ is unique for a specified μ. Note also that $\lambda_s \perp \lambda_a$.

The part of the theorem that deals with the Radon-Nikodym derivative is a bit more subtle. If you look at *every* measurable set E in σ-algebra $\mathcal{M}$, then there exists only one function h that accurately describes the relation-

ship between λ_a and μ over the whole σ-algebra. One of the points that must be satisfied is that λ_a and μ must be defined over the *same* σ-algebra. Kolmogorov and Fomin [141] point out that the Radon-Nikodym theorem only establishes the existence of the derivative $h = \frac{d\lambda_a}{d\mu}$, but does not tell how to compute it. They refer to Shilov and Gurevich (chapter 10) [242] for an explicit procedure for evaluating this derivative at a point $x_0 \in X$ by calculating the limit $\lim_{\epsilon \to 0} \frac{\lambda_a(E_\epsilon)}{\mu(E_\epsilon)}$ where $\{E_\epsilon\}$ is a system of sets converging to the point x_0 as $\epsilon \to 0$ in a suitably defined sense. In a very generalized way, this might define a sequence of sets such that for $\epsilon_k < \epsilon_{k-1}$ we have $E_k \subset E_{k-1}$ subject to the condition that $x_0 \in E_k$. In the case of a function f defined on $\mathbf{R}$, there is an explicit procedure for finding the derivative of f at a point x_0 given by

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

There are some handy rules for working with the Radon-Nikodym derivative. They are very similar to the rules for working with common derivatives. The primary difference is the explicit statement of conditions under which the rules work. The following are given by Phillips (p. 429) [207].

Theorem 2 *Manipulation Rules for the Radon-Nikodym Derivative.*

1. If $a, b \in \mathbf{R}^+$, $\nu \ll \mu$, and $\lambda \ll \mu$, then

$$\frac{d(a\nu + b\nu)}{d\mu} = a\frac{d\nu}{d\mu} + b\frac{d\lambda}{d\mu}$$

Note that both measures ν and λ are dominated by the same measure μ. It is not strictly correct to call $\frac{d}{d\mu}$ an operator. The technical point here is that $\frac{d}{d\mu}$ is meaningless. However, if you did consider it to be an operator, this shows that the operator is linear.

2. If $\nu \ll \mu$ and $\mu \ll \lambda$, then $\nu \ll \lambda$. The relation $\ll$ is transitive.

3. Given measures ν, μ, and λ such that $\nu \ll \mu$ and $\mu \ll \lambda$, then there is a chain rule

$$\frac{d\nu}{d\lambda} = \left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\lambda}\right)$$

Note that the measure in the denominator of each term dominates the corresponding term in the numerator.

4. If $\nu \ll \mu$ and $\mu \ll \nu$, then

$$\left(\frac{d\nu}{d\mu}\right)^{-1} = \left(\frac{d\mu}{d\nu}\right)$$

What does it mean for $\nu \ll \mu$ and $\mu \ll \nu$? It means that measures ν and μ have the same region of support in X, or equivalently ν and μ are concentrated on the same set.

A very interesting note is that if ν and μ are both measures defined on the same σ-algebra $\mathcal{M}$, then $\nu + \mu$ is also a measure defined on σ-algebra $\mathcal{M}$. Further, this new measure $\nu + \mu$ dominates both ν and μ. When we specialize our discussion to probability measures, then if ν and μ are probability measures,

then $\alpha\nu + \beta\mu$ is also a probability measure when $\alpha + \beta = 1$. In general, any convex sum $S = \sum_{k=1}^{n} \alpha_k \mu_k$, $\sum_{k=1}^{n} \alpha_k = 1$, of probability measures $\{\mu_k\}_{k=1}^{n}$ defined on the same σ-algebra $\mathcal{M}$ is also a probability measure, and that convex sum dominates each individual probability measure. It also dominates any convex sum formed from a subset of those probability measures.

4.2.3 Kiefer and Wolfowitz Development of Maximum Likelihood Estimators

We are now prepared to consider the work of Kiefer and Wolfowitz [140]. In examining the source literature, the reader will notice that Kiefer and Wolfowitz denote the parameter space by $\Omega \times \Gamma$ where I have only used Γ. They used the more structured space definition to facilitate their proof of consistency. Their level of detail is not required for the development of the following ideas.

Recall that a likelihood function is the conditional joint distribution of a collection of random samples for a given underlying distribution. The usual case of interest is where the underlying distribution is the unknown being sought. When the random samples are assumed independent and identically distributed, the likelihood function is the respective product of the marginal conditional distributions. We consider two classes of maximum likelihood estimators which are distinguished by the existence or non-existence of some dominating measure μ.

Maximum Likelihood Estimators when a Dominating Measure Exists

Let a dominating measure μ exist. This assumption distinguishes the following generalizations from ones that require use of the Lebesgue-Radon-Nikodym derivative.

Maximum Likelihood Estimator (MLE)

For a given random sample of size n, such a likelihood function can be expressed by

$$L(z_1, \cdots, z_n \mid \gamma) = \prod_{i=1}^{n} f(z_n \mid \gamma)$$

where γ is the underlying distribution, which we recall is a measurable function. We are interested in a sequence of μ-measurable functions $\{\hat{\gamma}\}$ such that

$$L(z_1, \cdots, z_n \mid \hat{\gamma}(z_1, \cdots, z_n)) \geq \sup\{L(z_1, \cdots, z_n \mid \gamma), \gamma \in \Gamma\}$$

for almost all $(z_1, \cdots, z_n)$ with respect to measure μ, and for all nonnegative integers $n \in \mathbf{N}$.

Let $z_n = (z_1, \cdots, z_n)$ and consider

$$\sup\{L(z_1, \cdots, z_n \mid \gamma), \gamma \in \Gamma\} \underset{=}{\text{def}} \sup\{L(z_n \mid \gamma), \gamma \in \Gamma\}$$

where L is a mapping from the product space $Z \times \Gamma$ into some space Y. The supremum is taken in Y. The finiteness of L for all (γ, n) implies that the

supremum of L is also finite. Therefore, I can have different sequences of $\gamma \in \Gamma$ that produce convergent sequences of L to its supremum, as shown in figure 4.2.

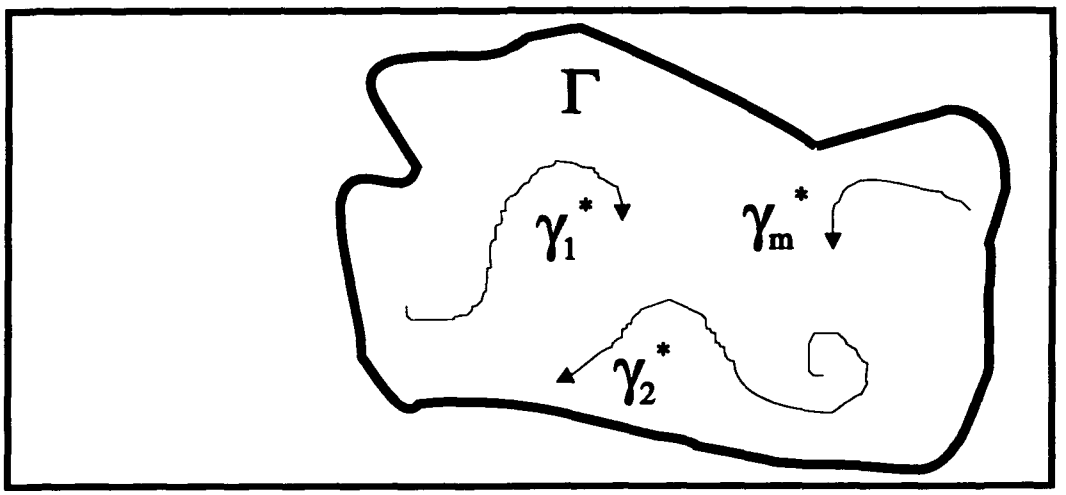


Figure 4.2. Maximum Likelihood Estimate (MLE) Convergent Sequences of L

There is no guarantee that the sequences $\{L_{n,k}\}$ which have a common supremum are produced by a unique sequence $\{\gamma_{n,k}\}$. For some fixed value of n, we can observe the following.

Supremum, L Sequence	Parameter Sequence
$\sup\{L(z_n \mid \gamma_{11}), \cdots, L(z_n \mid \gamma_{1k}), \cdots$	$\{\gamma_{1k}\} \rightarrow \gamma_1^*$
$= \sup\{L(z_n \mid \gamma_{21}), \cdots, L(z_n \mid \gamma_{2k}), \cdots$	$\{\gamma_{2k}\} \rightarrow \gamma_2^*$
$\vdots$	$\vdots$
$= \sup\{L(z_n \mid \gamma_{m21}), \cdots, L(z_n \mid \gamma_{mk}), \cdots$	$\{\gamma_{mk}\} \rightarrow \gamma_m^*$
$\vdots$	$\vdots$

So, we get *a* maximum likelihood estimator, not necessarily a unique maximum likelihood estimator. It is possible that γ_m^* is not contained within the set Γ of allowable distributions or measurable functions. If there is *no* γ_m^* contained in Γ, then we say that the maximum likelihood estimator does not exist.

Modified Maximum Likelihood Estimator (MMLE)

This is an approach to extend the concept of a maximum likelihood estimator to increase the number of cases for which a maximum likelihood estimator exists. As with the maximum likelihood estimator, we seek to find $\sup\{L(z_n \mid \gamma),\ \gamma \in \Gamma\}$. The supremum is taken of L in the set Y.

We are interested in a sequence of μ-measurable functions $\{\hat{\gamma}\}$ such that for some $0 < c < 1$, we have

$$L(z_1, \cdots, z_n \mid \hat{\gamma}(z_1, \cdots, z_n)) \geq c \cdot \sup\{L(z_1, \cdots, z_n \mid \gamma),\ \gamma \in \Gamma\}$$

for almost all $(z_1, \cdots, z_n)$ with respect to measure μ, and for all nonnegative integers $n \in \mathbf{N}$. When $c = 1$, this is the usual maximum likelihood estimator.

Consider looking at a number a little less than $W = \sup\{L(z_n \mid \gamma),\ \gamma \in \Gamma\}$, such as cW where $c \in (0,1)$. Then for c sufficiently small, we hope to find $\hat{\gamma}_n \in \Gamma$ such that $L(z_n \mid \hat{\gamma}_n) \geq cW$. In essence, we are defining a distance between $L(z \mid \gamma_1)$ and $L(z \mid \gamma_2)$. Call it $\rho(L_1, L_2)$. Conceptually, we want to find those L_2 having $\gamma \in \Gamma$ such that $\rho(L_1, L_2) < \epsilon$ where L_1 is the supremum

of $\{L(z_n \mid \gamma), \gamma \in \Gamma\}$ for some $\epsilon > 0$. Figure 4.3 illustrates the concept. It shows a region of Γ such that the maximum likelihood estimator is not in $A \subset \Gamma$, but $\rho(L(z_n \mid \gamma^*), L(z_n \mid \gamma_A)) < \epsilon$.

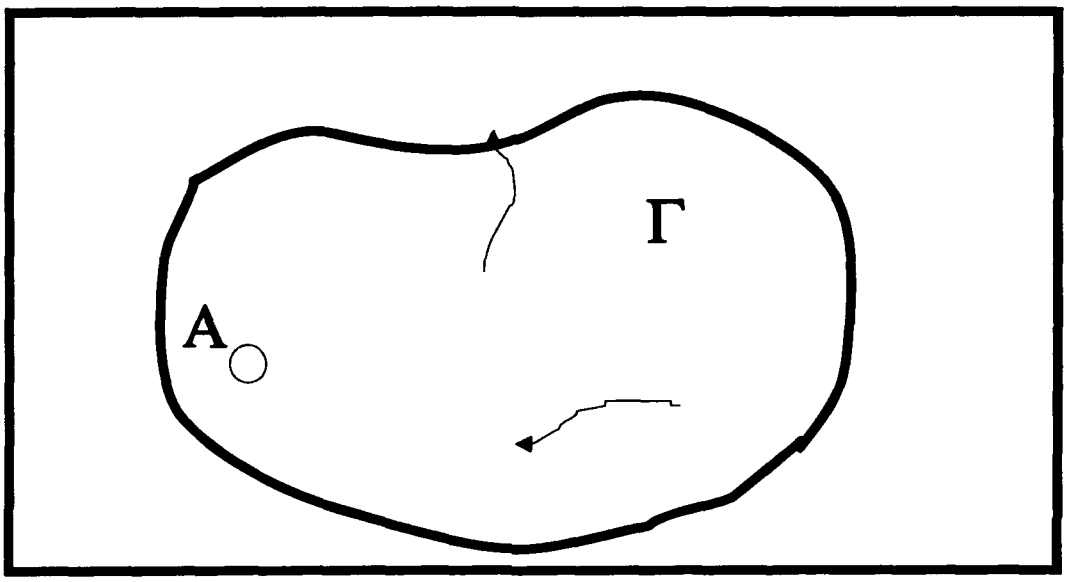


Figure 4.3. Modified Maximum Likelihood Estimator (MMLE) Convergent Sequences of L

The modified maximum likelihood estimators found in this way are not necessarily in the neighborhood of a maximum likelihood estimator when a maximum likelihood estimator exists, but a maximum likelihood estimator will always have a modified maximum likelihood estimate. For parameterized distributions, it is possible that a modified maximum likelihood estimator γ^* could be at a considerable distance (by some suitably chosen distance function) from *any* maximum likelihood estimate $\hat{\gamma}$.

Neighborhood Maximum Likelihood Estimator (NMLE)

A neighborhood maximum likelihood estimator is a sequence of μ-measurable functions $\{\gamma_n^*\}$ satisfying

$$\sup\{L(z_1, \cdots, z_n \mid \gamma),\ \gamma \in \Gamma,\ \delta(\gamma, \gamma_n^*(z_1, \cdots, z_n)) < \epsilon_n\}$$

for almost all $(z_1, \cdots, z_n)$ with respect to measure μ for a sequence of $\{\epsilon_n\}$ where $\epsilon_n > 0$ and $\epsilon_n \to 0$.

Again, let L be a mapping from the product space $Z \times \Gamma$ into Y. As before, a whole set of parameter values can be obtained that produce the same sup L. Call these $\{\hat{\gamma}_i\}_{i=1}^m$ for a fixed n. Then we get the following.

Supremum, L Sequence	Parameter Sequence
$\sup\{L(z_n \mid \gamma_{11}), \cdots, L(z_n \mid \gamma_{1k}), \cdots$	$\lim_{k\to\infty}\{\gamma_{1k}\} \to \hat{\gamma}_1$
$\vdots$	$\vdots$
$\sup\{L(z_n \mid \gamma_{m1}), \cdots, L(z_n \mid \gamma_{mk}), \cdots$	$\lim_{k\to\infty}\{\gamma_{mk}\} \to \hat{\gamma}_m$

The concept is illustrated in figure 4.4.

Pick some $\epsilon_n > 0$, and define a distance function $\delta(\gamma_1, \gamma_2)$. Then γ_m^* is any γ within distance ϵ_n of $\hat{\gamma}_m$. Then there is a family $\{\gamma_{n,m}^*\}$ of neighborhood maximum likelihood estimators, just as there was a family of modified or traditional maximum likelihood estimators in the previous examples. What has been gained is that the neighborhood maximum likelihood estimator exists.

Even though $\hat{\gamma}$ might be outside the space Γ of allowed parameters, γ^* can be chosen in Γ. When you find $\sup\{L(z_n \mid \gamma),\ \gamma \in \Gamma,\ \delta(\gamma, \gamma^*) < \epsilon_n\}$, you

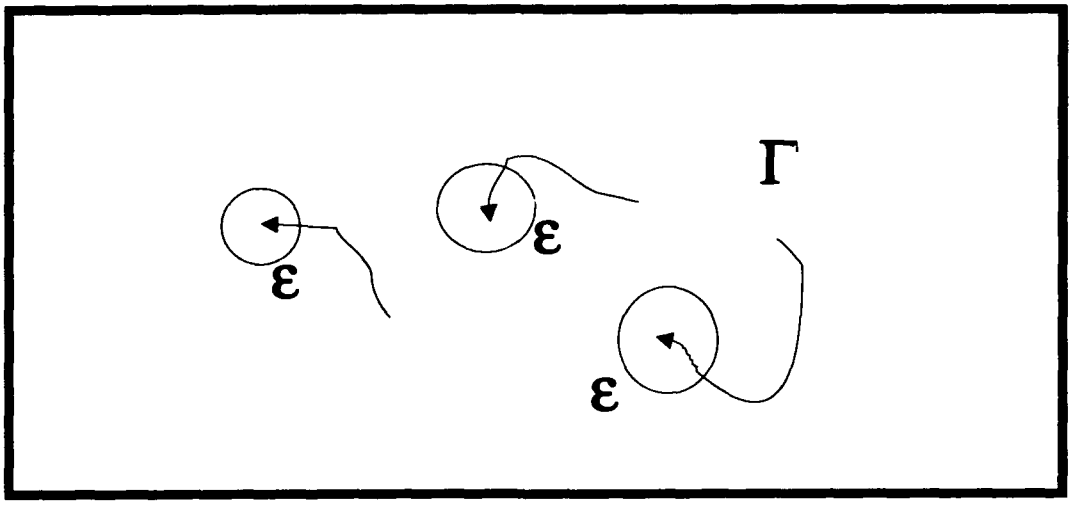


Figure 4.4. Neighborhood Maximum Likelihood Estimator (NMLE) Convergent Sequences of L

ensure you have located a maximum likelihood estimator by constraining this to equal $\sup\{L(z_n \mid \gamma), \gamma \in \Gamma\}$.

Maximum Likelihood Estimators Without Requiring Existence of a Dominating Measure

I seek to generalize the concepts above to ensure the existence of an estimator in its generalized form. For this, we turn to the Radon-Nikodym derivative as a generalization of a density function.

Generalized Maximum Likelihood Estimator (GMLE)

Following Johansen [123], let P_1 and P_2 be members of a non-dominated family of probability measures $\mathcal{P}$. Thus, P_1 and P_2 are measures. Further, there is no measure $\lambda \in \mathcal{P}$ that dominates all the other measures $P_k \in \mathcal{P}$. Recall that if $P_k \ll \lambda$, then everywhere $\lambda = 0$, we also require $P_k = 0$ for every measurable set E belonging to sigma-algebra $\mathcal{M}$. We know, however, that if we sum any two measures, the sum dominates the individual measures. Thus $P_1 \ll P_1 + P_2$ and also $P_2 \ll P_1 + P_2$. Define the Radon-Nikodym derivative

$$r(z_n, P_1, P_2) = \frac{dP_1}{d(P_1 + P_2)}(z_n)$$

The term $\frac{dP_1}{d(P_1+P_2)}(z_n)$ is the Radon-Nikodym derivative of the measure P_1 with respect to the dominating measure $(P_1 + P_2)$ evaluated at the point z_n. Then define $\hat{P}$ as the generalized maximum likelihood estimator if, for arbitrary fixed z_n, the condition $r(z_n, \hat{P}, P) \geq r(z_n, P, \hat{P})$ is satisfied for all $P \in \mathcal{P}$. This says that $\hat{P}$ is the generalized maximum likelihood estimator if

$$\frac{d\hat{P}}{d(P + \hat{P})}(z_n) \geq \frac{dP}{d(P + \hat{P})}(z_n) \tag{4.1}$$

for all $P \in \mathcal{P}$.

So, we are searching over the space of all allowable probability measures for the one that maximizes the Radon-Nikodym derivative, when taken with

respect to the pair-wise sum of the maximizing measure and each other allowable measure. Johansen [123] notes that when $\mathcal{P}$ is dominated by σ-finite measure μ, then equation 4.1 is equivalent to the usual definition of a maximum likelihood estimator.

The following are some useful relevant observations made by Kundu [157]. Suppose that $r(z, \hat{P}, P) \geq r(z, P, \hat{P})$ and $P + \hat{P} \ll \mu$. Now perform a change of variables. Let

$$\hat{P}(E) = \int_E r(z, \hat{P}, P) d(P + \hat{P})$$

and let

$$P(E) = \int_E r(z, P, \hat{P}) d(P + \hat{P})$$

Then

$$\hat{P}(E) = \int_E r(z, \hat{P}, P) \frac{d(P + \hat{P})}{d\mu} d\mu$$

where $P + \hat{P} \ll \mu$. This equals

$$\hat{P}(E) = \int_E g(z, \hat{P}, P) d\mu$$

where

$$g(z, \hat{P}, P) = r(z, \hat{P}, P) \frac{d(P + \hat{P})}{d\mu}$$

and

$$P(E) = \int_E g(z, P, \hat{P}) d\mu$$

Therefore,

$$r(z, \hat{P}, P) \geq r(z, P, \hat{P})$$

implies

$$g(z, \hat{P}, P) \geq g(z, P, \hat{P})$$

and both imply $\hat{P}(E) \geq P(E)$.

C. R. Rao [216] provided the following useful observation. To see the relationship between Johansen's expression for generalized maximum likelihood estimator and that defined by Kiefer and Wolfowitz, first observe that if $r = \frac{dP_1}{d(P_1+P_2)}$, then

$$1 - r = 1 - \frac{dP_1}{d(P_1 + P_2)} = \frac{d(P_1 + P_2) - dP_1}{d(P_1 + P_2)} = \frac{dP_2}{d(P_1 + P_2)}$$

Then, the condition of Johansen that

$$\frac{d\hat{P}}{d(P + \hat{P})}(z) \geq \frac{dP}{d(P + \hat{P})}(z)$$

for all $P \in \mathcal{P}$ becomes

$$\frac{\frac{dP}{d(P+\hat{P})}(z_n)}{\frac{d\hat{P}}{d(P+\hat{P})}(z_r)} \leq 1$$

which implies

$$\frac{\frac{dP}{d(P+\hat{P})}(z_n)}{1 - \frac{dP}{d(P+\hat{P})}(z_n)} \leq 1$$

which implies

$$\frac{r(z, P, \hat{P})}{1 - r(z, P, \hat{P})} = \frac{r}{1 - r} \leq 1$$

for all $P \in \mathcal{P}$.

The benefit of the generalized maximum likelihood estimator is that it can handle the situation where there is no dominating measure. Using the

Radon-Nikodym derivative does not avoid the issues of existence of a maximum likelihood estimator, or convergence, or uniqueness. The same geometry as shown in figure 4.2 applies.

Generalized Modified Maximum Likelihood Estimator

As with the generalized maximum likelihood estimator (GMLE), we extend the definition of the modified maximum likelihood estimator by using the Radon-Nikodym derivative as a generalized density. The discussion about the existence and uniqueness of a modified maximum likelihood estimator also applies to the generalized modified maximum likelihood estimator (GMMLE).

Consider the form

$$\frac{d\hat{P}}{d(P+\hat{P})}(z_n) \geq c\frac{dP}{d(P+\hat{P})}(z_n)$$

for all $P \in \mathcal{P}$. This is equivalent to saying

$$\frac{c\frac{dP}{d(P+\hat{P})}(z_n)}{\frac{d\hat{P}}{d(P+\hat{P})}(z_n)} = \frac{cr(z_n, P, \hat{P})}{1 - r(z_n, P, \hat{P})} = cd(z_n, P, \hat{P}) \leq 1$$

for all $P \in \mathcal{P}$. In particular,

$$c\sup\{d(z_n, P, \hat{P}), P \in \mathcal{P}\} \leq 1$$

When $c = 1$, this is the generalized maximum likelihood estimator.

Generalized Neighborhood Maximum Likelihood Estimator

The concept of the Generalized Neighborhood Maximum Likelihood Estimator (GNMLE) is to find "maximum likelihood estimators", and choose an

estimator whose distance is less than some ϵ within the allowable parameter space or family of distributions.

In using the Radon-Nikodym derivative as a generalized density, the procedure becomes:

1. Define a distance function $\delta(P_1, P_2)$.

2. Choose $\epsilon > 0$.

3. Find the set $\{P_m^*\}$ of functions which possibly are not within the allowable space $\mathcal{P}$ which satisfy

$$sup\{d(z_n, P, P^*), P \in \mathcal{P}\} \leq 1$$

4. Pick a $\hat{P}_m$ corresponding to each P_m^* within $\mathcal{P}$ such that $\delta(\hat{P}_m, P_m^*) < \epsilon$.

If P^* is the function found by

$$\sup\{d(z_n, P, P^*), P \in \mathcal{P}\} \leq 1$$

then the generalized neighborhood maximum likelihood estimator $\hat{P}$ satisfies

$$\sup\{r(z_n, P, P^*),\ P \in \mathcal{P}, \delta(P, \hat{P}) < \epsilon\} = \sup\{r(z_n, P, P^*),\ P \in \mathcal{P}\}$$

Again, the discussion about the existence and uniqueness regarding the neighborhood maximum likelihood also applied to the generalized neighborhood maximum likelihood estimator.

4.2.4 Uniqueness of the Maximum Likelihood Estimator

A question was raised about the uniqueness of the maximum likelihood estimate. The suggestion that for general sets of distributions that a maximum likelihood estimator might not be unique is pictorially presented in figure 4.1. C. R. Rao suggested that when the random sample constitutes the whole sample space that the maximum likelihood estimator would be unique. In general, the method of maximum likelihood does not produce a unique estimator. However, when the full sample space is included in the formulation of the likelihood function, then the maximum *likelihood estimator is unique* almost everywhere.

Counterexample to Uniqueness

Hogg and Craig (p. 207, problem 6.3) [109] provides a counterexample. Let $x_1, x_2, \cdots, x_n$ be a random sample of a distribution with density function $f(x;\theta) = 1$ where $\theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}$, for $-\infty < \theta < \infty$, and $f(x;\theta) = 0$ elsewhere. Let $\{x_1, x_2, \cdots, x_n\}$ be a proper subset of the full sample space. Then let $y_1 < y_2 < \cdots < y_n$ be the order statistic from this random sample. Then every statistic $u(x_1, x_2, \cdots, x_n)$ such that

$$y_n - \frac{1}{2} \leq u(x_1, x_2, \cdots, x_n) \leq y_1 + \frac{1}{2}$$

is a maximum likelihood estimator of θ. In particular,

$$(4y_1 + 2y_n + 1)/6$$

$$(y_1 + y_n)/2$$

and

$$(2y_1 + 4y_n - 1)/6$$

are three such statistics. Thus, uniqueness is not in general a property of a maximum likelihood estimator.

When the Random Sample is the Full Space

Recall from the Lebesgue-Radon-Nikodým theorem that when μ is a positive σ-finite measure on a σ-algebra $\mathcal{M}$ in a set X, and λ is a complex measure on $\mathcal{M}$, then there is a unique $a.e.[\mu]$ function $h \in L^1(\mu)$ such that $\lambda_a(E) = \int_E h\, d\mu$ for every set $E \in \mathcal{M}$, where $\lambda = \lambda_a + \lambda_s$, $\lambda_a \ll \mu$, and $\lambda_s \perp \mu$. This means that if two functions h_1 and h_2 satisfy this, then they differ only on a set of μ-measure zero, i.e. $\mu\{x : h_1 \neq h_2\} = 0$.

When the set E is the whole sample space, then $\lambda_a(E) = \lambda_a(X) = 1$ when $(X, \mathcal{M})$ is a probability space. Thus $\int_X h\, d\mu = 1$. When μ is taken to be Lebesgue measure of the appropriate dimension and X is Euclidean, then h is our probability density function in the usual sense and the measure is often denoted by m.

If $h(x,\theta)$ is a parameterized family of density functions, consider the collection of all θ_α such that $\int_X h(x,\theta_\alpha)dm = 1$. Then $h(x,\theta_{\alpha_1}) = h(x,\theta_{\alpha_2})$ $a.e.[m]$. In the general case, this does not require $\theta_{\alpha_1} = \theta_{\alpha_2}$. To assert uniqueness, more must be known about the family of density functions under consideration. For example, we know from Bickel and Doksum (p. 106, theorem 3.3.2) [40] that the exponential family given by

$$f(x;\theta) = \exp\{\sum_{i=1}^{k} c_i(\theta)T_i(x) + d(\theta) + S(x)\}$$

where $x \in A, \theta \in \Theta$, with C denoting the interior range of $(c_1(\theta),\cdots,c_k(\theta))$ has a unique maximum likelihood estimator of θ if $\mathcal{E}\{T_i(x)\} = T_i(x)$ for $i = 1,\cdots,k$, has a solution $\hat{\theta}(x) = (\hat{\theta}_1(x),\cdots,\hat{\theta}_k(x))$ for which

$$(c_1(\hat{\theta}(x)),\cdots,c_k(\hat{\theta}(x)) \in C$$

Thus, if we sufficiently restrict the allowable set of functions, we can achieve uniqueness, but uniqueness is not automatically a property.

Not everything that needs to be recorded has been recorded here. In particular, some thought is needed with respect to singular distributions and what it means in terms of allowable sets in $\mathcal{M}$ as well as the implications for choosing μ. This question is relevant to this thesis topic, but has not been pursued.

4.3 Specific Techniques

Most workers dealing with order estimation assume an information-theoretic approach. Techniques based on this approach have the advantage that we know how to do the computations today to get answers. Some very nice analytical surveys of techniques have appeared from time to time, although they are being developed almost as fast as they can be printed. Because these are attractive alternatives to the work in my thesis, they are cataloged here for the reader's use. Some of these had their birth in the study of univariate real time series. There are also techniques listed here that use approaches other than information-theoretic. Many of the below techniques have been discussed in the context of a line array with equally spaced elements, using the spatial analog to sampling a stochastic sequence indexed by time.

Pukkila and Krishnaiah [211] report that most of the proposed information-theoretic order determination criteria for ARMA(p, q) models can be expressed in the form of equation 4.2. The word *ARIMA* should not be a distractor. That was the motivating context of the discussion by Pukkila and Krishnaiah. If you prefer, let $q = 0$ to apply this to an autoregressive problem which has a spatial analog with the equally spaced line array. Even more basic than that, the criteria of the form discussed in this paper are derived from a basic information-theoretic approach. The number $(p + q)$ is merely the total number of parameters in the model. It arises as the degrees of freedom of

a χ^2 distributed random variable justified by a large sample approximation used to satisfy application of the Central Limit Theorem in statistics. Akaike [12] discusses application of this statistic to factor analysis, principal component analysis, analysis of variance, and multiple regression, in addition to autoregression of time series which electrical engineers are familiar with.

$$\delta(p,q) = n \log \hat{\sigma}^2 + (p+q)g(n) \tag{4.2}$$

You recognize that $\hat{\sigma}^2$ is the maximum likelihood estimate or its approximation for the residual variance σ^2. The term $(p+q)g(n)$ is a nonnegative penalty term which increases as the number of parameters increases. It is noted that the term $n \log \hat{\sigma}^2$ tends to decrease as the number of parameters increases. The function $g(n)$ produces other criteria which you may recognize. When $g(n) = 2$, we get the $AIC(p,q)$ criterion. The $BIC(p,q)$ criterion is obtained by selecting $g(n) = \log n$. The HQ criterion is obtained by $g(n) = c \log \log n$ where c is a specified constant. EDC is obtained by $q = 0$ and $g(n) = \gamma(n)$ where $\gamma(n)$ is a sequence of positive numbers such that

$$\lim_{n\to\infty} \gamma(n)/n = 0, \qquad \lim_{n\to\infty} \gamma(n) = \infty$$

A variation on this is obtained by selecting $\gamma(n)$ such that

$$\lim_{n\to\infty} \gamma(n)/n = 0, \qquad \lim_{n\to\infty} \gamma(n)/(\log \log n) = \infty$$

They call attention to a survey of different univariate order determination by de Grooijer, Abraham, Gould, and Robinson [69].

4.3.1 Akaike Information Criterion (AIC)

The definition given by Akaike in his December 1974 paper [16] is given in equation 4.3.

$$AIC = -2\max\{\log(jpdf)\} + 2(df) \tag{4.3}$$

The term *jpdf* is the joint probability density function, where you choose the model yielding the minimum AIC. It is derived using the Kullback-Leibler mean information measure [155] (pp. 26-27). The requirements and assumptions are: (1) the distribution must be a regular member of the exponential class in the sense of Hogg and Craig (p.357-358)[109], (2) large sample case (see p. 718, left, bottom[16]), (3) the third and higher order terms of a Taylor series expansion are dropped (see p. 718, right, middle[16]), (4) $\int f'(x,\theta_0)dx = 0$ and $f''(x,\theta_0)dx = 0$ (see Kullback, p.27, item3[155]), and (5) AIC is computed for each model considered. Parzen [203] defines AIC as the value of m minimizing

$$AIC(m) = \log\hat{\sigma}_m^2 + 2\frac{m}{T}$$

where $\hat{\sigma}_m^2$ is the estimator of the mean-square prediction error σ_m^2, and T is the total number of samples. Cremona and Brandon [63] give the following expressions for AIC :

$$AIC(M) = N\ \ln V_N + 2p$$

or

$$W_N = N\ \ln V_N + \varphi(N,p)$$

where $\varphi(N,p) = 2p$. The term V_N is the minimum of some loss function $V_N(\theta,\Lambda)$ (quadratic error criterion or likelihood criterion) and M is the model order and N is the number of samples. The quantity p is the size of the observation vector.

Note that we get AIC by selecting $g(n) = 2$ in equation 4.2. Pukkila and Krishnaiah [211] report that Shibata [241] proved that AIC is not a statistically consistent estimator for the order of a univariate autoregressive model. Instead, the AIC criterion tends to overestimate the order of an AR(p) model. Similarly, AIC does not produce a consistent estimate for the order of an ARMA(p,q) model [101]. Although Rissanen [226] regards consistency an generally necessary property for any criterion, he notes that consistency does not in itself guarantee good estimation results for small samples.

Kashyap [129] was one of the first to bring serious challenge to AIC. He showed that AIC was not statistically consistent. For the AIC rule, the probability of error is not less than 0.156 even when n tends to infinity. Kashyap recommends that attention be restricted to the class of consistent decision rules. He proposes a consistent decision function.

In 1983, Wax and Kailath [278] proposed an alternative for the number of free adjusted parameters within a model to be $k(2p-k)+1$ where k is the test order and p is the size of the observation vector which is sampled N times. These vectors are assumed to be independent and identically distributed

according to the real multivariate normal distribution $N_p(0, R)$. With this adjustment, AIC is modified as shown in equation 4.4. They observed that AIC yields an inconsistent estimate that tends, in the large-sample limit, to overestimate the true rank.

$$AIC(k) = -2\log\left[\frac{\prod_{i=k+1}^{p} l_i^2}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p} l_i^2\right)^{p-k}}\right]^N + 2k(2p-k) \qquad (4.4)$$

A few words about consistency are in order at this point because of the wide-spread criticism of Akaike's work. "Consistency" in statistics is a technical term. For an estimator that depends on the sample size n, then it is called consistent if its expected value is unbiased when n tends to infinity. Cochran (pp. 21-22)[55] has the following to say about estimators and consistency.

> The precision of any estimate made from a sample depends both on the method by which the estimate is calculated from the sample data and on the plan of sampling. ... When studying any formula that is presented, the reader should make sure that he or she knows the specific method of estimation for which the formula has been established. ··· [In the context of sampling theory,] a method of estimation is called *consistent* if the estimate becomes exactly equal to the population value when $n = N$, that is, when the sample consists of the whole population. ··· Consistency is a desirable property of estimators. On the other hand, *an inconsis-*

> *tent estimator is not necessarily useless, since it may give satisfactory precision when n is small compared to N.* $\cdots$ In classical statistics, an estimator is called consistent if the probability that it is in error by more than any given amount tends to zero as the sample becomes large.

Bickel and Doksum (pp. 134, 141, 225)[40] concur with this remark, and have the following to say about various kinds of estimators. The notions of consistency, asymptotic mean, variance, and unbiasedness are the properties of the *sequence* of the estimates $\{T_n(x_1, \cdots, x_n)\}$ for $n \geq 1$, not of any single T_n. These are properties of the *method* of maximum likelihood, not of the maximum likelihood estimate for a particular sample size. ... Small sample studies comparing the behavior of uniformly minimum variance unbiased estimators (UMVU) and MLEs are inconclusive. Simple examples in which there are many nuisance parameters are known for which MLEs behave very badly even for large samples. Neither MLEs nor UMVU estimates are satisfactory in general if one takes a Bayesian or minimax point of view. Nor are they necessarily robust. ... Likelihood ratio tests are based on heuristic grounds.

On this basis, there is insufficient evidence to discredit Akaike's work. We still have work to do for the small sample case.

4.3.2 Bayesian Information Criterion (BIC)

Pukkila and Krishnaiah [211] credit Schwarz [239] and Rissanen [224] for independently developing BIC starting from different points. BIC is defined in equation 4.5. This equation is obtained by letting $g(n) = \log n$ in equation 4.2. BIC produces a consistent estimate $(\hat{p}, \hat{q})$ for the order of an ARMA model.

$$BIC(p,q) = n \log \hat{\sigma}^2 + (p+q) \log n \tag{4.5}$$

4.3.3 Kashyap Information Criterion (KIC)

This is a variant of AIC. This discussion is based on [129]. Let the estimate of the unknown order m_0 based on Y_N be given by

$$m^* = \arg \left[\min_{m \in (1,r)} d_m(Y_N) \right]$$

where

$$d_m(Y_N) = N \ln \rho_m^* + m f(N)$$

The quantity ρ_m^* is the residual variance for the fitted autoregressive model having m lag terms. This term can be recursively computed from Y_N. See references [71][169]. Deterministic function $f(N)$ satisfies $f(N) > 0$, $f(N) \rightarrow \infty$, and $f(n)/N \rightarrow 0$.

4.3.4 Hannan-Quinn (HQ)

Pukkila and Krishnaiah [211] cite Hannan and Quinn [101] as the source for the HQ criterion given in equation 4.6 where c is a constant to be specified. This equation is obtained by letting

$$g(n) = c \log \log n$$

in equation 4.2. Select a constant $c > 2$ to guarantee a strongly consistent order estimate.

$$HQ(p,q) = n \log \hat{\sigma}^2 + (p+q)c \log \log n \tag{4.6}$$

4.3.5 Efficient Detection Criterion (EDC)

Zhao, Krishnaiah, and Bai [297] proposed the procedure for the white noise case now known as the Efficient Detection Criterion (EDC). Efficiency is a technical term in statistics. An estimator is called *efficient* if the Cramér-Rao lower bound is achieved. Zhao, Krishnaiah, and Bai [298] extended that work to the colored noise case for the signals and noise having independent real Wishart covariance matrices. They considered the asymptotic case. Bai, Krishnaiah, and Zhao [36] define EDC as follows. Let $x(t) = As(t) + n(t)$ where the column signal vector $s(t)$ and the column noise vector $n(t)$ are complex random vectors distributed independently with mean 0. Let the matrix $X = [x(t_1), \cdots, x(t_n)]$ be the sample of size n of the process $x(t)$. The covari-

ance of $s(t)$ is given by Ψ, and the covariance of $n(t)$ is given by $\sigma^2 I_p$ where I_p is a $p \times p$ identity matrix. The matrix $A = [A(\phi_1), \cdots, A(\phi_q)]$ is a complex vector of unknown parameters associated with the i^{th} signal. The number of unknown parameters for each signal is assumed known. Let the eigenvalues of Σ be $\lambda_1^2 \geq \cdots \geq \lambda_p^2$. Let S_n be the maximum likelihood estimator of Σ where $nS_n = XX^H$, and let the sample eigenvalues of S_n be given by $l_1^2, \cdots, l_p^2$. Let H_q be the hypothesis that the number of signals is equal to q. Thus

$$H_q : \lambda_1^2 \geq \cdots \lambda_q^2 > \lambda_{q+1}^2 = \cdots = \lambda_p^2 = \sigma^2$$

When σ^2 is unknown and $\{x(t_i)\}_1^n$ are independently distributed as complex normal, the logarithm of the likelihood ratio test statistic for H_q is given by

$$L(q) = n\left\{\left(\sum_{i-q+1}^{p} \log l_i^2\right) - (p-q)\log\left(\frac{1}{p-q}\sum_{i=q+1}^{p} l_i^2\right)\right\}$$

Then EDC is given by equation 4.7.

$$EDC(k, C(n)) = -2L(k) + \nu(k,p)C(n) \tag{4.7}$$

In equation 4.7

$$\nu(k,p) = k(2p-k+1)+1$$

is the number of free parameters when H_k is true. Then the estimate $\hat{q}$ of q is the value of $\hat{q}$ that satisfies equation 4.8

$$EDC(\hat{q}, C(n)) = \min\{EDC(0, C(n)), \cdots, EDC(p-1, C(n))\} \tag{4.8}$$

The quantity $C(n)$ is chosen so that it satisfies the following conditions: 1) $\lim[C(n)/n] = 0$ and 2) $\lim_{n\to\infty}[C(n)/\log\log n] = \infty$. When σ^2 is known then it can be assumed to be unity without loss of generality. Then EDC^* is given by equation 4.9.

$$EDC^*(\hat{q}, C(n)) = \min\{EDC^*(0, C(n)), \cdots, EDC^*(p-1, C(n))\} \qquad (4.9)$$

In equation 4.9, the individual entries over which the minimum is taken is given by equation 4.10.

$$EDC^*(k, C(n)) = -2L^*(k) + \nu^*(k, p)C(n), \qquad (4.10)$$

In computing the term L^*, τ is the number of sample eigenvalues l_i greater than unity where

$$L^*(k) = n \sum_{i=1+\min(\tau,k)}^{p} (\log l_i^2 + 1 - l_i^2)$$

The 1989 paper [36] gives bounds under certain conditions on the probability of a wrong decision. In this paper, Bai et al. point out that the estimator is a statistically consistent estimator, the rate of convergence of the estimate of the number of signals to the true value is rapid, and no threshold value is required to form the estimator. This paper is a good entry point into the literature on information theoretic approaches.

4.3.6 White Noise Tests (T_1, $TAIC$, $TBIC$, THQ)

Pukkila and Krishnaiah [211] consider the order determination problem for real-valued autoregressive (AR) models and using concepts motivated by Box and Jenkins [42]. Testing the adequacy of a fitted model is based on the estimated autocorrelation structure of the residual series from the estimated model. Starting from a simple, parsimoniously parameterized model, a model builder adds new parameters until the residual series is close enough to a white noise. Pukkila and Krishnaiah accomplish this by creating a family of test statistics built from the forms of AIC, BIC, and HQ.

For the autoregressive (AR) model, equation 4.2 is minimized for $p = 0, 1, \cdots, p^*$ where p^* is the largest model order we are willing to consider, and the quantity $q = 0$ is used to restrict the case to the AR model. They use the Hannan and Quinn estimator for the AR model residual variance given by

$$\hat{\sigma}^2 = c(0)\left(1 - \sum_{k=1}^{p} \hat{\phi}_k r(k)\right)$$

where $\hat{\phi}_1, \cdots, \hat{\phi}_p$ are the Yule-Walker estimates of the autoregressive coefficients $\{\phi_k\}_1^p$ and $\{r(k)\}_1^p$ are the autocorrelations. The autocorrelations are computed by $r(k) = c(k)/c(0)$ where

$$c(k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

for $1 \leq k \leq p$. The $\{\hat{\phi}_k\}_1^p$ are obtained by solving the Yule-Walker equation

4.11.

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(k-1) \\ r(1) & r(0) & & r(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(k-1) & r(k-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_k \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(k) \end{bmatrix} \tag{4.11}$$

The test statistic is then given by equation 4.12.

$$T_1(p^*) = \min_{1 \le p \le p^*} \left\{ 0, n \log \left(1 - \sum_{k=1}^{p} \hat{\phi}_k r(k) \right) + pg(n) \right\} \tag{4.12}$$

If $T_1(p^*) < 0$ then reject $H_0 : \{x_t$ is generated by an AR(0) process$\}$ in favor of $H_1 : \{x_t$ is generated by an AR(k) process where $k > 0\}$. To use a traditional order estimation criterion for $\delta(p)$ substitute the corresponding expression for $g(n)$. Thus, to obtain $TAIC(p^*)$ corresponding to AIC, select $g(n) = 2$. Similarly, choose $g(n) = \log n$ for $TBIC(p^*)$ and choose

$$g(n) = c \log \log n$$

to get $THQ(p^*)$. Pukkila and Krishnaiah also provide the asymptotic values of the significance levels $\alpha(n)$ and lower bounds for the power functions for these proposed tests.

4.3.7 Minimum Description Length (MDL)

The comments regarding MDL are based on [224][225][226][81].

$$(\hat{n}, \hat{\theta}) = \arg \min_{n,\theta} \left[-\log p(x;\theta) + \frac{1}{2} n \log N \right] = \arg \min_{n,\theta} I_\theta(x)$$

The number of parameters in parameter vector θ is n. N is the length of the observation sequence. $I_\theta(x)$ is the information of the sequence x with respect to the given probability distribution family. Feder's article is very readable.

The reader is encouraged to consult reference [225] which generalizes MDL so that it is invariant with respect to all linear coordinate transformations. Rissanen does not compromise on technical quality, his writing is very clear, and he accompanies his developments with insightful remarks.

Define

$$\log^*(y) = \log y + \log\log y + \ldots$$

and

$$\|\theta\|_{M(\theta)} = \sqrt{< \theta,\ M(\theta)\theta >}$$

where $< \bullet, \bullet >$ is the inner product of the k-component parameter vector θ and the product of the information matrix $M(\theta) = n \times I_\theta$ times θ. The value of k is the model order. The information matrix I_θ is defined by

$$I_\theta = \frac{1}{N} E\left\{ \left(\frac{\partial \log P(y \mid \theta)}{\partial \theta} \right) \left(\frac{\partial \log P(y \mid \theta)}{\partial \theta} \right)^T \right\}$$

Then the MDL criterion is given by equation 4.13.

$$-\log P(y, \theta) = -\log P(y \mid \theta) + \log^* \left(\|\theta\|_{M(\theta)} \right) \tag{4.13}$$

Rissanen calls $-\log P(y, \theta)$ the joint ideal code length which is to be minimized as a function of model order k.

As discussed earlier, in 1983 Wax and Kailath [278] proposed an alternative for the number of free adjusted parameters within a model to be $k(2p-k)+1$ where k is the test order and p is the size of the observation vector which is sampled N times. These vectors are assumed to be independent and identically distributed according to the real multivariate normal distribution $N_p(0,R)$. With this adjustment, MDL is modified as shown in equation 4.14. They observed that MDL is a consistent estimator of true system rank.

$$MDL(k) = -\log\left[\frac{\prod_{i=k+1}^{p} l_i}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p} l_i\right)^{p-k}}\right]^N + \frac{1}{2}k(2p-k)\log N \qquad (4.14)$$

In 1985, Wax and Kailath [280] apply MDL to the problem of estimating the number of signals in a multi-channel time series. In this paper, they generalize earlier proofs that MDL is a consistent estimator.

4.3.8 Wang's Sphericity Test

Wang's Sphericity Test is my name for the test Wang and Kaveh proposed in their 1986 paper (equations 6 through 8)[275]. Let $R = E\{S_t S_s^H\}$ be the covariance matrix of signals. Let $W = ARA^H + \sigma_n^2 I$ be estimated by $\hat{W}$ with sample eigenvalues $\{l_1^2, \cdots, l_M^2\}$ where $l_i^2 > l_{i+1}^2$. The estimate $\hat{d}$ of the number of sources d minimizes the quantity in equation 4.15.

$$\Lambda(d,p,k) \stackrel{\text{def}}{=} k(M-d)\log\left(\frac{\frac{1}{M-d}\sum_{i=d+1}^{M} l_i^2}{\left(\prod_{i=d+1}^{M} l_i^2\right)^{1/(M-d)}}\right) + p(d,k) \qquad (4.15)$$

The quantity $p(d,k)$ is a chosen penalty function for the overdetermination of d. If you chose

$$p(d,k) = d(2M - d)$$

you have AIC. If you chose

$$p(d,k) = \frac{1}{2}d(2M - d)\log k$$

you have MDL. Using these, Wang and Kaveh examined the probabilities of underestimating and overestimating the number of sources for the cases of up to two closely spaced sources in spatially white noise. Wang and Kaveh applied their findings to narrowband systems. In a 1987 paper [276], they continued their study and applied it to wide band systems. In both these papers they studied the asymptotic case.

4.3.9 Finite Markov Chain Maximum Entropy Order Estimator (FMCME)

This reviews work reported primarily in [175]. Consider a discrete-time k^{th}-order Markov process where each random variable x_i takes on values in a finite set A. A k^{th}-order Markov process is one where the probability of the occurrence of x_i depends on the preceding k samples $\{x_{i-1}, x_{i-2}, \cdots x_{i-k}\}$, but not on the preceding $k+1$ samples. The goal is to estimate k as accurately as possible. To measure accuracy, the following performance criterion is used.

Among all estimators $\hat{k}$ for which the overestimation probability $P_k(\hat{k} > k)$ decays faster than $2^{-\lambda n}$ (for a specified $\lambda > 0$) uniformly for any Markovian probability measure P_k of order k, find an estimator that minimizes the underestimation probability $P_k(\hat{k} < k)$ uniformly for every P_k.

Let $x = (x_1, \cdots, x_n) \in A^n$ be an observed sequence from the unknown k^{th}-order Markov process. Let s_i at time i specify the state of the Markov source that governs when sample x_i is drawn. Thus

$$s_i = (x_{i-1}, \cdots, x_{i-k}) \in A^k$$

Let u be an arbitrary member of A, and let s be an arbitrary member of A^k. Define the delta function

$$\delta(x_i, u, s_i, s) = \delta(x_i, u)\delta(s_i, s)$$

where δ is one when the arguments are equal, and δ is zero otherwise. Let k_0 be a finite integer which is an upper bound for the true order k. There are two versions of FMCME. Version k^* applies when the value of k_0 is known. Version k^{**} applies when k_0 is unknown.

Version k^* is the estimated order you seek.

$$k^* = \min\{j : H(q_x^j) - H(q_x^{k_0}) \leq \lambda\}$$

where

$$H(q_x^k) \stackrel{\text{def}}{=} -\sum_{s \in A^k} q_x^k(s) \sum_{u \in A} q_x^k(u \mid s) \log q_x^k(u \mid s)$$

$$q_x^k(u,s) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\delta(x_i,u,s_i,s)$$

$$q_x^k(s) \stackrel{\text{def}}{=} \sum_{u\in A} q_x^k(u,s)$$

$$q_x^k(u \mid s) = \begin{cases} q_x^k(u,s)/q_x^k(s), & q_x^k(s) > 0 \\ 0, & q_x^k(s) = 0 \end{cases}$$

Using Rissanen's MDL, the estimator k^* is asymptotically equivalent to

$$\min\left\{j; \frac{1}{n}MDL(j) - \frac{1}{n}MDL(k_0) < \lambda\right\}$$

Version k^{**} applies when k_0 is unknown. It is based on the LZ data compression algorithm described in reference [300]. The LZ code word length of x is U_{LZ} which is computed by the algorithm. The unknown term $H(q_x^{k_0})$ in the expression for k^* is approximated by the normalized LZ code word length function.

$$k^{**} \stackrel{\text{def}}{=} \min\left\{j : H(q_x^j) - \frac{1}{n}U_{LZ}(x) \le \lambda\right\}$$

By applying the theory of large deviations, the estimator k^* has been extended in reference [176] to exponential families. This is applicable to the Gaussian linear regression model and the autoregressive (AR) model.

4.3.10 Coherent MDL

This approach, reported in [281], yielded two test statistics. The first is most suitable for the detection-only problem. The second is for the joint detection and estimation problem. In this approach the signal is considered as unknown constants without an assumed stochastic model. The motivation for this approach is that previous approaches were not applicable to the case of a fully correlated signal, such as occurs with a specular multipath situation. Both approaches were proven to produce statistically consistent estimators. For the detection problem, the MDL estimator for the number of sources is given in equation 4.16.

$$\hat{k}_{MDLB} = \arg \min_{k \in \{0,\cdots,p-1\}} MDLB(k) \tag{4.16}$$

where

$$MDLB(k) = M(p-k) \log \left(\frac{\frac{1}{p-k} \sum_{i=1}^{p-k} l_i^2(\hat{\theta}^{(k)})}{\left[\prod_{i-1}^{p-k} l_i^2(\hat{\theta}^{(k)}) \right]^{1/(p-k)}} \right) + \frac{1}{2} k(2p-k+1) \log M \tag{4.17}$$

with $\hat{\theta}^{(k)}$ given by

$$\hat{\theta}^{(k)} = \arg \min_{\theta^{(k)}} \left\{ \log \left[\frac{\frac{1}{(p-k)} \sum_{i=1}^{p-k} l_i^2(\theta^{(k)})}{\left(\prod_{i=1}^{p-k} l_i(\theta^{(k)}) \right)^{1/(p-k)}} \right] \right\} \tag{4.18}$$

The combined detection-estimation estimator of the number of sources is given in equation 4.19.

$$\hat{k}_{MDLC} = \arg \min_{k \in \{0,\cdots,p-1\}} MDLC(k) \tag{4.19}$$

where

$$MDLC(k) = M(p-k)\log\left[\frac{\frac{1}{(p-k)}\sum_{i=1}^{p-k} l_i^2(\hat{\theta}^{(k)})}{\left(\prod_{i=1}^{p-k} l_i(\hat{\theta}^{(k)})\right)^{1/(p-k)}}\right] + \frac{1}{2}k(2p-k+1)\log M \quad (4.20)$$

with $\hat{\theta}^{(k)}$ given by

$$\hat{\theta}^{(k)} = \arg\min_{\theta^{(k)}} \sum_{i=1}^{p-k} l_i^2(\theta^{(k)}) \quad (4.21)$$

4.3.11 Maximum Likelihood (ML)

This is information taken from [81]. Assume that the desired probability distribution $p(\bullet)$ belongs to a parameterized distribution family P_Θ indexed by parameter vector $\theta \in \Theta$. Then the maximum likelihood criterion will choose $\hat{p}$ from P_Θ by

$$\hat{p} = \arg\max_{p \in P_\Theta} \log p(x) \text{ or } \hat{\theta} = \arg\max_{\theta \in \Theta} \log p(x;\theta)$$

4.3.12 Maximum Entropy (ME)

Let the desired probability distribution $p(\bullet)$ belong to a set of distributions P where

$$P = \{p(x) \mid E_p[g(x)] = \bar{g}\}$$

such that $\bar{g}$ is known. The given averages are the only information available. Then the chosen distribution is $\hat{p}$ where

$$\hat{p} = \arg\max_{p \in P} H(p) = \arg\max_{p \in P}\left[-\int_x p(x)\log p(x)dx\right]$$

The reader is strongly recommended to read [81] for the remarkably clear presentation.

Miller and Snyder [182] remark that the probability density maximizing entropy is identical to the conditional density of the complete data given the incomplete data. This equivalence comes from viewing the measurements as specifying the domain over which the density is defined. The identity between the maximum entropy and the conditional density comes from the fact that the maximum-likelihood estimates may be obtained via a joint maximization (minimization) of the entropy function (Kullback-Liebler divergence).

4.3.13 Criterion Autoregressive Transfer Function

The Criterion Autoregressive Transfer (CAT) function approach is reported in [203].

$$CAT(m) = 1 - \frac{\hat{\sigma}_{\infty}^2}{\ddot{\sigma}_m^2} + \frac{m}{T}$$

where $\hat{\sigma}_{\infty}^2$ is the estimator for the mean-square prediction error σ_{∞}^2 of an infinite order autoregressive model AR(∞). The quantity $\ddot{\sigma}_m^2$ in the denominator is defined as $\ddot{\sigma}_m^2 = \frac{T}{T-m}\hat{\sigma}_m^2$, which is the unbiased estimator for σ_m^2. The value $\hat{m}$ minimizing $CAT(m)$ is chosen not as the order of an autoregressive model chosen to fit the observed time series, but as the order of an autoregressive estimator of the infinite order autoregressive transfer function (ARTF) $g_{\infty}(\bullet)$.

4.3.14 Final Prediction Error Criterion (FPE)

This is taken from [63].

$$FPE(M) = V_N \frac{1 + \frac{p}{N}}{1 - \frac{p}{N}}$$

Rissanen [226] credits Davisson [66] with the first statement of "final prediction error." Söderström [250] credits Akaike with proposing FPE in 1969 in reference [3].

4.3.15 Weak Parameter Criterion (WPC)

Broersen's 1985 paper [46] suggests that weak parameters should be removed if the squares of their estimates are less than twice the expectation for a white noise signal. The measure 2 for significance is derived from asymptotic conditions. WPC is based on the same principles as Mallows' C_p, FPE, and AIC. Choose the value of M as model order which minimizes $WPC(M)$.

$$WPC(M) = S_M^2 / \left(\prod_{j=0}^{M} (1 - 2v_j) \right)$$

In this expression, $v_0 = 0$. When Yule-Walker estimates are used for model reflection coefficients then

$$v_j = (N - j)/[N(N + 2)]$$

When Burg estimates are used for model reflection coefficients then $v_j = 1/(N - j + 1)$. The quantity S_M^2 is the residual reduction by adding reflection

coefficients, and it can be described by $S_M^2 = \prod_{i=1}^{M}(1 - k_i^2)$ where the set $\{k_i\}_1^M$ are the reflection estimated coefficients.

4.3.16 Singular Value Plot Criterion (R_t)

This information came from [63].

$$R_t = 1 - \frac{\sum_{k=0}^{N} \epsilon_k^T \epsilon_k}{\sum_{k=0}^{N} y_k^T y_k}$$

where $-\infty \leq R_t \leq +1$.

4.3.17 C_p Criterion

The C_p criterion is part of the early training of statisticians. It is discussed in [170][190][16][46].

$$C_p = (\hat{\sigma}^2)^{-1}(L_p) - N + 2p$$

or equivalently

$$C_p = \frac{SSE_p}{\hat{\sigma}^2} - (n - 2p)$$

where $\hat{\sigma}^2 = MSE(x_1, \cdots, x_{P-1})$.

The quantity p is the number of parameters in a regression model. $L_p = SSE_p$ is the residual sum of squares. N is the number of samples. The quantity $\hat{\sigma}^2$ is the estimate of σ^2 based on a model that includes all the parameters. The number $P - 1$ is the number of all the potential independent variables, assumed to have been carefully chosen to yield an unbiased estimate $\hat{\sigma}^2$ of

σ^2. The notations SSE_p and MSE are common in the regression and linear models statistical literature.

4.3.18 Bayesian Quickest Decision

This approach includes a penalty on the classical Bayes wrong decision cost function for delays in detecting a signal. The minimization of the average risk function leads to the optimum decision regions. A more detailed description of this approach would essentially repeat the original paper, so the interested reader is referred to original works by Bouvet [41]. This paper should be read together with Pelkowitz and Schwartz' 1987 paper [206].

4.3.19 Quickest Detection Sample Size

This method was proposed in [206]. The goal is to find the sample size M that minimizes the mean time to detection M_D for detecting a sudden change in the statistics of an observed process for a given mean time between false alarms $M_F = (\text{False Alarm Rate})^{-1}$. Let β be the single-sample signal-to-noise ratio. This paper shows that for small β and large M_F that the optimum sample size M and the system performance depends on M_F and β only through the product $\beta\sqrt{M_F}$. Graphs are provided in the paper for choosing parameters for the solution.

Let

$$R = \frac{M_D}{M_F} = \frac{\text{mean number of samples to detection}}{\text{mean number of samples between false alarms}}$$

and let λ be the detection threshold which is a function of the given test statistic, the data sample size M, and probability of false alarm α. Call R the average sample size ratio. Let the stationary noise process have mean μ_0 and variance σ_0^2. Let $F(x)$ be the cumulative distribution function of the received random process, and let $\Phi(x) = 1 - F(x)$. Let $r_o(m)$ be the normalized autocorrelation function of the stationary noise defined by

$$r_o(m) = \frac{1}{\sigma_0^2} E\left\{[x_0(n) - \mu_0]\,[x_0(n+m) - \mu_0]\right\} = r_0(-m)$$

and let

$$\gamma_0 = \sum_{m=-L}^{L} r_0(m) = 1 + 2\sum_{m=1}^{L} r_0(m)$$

Under the conditions that signal strength $\sigma \rightarrow 0$, signal-to-noise ratio $\beta(\rho) \rightarrow 0$, mean number of samples between false alarms $M_F \rightarrow \infty$, and $\beta(\rho)\sqrt{M_F}$ is some fixed constant Ψ_F, then the limiting values of the average sample size ratio R and detection threshold λ are given by

$$R(\alpha, \Psi_F) = \alpha\left\{\frac{1}{2} + \frac{h(\alpha; \Psi_M\sqrt{\alpha})}{\Phi\left[\Phi^{-1}(\alpha) - \Psi_F\sqrt{\alpha}\right]}\right\}$$

and

$$\lambda(\alpha; M_F) = \alpha M_F \mu_0 + \Phi^{-1}(\alpha)\sigma_0\sqrt{\gamma_0 \alpha M_F}$$

4.3.20 Likelihood Ratio Test

Söderström [250] gives the likelihood ratio test statistic as equation 4.22 for testing between model M_1 and model M_2. Model M_1 is chosen if λ is close to 1.

$$\lambda = \frac{\sup_{\theta,\Lambda;\theta\in M_1} L(\theta,\Lambda)}{\sup_{\theta,\Lambda;\theta\in M_2} L(\theta,\Lambda)} \tag{4.22}$$

Wilks [289] was the first to propose this statistic in 1938.

4.3.21 Guttman Lower Bound Criterion

This criterion, discussed in [111][99], recommends retaining all of the principal components that contribute more total variance than does the typical normalized time series. Richman [222] notes that it is safer to choose too many components than to choose fewer components than are suggested by such criteria. (Note that in the adaptive filtering context, we know that choosing a model order that is too large can lead to an unstable filter.)

4.3.22 Other Significance Tests

Horel [111] cites other significance tests which have escaped the electrical engineering literature. These tests are given in [37][201][194][191]. Testing of complex principal components is a part of geophysical data analysis. Söderström [250] discusses the use of the F-test for comparing two models.

4.4 Comparisons and Evaluations

Hipel [107] reported on the use of AIC in the context of geophysical time series. This is a broad ranging paper with an extensive bibliography. Hipel states the AIC formula, discusses its use in ARMA models for order determination, discusses model construction, alternatives to AIC, and some disadvantages of AIC. Alternatives include the maximum χ^2 method, Parzen's CAT, Gray's D-statistic, Mallows' C_p statistic, and Sawa's BIC statistic. He also discusses Akaike's $MAICE$ and final prediction error (FPE) technique.

Some disadvantages of the AIC and the other automatic selection criteria are that an overall statistic tends to cover up much of the information in the data and the practitioner may lose his sense of feeling for the inherent characteristics of the time series if he bases his decisions solely upon one statistic. However, when $MAICE$ is used in conjunction with the three stages of model construction, there is no doubt that $MAICE$ greatly improves the modeling process.

Söderström [250] observed that AIC and FPE are asymptotically equivalent to an F-test. Kundu [158] compared simulation results of several information-theoretic criterion (AIC, MDL, and EDC) and Cross Validation. He observed that AIC and Cross Validation perform quite well for small samples and large error standard deviation, and noted that the small sample properties of MDL and EDC have not been investigated fully. When the radian frequency of two

signals are close, then the Cross Validation approach performs better than any other method.

Wang and Kaveh [275] compared the asymptotic performance of AIC and MDL as part of a study of a generalized information theoretic order determination method that subsumes those two methods as applied to the case of an array of M sensors. They concluded for cases of up to two closely spaced sources in spatially white noise, that Rissanen's MDL penalty function was shown to result in a larger probability of underestimating but smaller probability of overestimating the number of sources in comparison to Akaike's AIC penalty function.

Zhang, Wong, Yip, and Reilly [296] did a statistical theory and simulation comparison of AIC and MDL. They concluded that AIC is more efficient in reducing the probability of missing a detection than the MDL criterion. On the other hand, for a moderate number of snapshots, the probability of false alarm using the MDL criterion approaches zero whereas that for the AIC remains constant. The MDL criterion is more efficient in reducing the probability of false alarm than the AIC. The choice of the penalty term by AIC emphasizes better performance under relatively lower SNR or smaller number of snapshots (or both) at the expense of being inconsistent. The penalty term adopted by MDL emphasizes the performance when the number of snapshots is large, sacrificing the performance at relatively lower SNR or

smaller number of snapshots (or both). They cite Chen, Reilly, and Wong [54] to remark that the penalty function can be adjusted to obtain a criterion whose performance best satisfies the chosen goal. The choice depends on the number of snapshots and the signal-to-noise ratio. Under low SNR, both AIC and MDL necessitate a large number of snapshots. The authors show in another paper that the performance of both criteria can be improved by choosing a more appropriate log-likelihood function [292].

Cremona and Brandon [63] remark that statistical tests (χ^2), AIC criterion, and FPE criterion tests are restricted to recursive minimum prediction error methods. Independent of their good estimation, they are statistically based: they are partially subjective techniques because they use the asymptotic property of the estimates on which to base the model order estimation strategy.

In this chapter, an abstract setting via the Lebesgue-Radon-Nikodym derivative was provided to illustrate that, collectively, order determination and estimation are pieces of the same task of locating or discovering the distribution that best describes the sampled data. Most of the examples of methods for order estimation are variations of information-theoretic approaches. There are also approaches from the points of view of coding theory, maximum likelihood, maximum entropy, and classical regression methods of statistics. included in this review are reviews of comparisons among techniques.

The approach of this thesis is a classical Neyman-Pearson hypothesis testing approach. It requires knowledge of the density functions of distributions of interest and specification of the acceptable chance of error of a test. The mathematics for the small sample complex principal components analysis, the simplest of the multivariate cases relying on sampling from a complex vector normal distribution, has not previously been worked out. Many necessary pieces have. The next chapter reviews what I have learned about the existing background material.

Chapter 5

PREVIOUS WORK

The purpose of this chapter is to present material I have found which provides the necessary foundations for the development of the small sample complex principal components analysis approach for order determination. Three main areas will be reviewed: array processing, statistics, and mathematics.

Traditional lines of demarcation between disciplines become very inappropriate when studying the order identification problem in array processing. Motivated by the acoustic signal processing goals, the appropriate locus of solutions lie beyond the traditional mathematical training of engineers and statisticians, and is in research areas by specialists in mathematics and statistics. The history of development of the necessary mathematics reveals that much of the important mathematical theory has been developed by application scientists. What is considered pure or abstract mathematics by most engineers and statisticians truly forms the working set of knowledge necessary for making headway in the solution of practical array processing problems.

With this in mind, I have rather artificially clustered historical work as follows. Under the title of "array processing" I have collected works drawn primarily from the acoustics, ocean engineering, and electrical engineering literature. These works deal primarily with exploration of different principles and algorithms. Material collected under the heading of "statistics" is further

partitioned into "eigenvalue distribution and testing" and "complex statistics other than eigenvalue testing". The set called "eigenvalue distribution and testing" discusses work done with respect to eigenvalues of both real and complex Wishart matrices. It focuses on the form of the test statistics, the distributions of the eigenvalues, and the distributions of the test statistics. The material collected under "complex statistics other than eigenvalue testing" refers to the body of literature that forms the supporting background theory for eventually developing the necessary tests and test statistic distributions. Under the final grouping of "mathematics" I have included material related to the development of zonal polynomials and hypergeometric functions of matrix argument which is presented independent of the context of statistics. This is necessarily set in the context of group representation theory which provides the foundation for these functions.

Not mentioned, yet present in the background, is the vast body of knowledge collected under the subject of Lie theory. There is some artificiality here because much of the ancestral work is by physicists and statisticians seeking answers to the eigenvalue testing problem. There is a lot of interplay between these groupings. With just a little exposure to the literature, one can see that the overlap is tremendous.

5.1 Array Processing

A well written brief tutorial review of beamforming methods was presented by Johnson [125] as an invited paper for the *Proceedings of the IEEE*, as pointed out in the introduction. One of the methods he discussed was that of the Maximum Likelihood Method (MLM). He references three articles on the subject [76][51][52]. The approach is to find the steering vector a which yields the minimum beam energy $a^H Ra$ subject to the constraint that $a^H b = 1$, where b represents an ideal plane wave corresponding to the desired direction-of-look and R represents the spatial correlation matrix. The solution is $a = \frac{R^{-1}b}{b^H R^{-1}b}$.

Another approach is the eigenspace approach. The idea of an eigenspace approach to signal processing is not new. In particular, the principal component analysis approach is now considered classic. Priestley et al. [210] discussed the application of principal component analysis and factor analysis to multivariate systems for the purpose of dimensionality reduction. They chose as their goal to obtain the best r-dimensional representation of the system output vector $Y(t)$. Their method is as follows. Apply the discrete Fourier transform to $Y(t)$, obtaining $Y(\omega)$, and then obtain eigenvalue decompositions of the resulting frequency-dependent covariance matrices. Process $Y(\omega)$ with the eigenvectors corresponding to the r largest eigenvalues, obtaining r-

dimensional frequency domain principal components $Y_r(\omega)$. The r-dimensional time domain output $Y_r(t)$ is obtained by an inverse discrete Fourier transform. The authors point out that there may be an aliasing problem, as discussed by Haggan and Priestly [100] who successfully applied the method to a real system. The issue of order estimation was not discussed.

Schmidt discussed the MUSIC (MUltiple SIgnal Classification) theory in March 1986 [238]. Consider a sonar array with m elements. Let the noise at these elements be given by the column vector w where $w^T = (w_1, \cdots, w_m)$. Let d be the number of signals $\{f_i\}_{i=1}^d$, independent of the noise. Denote the set of signals by the column vector f defined by $f^T = (f_1, \cdots, f_d)$. Each signal f_i has its own beamformer parameter index (which we usually associate with direction of arrival θ_i). The transfer function of the beamformer on the set of signals is given by the matrix $A = [a(\theta_1), \cdots, a(\theta_d)]$ where each $a(\theta_i)$ describes the response of the beamformer to a signal coming from direction θ_i. It is assumed that the beamformer function $a(\theta)$ is known for all θ. For this reason, for a collection of specific desired look-directions, the matrix A defines a set of vectors that form a basis (in the sense of linear algebra) for the space in which signals processed by those beams can be described. Therefore, we can use matrix A to form an orthonormal basis for the space containing the signals. All of the signals and some of the noise processed by the beamformer will be contained in this space. The orthogonal complement of this space will

contain only noise. The beamformer output of signal plus noise is given by the column vector $x = Af + w$. Let the signal covariance matrix P be defined by $P = \mathcal{E}\left\{ff^H\right\}$, and let the noise covariance matrix $\lambda^2 S_0$ be defined by $\lambda^2 S_0 = \mathcal{E}\left\{ww^H\right\}$. (If you compare this to Schmidt's paper, you will notice I use λ as a singular value throughout, and thus λ^2 is the eigenvalue.) The expected value of the covariance matrix of the beamformer output is given by $S = APA^H + \lambda^2 S_0$. The eigenvalues of S and of $(S - \lambda^2_{\min} S_0)$ differ by $\lambda^2_{\min}$. The multiplicity of $\lambda^2_{\min}$ in matrix S or the multiplicity of the zero eigenvalue in $(S - \lambda^2_{\min} S_0)$ tells us the dimension of the space containing only noise, and therefore we also know the dimension of the space containing the signals. The problem is stated in terms of examining the roots of the characteristic equation

$$\det\left(APA^H\right) = \det\left(S - \lambda^2_{\min} S_0\right) = 0$$

The paired sets of eigenvalues and eigenvectors $(\lambda_i^2, q_i)_{S_0}$ are called eigensolutions of S with respect to S_0. Another terminology used is that these are eigensolutions of S in the metric of S_0. Schmidt notes that the eigensolutions satisfy the relationships $Sq_i = \lambda_i^2 S_o q_i$ and $APA^H q_i = (\lambda_i^2 - \lambda^2_{\min}) S_0 q_i$. So, the goal is to construct a test to see how many of the smallest $\{\lambda_i^2\}_1^m$ are equal. Various authors have chosen different approaches to identifying this multiplicity, including the selection of the estimator of APA^H upon which to base tests. Note that the maximum dimension of APA^H is $\min(d, m)$. The dimension can be reduced by singularity of P.

Kaveh and Barabell [130] credit Kumaresan and Tufts [156] with the Minimum Norm method. Kumaresan and Tufts considered a line array. Kumaresan and Tufts' description is repeated here. The author's variables are renamed to make comparison with the MUSIC algorithm easier. Let the number of elements of this array be m. Assume a known number of sources, which we will call d. Let the $m \times m$ signal-plus-noise covariance matrix for the beamformer output S be estimated by R. Let R have the eigenvalue decomposition $R = PL^2P^H$ corresponding to the eigenvalue decomposition of S given by $S = Q\Lambda^2Q^H$. Let $a(\theta_k) = a_k$ be the direction vector associated with source number k having direction-of-arrival at the array at an angle related to θ_k. The problem is to estimate a_k. If a vector $b = [b_1, \cdots, b_m]$ has the property that $a_k^H b = 0$ for each source k, then a polynomial $D(z) = \sum_{k=1}^{m} b_k z^{-k+1}$ has roots at values of z corresponding to the $\{\theta_k\}$. The $m - d + 1$ eigenvectors $\{q_k\}_{d+1}^{m}$ of S corresponding to the noise eigenvalues $\{\lambda_k^2\}_{d+1}^{m}$ have this property. This is approximately true for the sample eigenvectors $\{p_k\}_{d+1}^{m}$ computed from R corresponding to the noise subspace.

The goal is to find b spanning the whole noise subspace of R. The source of the name "Minimum-Norm" comes from the following criterion. Its Euclidean length (its norm) is required to be minimized. To make the solution unique, the first element is constrained to be unity.

Partition the sample eigenvectors $P = [P_S, P_N]$ into the set

$$P_S = (p_1, \cdots, p_d) = \begin{bmatrix} g^T \\ \\ P_{[S]} \end{bmatrix}$$

corresponding to the signal subspace, and

$$P_N = (p_{d+1}, \cdots, p_m) = \begin{bmatrix} c^T \\ \\ P_{[N]} \end{bmatrix}$$

corresponding to the noise subspace. The vectors g^T and c^T are the top rows of their respective matrices. The matrices $P_{[S]}$ and $P_{[N]}$ are merely the remainder of their respective partitions. The solution is given by

$$b = \begin{bmatrix} 1 \\ \\ \frac{P_{[S]}g^*}{1-g^Hg} \end{bmatrix} = \begin{bmatrix} 1 \\ \\ c^Hc \end{bmatrix}$$

where the top element of b is unity.

Their theory is developed based on partitioning the eigenstructure of the underlying covariance matrix of the beamformer output (not the sample covariance matrix). The discussion implied that the simulated sample covariance matrix was decomposed. The number of signals was assumed.

Kaveh and Barabell [130] evaluated the asymptotic statistical performance of the MUSIC algorithm and the Minimum Norm algorithms in April 1986 against closely spaced narrowband plane waves. The Minimum Norm null-spectrum had a smaller bias at a source angle compared to the MUSIC null-spectrum. In a simulation, a fixed resolution threshold was achieved at a

lower signal-to-noise ratio for the Minimum Norm method than by MUSIC. Moghaddamjoo reported simulation results in [184]. It was shown that if at least one signal eigenvalue is close to the noise-related eigenvalues, then the associated eigenvector will have significant errors which translate into a significant bearing error. This is to be expected for a low signal-to-noise ratio case.

Wax, Shan, and Kailath [279] discussed eigenstructure methods for beamforming for both the narrowband and wide band cases. They specialized their treatment to a line array of equally spaced sensors.

Each of the m sensors feeds a delay line with p registers. Each independent sample thus contains mp pieces of data. The number of sources d is unknown. Under Gaussian assumptions, the statistic for testing

$$H_k : \lambda_{k+1}^2 = \cdots = \lambda_{mp}^2$$

is given by Anderson's likelihood ratio given here as equation 5.1 where the $\{l_i^2\}_1^{mp}$ are the eigenvalues of the sample covariance matrix. Wax et al. discussed the application of the asymptotic case for the statistic distribution.

$$Q_k = \left[\frac{\prod_{i=k+1}^{mp} l_i^2}{\left(\frac{1}{mp-k} \sum_{i=k+1}^{mp} l_i^2 \right)^{mp-k}} \right]^N \tag{5.1}$$

Scharf [237] proposed looking at the quantity he calls divergence, which is the difference between the expected values of the log likelihood ratio test

statistic under the null and alternate hypotheses. Based on this, he points out that it is the sum $\lambda_n^2 + \lambda_n^{-2} = 2\cosh(2\ln\lambda_n)$ which determines the contribution of an eigenvalue to divergence, not the value of λ_n^2. He presents an algorithm that selects the dominant eigenvalues. He also notes that each eigenvalue satisfies the generalized eigen problem $(R_1 - \lambda^2 R_0)x = 0$ where R_0 is the covariance matrix under the null hypothesis and R_1 is the covariance under the alternative. This is not restricted to the case of $R_1 = R_0 + R_s$ where R_s is the signal covariance; however, that is the usual assignment.

Friedlander presented an eigenspace approach to interference cancellation in his nicely written December 1988 paper [88]. The key to his approach is constructing a weighting vector W such that W lies in the signal subspace and is orthogonal to the interference component of the array manifold. The array manifold $a(\gamma)$ is defined to be that portion of the factorization of the array response function which is due to the geometry of the array (function of the time delay from each array element to a reference point) and the steering direction γ. The signal subspace is defined to be the set of vectors in the array manifold associated with each signal and interference source, $[a(\gamma_{signal}), a(\gamma_{source2}), \cdots, a(\gamma_{sourceP})]$. This same subspace is also spanned by the eigenvectors associated with the largest eigenvalues of the covariance matrix of the received signals.

The innovation of this paper is finding a way around having to know $a(\gamma)$

of each interfering source. To overcome this, he considers a cost function based on the spectral characteristics of the desired signal. The method will not work directly where there is a coherent multipath present, but a possible modification is proposed. Although his paper is written in terms of eigenvalue decompositions, he makes it clear that use of the singular value decomposition is a related idea.

Fuchs [89] also discusses an eigenspace approach in that same journal issue. He bases his approach on a matrix perturbation analysis. Lee and Wengrovitz [163] studied the ability of MUSIC to separate closely spaced sources when a beamforming preprocessor is used. It was shown that this technique, called Beamspace MUSIC, performed better than the Minimum Norm technique. The key is to reduce the noise subspace. They also suggested that a beamforming preprocessor improves the performance of the Minimum Norm algorithm.

5.2 Statistics

5.2.1 Eigenvalue Distributions and Testing

In statistics, the problem being considered is known as the part of the Exact Principal Components Analysis (PCA) problem for the complex variables case. The inner product of a data vector with the k^{th} eigenvector of the sample

covariance matrix is the k^{th} *Principal Component.* The sample variance of the k^{th} principal component is the corresponding sample eigenvalue l_k^2. Eigenvalues go by several names in the literature. They are also known as characteristic roots and as latent roots. A very important fact [197] is that the eigenvalues of a sample covariance matrix are all different, with probability 1. A wonderful introduction to Principal Component Analysis is given in Chapter 8 of reference [186]. Sections of special interest are 8.3 (geometrical meaning of principal components) and 8.7 (sampling properties of principal components).

In geophysics and meteorology, principal components are known as Empirical Orthogonal Functions. A solution to the problem consists of several parts. The first and easiest part is the specification of test statistics. The next part is obtaining the distribution of the test statistics. Closed form solutions are desired but not always obtainable. Sometimes they are obtainable with great effort or clever tricks. Often the density of a desired distribution is the marginal density of some obtainable joint distribution that is difficult to integrate.

Von Storch and Hannoschöck [261] discussed estimating principal components in the small sample case in the context of meteorology. Their conclusions are important enough to repeat.

1. The sample eigenvalue l^2 is a considerably biased estimator of the true eigenvalue λ^2. The bias is positive for the largest λ^2; the bias is negative for

the smallest eigenvalue. It is of the order of $1/m$ where m is the number of independent samples. The variances of l^2 is the order of $1/m$ too.

2. By means of correction methods, unbiased eigenvalue estimators are constructed. However, the decrease of the bias is accompanied by an increase of the estimator's variance. For the largest eigenvalue, at least, the Jackknife yields favorable results.

3. The following comments are in the context of estimated second moments of generalized Fourier coefficients of a fixed set of principal components. On average, for small i (large i) the sample eigenvalue l_i^2 will overestimate (underestimate) the variance expressed by the corresponding principal component considerably. The covariances are generally not negligible. This means that the independence of parameter covariance matrix eigenvector coefficients cannot be transferred to principal component coefficients derived from the sample covariance matrix.

Kshirsagar [154] (p. 58) gives a fascinating review of the history of the derivation of the real Wishart distribution. He says the case of $p = 2$ was first derived by Fisher in 1915 [83], and that Wishart did it for $p = n$ in 1928 [290]. It was in 1935 (almost yesterday, when my father was 17) that Fisher published his paper [84] on the density and cumulative distribution functions for the univariate χ^2 and t distribution. In 1937, Hoel [108] derived approximations for the distributions for the generalized variance (the determinant of the covari-

ance matrix), one for the case when samples are not too small, and the other for large samples. Two early papers on principal components are by Hotelling [113] in 1933 and by Girshick [90] in 1936, both of which are referenced in one of the early works in the distribution of sample eigenvalues by Girshick [91] in 1939. In this 1939 paper, he derives the asymptotic distribution for the sample eigenvalues of a real Wishart matrix, as well as other quantities. Let σ_{ij} be the population covariance between random variables x_i and x_j in multivariate normal random vector $x^T = (x_1, \cdots, x_p)$. The fundamental equation derived by Girshick in this paper is his equation (3.11),

$$\mathcal{E}\{d\sigma_{ij} d\sigma_{km}\} = \frac{1}{n}(\sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk})$$

From this equation, he produces his other results. Specifically, the variance of the sample eigenvalue l_k^2 is given by $\text{var}(l_k^2) = \frac{2}{n}\lambda_k^4$ where n is the number of samples from which the estimate is derived. (When you compare the formulae written here, remember that in this paper, the $\{l_k\}_1^p$ are estimates of the singular values $\{\lambda_k\}_1^p$.) The set of quantities

$$\left\{ \frac{l_k^2 - \lambda_k^2}{\sigma_{l_k}} = \frac{l_k^2 - \lambda_k^2}{\lambda_k^2\sqrt{\frac{2}{n}}} \right\}_1^p$$

is distributed asymptotically $N_p(0, I_p)$. By a clever insight, Girshick considers the quantity $\log l_k^2$ as a way to eliminate the population eigenvalue λ_k^2. By applying a Taylor series expansion and ignoring higher order terms, he finds the asymptotic variance of $\log(l_k^2)$, which is given by $\text{var}(\log l_k^2) = \frac{2}{n}$. As

an aside, Girshick uses the following convention we have come to associate with Einstein. A repeated subscript in the same term stands for summation. If repeated subscripts appearing in a term are not to be summed, they are placed in brackets following the expression in which they appear.

Lawley [161] studied tests involving the latent roots of sample covariance and correlation matrices in 1956. His interest was in those cases where the effects of the k largest latent roots have been removed, and he tested the hypothesis that the re ning roots are equal. The Principal Components Analysis problem for the raw covariance matrix was solved by T. W. Anderson in his 1963 paper [24] which has become a classic paper in the statistics literature. He gives a test of significance on eigenvalues for the large sample case where the data is sampled from the real multivariate normal distribution. In the immediately following article, Lawley [162] extended Anderson's result to test a set of correlation coefficients for equality. It was solved for the large sample complex multivariate normal distribution case by R. P. Gupta [98] in 1965 who purposefully paralleled Anderson's derivations and used the same notation as closely as possible. Work on the asymptotic cases has been continued by Tyler [269][270].

The solution for the Exact PCA problem was considered intractable for a long time, as it is often true that small sample cases are much more difficult than the corresponding asymptotic cases. That is why the asymptotic cases

are studied. This statement provides an opportunity to establish an important and easy to miss point. The label "asymptotic" is ambiguous because it is used in two different ways in the technical literature. The commonly assumed meaning is the large sample case, obtained by letting the number of samples tend to infinity. The second meaning is related to the number of terms carried in the expansion of an exact or approximation expression. In this second case, a small number of terms may yield a more accurate approximation than a large number of terms. This point is nicely discussed in Keener's text [131] (p. 425) as follows. He defines $f_n(z)$ to be asymptotic to $f(z)$ if

$$\lim_{z \to \infty} \mid z^n(f_n(z) - f(z)) \mid = 0$$

for a fixed value of n. This concept has nothing at all to do with convergence since finding a good approximation does not require taking more terms. A series can be asymptotic even though it may be divergent. In fact, asymptotic series are often divergent, so taking more terms is not simply more work, it is actually damaging.

Progress in attacking the small sample cases was motivated by two seminal works by James [118][120]. In 1960, he found the sampling distributions of the eigenvalues of the covariance matrix from a sample of the real multivariate normal distribution. He relied on representation theory of the linear group. In 1964, James extended his results to include distributions derived from the complex normal distribution, as well as other forms that are related to the

multivariate normal distribution. He developed his results through the use of zonal polynomials of matrix argument, and expressed his results in terms of hypergeometric functions of one and two matrix arguments. He did his work for the case of real variables. Based on similarity of forms, he summarily wrote down the results for the complex case without proof. In 1966, James [121] applied his work to principal components in the case of a sample covariance matrix of real variables. An interesting observation he made concerns the effect of extreme roots on the likelihood ratio of other adjacent roots. Suppose that the ratios of the root l_j^2 to the adjacent roots l_i^2, l_{i+1}^2 are both much less than 1 or both much greater than 1. Then the j^{th} root influences the likelihood of the other two by a factor

$$[(l_i^2 - l_j^2)(l_{i+1}^2 - l_j^2)]^{-1/2}$$

Muirhead [187] elaborated on James' work, collecting many of the ideas into the setting of studying distribution theory for real multivariate analysis. Muirhead's book is the natural descendent of Anderson's classic text. Muirhead produced the first comprehensive text on multivariate distribution theory incorporating zonal polynomials, hypergeometric functions of matrix argument, and application of exterior products. The importance of this development is its application to the derivation of noncentral distributions needed to evaluate the power of test statistics.

Krishnaiah was very active in developing exact and asymptotic distribu-

tions of eigenvalues and their tests based on both real and complex Wishart matrices, often expressing results in terms of zonal polynomials or using zonal polynomials in his proofs. Much of this work was done through the Aerospace Research Laboratories of the United States Air Force at Wright-Patterson Air Force Base. In 1969, Krishnaiah and Waikar [144] reported on tests of eigenvalues from a real Wishart matrix based on Roy's union-intersection principle. Effectively, the null hypothesis is

$$H : \lambda_1^2 = \lambda_2^2 = \cdots = \lambda_p^2$$

Five different alternative hypotheses were derived.

A_1 : $\lambda_1^2 > \lambda_2^2 > \cdots > \lambda_p^2$

A_2 : $(\lambda_1^2 > \lambda_p^2) \cup (\lambda_2^2 > \lambda_p^2) \cup \cdots \cup (\lambda_{p-1}^2 > \lambda_p^2)$

A_3 : $(\lambda_1^2 > \lambda_2^2) \cup (\lambda_1^2 > \lambda_3^2) \cup \cdots \cup (\lambda_1^2 > \lambda_p^2)$

A_4 : $(\lambda_1^2 \neq \bar{\lambda}^2) \cup (\lambda_2^2 \neq \bar{\lambda}^2) \cup \cdots \cup (\lambda_{p-1}^2 \neq \bar{\lambda}^2)$

A_5 : $(\lambda_1^2 > \lambda_2^2) \cup (\lambda_1^2 > \lambda_3^2) \cup \cdots \cup (\lambda_1^2 > \lambda_p^2) \cup (\lambda_2^2 > \lambda_3^2) \cup \cdots \cup (\lambda_{p-1}^2 > \lambda_p^2)$

The joint densities for these tests were provided for the case of the real Wishart distribution, expressed in terms of the hypergeometric function of two matrix variables and in terms of normalized zonal polynomials. One of these is generalized in Muirhead's derivation [187] of his Theorem 3.2.20. The different cases are for different test statistics and alternative hypotheses. Krishnaiah and Waikar also work out the asymptotic cases (large sample size). Three months later, Krishnaiah and Waikar [145] further developed the test against

alternative A_5 by finding the density function for the test statistic $\frac{l_1^2}{l_p^2}$.

In 1970, Krishnaiah and Chang [146] reported on the exact distribution of the smallest root of the real Wishart distribution $W_p(n, nI_p)$ where they require the number $(n-p-1)/2$ to be an integer. They accomplish this by changing variables from the sample eigenvalues $(l_1^2, \ldots, l_p^2)$ to $(g_1, \ldots, g_{p-1}, \theta_p)$ where $g_i = \frac{l_i^2}{l_p^2}$ and $\theta_p = l_p^2$, and then integrating out the g_i. The result is expressed in terms of zonal polynomials. Three months later, Krishnaiah and Waikar [147] reported on the cumulative distribution function of the intermediate eigenvalue l_r^2 of the real Wishart matrix distributed as $W_p(n, I_p)$. The results are reported in an integral form. They assume l_1^2 is known, and they look at

$$P\left\{l_r^2 < x\right\} = P\left\{l_{r-1}^2 < x\right\} - P\left\{l_1^2 < \cdots < l_{r-1}^2 < x < l_r^2 < \cdots < l_p^2\right\}$$

Lemma 2.1 of [147] is referenced in later reports. In a separate simultaneous report [148], they show how to evaluate

$$P\left\{X_1 < l_r^2 < l_s^2 < X_2\right\}$$

for the real variables case. This is expressed as the sum of four probabilities that are characterized by the details of the end points of evaluation. The message is to consider the different combinations suggested by the set of inequalities given in equation 5.2.

$$\begin{bmatrix} l_{r-1}^2 \\ X_1 \end{bmatrix} < l_r^2 < l_s^2 < \begin{bmatrix} l_{s+1}^2 \\ X_2 \end{bmatrix} \tag{5.2}$$

Two months later, Waikar, Chang, and Krishnaiah [272] extended the work to find the joint density function of any few unordered roots of a noncentral complex Wishart matrix. Without loss of generality, they consider the first r roots. The case for the central distribution was worked out by Wigner [286]. Waikar et al. used assumptions on the structure of matrix A that are different than those based on Goodman's work in relating complex and real Gaussian distributions. Thus, some care is needed in using results by one author in the results of another author.

In the Fall of 1971, Krishnaiah and Waikar [149] reported on the distribution of arbitrary consecutive ordered roots of the real Wishart matrix. This work includes the marginal density function and the cumulative distribution functions. Results are reported in integral form. In 1972, Davis [65] reported on the ratios of individual eigenvalues to the trace of a Wishart matrix. See also the work by Khatri [139] on the exact finite series distribution of the smallest or the largest eigenvalue. In that same year, Waikar, Chang and Krishnaiah [273] derived expressions for the joint densities of any few unordered roots of the noncentral complex Wishart matrix (as well as for three other matrices).

In 1973, Krishnaiah [150] continued the study of eigenvalues of complex random matrices by deriving the exact distributions of some test statistics based on eigenvalues of the matrix $Z = A(A + B)^{-1}$ where $A \sim \mathbf{C}W_p(n, \Sigma_1)$

and $B \sim \mathbf{C}W_p(m, \Sigma_2)$. Krishnaiah computed the joint density function of $(\frac{l_1^2}{l_p^2}, \frac{l_2^2}{l_p^2}, \cdots, \frac{l_{p-1}^2}{l_p^2})$, leaving the result as an integral and a product of sums of normalized zonal polynomials. He likewise computed the joint density function where the smallest sample eigenvalue l_p^2 is replaced by the sum of the sample eigenvalues in the denominators.

In 1974, Krishnaiah and Shuurmann [151] derived expressions for the distributions of the ratios of the intermediate roots to the trace of the real Wishart matrix, and the intermediate roots of the real Wishart matrix. They obtained a relationship between the Laplace transformations of the ratios of the individual roots to the trace of the complex Wishart matrix $\mathbf{C}W_p(\frac{n+p-1}{2}, I_p^2)$ and the distributions of the individual roots of this matrix. Using this relationship and expressions for the densities of the individual roots of the complex Wishart matrix, they obtained expressions for the distributions of the ratios of the individual roots to the trace of that matrix.

In 1976, Krishnaiah [152](pp. 26-27) proposed two more tests of interest when you know in advance that $\lambda_i^2 \neq \lambda_j^2$. The first test is for $H_{ij} : \lambda_i^2 < d\,\lambda_j^2$ for $d > 1$ against the alternative $A_{ij} : \lambda_i^2 > \lambda_j^2$ where $i > j$. Hypothesis H_{ij} is not rejected if $l_i^2/d\,l_j^2 < c_\alpha$ where

$$\Pr\left\{l_p^2/l_1^2 \leq d\,c_\alpha \mid \lambda_p^2 < d\,\lambda_1^2\right\} = (1-\alpha)$$

The second test is for $H_{ij} : \lambda_i^2 - \lambda_j^2 < d$ for $d > 0$ against the alternative

$A_{ij} : \lambda_i^2 - \lambda_j^2 > d$. Hypothesis H_{ij} is not rejected if $(l_i^2 - l_j^2 - d) < c_\alpha$ where

$$\Pr\left\{l_p^2 - l_1^2 \leq d + c_\alpha \mid \lambda_p^2 - \lambda_1^2 < d\right\} = (1 - \alpha)$$

The untimely and premature death due to cancer of the great statistician Paruchuri R. Krishnaiah on 01 August 1987 in Pittsburgh, Pennsylvania interrupted brilliant progress on these difficult problems.

In 1984, Jolicoeur [126] proposed a test about the direction of multivariate normal principal axes for the small sample case. Let S be a real-valued sample covariance matrix with normalized eigenvectors Γ having row vectors γ_i as the direction cosines of the i^{th} principal axes. Then the statistic

$$\left(\frac{N-p}{p-1}\right)\left(\gamma_i S \gamma_i^T \gamma_i S^{-1} \gamma_i^T - 1\right)$$

is distributed according to the F distribution with $p-1$ and $N-p$ degrees of freedom, $F_{(p-1,N-p)}$.

Konstantinides and Yao [142] reviewed criteria used to test the effective rank $t \leq n$ of an observed real matrix X by using the singular values. They critiqued the following test criteria and performed a perturbation analysis on the real matrix model $X = A + E$.

$$l_1^2 \geq l_2^2 \geq \cdots \geq l_t^2 > \delta_1 \geq l_{t+1}^2 \geq \cdots \geq l_n^2 \tag{5.3}$$

$$\frac{l_t^2}{l_1^2} > \delta_2 > \frac{l_{t+1}^2}{l_1^2} \tag{5.4}$$

$$l_t^2 \gg l_{t+1}^2 \tag{5.5}$$

$$l_{t+1}^2 + l_{t+2}^2 + \cdots + l_n^2 < \delta_4 \tag{5.6}$$

$$\frac{l_1^2 + l_2^2 + \cdots + l_t^2}{l_1^2 + l_2^2 + \cdots + l_n^2} > \delta_5 \tag{5.7}$$

Konstantinides and Yao also reported the following interesting theorems.

Theorem 3 *Let A be any real-valued $m \times n$ matrix $A = (a_1, \cdots, a_n)$. Let $\|A\|_F$ be the Frobenious norm of A defined as the square root of the sum of the squares of each element of A. Let the 2-norm $\|A\|_2 = \max\left(\frac{\|Ax\|_2}{\|x\|_2}\right)$ where the 2-norm of x is the square root of the sum of the squares of the elements of vector x. Then the following inequalities are valid:* $\max \mid a_{ij} \mid \leq \max \|a_j\|_2 \leq \|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|a_j\|_2 \leq \sqrt{mn} \max \mid a_{ij} \mid$.

Theorem 4 *Let A, B, and E be $m \times n$ real-valued matrices with $B = A + E$. Denote their respective singular values by α_i, β_i, and ϵ_i where $1 \leq i \leq k \leq \min(m, n)$, each set labeled in non-increasing order. Then $\mid \beta_i - \alpha_i \mid \leq \epsilon_i = \|E\|_2$ where $1 \leq i \leq k$.*

Theorem 5 *Let A, B, and E be $m \times n$ matrices with the 2-norm of E denoted by ϵ_1. If $\alpha_r > 2\epsilon_i$, then $\beta > \epsilon_1 \geq \beta_{r+1}$, and B is said to have effective rank of r.*

Horel presented a good practical review of complex principal component analysis [111]. One important property of complex principal component (CPC) analysis in the time domain mentioned is that since correlations between time

series are heavily weighted by periods during which the amplitudes of the time series are large, more weight is given to sharp transitions and noisy spikes than to periods during which the signal varies slowly. The Hilbert transform does not act as a low-pass filter upon the data. It contains as much energy due to noise as the original data and it may redistribute the noise to different parts of the time series. To minimize this problem, the filter weights $W(\omega)$ can be chosen so as to apply a low-pass filter to both the original data and its Hilbert transform prior to further computations.

The phase of the principal components is ambiguous. This indeterminacy becomes important when the researcher wishes to compare complex principal components obtained from independent data sets. In such cases, it is impossible to determine lead-lag relationships between the independent complex principal components by simply computing their cross-correlation since the phase of each complex principal component is known only to within an arbitrary constant.

In looking at time series, the real and imaginary parts of complex principal components are not Hilbert transforms of one another. They do not necessarily explain the same amount of variance in each frequency band and thus the real part of the complex principal components does not contain all the relevant information. Frequency domain principal component analysis does not suffer from this problem because in that approach the principal component is a real

time series.

Wong, Zhang, Reilly, and Yip proposed new estimates for sample eigenvalues in 1990. They account for the bias in the estimation of eigenvalues from looking at the eigenvalues of the sample covariance matrix. Let $\{\hat{\lambda}_i^2\}_1^M$ be the revised estimates of the corresponding population eigenvalues. These are computed in equation 5.8, with the estimated variance given in equation 5.9.

$$\hat{\lambda}_m^2 = l_m^2 - \frac{\hat{\lambda}_m^2}{N} \sum_{\substack{i=1 \\ i \neq m}}^{k} \frac{\hat{\lambda}_i^2}{(\hat{\lambda}_m^2 - \hat{\lambda}_i^2)} - \frac{M-k}{N} \frac{\hat{\lambda}_m^2 \hat{\sigma}_v^2}{(\hat{\lambda}_m^2 - \hat{\sigma}_v^2)}$$

where

$$m = 1, \cdots, k \tag{5.8}$$

$$\hat{\sigma}_v^2 = \frac{1}{M-k} \sum_{i=k+1}^{M} l_i^2 + \frac{1}{N} \sum_{i=1}^{k} \frac{\hat{\lambda}_v^2 \hat{\sigma}_v^2}{(\hat{\lambda}_i^2 - \hat{\sigma}_v^2)} \tag{5.9}$$

The good idea is that these provide a correction to the estimated eigenvalues that accounts for the effects of other eigenvalues. On the other hand, these estimates no longer obey the simpler joint distribution which makes finding the distribution of relevant test statistics more difficult. The desirability of making these corrections depends on what you want to use the answer for. The point here is that the $\{l_i^2\}$ might be biased estimates of the underlying population eigenvalues, but even more importantly they are statistics whose distribution we know.

5.2.2 Complex Statistics Other than Eigenvalue Testing

Development of joint distributions of sample eigenvalues and related test statistics requires a supporting body of distributional results. The literature regarding complex multivariate statistics is sparse and isolated.

The study of statistics of complex variables is still so young that fundamental results are still in dispute. Some of the results I need simply are not in the many references I consulted. For this reason, I have undertaken a systematic development of fundamental properties and distributions related to complex multivariate random variables. This section reviews the literature I have found on the subject.

Working with complex variables in the context of statistics dates at least as early as the renaissance of statistics in the 1930s. Ingham published a paper in 1933 [114] evaluating the integral

$$\left(\frac{1}{2\pi}\right)^{p(p+1)/2} \int_{\mathbf{R}^{p(p+1)/2}} \exp[-i \operatorname{tr}(CT)][\det(A - iT)]^{-k}(dT)$$

where A is positive definite real and C and T are real symmetric. The univariate complex Gaussian distribution was first introduced by Wooding [293] in 1956. He looked at the complex Fourier series

$$z(t) = \sum_j (a_j - ib_j) \exp[i\theta_j(t)]$$

where θ_j, a_j, and b_j are real-valued coefficients. He followed some of the

work previously done with pre-envelopes and analytic signals by S. O. Rice [220][221]. This work was extended by Dugundji [70] in 1958. Wooding applied the Hilbert transform to a real signal $x(t)$ and formed a complex variable

$$z(t) = x(t) + i\hat{x}(t)$$

The Hilbert transform is defined by

$$\hat{x}(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} x(t-\sigma)\frac{d\sigma}{\sigma}$$

with its inverse given by

$$x(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \hat{x}(t+\sigma)\frac{d\sigma}{\sigma}$$

The *P.V.* before the integral sign signifies that the Cauchy Principal Value is used in doing the evaluation. Some references use a bar through the integral instead of writing *P.V.*. As acknowledged by Wooding, the notion of a stochastic process as being complex Gaussian was not new. Root and Pitcher [227] mention it in their paper in 1955.

The next major contributor to the theory of the multivariate complex normal distribution and related statistics was N. R. Goodman [92][93]. Goodman demonstrated the relationships between complex and real vector variables. He gave explicit expression to how various properties are related. He showed that multiplication of complex scalars of the form $z = x + iy$ is the same as multiplication of matrices of the form $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$. If you replace each element

in an $n \times n$ complex matrix with the corresponding 2×2 matrix, you then have a real-valued $2n \times 2n$ matrix that acts the same under multiplication and addition as the complex matrix, except that much more computational effort is required. He proved other algebraic results as well. In statistics, he stated the density function of the zero mean vector complex normal distribution with covariance matrix Σ and derived its characteristic function. He derived the density function for the central complex Wishart distribution and the characteristic function of a distribution related to the central complex Wishart distribution. He also derived the density function of the Hermitian square root upper triangular matrix of a Wishart matrix, where $W = T^H T$. In the companion paper, Goodman derived the distribution of $\det(W)$.

In reference [135], Khatri cited Wishart's 1948 effort in Biometrika to catalog many different methods of deriving the real Wishart distribution. Khatri said these methods use different kinds of tools like transformations, direct integration, characteristic function, and inversion theorem, geometrical method, induction, rectangular coordinates, random orthogonal transformations, orthogonal groups, etc. He pointed to Kshirsagar's Bartlett decomposition of a Wishart matrix, and also produced his own derivation which involved a partitioning scheme. In 1963, Khatri published a paper [136] discussing conditions under which a second degree polynomial in elements of a real matrix normal variable would be Wishart. He also discussed issues of independence of real

vector normal variable sample mean and sample covariance matrix. As an epilogue, he remarked that the results also hold for the complex case with the appropriate changes. In 1965, Srivastava [256] published his important paper on the complex Wishart distribution, which included a powerful generalization for finding the density function of any random variable $A = BB^H$ when the density of the random variable B depends on B only through the form BB^H. Shortly thereafter, in the same year, Khatri [137] published a comprehensive review of classical statistical analysis based on the vector complex normal distribution. A paper published by Tan in 1968 in the Tamkang Journal of Mathematics [266] gives an extensive development of distribution theory related to the complex normal distribution. It has been relatively unnoticed because it was published in Taipei. It deserves much wider recognition. Krishnaiah [152] updated this review in his comprehensive paper of 1976. Srivastava and Khatri's book [257] on multivariate statistics in 1979 treats the complex case where it can do so profitably without destroying the flow of the material. They include complex matrices in their theorems on matrix theory.

There are two texts devoted to the statistics of complex variables, both by Kenneth S. Miller. Miller's interest in statistics of complex variables is a natural extension of his earlier work [177] which has wide application to a more traditional treatment of signal processing restricted to real variables. That text includes discussions on topics such as the Generalized Rayleigh distribution,

Rice variates, Whittaker functions, envelope detection, Cramer-Rao bounds, Wiener-Khintchine relations, and passage of Gaussian noise through a linear filter. Miller's 1974 book [180] deals with complex stochastic processes. The next book [181] develops the theory of hypothesis testing using univariate and bivariate complex Gaussian variables. It begins with a review of Neyman-Pearson testing. Throughout, it works with the bivariate complex normal distribution, $\mathbf{C}N_2(c, R)$. He derives the bivariate complex Wishart density function $\mathbf{C}W_2(n, R)$ and references Goodman [92] for the density function of $\mathbf{C}W_p(n, R)$. He addresses groups of transformations, functions invariant with respect to a group, and functions that are maximal invariant. He observes that uniformly most powerful (UMP) tests do not abound, but sometimes it is possible to find UMP invariant tests with respect to some group G. He also recommends further restriction to the class of unbiased tests.

Compared to other areas of statistics, the literature on statistics of complex variables appears sparse. Saxena provided a nice annotated bibliography of 60 references subtending Rice's 1944 paper [220] through 1976. The largest number of references in any one year was 8 in 1972. The early works are application oriented. A short small spurt of work began with Wooding's 1956 paper. Work resumed in 1963 which motivated about 7 years of work. Another increase in productivity began in 1970 which lasted 4 years. No papers were published in 1974, one in 1975, and two in 1976, which was the last year

included in the bibliography.

The following references either are from the period 1977-1991 or are earlier references I have located which were not cited in Saxena's paper.

Freedman and Lane [87] reported in 1980 that the first $n-1$ Fourier coefficients of the discrete Fourier transform of n independent complex normal variables are independent identical complex normal random variables.

Fang and Krishnaiah [79] published the asymptotic distributions of functions of eigenvalues of the complex noncentral Wishart matrix via perturbation theory in 1981.

Singh and Pillai [245] reported on the exact non-null distribution of Wilks' L_{vc} criterion in the complex case for testing the hypothesis

$$H : \Sigma = \sigma^2[(1-\rho)I + \rho ee^T]$$

where $\sigma > 0$ and ρ are unknown against the alternative hypothesis of inequality. The vector e is a vector of all ones, $e^T = (1, \cdots, 1)$.

Khatri [138] derived a test to determine if a complex Wishart matrix could be a real Wishart matrix. Andersson and Perlman [29] derived tests to determine if a p-dimensional sample complex covariance matrix could have come from a p-dimensional real multivariate distribution, and to test if a 2p-dimensional sample real covariance matrix could be considered to have the structure of a p-dimensional complex multivariate distribution.

B. N. Nagarsenker, P. B. Nagarsenker, and Quinn [189] derived an asymp-

totic expansion of the non-central distribution of Wilks' statistic for the complex Gaussian case. Wilks' Λ statistic is given by $\Lambda = \left[\frac{\det(A)}{\det(A+B)}\right]^N$ where $A \sim \mathbf{C}W_p(n, \Sigma, 0)$ and $B \sim \mathbf{C}W_p(m, \Sigma, \delta)$ where $\delta = \mu\mu^H\Sigma^{-1}$. A nice review of the life and works of Wilks, with insightful comments on his results, is found in Anderson [25].

Patil et al. published an encyclopedic dictionary of multivariate distributions [205] in 1984 which includes those defined of the field of complex numbers. A wonderful feature of this dictionary series is that is makes explicit the relationship between various distributions. This is an excellent entry point into the literature on distributions.

5.3 Zonal Polynomials, Hypergeometric Functions, Group Representation Theory

The purpose of this section is to review the development of zonal polynomials. Zonal polynomials are the key to developing the joint density function of sample eigenvalues of a complex Wishart matrix. The eigenvalues examined in this thesis follow that distribution.

The distribution for the case of the real Wishart matrix was derived by James [120] in 1964. He also wrote down the result by inspection for the case of a complex symmetric matrix, without derivation.

As of 1987, zonal polynomials had only been developed for the case of the real symmetric matrix and the two matrix argument case of a real symmetric matrix and real symmetric positive definite matrix [188]. Gross and Richards developed zonal polynomials for the case of complex Hermitian matrices in 1987 [96]. A contribution of this thesis is the application of their work to the distribution of sample eigenvalues of a complex matrix.

The reason we need to even think about zonal polynomials, hypergeometric functions, and group representation theory is because of the need to evaluate the integral $\int_{\mathbf{U}(p)} \text{etr}(-\Sigma^{-1}U^H AU)dU$ where the integral is taken over the set of all $p \times p$ unitary matrices. The function $\text{etr}(X)$ is a standard notation for $\exp(\text{tr}(X))$ in the literature and texts dealing with distribution theory in multivariate analysis.

Zonal polynomials are important to the study of the distribution of eigenvalues of a Wishart matrix. Takemura [265] has recorded a wonderful history of the development through 1984, from which I have taken many of the comments made below. Some of the works of James were briefly described in the earlier section on eigenvalue testing, yet the history of the development of zonal polynomials rests on these same works. James is often cited as the prime motivator for work with zonal polynomials.

In 1960, James published his paper [118] on the distributions of eigenvalues using representation theory of the linear group. His results are given in terms

of zonal polynomials. Much of the interest in zonal polynomials since 1964 has been a direct result of the application to multivariate statistics and the paper by James [120]. In that paper, he generalized his previous work and discussed a general method for calculating the zonal polynomial. Until the mid-1980s, work on zonal polynomials has been done primarily by statisticians. Since then, mathematicians have started to examine zonal polynomials in the context of more general structures which has led to new results and powerful generalizations.

Zonal polynomials form a subset of spherical functions. They are homogeneous harmonic polynomials defined on the surface of a multidimensional sphere. Zonal polynomials are orthogonal functions on n-dimensional spheres. You can think of them as generalized Legendre polynomials. In 3-dimensional space, in fact, zonal polynomials are directly proportional to Legendre polynomials [251].

Another early worker in this area is Constantine, who worked with James at least as early as 1958 [56]. In his 1963 paper [57], he worked in terms of complex symmetric matrices (not the same thing as Hermitian matrices), and defined the hypergeometric function of complex matrix argument as a function of zonal polynomials of complex symmetric matrix argument. With this, he derived the density function of the noncentral real Wishart matrix. He also found the moments of the determinant of a noncentral real Wishart

matrix. In his 1966 paper [58], Constantine defined a generalized Laguerre polynomial of complex symmetric matrix argument which, in turn, is defined in terms of zonal polynomials of complex symmetric matrix argument. With these, he finds the distribution of the generalized Hotelling's T_0^2 statistic where $T_0^2 = \mathrm{tr}(AB^{-1})$. Matrix A is a real noncentral Wishart matrix distributed as $A \sim W_p(n, \Sigma, \Omega)$ and real central Wishart matrix B is distributed as $B \sim W_p(m, \Sigma)$. When the multivariate normal distribution underlying A has mean vector μ, then the noncentrality parameter Ω is defined by $\Omega = \mu\mu^H$. The 1976 paper by Constantine and Muirhead [59] presents asymptotic expansions for distributions for several very important matrices, including $A(A+B)^{-1}$, for some or all of the eigenvalues of Ω large, which can be thought of as a generalized signal-to-(signal plus noise) ratio where A and B are defined above. They also develop asymptotic distributions for $\frac{1}{m}B$ and BC^{-1} where $\mathbf{C} \sim W_p(k, \Sigma)$. As in earlier papers, these results are developed in terms of hypergeometric functions of matrix argument.

By 1982, the importance of zonal polynomials to the development of distributions in multivariate statistics became recognized. Muirhead [187] (the student of both James and Constantine) published his text which included a major chapter devoted to zonal polynomials. Muirhead develops zonal polynomials as a solution to the partial differential equation

$$\Delta_y Z_\kappa(y) = [\rho_\kappa + k(m-1)]Z_\kappa(y)$$

where Δ_y is the differential operator, called the Laplace-Beltrami operator, defined by

$$\Delta_y = \sum_{i=1}^{m} y_i^2 \frac{\partial^2}{\partial y_i^2} + \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i}$$

where $\rho_\kappa = \sum_{i=1}^{m} k_i(k_i - i)$ and $\kappa = (k_1, \cdots k_m)$ such that $k = k_1 + k_m$. It has become traditional to use the Greek kappa (κ), the Latin letter k, and its subscripted partitions k_i even though the opportunity for ambiguity after copying exists. Muirhead provides a recurrence relation for computing the coefficients of the zonal polynomials. He also sketches the group representation theory development of zonal polynomials used by James.

In 1984, the Institute of Mathematical Statistics (IMS) published Takemura's monograph on the subject. This was only the fourth monograph IMS published on any subject. Takemura defines zonal polynomials as symmetric homogeneous polynomials on the eigenvalues of a symmetric matrix. He writes down the definition and properties of complex zonal polynomials without proof since the proofs for the real and complex cases are the same for his development. He remarks that complex zonal polynomials are simpler than real zonal polynomials, noting that the complex zonal polynomials are the same as homogeneous symmetric polynomials called the Schur functions. The explicit relationship is given by Saw's generating function introduced by Farrell [80]. Takemura shows these to be the same via the uniqueness property of the triangular decomposition of a positive definite symmetric matrix. Note

that it is possible to have a complex symmetric matrix, which is different than an Hermitian matrix. He also writes down the density function for the zero mean complex vector normal and the central complex Wishart distributions. In his Chapter 5, Takemura uses the symbol $\sim$ to denote complexification of a theorem established for the case of real variables. Takemura uses that symbol over variables to indicate they apply to the complex case. Prior to the work by Gross and Richards, Takemura's development of complex zonal polynomials was the most complete I have found in the literature.

The development of the theory of zonal polynomials has proceeded simultaneously from a traditional physics and special functions point of view as represented by Stein and Weiss [258], and from a mathematician's point of view as represented by Gross and Richards [96]. The nicest introduction to zonal polynomials from an engineer's point of view is Stein and Weiss' book. Its work was done without reference to James' work. Stein and Weiss work in the field of real numbers and use differentiation, so application to the complex field must proceed cautiously. They do not develop the splitting theorem needed for this thesis.

Gross and Richard's work is a development of the theory of hypergeometric functions of matrix argument, firmly rooted in group representation theory, that simultaneously treats the case of real, complex Hermitian, and quaternionic variables. Of importance to this work, Gross and Richards provided a

development of the splitting property for zonal polynomials in the context of complex variables. In the development, they assume Hermitian matrices for the complex case rather than complex symmetric matrices. This is evident by application of the unitary group. It is more mathematically motivated, and less applied, than the work by James. Compared to the work by Stein and Weiss, Gross and Richards do not include the specification of a reference point on the sphere. This prevents drawing observations about coordinate transformations made clear in the approach by Stein and Weiss. Closing the connections between these two works is a valuable task that needs yet to be done.

Gross and Richards published a continuation [97] of their studies in 1989 which introduces the concept of total positivity in the context of spherical series and hypergeometric functions of matrix argument. They point out that the spherical function known by mathematicians is the zonal polynomial known by statisticians. They also remark that up to scalar multiples, the spherical functions coincide with the Schur functions. They show the Euler integral

$$Z_m(t) = \frac{\Gamma_n(b)}{\Gamma_n(b-a)\Gamma_n(a)} \frac{[b]_m}{[a]_m} \int_{0<r<1} Z_m(rt)[\det(R)]^{a-n}[\det(1-r)]^{b-a-n} dr$$

where

$$\mathrm{Re}(a) > n - 1$$

$$\mathrm{Re}(b - a) > n - 1$$

$$t \in S_n$$

r : Hermitian matrices whose eigenvalues are between 0 and 1

$$\Gamma_n(a) = \pi^{n(n-1)/2} \prod_{i=1}^{n} \Gamma(a - i + 1)$$

as an example of a reproducing integral formula. It would be good to look at this in the context of section 2.2 and chapter 3 of Fowler's thesis [86] on Reproducing Kernel Hilbert Space because Krantz [143] showed that zonal polynomials are reproducing kernels.

Chapter 6

STATISTICAL TESTS

This chapter provides distributional results for test statistics that examine sample eigenvalues to gain understanding of underlying parameter eigenvalues. It is in this chapter that the thesis topic is most directly addressed. You will observe that I have only answered special cases of the thesis question.

Tests of greatest interest take a set of samples and form one statistic upon which decisions are based. These make the most efficient use of the data, but they also have sampling distributions that are very difficult to compute. A compromise is to partition the data set into independent sets, and form a test statistic from these sets. This approach does not make efficient use of the data. One version of this approach results in a test statistic that is easy to compute and has a sampling distribution represented by a function that is a standard function that statisticians work with.

Several approaches are presented in this chapter. The first approach arbitrarily partitions the data into two independent sets and forms an F-statistic from the ratio of independent sums of the sample eigenvalues. A second approach partitions the data into one block assumed to be noise-only and another partition that possibly contains a signal. The result is the joint distribution of the sample eigenvalues of the signal-plus-noise sample covariance matrix. This is the form of Schmidt's MUSIC problem [238].

A third approach is to work with data transformed into the form of $W = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ and obtaining the distribution of the statistic $x = \frac{a}{b}$. This requires assuming a population covariance matrix $\sum$ which can be part of the null hypothesis of a test. The ratio of sample eigenvalue sums or averages belong to this class.

A fourth approach has its origin as a maximum likelihood ratio test statistic which requires only partial knowledge of the population covariance matrix. For the real variables case, the asymptotic distribution of this sphericity test was derived by Anderson [24]. I have provided the joint density of this statistic with some nuisance variables. For the case of $p = 2$, I have provided the density and cumulative distribution function.

The last approach I examined, and the one of greatest interest in the general case, involves simple transformations of the sampling distribution of eigenvalues. I have assumed that the special case of the sample eigenvalues D having a joint distribution $\mathbf{C}W_p(n, \Lambda^2)$. The statistics for which I computed distributions were motivated by Krishnaiah's works (which will be referenced in their respective discussions). This section is the culmination of the supporting work in the appendices, both of the complex variables and zonal polynomial theory. There is still a great deal of work left to extend these to the general case.

6.1 Tests Based on Two Independent Sets of Samples

6.1.1 F-Statistic from Ratio of Independent Sums of Sample Eigenvalues

Consider the following procedure. Assume that all samples are statistically independent. Then it is possible to define an arbitrary partition of the sample set, splitting it into two sets. Form a statistic within each of the partitions and then compare the statistics. For example, let the statistic in one set be the sum of the m_1 largest sample eigenvalues, and let the statistic in the other set be the sum of the m_2 smallest sample eigenvalues. Because the two statistics were obtained from independent samples, the statistics themselves are independent. We know that linear combinations of sample eigenvalues yields a chi-square random variable. The independence of these statistics gives us hope that an F-statistic can be formed. A benefit is that the F-distribution is one of the most widely known and used distributions in statistics. Its properties have long been known.

Recall that if x_1 has the non-central chi-square distribution with parame-

ters ν_1 and δ_1, x_2 has the non-central chi-square distribution with parameters ν_2 and δ_2, and if x_1 and x_2 are independent, then $y = (\nu_2 x_1)/(\nu_1 x_2)$ has the doubly non-central F-distribution with parameters ν_1, ν_2, δ_1, and δ_2. The parameters ν_1 and ν_2 are usually called "degrees of freedom".

Theorem 6 *Let $W_1 \sim \mathbf{C}W_p(n_1, \Sigma_1, \delta_1)$ and $W_2 \sim \mathbf{C}W_q(n_2, \Sigma_2, \delta_2)$ be independent complex Wishart random variables. Let c_1 be a $p \times 1$ vector of known fixed constants, and let c_2 be a $q \times 1$ vector of known constants. Then*

$$F = \frac{\frac{2c_1^H W_1 c_1}{2n_1 c_1^H \Sigma_1 c_1}}{\frac{2c_2^H W_2 c_2}{2n_2 c_2^H \Sigma_2 c_2}} = \frac{n_2 c_1^H W_1 c_1 c_2^H \Sigma_2 c_2}{n_1 c_2^H W_2 c_2 c_1^H \Sigma_1 c_1} \sim dncF\left(2n_1, 2n_2, \frac{2c_1^H \delta_1 c_1}{c_1^H \Sigma_1 c_1}, \frac{2c_2^H \delta_2 c_2}{c_2^H \Sigma_2 c_2}\right)$$

Proof. By theorem 54,

$$c_1^H W_1 c_1 \sim \mathbf{C}W_1(n_1, c_1^H \Sigma_1 c_1, c_1^H \delta_1 c_1)$$

and

$$c_2^H W_2 c_2 \sim \mathbf{C}W_1(n_2, c_2^H \Sigma_2 c_2, c_2^H \delta_2 c_2)$$

Let $c_1^H W_1 c_1$ and $c_2^H W_2 c_2$ be positive. This is satisfied if Σ_1 and Σ_2 are positive definite. Then by theorem 53 we know

$$\frac{2c_1^H W_1 c_1}{c_1^H \Sigma_1 c_1} \sim \chi^2_{2n_1}\left(\frac{2c_1^H \delta_1 c_1}{c_1^H \Sigma_1 c_1}\right)$$

and

$$\frac{2c_2^H W_2 c_2}{c_2^H \Sigma_2 c_2} \sim \chi^2_{2n_2}\left(\frac{2c_2^H \delta_2 c_2}{c_2^H \Sigma_2 c_2}\right)$$

Taking the ratio of these terms, each divided by their respective degrees of freedom, gives us the doubly non-central F-distributed random variable F. □

Patil et al. (pp. 142-143) [204] catalog the doubly non-central F-distribution

$$dncF(\nu_1, \nu_2, \delta_1, \delta_2)$$

A random variable x has the doubly non-central F-distribution with parameters ν_1, ν_2, δ_1, and δ_2 if its probability density function is

$$f(x) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{1}{x}\left(\frac{\nu_1 x}{\nu_1 x + \nu_2}\right)^{\frac{\nu_1}{2}+j}\left(\frac{\nu_2}{\nu_1 x + \nu_2}\right)^{\frac{\nu_2}{2}+k}$$

$$\times \frac{\left(\frac{\delta_1}{2}\right)^j \left(\frac{\delta_2}{2}\right)^k}{j!k!B\left(\frac{\nu_1}{2}+j, \frac{\nu_2}{2}+k\right)} \exp\left(-\left\{\frac{\delta_1+\delta_2}{2}\right\}\right)$$

where $x > 0$. The numbers ν_1, ν_2 are positive integers, and $\delta_1, \delta_2 \geq 0$. The function

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx$$

$p > 0$, $q > 0$ is the beta function. A famous identity is $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ where Γ is the gamma function.

Under the null hypothesis $H_0 : c_1^H \Sigma_1 c_1 = c_2^H \Sigma_2 c_2$, the density function of the test statistic is

$$F = \frac{\frac{2c_1^H W_1 c_1}{2n_1 c_1^H \Sigma_1 c_1}}{\frac{2c_2^H W_2 c_2}{2n_2 c_2^H \Sigma_2 c_2}} = \frac{n_2 c_1^H W_1 c_1 c_2^H \Sigma_2 c_2}{n_1 c_2^H W_2 c_2 c_1^H \Sigma_1 c_1} = \frac{n_2 c_1^H W_1 c_1}{n_1 c_2^H W_2 c_2}$$

$$\sim dncF\left(2n_1, 2n_2, \frac{2c_1^H \delta_1 c_1}{c_1^H \Sigma_1 c_1}, \frac{2c_2^H \delta_2 c_2}{c_1^H \Sigma_1 c_1}\right)$$

This test allows you to compare special linear combinations of elements of the complex Wishart matrix. It is particularly useful if you want to compare any two elements on the main diagonal of the complex Wishart matrix. Establish

a null hypothesis (or default assumption) that the two sample special linear combinations $n_2 c_1^H W_1 c_1$ and $n_1 c_2^H W_2 c_2$ are really the same. Form the test statistic F. If the null hypothesis is rejected at your chosen α level of significance, then you conclude the alternate hypothesis $H_a : c_1^H \Sigma_1 c_1 \neq c_2^H \Sigma_2 c_2$. This test can be applied sequentially to discover the order of the underlying system.

Corollary 1 *Let* $W_1 \sim \mathbf{C}W_p(n, \Sigma_1, \delta_1)$ *and* $W_2 \sim \mathbf{C}W_q(m, \Sigma_2, \delta_2)$ *be independent complex Wishart random variables. Let* c_1 *be a* $p \times 1$ *vector of known fixed constants, and let* c_2 *be a* $q \times 1$ *vector of known fixed constants. Let* $W_1 = U_1 L_1^2 U_1^H$ *and* $W_2 = U_2 L_2^2 U_2^H$ *be the eigenvalue decompositions of* W_1 *and* W_2*, respectively. Then*

$$F = \frac{\frac{2c_1^H L_1^2 c_1}{2n_1 c_1^H U_1^H \Sigma_1 U_1 c_1}}{\frac{2c_2^H L_2^2 c_2}{2n_2 c_2^H U_2^H \Sigma_2 U_2 c_2}} = \frac{n_2 c_1^H L_1^2 c_1 c_2^H U_2^H \Sigma_2 U_2 c_2}{n_1 c_2^H L_2^2 c_2 c_1^H U_1^H \Sigma_1 U_1 c_1}$$

$$\sim dncF\left(2n_1, 2n_2, \frac{2c_1^H U_1^H \delta_1 U_1 c_1}{c_1^H U_1^H \Sigma_1 U_1 c_1}, \frac{2c_2^H U_2^H \delta_2 U_2 c_2}{c_2^H U_2^H \Sigma_2 U_2 c_2}\right)$$

Proof. From theorem 53 we know that

$$\frac{2c_1^H L_1^2 c_1}{c_1^H U_1^H \Sigma_1 U_1 c_1} \sim \chi^2_{2n_1}\left(\frac{2c_1^H U_1^H \delta_1 U_1 c_1}{c_1^H U_1^H \Sigma_1 U_1 c_1}\right)$$

and

$$\frac{2c_2^H L_2^2 c_2}{c_2^H U_2^H \Sigma_2 U_2 c_2} \sim \chi^2_{2n_2}\left(\frac{2c_2^H U_2^H \delta_2 U_2 c_2}{c_2^H U_2^H \Sigma_2 U_2 c_2}\right)$$

Taking the ratio, each divided by its respective degrees of freedom, gives us a doubly non-central F-distributed random variable. □

A common practical situation is where W_1 and W_2 arise out of a sampling of a common complex vector normal distribution with $\Sigma_1 = \Sigma_2$. For W_1 and W_2 obtained from independent samples, we know $W_1 \neq W_2$ with probability 1. Thus $U_1 \neq U_2$, and we have no hope of finding a distribution of the ratios independent from Σ.

6.1.2 Density of Eigenvalues of Sample Signal Plus Noise Covariance Matrix with Respect to Independent Sample Noise-Only Covariance Matrix

Theorem 7 *Let* $A_1 \sim \mathbf{C}W_p(m, \Sigma)$ *and* $B_1 \sim \mathbf{C}W_p(n, \Sigma)$ *where* $m, n > p$. *Then the joint density of the unordered roots of* $\det(A_1 - l^2 B_1) = 0$, *which we sort for testing, is*

$$f(L^2) = p! g(L^2)$$

$$= p! c_2 \left[\prod_{i=1}^{p} l_i^{2(m-p)} (1 + l_i^2)^{2(p-i-1)-(m+n)} \right] \left(1 + l_p^2\right)^2 \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dL^2)$$

where c_2 *is defined by*

$$c_2 = \frac{\pi^{p(p-1)} \mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}$$

This is a complexification of Anderson's theorem 13.2.2 (pp. 522-530) [26].

Discussion. In the context of signal processing, the matrix A_1 can be taken to be the sample covariance matrix of a deterministic signal plus random noise

measured during the time period of interest. The matrix B_1 can be the sample covariance matrix measured during a period when signal is assumed to be absent, but the noise remains the same as when the signal was measured. This theorem then gives the density of the eigenvalues of the sample signal-plus-noise covariance matrix with respect to an independently measured sample noise-only covariance matrix. The number of samples taken to estimate the covariance matrix are allowed to be different. Note that when B_1 is nonsingular, then the result is also the joint density of the roots of $\det(A_1 B_1^{-1} - l^2 I) = 0$ or variations on $\det(B_1^{-H/2} A_1 B_1^{-1/2} - l^2 I) = 0$. Thus, $A_1 B_1^{-1}$, $B_1^{-H/2} A_1 B_1^{-1/2}$, $B_1^{-1/2} A_1 B_1^{-H/2}$, $B_1^{-1/2} A_1 B_1^{-T/2}$, $B_1^{-T/2} A_1 B_1^{-1/2}$, or $B_1^{-1/2} A_1 B_1^{-1/2}$ (depending on the factorization theorem you use) has the interpretation of a generalized (signal-plus-noise) to noise ratio.

Compare the problem being treated here with the work on MUSIC by Schmidt [238]. For this theorem to apply, we need the population covariance matrix to be the same for the two sampled matrices under the null hypothesis. Deflate the sample covariance matrix of the signal-plus-noise by the eigenvalues thought to be due to a signal component ($l_i^2 - l_{\min}^2$). Call this deflated matrix A_1. If the noise-only component truly has been removed, then none of the eigenvalues of A_1 should be $\lambda_{\min}^2$. Under the null hypothesis that $H_0 : P = 0$, the A_1 here is the $S = APA^H + \lambda^2 S_0$ of Schmidt and the $l^2 B_1$ here is the $\lambda_{\min}^2 S_0$ of Schmidt. Note that no deflation is required for the initial detection

in absence of interference problem under the null hypothesis that there is no signal.

Proof. The proof presented here parallels the proof Anderson provided for the case of real variables where the proper Jacobians have been substituted and other appropriate modifications made. The strategy is to find the joint distribution of an intermediate matrix E and the roots of $\det[A-f(A+B)] = 0$ where f is a scalar. Then, observe that E and $F = \text{diag}(f_1, \cdots, f_p)$ are statistically independent. Find the density of E, and divide into the joint density to obtain the density of F. Change variables from F to $L^2 = \text{diag}(l_1^2, \cdots, l_p^2)$ to obtain the density of the roots of $\det(A_1 - l^2 B_1) = 0$. A difference from Anderson's work is my consideration of unordered versus ordered eigenvalues and the process by which the sorted eigenvalues are obtained. The algebra is straight forward, but the original choice of the changes of variables (which I copied from Anderson) that allows the solution to be obtained requires uncommon insight.

Begin with the general eigenvalue problem

$$A_1 x_1 = l^2 B_1 x_1 \tag{6.1}$$

where l^2 is the eigenvalue and x_1 is the associated eigenvector of A_1 with respect to (or in the metric of) B_1. The first simplification is a transformation to standardize the covariance to the identity matrix. Choose matrix C so that $C\Sigma C^H = I$. Let $A = CA_1C^H$ and $B = CB_1C^H$. By theorem 54, $A \sim$

$\mathbf{C}W_p(m, I)$ and $B \sim \mathbf{C}W_p(n, I)$. Also note that

$$\det(A - l^2 B) = \det(CA_1C^H - l^2 CB_1C^H) = \det[C(A_1 - l^2 B_1)C^H]$$

$$= \det(C)\det(A_1 - l^2 B_1)\det(C^H) = 0$$

implies $\det(A_1 - l^2 B_1) = 0$ when $\det(C) \neq 0$. So, the eigenvalues of A with respect to B are also the eigenvalues of A_1 with respect to B_1. If we premultiply $(A - l^2 B)x = 0$ by C^{-1} we observe that

$$0 = C^{-1}(A - l^2 B)x = (C^{-1}CA_1C^H - l^2 C^{-1}CB_1C^H)x = (A_1 - l^2 B_1)C^H x$$

Thus $x_1 = C^H x$ relates the eigenvectors.

Now for the trick. Consider the eigenvalues $\{f_i\}_1^p$ that satisfy

$$\det[A - f(A + B)] = 0$$

and the eigenvectors $\{y_i\}_1^p$ satisfying

$$[A - f_i(A + B)]y_i = 0 \tag{6.2}$$

Observe that when $f_i \neq 1$ this can be written $\left[A - \frac{f_i}{1-f_i}B\right] y_i = 0$. So, the eigenvalues of equation 6.1 are related by $l_i^2 = \frac{f_i}{1-f_i}$.

To proceed, establish an ordering on the eigenvalues $\{f_i\}_1^p$. This ordering will determine the ordering of associated eigenvectors $\{y_i\}_1^p$ to establish the matrix Y. As far as the derivation is concerned, it does not care what ordering you choose as long as it remains fixed for the remainder of the derivation. We

will observe from equation 6.3 that the order of the eigenvalues affects $f(F)$ by looking at the $\prod_{i<j}^{p}(f_i - f_j)^2$ term. This differs from Anderson's work

I have control over the ordering by the algorithms used to extract eigenvalues. Given any selection of eigenvalues and associated eigenvectors, I form our ordered set by sorting the eigenvalues. Thus, I actually want to end up with $p!F(f)$ since I actually am concerned with the case of the unordered eigenvalues which are then sorted. See section E.6. Okamoto [197] shows us that the probability of two roots being equal is zero. Define

$$F = \begin{pmatrix} f_1 & & \\ & \ddots & \\ & & f_p \end{pmatrix}$$

and $Y = (y_1, \cdots y_p)$. Then equation 6.2 can be rewritten as $AY = (A+B)YF$. Suppose that $Y^H(A+B)Y = I$. Then

$$Y^H AY = Y^H(A+B)YF = F$$

Multiplying by $(Y^H)^{-1}$ and Y^{-1} we see $A+B = Y^{-H}Y^{-1}$ and $A = Y^{-H}FY^{-1}$. The next simplification is to let $E = Y^{-1}$. Then

$$A + B = E^H E = G$$

$$A = E^H F E$$

$$B = E^H E - E^H F E = E^H(I - F)E$$

Now the known variables (A, B) are in terms of the variable I want (F) and a nuisance variable (E).

Recall that the eigenvectors are unique, apart from a scale factor. The restriction of

$$Y^H(A+B)Y = Y^H GY = I$$

determines Y up to a phase factor, where $Y = (y_1, \cdots, y_p)$. Consider

$$Y_* = (e^{i\theta_1}y_1, e^{i\theta_2}y_2, \cdots, e^{i\theta_p}y_p)$$

Then

$$Y_*^H GY_* = \begin{pmatrix} y_1^H Gy_1 & e^{-i(\theta_1-\theta_2)}y_1^H Gy_2 & \cdots & e^{-i(\theta_1-\theta_p)}y_1^H Gy_p \\ e^{i(\theta_1-\theta_2)}y_2^H Gy_1 & y_2^H Gy_2 & \cdots & e^{-i(\theta_2-\theta_p)}y_2^H Gy_p \\ \vdots & \vdots & \ddots & \vdots \\ e^{i(\theta_1-\theta_p)}y_p^H Gy_1 & e^{i(\theta_2-\theta_p)}y_p^H Gy_2 & \cdots & y_p^H Gy_p \end{pmatrix}$$

Because $y_i^H Gy_j = \delta_{ij}$, we know that

$$Y_*^H GY_* = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I_p$$

So, each eigenvector can be multiplied by a constant phase shift and still satisfy its orthogonality relation. From $E = Y^{-1}$ we know $EY = I$. Let

$$E = \begin{pmatrix} e^{i\omega_1}\epsilon_1 \\ \vdots \\ e^{i\omega_p}\epsilon_p \end{pmatrix}$$

Then

$$EY_* = \begin{pmatrix} e^{i(\omega_1+\theta_1)}\epsilon_1 y_1 & \cdots & e^{i(\omega_1+\theta_p)}\epsilon_1 y_p \\ e^{i(\omega_2+\theta_1)}\epsilon_2 y_1 & \cdots & e^{i(\omega_2+\theta_p)}\epsilon_2 y_p \\ \vdots & & \vdots \\ e^{i(\omega_p+\theta_1)}\epsilon_p y_1 & \cdots & e^{i(\omega_p+\theta_p)}\epsilon_p y_p \end{pmatrix}$$

$$= \begin{pmatrix} e^{i(\omega_1+\theta_1)} & & & \\ & e^{i(\omega_2+\theta_2)} & & \\ & & \ddots & \\ & & & e^{i(\omega_p+\theta_p)} \end{pmatrix}$$

To make $EY_* = I_p$ be satisfied, we merely choose $\omega_i = -\theta_i$. This defines the relationship uniquely between E and Y. However, we still have to fix the value of Y. The reason we have to fix it is so that the transformation between (A, B) and (E, F) is unique. For this reason, we choose ω_k so that $e^{i\omega_k}\epsilon_{k1} \geq 0$. We can always do this.

Now we want to evaluate the Jacobian of the transformation $J[(A, B) \rightarrow (E, F)]$. Let's summarize our transformations. $A = E^H F E$ implies

$$(dA) = (dE^H)FE + E^H(dF)E + E^H F(dE)$$

The transformation $G = E^H E$ implies

$$(dG) = (dE^H)E + E^H(dE)$$

Multiply by E^{-H} and E^{-1} to obtain $(d\bar{A})$ and $(d\bar{G})$ as follows.

$$(d\bar{A}) = (E^H)^{-1}(dA)E^{-1} = E^{-H}(dE^H)F + dF + F(dE)E^{-1}$$

$$(d\bar{G}) = (E^H)^{-1}(dG)E^{-1} = E^{-H}(dE^H) + (dE)E^{-1}$$

Let $(dW) = (dE)E^{-1}$. Then

$$(d\bar{A}) = (dW)^H F + dF + F(dW)$$

and

$$(d\bar{G}) = (dW)^H + (dW)$$

Stringing these all together, we find the joint distribution of (A, B) in terms of the joint distribution of (E, F).

$$\begin{aligned} f(A,B) &= f(E,F)J_1[(A,B) \to (A,G)] \\ &\times J_2[(A,G) \to (\bar{A},\bar{G})]J_3[(\bar{A},\bar{G}) \to (W,F)] \\ &\times J_4[(W,F) \to (E,F)] \end{aligned}$$

Evaluating J_1, we have the relations

$$\left.\begin{array}{rcl} A &=& A \\ G &=& A+B \end{array}\right\} \Leftrightarrow \left\{\begin{array}{rcl} A &=& A \\ B &=& G-A \end{array}\right.$$

$$\det \begin{pmatrix} \frac{\partial A}{\partial A} & \frac{\partial (G-A)}{\partial A} \\ \frac{\partial A}{\partial G} & \frac{\partial (G-A)}{\partial G} \end{pmatrix} = \det \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} = 1$$

where $\frac{\partial A}{\partial G}$ means the matrix formed by

$$\frac{\partial A}{\partial G} = \begin{pmatrix} \frac{\partial a_{11R}}{\partial g_{11R}} & \frac{\partial a_{11I}}{\partial g_{11R}} & \cdots & \frac{\partial a_{1pI}}{\partial g_{11R}} & \frac{\partial a_{21R}}{\partial g_{11R}} & \cdots & \frac{\partial a_{ppI}}{\partial g_{11R}} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial a_{11R}}{\partial g_{ppI}} & \cdots & \cdots & \cdots & \cdots & \cdots & \frac{\partial a_{ppI}}{\partial g_{ppI}} \end{pmatrix}$$

Evaluating J_2, we note that since A and $\bar{A}$ are functionally independent of G and $\bar{G}$, we can write

$$J[(A,G) \rightarrow (\bar{A},\bar{G})] = J[(dA,dG) \rightarrow (d\bar{A},d\bar{G})]$$

$$= J[dA \rightarrow d\bar{A}]J[dG \rightarrow d\bar{G}]$$

Since $A_1 \sim \mathbf{C}W_p(m,\Sigma)$ we know $A_1 = A_1^H$. From the transformation $A = CA_1C^H$, we also know $A = A^H$. Similarly, $B = B^H$. From theorem 38,

$$J[dA \rightarrow d\bar{A}] = \left|\det E^{-1}\right|^{-2p}$$

and

$$J[dG \rightarrow d\bar{G}] = \left|\det E^{-1}\right|^{-2p}$$

This means

$$J[(d\bar{A},d\bar{G}) \rightarrow (\bar{A},\bar{G})] = |\det E|^{4p}$$

Evaluating

$$J_3[(\bar{A},\bar{G}) \rightarrow (W,F)] = J[(d\bar{A},d\bar{G}) \rightarrow (dW,dF)] = \left|\frac{\partial(\bar{A},\bar{G})}{\partial(W,F)}\right|$$

is a bit trickier.

$$d\bar{a}_{ii} = df_i + f_i(dw_{ii})^* + f_i(dw_{ii}) = df_i + 2f_i\,\mathrm{Re}(dw_{ii})$$

$$d\bar{a}_{ij} = f_j(dw_{ji})^* + f_i(dw_{ij}) \quad , i < j$$

$$d\bar{g}_{ii} = 2\,\mathrm{Re}(dw_{ii})$$

$$d\bar{g}_{ij} = (dw_{ji})^* + (dw_{ij}) \quad , i < j$$

Note that since Y is not Hermitian, then neither are E or dW. However, G, dG, and $d\bar{G}$ are Hermitian by construction. We separate the real and imaginary parts to compute the Jacobian for the transformation of variables. Note that F is real. The subscripts R and I, to follow next, refer to the real and imaginary parts of the variables.

$$d\bar{a}_{ii} = df_i + 2f_i(dw_{iiR})$$

$$d\bar{a}_{ijR} = f_j(dw_{jiR}) + f_i(dw_{ijR}) \quad , i < j$$

$$d\bar{a}_{ijI} = -f_j(dw_{jiI}) + f_i(dw_{ijI}) \quad , i < j$$

$$d\bar{g}_{ii} = 2(dw_{iiR})$$

$$d\bar{g}_{ijR} = (dw_{jiR}) + (dw_{ijR}) \quad , i < j$$

$$d\bar{g}_{ijI} = -(dw_{jiI}) + (dw_{ijI}) \quad , i < j$$

To compute the Jacobian more easily, define two matrices M and N, as found in Anderson [26](p. 527).

$$M = \begin{pmatrix} f_1 I_{p-1} & & & \\ & f_2 I_{p-2} & & \\ & & \ddots & \\ & & & f_{p-1} I_1 \end{pmatrix}$$

and

$$N = \begin{pmatrix} f_2 & & & & & & \\ & \ddots & & & & & \\ & & f_p & & & & \\ & & & f_3 & & & \\ & & & & \ddots & & \\ & & & & & f_p & \\ & & & & & & \ddots \\ & & & & & & f_p \end{pmatrix}$$

Note that

$$\det(M - N) = \prod_{i<j}^{p}(f_i - f_j)$$

$$= (f_1 - f_2)(f_1 - f_3)\cdots(f_1 - f_p)(f_2 - f_3)\cdots(f_2 - f_p)\cdots(f_{p-1} - f_p)$$

We recognize this as a Vandermonde determinant. Graybill [95] p. 266., tells us that the corresponding Vandermonde matrix is

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ f_1 & f_2 & f_3 & \cdots & f_p \\ f_1^2 & f_2^2 & f_3^2 & \cdots & f_p^2 \\ f_1^3 & f_2^3 & f_3^3 & \cdots & f_p^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_1^{p-1} & f_2^{p-1} & f_3^{p-1} & \cdots & f_p^{p-1} \end{pmatrix}$$

We thus seek the determinant of the matrix given below for the linear change

of variables.

	$d\bar{a}_{ii}$	$d\bar{g}_{ii}$	$d\bar{a}_{ijR}\ (i<j)$	$d\bar{g}_{ijR}\ (i<j)$	$d\bar{a}_{ijI}\ (i<j)$	$d\bar{g}_{ijI}\ (i<j)$
df_i	I	0	0	0	0	0
dw_{iiR}	$2F$	$2I$	0	0	0	0
$dw_{ijR}\ (i<j)$	0	0	M	I	0	0
$dw_{ijR}\ (i>j)$	0	0	N	I	0	0
$dw_{ijI}\ (i<j)$	0	0	0	0	M	I
$dw_{ijI}\ (i>j)$	0	0	0	0	$-N$	$-I$

The determinant of this matrix is

$$\det\begin{pmatrix} I & 0 \\ 2F & 2I \end{pmatrix}\det\begin{pmatrix} M & I \\ N & I \end{pmatrix}\det\begin{pmatrix} M & I \\ -N & -I \end{pmatrix} = 2^p \det(M-N)\det(N-M)$$

$$= (-1)^{p(p-1)/2}2^p[\det(M-N)]^2$$

where

$$\det(N-M) = \det[(-1)(M-N)] = (-1)^{p(p-1)/2}\det(M-N)$$

The Jacobian is the absolute value of this determinant, which is

$$J_3[(d\bar{A}, d\bar{G}) \to (dW, dF)] = 2^p \prod_{i<j}^{p}(f_i - f_j)^2$$

Compare this with Equation (2.9) of [137].

Finally, consider the Jacobian

$$J_4[(W,F) \to (E,F)] = J_4[(dW,dF) \to (dE,dF)]$$

Recall that $(dW) = (dE)E^{-1}$, and that E is not a matrix with special structure. Therefore

$$J_4[(dW, dF) \to (dE, dF)] = J[(dW) \to (dE)] = \left|\det(E^{-1})\right|^{2p}$$

by taking the transpose of theorem 31.

Now put it all together. The joint density $f(A, B)$ of random variables A and B is given by

$$f(A,B) = f(E,F)J_1J_2J_3J_4 = f(E,F) \times 1 \times |\det E|^{4p} 2^p \prod_{i<j}^{p}(f_i - f_j)^2 |\det E|^{-2p}$$

$$= f(E,F) |\det E|^{2p} 2^p \prod_{i<j}^{p}(f_i - f_j)^2$$

Thus

$$J[(A,B) \to (E,F)] = |\det E|^{2p} 2^p \prod_{i<j}^{p}(f_i - f_j)^2$$

is the Jacobian of the transformation $A = E^H F E$ and $B = E^H(I-F)E$ where E is a $p \times p$ complex matrix without special structure and F is a $p \times p$ diagonal real matrix where the ordering of individual eigenvalues is fixed and arbitrary.

Now we introduce the dependence of the distribution of A and B. Recall that A and B are statistically independent and $A \sim \mathbf{C}W_p(m, I)$ and $B \sim \mathbf{C}W_p(n, I)$. Thus the joint density of A and B is

$$f(A,B) = \frac{1}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} |\det A|^{m-p} |\det B|^{n-p} \operatorname{etr}[-(A+B)]$$

Therefore the joint density of E and F is

$$g(E,F) = \frac{1}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} \left|\det(E^H F E)\right|^{m-p} \left|\det[E^H(I-F)E]\right|^{n-p} \times$$

$$\times \operatorname{etr}[-(E^H E)] 2^p |\det E|^{2p} \prod_{i<j}^{p} (f_i - f_j)^2$$

Examine the determinants. We want to be able to rewrite $g(E, F)$ as $g_1(E)g_2(F)$ with $g_1(E)$ of a form that is easy to integrate. This would leave us with a function only of F.

$$\det(E^H F E) = \det(E^H)\det(F)\det(E) = \det(F)\det(E^H E) = \det(E^H E)\prod_{i=1}^{p} f_i$$

$$\det[E^H(I - F)E] = \det(I - F)\det(E^H E) = \det(E^H E)\prod_{i=1}^{p}(1 - f_i)$$

$$|\det(E)|^2 = \det(E^H E)$$

Substituting into $g(E, F)$ we obtain

$$g(E, F) = \frac{2^p}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)}[\det(E^H E)]^{m+n-p} \times$$

$$\times \left[\prod_{i=1}^{p} f_i^{m-p}(1 - f_i)^{n-p}\right]\left[\prod_{i<j}^{p}(f_i - f_j)^2\right] \operatorname{etr}(-E^H E)$$

By the factorization theorem, we know that E and F are independent. Notice that $\det(E^H E)$ and $\operatorname{etr}(-E^H E)$ are "generalized even" functions of E (see definition 84).

If we had not restricted $e_{i1} \geq 0$ then we would recognize that

$$\int [\det(E^H E)]^{m+n-p} \frac{1}{\pi^{p^2}} \operatorname{etr}(-E^H E)(dE)$$

is the expected value of $[\det(E^H E)]^{m+n-p}$ when E is distributed by $\mathbf{C}N_{p,p}(0, I, I)$.

Consider

$$E^H E = (e^{-i\omega_1}\epsilon_1^H, \cdots, e^{-i\omega_p}\epsilon_p^H) \begin{pmatrix} e^{i\omega_1}\epsilon_1 \\ \vdots \\ e^{i\omega_p}\epsilon_p \end{pmatrix} = \epsilon_1^H\epsilon_1 + \cdots + \epsilon_p^H\epsilon_p$$

Thus $E^H E$ is invariant as a function of $\{\omega_k\}_1^p$. Therefore, when

$$[\det(E^H E)]^a \operatorname{etr}(-E^H E)$$

is integrated over E where we restrict ω_k so that $e^{i\omega_k}\epsilon_{k1} \geq 0$, we get the same answer as when we integrate

$$\left(\frac{1}{2\pi}\right)^p [\det(E^H E)]^a \operatorname{etr}(-E^H E)$$

without restriction on ω_k. For each k, we observe

$$\int_0^{2\pi} d\omega_k = 2\pi, \quad 1 \leq k \leq p$$

Thus we know to consider

$$\pi^{p^2}\left(\frac{1}{2\pi}\right)^p \int [\det(E^H E)]^{m+n-p} \left(\frac{1}{\pi^{p^2}}\right) \operatorname{etr}(-E^H E)(dE)$$

Now, since we want to evaluate this integral, we consider E as being distributed as $\mathbf{C}N_{p,p}(0, I, I)$. When this is true, we know from the definition of the complex Wishart distribution that

$$G = E^H E \sim \mathbf{C}W_p(p, I)$$

By theorem 79, we know that

$$\mathcal{E}\{|G|^{m+n-p}\} = \frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(p)}$$

Thus the integral above gives us the scaled expected value

$$\frac{\pi^{p^2}\left(\frac{1}{2\pi}\right)^p \mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(p)} = \frac{2^{-p}\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(p)}$$

When we consider the integration when e_{i1} is not restricted, we find

$$\int \frac{2^{-p}\pi^{p(p-1)}[\det(E^HE)]^{m+n-p}\pi^{-p^2}\operatorname{etr}(-E^HE)(dE)}{2^{-p}\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)/\mathbf{C}\Gamma_p(p)}$$

$$= \int \frac{\mathbf{C}\Gamma_p(p)[\det(E^HE)]^{m+n-p}\operatorname{etr}(-E^HE)(dE)}{\pi^{p^2}\mathbf{C}\Gamma_p(m+n)} = \int f_1(E)(dE)$$

$f_1(E)$ is a density function when e_{i1} is not restricted. We want the density of E when $e_{i1} \geq 0$. Recall that since

$$[\det(E^HE)]^a \operatorname{etr}(-E^HE)$$

is a generalized even function, we multiplied our function by $\left(\frac{1}{2\pi}\right)^p$ and extended the region of integration. To recover the desired density function, we want $f(E) = (2\pi)^p f_1(E)$. Thus

$$f(E) = \frac{2^p\mathbf{C}\Gamma_p(p)[\det(E^HE)]^{m+n-p}\operatorname{etr}(-E^HE)}{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}(dE)$$

is the density of E when $e_{i1} \geq 0$.

Since E and F are statistically independent, we find $f(F)$ by dividing the joint density $g(E,F)$ by $f(E)$ as follows.

$$f(F) = \frac{g(E,F)}{f(E)}$$

$$= \frac{\frac{2^p}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)}[\det(E^H E)]^{m+n-p}\,\mathrm{etr}(-E^H E)}{\frac{2^p \mathbf{C}\Gamma_p(p)}{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}[\det(E^H E)]^{m+n-p}\,\mathrm{etr}(-E^H E)} \times$$

$$\times \left[\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right]\left[\prod_{i<j}^{p}(f_i - f_j)^2\right]$$

Simplifying, we get

$$f(F) = \frac{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\left[\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right]\left[\prod_{i<j}^{p}(f_i - f_j)^2\right] \quad (6.3)$$

where $1 > f_i > 0$ for each i since $f = \frac{l^2}{1+l^2}$ in the original problem and the ordering of the $\{f_i\}$ is as fixed earlier in the derivation. Strict inequality is specified because Okamoto [197] showed that the probability of two sample eigenvalues being equal is zero.

The density of l_i^2 is obtained from $f(F)$ using

$$f_i = \frac{l_i^2}{1+l_i^2}$$

Following Anderson (p. 530) [26], we note that:

$$\frac{df_i}{dl_i^2} = \frac{d}{dl_i^2} l_i^2(1+l_i^2)^{-1} = (1+l_i^2)^{-1} - l_i^2(1+l_i^2)^{-2}$$

$$= [(1+l_i^2) - l_i^2](1+l_i^2)^{-2} = \frac{1}{(1+l_i^2)^2}$$

Thus

$$J[F \to L^2] = \prod_{i=1}^{p}\left(\frac{1}{1+l_i^2}\right)^2$$

We also note that

$$(f_i - f_j)^2 = \left[\frac{l_i^2}{1+l_i^2} - \frac{l_j^2}{1+l_j^2}\right]^2 = \left[\frac{l_i^2 - l_j^2}{(1+l_i^2)(1+l_j^2)}\right]^2$$

and

$$1 - f_i = 1 - \frac{l_i^2}{1 + l_i^2} = \frac{1}{1 + l_i^2}$$

Note that

$$\prod_{i<j}^{p} \frac{l_i^2 - l_j^2}{(1 + l_i^2)(1 + l_j^2)} = \frac{\prod_{i<j}^{p} (l_i^2 - l_j^2)}{(1 + l_p^2)^{p-1} \prod_{i=1}^{p-1} (1 + l_i^2)^i}$$

Performing the change of variables, we get

$$g(L^2) = f\left(\frac{l_i^2}{1 + l_i^2}\right) J(F \to L^2)$$

$$= C \left[\prod_{i=1}^{p} \left(\frac{l_i^2}{1 + l_i^2}\right)^{m-p} \left(\frac{1}{1 + l_i^2}\right)^{n-p}\right] \left[\prod_{i<j}^{p} \frac{l_i^2 - l_j^2}{(1 + l_i^2)(1 + l_j^2)}\right]^2 \times$$

$$\times \left[\prod_{i=1}^{p} \left(\frac{1}{1 + l_i^2}\right)^2\right] (dL^2)$$

$$= C \left[\prod_{i=1}^{p} l_i^{2(m-p)} \left(\frac{1}{1 + l_i^2}\right)^{m+n-2p}\right] \left[\frac{1}{(1 + l_p^2)^{2(p-1)} \prod_{i=1}^{p-1} (1 + l_i^2)^{2i}}\right] \times$$

$$\times \left[\prod_{i=1}^{p} \left(\frac{1}{1 + l_i^2}\right)^2\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dL^2)$$

At this point we note that $2p - m - n - 2 - 2i - 2(p - i - 1) - (m + n)$. Then

$$g(L^2) = C \left[\prod_{i=1}^{p} l_i^{2(m-p)} \left(1 + l_i^2\right)^{2(p-i-1)-2(m+n)}\right] \left(1 + l_p^2\right)^2$$

$$\times \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dL^2)$$

where

$$C = \frac{\pi^{p(p-1)} \Gamma_p(m + n)}{\Gamma_p(m) \Gamma_p(n) \Gamma_p(p)}$$

Putting it all together, the joint density of the sample eigenvalues $\{l_i^2\}$ which satisfy $\det(A_1 - l^2 B_1) = 0$ where $A_1 \sim \mathbf{C}W_p(m, \Sigma)$ and $B_1 \sim \mathbf{C}W_p(n, \Sigma)$ is

$$g(L^2) = \frac{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\prod_{i=1}^{p} \frac{l_i^{2(m-p)}(1+l_i^2)^{2(p-i)}}{(1+l_i^2)^{m+n+2}}\right] \left(1+l_p^2\right)^2 \times \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dL^2)$$

where $L^2 = \operatorname{diag}(l_1^2, \cdots, l_p^2)$ and

$$\prod_{i<j}^{p}(l_i^2 - l_j^2) = (l_1^2 - l_2^2)(l_1^2 - l_3^2)(l_2^2 - l_3^2)(l_1^2 - l_4^2)(l_2^2 - l_4^2)(l_3^2 - l_4^2)\cdots(l_{p-1}^2 - l_p^2)$$

A few notational simplifications can be made. Since the l_i^2 are the generalized eigenvalues of A with respect to B, if B is nonsingular we observe that if $\det(A - l^2 B) = 0$ then $\det(AB^{-1} - l^2 I) = 0$. Thus $\{l_i^2\}_1^p$ are the eigenvalues of AB^{-1}. By theorem 115, there is a unitary matrix U so that $L^2 = U^H AB^{-1}U$. Also,

$$L^2 + I = U^H AB^{-1}U + I = U^H AB^{-1}U + U^H IU = U^H(AB^{-1} + I)U$$

We observe

$$\prod_{i=1}^{p} l_i^2 = \det(L^2) = \det(AB^{-1})$$

and

$$\prod_{i=1}^{p}(1+l_i^2) = \det(L^2 + I) = \det(AB^{-1} + I)$$

Substituting into our density function yields

$$g(L^2) = \frac{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \frac{[\det(AB^{-1})]^{m-p}[\det(AB^{-1}+I)]^{2(p-1)}}{[\det(AB^{-1}+I)]^{m+n}}$$

$$\times \left[(1+l_p^2)^2\right] \left[\frac{\prod_{i<j}^{p}(l_i^2-l_j^2)^2}{\prod_{i=1}^{p}(1+l_i^2)^{2i}}\right] (dL^2)$$

The final form, $f(L^2) = p!g(L^2)$, accounts for the procedure of sorting the randomly selected unordered eigenvalues. This final form is the density function for the set of sorted sample eigenvalues. This is the appropriate density for deriving the density of test statistics based on these sample eigenvalues. This is the starting point for determining the number of significant sources for the MUSIC algorithm.

6.2 Tests Based on One Set of Samples

In general, the distribution of a test statistic formed from dependent random variables will be more complicated to evaluate than the distribution of a test statistic formed from independent random variables. When the random variables are dependent, then you must know or assume information about that dependency.

6.2.1 Tests that Require Specifying the Population Covariance Matrix

In this section, I am interested in finding the density function of the ratio of linear combinations of sample eigenvalues of a $p \times p$ complex Wishart matrix.

The test considered here requires that the population covariance matrix Σ be specified. This differs from a test where Σ might cancel out in the forming of the density function of the test statistic. As an intermediate result, consider $W \sim \mathbf{C}W_2(n, \Sigma)$ where $W = \begin{pmatrix} a & 0 \\ o & b \end{pmatrix}$. Then, apply the expression for the density function for the complex Wishart distribution.

Theorem 8 *Let* $W \sim \mathbf{C}W_2(n, \Sigma)$ *where* $W = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ *and* $n > 1$. *Let* $x = \frac{a}{b}$. *Then the density function of* x *is given by*

$$g(x) = \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{n-2}}{\pi\beta[(n-1),(n-2)]} \frac{|x|^{n-2}}{(x\Sigma_{22} + \Sigma_{11})^{2(n-1)}} \tag{6.4}$$

where $\beta(\bullet, \bullet)$ *is the beta function.*

Proof. We know that the density $f(W)$ is given by

$$f(W) = \frac{\left|\det \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\right|^{n-2} \operatorname{etr}\left[-\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\right]}{\left[\det \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]^n \mathbf{C}\Gamma_2(n)} \tag{6.5}$$

When we evaluate the determinants, we obtain Equation 6.6.

$$f(W) = \frac{|ab|^{n-2} \operatorname{etr}\left[-\left(\frac{1}{\Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{12}}\right) \begin{pmatrix} \Sigma_{22} & -\Sigma_{12} \\ -\Sigma_{21} & \Sigma_{11} \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\right]}{(\Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{12})^n \mathbf{C}\Gamma_2(n)} \tag{6.6}$$

$$= \frac{|ab|^{n-2}\,\mathrm{etr}\left[-\left(\frac{1}{\Sigma_{11}\Sigma_{22}-\Sigma_{21}\Sigma_{12}}\right)\begin{pmatrix} a\Sigma_{22} & -b\Sigma_{12} \\ -a\Sigma_{21} & b\Sigma_{11}\end{pmatrix}\right]}{(\Sigma_{11}\Sigma_{22}-\Sigma_{21}\Sigma_{12})^{n}\mathbf{C}\Gamma_2(n)} \tag{6.7}$$

$$= \frac{|ab|^{n-2}\exp\left(-\frac{a\Sigma_{22}+b\Sigma_{11}}{\Sigma_{11}\Sigma_{22}-\Sigma_{21}\Sigma_{12}}\right)}{(\Sigma_{11}\Sigma_{22}-\Sigma_{21}\Sigma_{12})^{n}\mathbf{C}\Gamma_2(n)} \tag{6.8}$$

Notice that since $\Sigma = \Sigma^H$, we have

$$f(W) = \frac{|ab|^{n-2}\exp\left(-\frac{a\Sigma_{22}+b\Sigma_{11}}{\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2}\right)}{\left(\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2\right)^{n}\mathbf{C}\Gamma_2(n)} = f(a,b) \tag{6.9}$$

We can make the change of variables $x = \frac{a}{b}$ and $y = b$. The inverse relations are given by $b = y$ and $a = xy$. The Jacobian of the transformation is given by

$$J = \det\begin{pmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial y} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial y}\end{pmatrix} = \det\begin{pmatrix} \frac{\partial xy}{\partial x} & \frac{\partial xy}{\partial y} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial y}\end{pmatrix} = \det\begin{pmatrix} y & x \\ 0 & 1\end{pmatrix} = y$$

Then

$$f(a,b) = f(xy,y)\,|J| = g(x,y) = g\left(\frac{a}{b},b\right)$$

$$= \frac{|xy^2|^{n-2}\exp\left[-\frac{xy\Sigma_{22}+y\Sigma_{11}}{\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2}\right]}{\left(\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2\right)^{n}\mathbf{C}\Gamma_2(n)}\,|y| \tag{6.10}$$

$$= \frac{|x|^{n-2}\,|y|^{2n-3}\exp\left[-\frac{y(x\Sigma_{22}+\Sigma_{11})}{\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2}\right]}{\left(\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2\right)^{n}\mathbf{C}\Gamma_2(n)} \tag{6.11}$$

Now integrate out y to find $g(\frac{a}{b})$. Since $b \geq 0$, we know that $0 \leq y \leq \infty$. Concentrating just on the function of y, evaluate the integral

$$I = \int_Y |y|^{2n-3}\exp(-hy)\,dy \tag{6.12}$$

where h is a function of x which is held constant during the integration, given by

$$h = \frac{x\Sigma_{22} + \Sigma_{11}}{\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2}$$

By corollary 48 we know

$$I = \int_0^\infty |y|^{2n-3} \exp(-hy)\, dy = (2n-3)!\, h^{-2(n-1)}$$

This gives us our expression for $g(x)$, where we recall that the beta function is $\beta(m,n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$ and $x \geq 0$.

$$g(x) = \frac{|x|^{n-2}\,(2n-3)!\,\left(\frac{x\Sigma_{22}+\Sigma_{11}}{\Sigma_{11}\Sigma_{22}-|\Sigma_{12}|^2}\right)^{-2(n-1)}}{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^n \mathbf{C}\Gamma_2(n)}$$

$$= \frac{|x|^{n-2}\,(2n-3)!}{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^n \pi(n-1)!(n-2)!} \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{2(n-1)}}{(x\Sigma_{22} + \Sigma_{11})^{2(n-1)}}$$

$$g(x) = \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{n-2}}{\pi\beta[(n-1),(n-2)]} \frac{|x|^{n-2}}{(x\Sigma_{22} + \Sigma_{11})^{2(n-1)}}$$

Theorem 9 *Let* $W \sim \mathbf{C}W_2(n,\Sigma)$ *where* $W = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, $a > 0$, $b > 0$, *and* $n > 1$. *Let* $x = \frac{a}{b}$. *Then the cumulative distribution function is given by*

$$F(x) = \Pr\{x > c\} = \int_c^\infty g(x)dx$$

$$= \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{n-2}}{\pi\beta\left[(n-1)(n-2)\right]} \left(\frac{1}{\Sigma_{22}^{n-1}}\right) \sum_{k=0}^{n-2} \left(\frac{1}{n+k-1}\right) \frac{(-\Sigma_{11})^k}{(c\Sigma_{22} + \Sigma_{11})^{n+k-1}}$$

This theorem is supplied by me.

Proof. By theorem 8 the density function is

$$g(x) = \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{n-2}}{\pi\beta\left[(n-1)(n-2)\right]} \frac{|x|^{n-2}}{(x\Sigma_{22} + \Sigma_{11})^{2(n-1)}}$$

Since $a > 0$ and $b > 0$, we know x is real and positive. This permits us to drop the absolute values signs. Then, apply theorem 144. We note that since $n > 1$ we can use the solution for $k < p - 1$ in that theorem statement. Then

$$\int_c^\infty \frac{x^{n-2}}{(x\Sigma_{22} + \Sigma_{11})^{2(n-1)}} dx$$

$$= \left(\frac{1}{\Sigma_{22}}\right)^{n-1} \sum_{k=0}^{n-2} \binom{n-2}{k} (-\Sigma_{11})^k \left(\frac{1}{1-n-k}\right) (x\Sigma_{22} + \Sigma_{11})^{1-n-k} \Bigg|_c^\infty$$

$$= -\left(\frac{1}{\Sigma_{22}}\right)^{n-1} \sum_{k=0}^{n-2} \binom{n-2}{k} (-\Sigma_{11})^k \left(\frac{1}{1-n-k}\right) (c\Sigma_{22} + \Sigma_{11})^{1-n-k}$$

$$= \left(\frac{1}{\Sigma_{22}}\right)^{n-1} \sum_{k=0}^{n-2} \binom{n-2}{k} \left(\frac{1}{n+k-1}\right) \frac{(-\Sigma_{11})^k}{(c\Sigma_{22} + \Sigma_{11})^{n+k-1}}$$

The full answer is

$$F(x) = \Pr\{x > c\}$$

$$= \frac{\left(\Sigma_{11}\Sigma_{22} - |\Sigma_{12}|^2\right)^{n-2}}{\pi\beta\left[(n-1)(n-2)\right]} \left(\frac{1}{\Sigma_{22}}\right)^{n-1} \sum_{k=0}^{n-2} \binom{n-2}{k} \left(\frac{1}{n+k-1}\right) \frac{(-\Sigma_{11})^k}{(c\Sigma_{22} + \Sigma_{11})^{n+k-1}}$$

□

Corollary 2 *Let $W = UL^2U^H$ be the eigenvalue decomposition of $W \sim CW_p(n, \Sigma, \delta)$. Then $U^HWU = L^2$ is distributed according to*

$$\mathbf{C}W_p(n, U^H\Sigma U, U^H\delta U)$$

Note that U consists of the eigenvectors of W, which generally are *not* the eigenvectors of Σ. Thus Σ is not diagonalized by this transformation. $\mathbf{C}W_p(n, U^H\Sigma U, U^H\delta U)$ is the distribution of the sample eigenvalues of $W = X^HX$ where X has the complex matrix normal distribution $\mathbf{C}N_{n,p}(\mu, I, \Sigma)$. This is an application of theorem 54.

Corollary 3 *Suppose that we define a* $p \times 2$ *matrix* $C = (c_1, c_2)$ *such that the Hadamard product* $c_1 \otimes c_2 = 0$. *Then look at* C^HL^2C. *Suppose*

$$C = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = (c_1, c_2). \tag{6.13}$$

Then

$$C^HL^2C = \begin{pmatrix} l_1^2 + l_2^2 & 0 \\ 0 & l_4^2 + l_5^2 \end{pmatrix}$$

This is now distributed according to the second order complex Wishart distribution,

$$C^HL^2C \sim \mathbf{C}W_2(n, C^HU^H\Sigma UC, C^HU^H\delta UC)$$

If we know that the mean of the complex multivariate normal distribution is zero then $\delta = 0$, *and the third term in the distribution notation is omitted.*

Let $a = c_1^H L^2 c_1 = l_1^2 + l_2^2$ and $b = c_2^H L^2 c_2 = l_4^2 + l_5^2$. Then Equation 6.4 is the probability density function of the ratio of two disjoint linear combinations of eigenvalues of the sample covariance matrix where the underlying data sample of size n is distributed according to the zero-mean vector complex normal distribution $\mathbf{C}N_p(0, \Sigma)$. The subscripted values Σ_{11}, Σ_{12}, and Σ_{22} refer to the partitions of $C^H U^H \Sigma U C$ and not to partitions of the original population covariance matrix Σ.

Although this density function is for a simple test statistic

$$T_{13} = \frac{c_1^H L^2 c_1}{c_2^H L^2 c_2}$$

interpreting the statistic is not as simple as a modification of this statistic. Instead, consider the average of sample eigenvalues that make up a and b. Let m_1 be the number of sample eigenvalues picked by $c_1^H L^2 c_1$, and let m_2 be the number of sample eigenvalues picked by $c_2^H L^2 c_2$. Look at the test statistic

$$T_{14} = \frac{\frac{1}{m_1} c_1^H L^2 c_1}{\frac{1}{m_2} c_2^H L^2 c_2} = \frac{m_2 c_1^H L^2 c_1}{m_1 c_2^H L^2 c_2}$$

When all the sample eigenvalues are equal then the test statistic $T_{14} = 1$. The further T_{14} is away from 1, the averages of the sets of sample eigenvalues are more different. Thus, when T_{14} is very close to 1, expect that saying "the corresponding population eigenvalues are all equal" to have a small chance of being in error.

So that the computation of the density of the test statistic is not altered, account for the weighting done in the averaging process in the vectors c_1 and

c_2. Define a $p \times 2$ matrix $C = (c_1, c_2)$ such that the elements of c_k pick out the sample eigenvalues of interest, and the nonzero entries are the reciprocal of the number of sample eigenvalues extracted. Let the sets of sample eigenvalues chosen be disjoint. Then the Hadamard product $c_1 \otimes c_2 = 0$. Look at $C^H L^2 C$. Suppose

$$C = \begin{pmatrix} \frac{1}{\sqrt{m_1}} & 0 \\ \frac{1}{\sqrt{m_1}} & 0 \\ 0 & 0 \\ 0 & \frac{1}{\sqrt{m_2}} \\ 0 & \frac{1}{\sqrt{m_2}} \end{pmatrix} = (c_1, c_2). \tag{6.14}$$

Then

$$C^H L^2 C = \begin{pmatrix} (l_1^2 + l_2^2)/m_1 & 0 \\ 0 & (l_4^2 + l_5^2)/m_2 \end{pmatrix}$$

For this simple example, $m_1 = m_2 = 2$. Then $T_{14} = \frac{c_1^H L^2 c_1}{c_2^H L^2 c_2}$.

Now we know $x \geq 1$ since $\hat{\lambda}_{m_1} > \hat{\lambda}_{p-m_2+1}$. We are interested in testing if x is significantly greater than 1. Let $c > 1$ be some critical threshold we want to test against. If c is a detection threshold, then this is the probability of detection for a signal-to-noise ratio of x in "linear" units. For $SNR = d = 10 \log x$, then $x = 10^{d/10} = 10^{SNR/10}$ for SNR given in dB. SNR here is interpreted in the sense of [49] with noise measured in the same bandwidth signal is measured in.

6.2.2 $F^{\#}$ in MUSIC

In this section, an statistic with an F-distribution is derived for examining the decomposition of a received signal-plus-noise plus noise data set as constructed with the MUltiple SIgnal Classification (MUSIC) technique. The motivation for this is the work by Schmidt (1986) [238] and Wax (1991) [282]. I will draw most heavily from the second paper to develop the assumptions. All of the distributional work is provided by me.

Let $a(\theta)$ be a $p \times 1$ steering vector for an array of p sensors in an array having a fixed arbitrary geometry. Assume that the signals from q sources arrive at the array. Each source s_i is coherently processed by a corresponding linear beamforming function $a(\theta_i)$. Assume that the stochastic signals are independent from the noise received at each sensor. This is the signal-aligned beamformer case of Monzingo and Miller [185].

Let

$$A_{[q]}(\Theta) = [a(\theta_1), a(\theta_2), \cdots, a(\theta_q)]$$

Then $A_{[q]}(\Theta)$ is a deterministic $p \times q$ complex matrix whose column vectors span the vector space which contains the signals. Note that some of the noise is also in this space. Let $s(t)$ be a $q \times 1$ vector of the signals at the array reference point at time t. Let $n(t)$ be a $p \times 1$ vector of the random noise appearing at time t at each sensor. Then, let the beamformer output for signals arriving at

the array reference point at time t be given by

$$x(t) = A_{[q]}(\Theta)\, s(t) + n(t)$$

Obtain independent samples from the sensors at M different times,

$$(t_1, t_2, \cdots, t_M)$$

Let

$$S_{q\times M} = [s(t_1), \cdots, s(t_M)]$$

$$N_{p\times M} = [n(t_1), \cdots, n(t_M)]$$

$$X_{p\times M} = [x(t_1), \cdots, x(t_M)]$$

Then

$$X = A_{[q]}(\Theta)\, S + N$$

To complete the problem description, we need to know something about the distributions of S and N. Let the noise matrix N be distributed according to the matrix complex normal distribution having a mean of zero and row covariance Σ. Thus $N \sim \mathbf{C}N_{p\times M}(0, \Sigma_{p\times p}, I_M)$. Let the signal matrix S be distributed according to the matrix complex normal distribution having a mean of zero and row covariance R. This is stated as $S \sim \mathbf{C}N_{q\times M}(0, R_{q\times q} I_M)$. By theorem 41, we know

$$A_{[q]}(\Theta)_{p\times q}\, S_{q\times M} \sim \mathbf{C}N_{p,M}\left(0, ARA^H, I_M\right)$$

We sum the independent random variables according to theorem 48 to get

$$X = A_{[q]}(\Theta)\, S + N \sim \mathbf{C}N_{p,M}\left(0 + 0, \Sigma + ARA^H, I_M + I_M\right)$$

$$= \mathbf{C}N_{p,M}\left(0, \Sigma + ARA^H, 2I_M\right)$$

Note that the presence of the scalar 2 came from the sum of the two column covariance matrices which each are I_M. By lemma 13 we know that the row and column covariance matrices are not unique. This is good here because scalar multiples commute between these matrices. Therefore,

$$X \sim \mathbf{C}N_{p,M}\left(0, 2(\Sigma + ARA^H), I_M\right)$$

We will need the column covariance matrix to be an identity matrix to form a complex Wishart distributed random variable.

The next step in MUSIC is to find an orthonormal basis for the space spanned by the beamformer when adjusted to coherently process signals with parameters $\theta_1, \cdots, \theta_q$ which we usually associate with direction (but this association does not strictly have to hold). We find the required orthonormal basis by performing a QR decomposition of $A_{[q]}(\Theta)$. Recall that Q is the orthonormal matrix obtainable by the inner product version of the Gram-Schmidt process. Because the symbol R is already in use, let the triangular matrix factor from the QR decomposition be T. They by proposition 67, $A = QT$ where $Q^HQ = I_q$ and T is an upper triangular $q \times q$ matrix with positive real elements on the diagonal. Alternately, we can apply proposition 71 to get $A = QT$ where T is a lower triangular $q \times q$ matrix.

The matrix Q is called *subunitary*, and it forms an orthonormal basis for the space spanned by the columns of the signal-directed beamformer $A_{[q]}(\Theta)$. We

can continue to construct vectors orthonormal to Q and mutually orthonormal to each other until we have a set of p vectors. These last $(p-q)$ vectors form an orthonormal basis for the space orthogonal to the space spanned by the columns of A. Then we have the $p \times p$ orthonormal matrix

$$G = [Q, V]$$

This matrix G is special. Observe that

$$G^H_{p\times p} X_{p\times M} = \begin{bmatrix} Q^H_{q\times p} \\ \\ V^H_{(p-q)\times p} \end{bmatrix} X_{p\times M} = \begin{bmatrix} (Q^H X)_{q\times M} \\ \\ (V^H X)_{(p-q)\times M} \end{bmatrix} = \begin{bmatrix} Y \\ \\ Z \end{bmatrix}$$

This has partitioned the rows of X into disjoint matrices Y and Z. Let us examine these more closely, recalling that $X = AS + N$.

Since Q spans the same space spanned by A, then all of the signal component lies in the space spanned by Q and none of the signal component lies in the space spanned by V. All of the data in the space spanned by V consists of only noise. It is in this sense that the space spanned by Q is called the "signal subspace", and the space spanned by V is called the "noise subspace". Caution: these designations are basically useful tags, but you must remember that *some* of the noise is also in the space spanned by Q. *Not all* of the noise is in the space spanned by V. This is consistent with hearing noise when we listen to a beamformer output.

Let us find the distributions for Y and Z. Once again, apply theorem 41

to get

$$G^H X \sim \mathbf{C}N_{p,M}\left(0, 2G^H(\Sigma + ARA^H)G, I_M\right)$$

Take a closer look at the column covariance matrix.

$$2G^H(\Sigma + ARA^H)G = 2\begin{pmatrix} Q^H \\ V^H \end{pmatrix}\left(\Sigma + ARA^H\right)\begin{pmatrix} Q & V \end{pmatrix}$$

$$= 2\begin{pmatrix} Q^H(\Sigma + ARA^H)Q & Q^H(\Sigma + ARA^H)V \\ V^H(\Sigma + ARA^H)Q & V^H(\Sigma + ARA^H)V \end{pmatrix}$$

Since all of the signal is projected by Q onto the space spanned by A, we know that $V^H = 0$ and $A^H V = 0$. Thus we get

$$2G^H(\Sigma + ARA^H)G = 2\begin{pmatrix} Q^H(\Sigma + ARA^H)Q & 0 \\ 0 & V^H\Sigma V \end{pmatrix}$$

where $Q^H_{p\times p}(\Sigma + ARA^H)_{p\times p}Q_{p\times q}$ is $q \times q$ and $V^H_{(p-q)\times p}\Sigma_{p\times p}V_{p\times(p-q)}$ is $(p-q) \times (p-q)$. Therefore

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix} \sim \mathbf{C}N_{p,M}\left(0, 2\begin{pmatrix} Q^H(\Sigma + ARA^H)Q & 0 \\ 0 & V^H\Sigma V \end{pmatrix}, I_M\right)$$

By theorem 43, Y and Z are independent. Since

$$\begin{pmatrix} I_q & 0_{q\times(p-q)} \end{pmatrix}\begin{pmatrix} Y_{q\times M} \\ Z_{(p-q)\times M} \end{pmatrix} = Y_{q\times M}$$

we again apply theorem 41 to show

$$Y \sim \mathbf{C}N_{q,M}\left(0, 2Q^H(\Sigma + ARA^H)Q, I_M\right)$$

Similarly,

$$\begin{pmatrix} 0_{(p-q)\times q} & I_{p-q} \end{pmatrix} \begin{pmatrix} Y_{q\times M} \\ Z_{(p-q)\times M} \end{pmatrix} = Z_{(p-q)\times M}$$

is distributed as

$$Z \sim \mathbf{C}N_{(p-q),M}\left(0, 2V^H \Sigma V, I_M\right)$$

To form complex Wishart random variables, the underlying matrix complex random variables must have independent rows. Apply corollary 9 to show

$$Y^H \sim \mathbf{C}N_{M,q}\left(0, I_M, 2Q^H(\Sigma + ARA^H)Q\right)$$

$$Z^H \sim \mathbf{C}N_{M,(p-q)}\left(0, I_M, 2V^H \Sigma V\right)$$

By definition 6 of the complex Wishart distribution,

$$W_Y = YY^H \sim \mathbf{C}W_q\left(M, 2Q^H(\Sigma + ARA^H)Q\right)$$

$$W_Z = ZZ^H \sim \mathbf{C}W_{p-q}\left(M, 2V^H \Sigma V\right)$$

Note that since Y and Z are independent, we know W_Y and W_Z are independent.

We now can take the ratio of arbitrary quadratic forms to obtain an F-distributed statistic. Apply theorem 6 where we observe that W_Y and W_Z are from *central* complex Wishart distributions. The noncentrality parameters δ_1 and δ_2 of the theorem are zero matrices. Thus, we obtain the ordinary F-distribution, without the complications of noncentrality. You can define C_1

and C_2 in the manner done in the previous theorem. Thus, the statistic

$$F = \frac{MC_1^H W_Y C_1 C_2^H 2V^H \Sigma V C_2}{MC_2^H W_Z C_2 C_1^H 2Q^H (\Sigma + ARA^H) Q C_1}$$

$$F = \frac{C_1^H W_Y C_1 C_2^H V^H \Sigma V C_2}{C_2^H W_Z C_2 C_1^H Q^H (\Sigma + ARA^H) Q C_1}$$

is distributed as

$$\sim dncF(2M.2M, 0, 0) = F(2M, 2M)$$

If the noise is modeled with $\Sigma = \sigma^2 I$, then

$$F = \frac{C_1^H W_Y C_1 \sigma^2 \|C_2\|^2}{C_2^H W_Z C_2 C_1^H Q^H (\sigma^2 I + ARA^H) Q C_1}$$

$$= \frac{C_1^H W_Y C_1 \sigma^2 \|C_2\|^2}{C_2^H W_Z C_2 \left[\sigma^2 \|C_1\|^2 + C_1^H Q^H ARA^H Q C_1\right]}$$

If you require $\|C_1\|^2 = 1$ and $\|C_2\|^2 = 1$ this further simplifies to

$$F = \frac{C_1^H W_Y C_1 \sigma^2}{C_2^H W_Z C_2 \left[\sigma^2 + C_1^H Q^H ARA^H Q C_1\right]}$$

Under the hypothesis that $q = 0$, then $W_Y = 0$, $R = 0$ and $F = 0$, a useless triviality. Under the hypothesis that $q = 1$, then W_Y and R are scalars. We get

$$F = \frac{W_Y \sigma^2}{C_2^H W_Z C_2 \left[\sigma^2 + R \|Q^H A\|^2\right]} \sim F(2M, 2M)$$

If $\|a(\theta)\| = 1$ then $Q = A$ and this further simplifies to

$$F = \frac{W_Y \sigma^2}{C_2^H W_Z C_2 \left[\sigma^2 + R\right]} \sim F(2M, 2M)$$

□

6.2.3 Test that Requires Only Partial Knowledge of the Covariance Matrix

The real variables version of following test statistic appears as Theorem 3.2.20 in [187] and it is nearly the same as the sphericity test given by Anderson. The form of the density is tedious to compute, and I have evaluated it completely only for the bivariate case.

Independence of Sphericity Test Statistic and the Trace Function

Theorem 10 *Let $A \sim \mathbf{C}W_p(n, \lambda^2 I_p)$ where $n \geq p$ is an integer. Then $u = \frac{\det A}{\left[\frac{1}{p}\operatorname{tr} A\right]^p}$ and $v = \operatorname{tr} A$ are independent. This is a complexification of Muirhead [187] theorem 3.2.20.*

Proof. Let $D = \operatorname{diag}(l_1^2, \cdots, l_p^2)$ contain the eigenvalues of A, and then by corollary 21

$$dF(D) = \frac{\pi^{p(p-1)}}{\lambda^{2pn}} \frac{\exp\left(-\frac{1}{\lambda^2}\sum_{i=1}^{p} l_i^2\right)}{\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \prod_{i=1}^{p} l_i^{2(n-p)} \prod_{i<j}^{p} (l_i^2 - l_j^2)^2 (dD)$$

Change variables from $(l_1^2, \cdots, l_p^2)$ to $(\eta, y_1, \cdots, y_{p-1})$ given by $\eta = \frac{1}{p}\sum_{i=1}^{p} l_i^2 = \frac{1}{p}\operatorname{tr} A$, and $y_i = l_i^2/\eta$ for $1 \leq i \leq p$. Note that $\sum_{i=1}^{p} y_i = p$. Then

$$u = \frac{\det A}{\left[\frac{1}{p}\operatorname{tr} A\right]^p} = \sum_{i=1}^{p} \frac{l_i^2}{\eta} = \prod_{i=1}^{p} y_i$$

Note that u is bounded on the closed interval $[0, 1]$.

Note that $l_i^2 \geq 0$ are all real numbers, which affects the form of the Jacobian. Unlike many other changes of variables in this thesis where Jacobians needed to account for complex variables, here we can use Jacobians we have computed for real variables. First, change variables from $(l_1^2, \cdots, l_p^2)$ to $(l_1^2, \cdots, l_{p-1}^2, \eta)$. The transformation matrix is the familiar

$$\begin{pmatrix} & & & 0 \\ & I_{p-1} & & 0 \\ & & & \vdots \\ & & & 0 \\ \frac{1}{p} & \frac{1}{p} & \cdots \frac{1}{p} & \frac{1}{p} \end{pmatrix}$$

which has determinant $\frac{1}{p}$. The Jacobian is

$$J[(l_1^2, \cdots, l_p^2) \rightarrow (l_1^2, \cdots, l_{p-1}^2, \eta)] = p$$

The first step in our change of variables is given below for

$$C = \frac{\pi^{p(p-1)}}{\lambda^{2pn} \mathbf{C}\Gamma_p(n) \mathbf{C}\Gamma_p(p)}$$

We obtain

$$dF(l_1^2, \cdots, l_{p-1}^2, \eta) = dF(D_1)$$

$$= C \exp(-\frac{p}{\lambda^2}\eta) \left[\prod_{i=1}^{p-1} l_i^{2(n-p)}\right] \left[p\eta - \sum_{i=1}^{p-1} l_i^2\right]^{n-p}$$

$$\times \left[\prod_{i<j<p} (l_i^2 - l_j^2)^2\right] \left[\prod_{i=1}^{p-1} \left(l_i^2 - p\eta + \sum_{j=1}^{p-1} l_j^2\right)^2\right] p(dD_1)$$

The transformation matrix from $(l_1^2, \cdots, l_{p-1}^2, \eta)$ to $(y_1, \cdots, y_{p-1}, \eta)$ is

$$\begin{pmatrix} \frac{1}{\eta} & & & \\ & \ddots & & \\ & & \frac{1}{\eta} & \\ & & & 1 \end{pmatrix}$$

which has determinant $\left(\frac{1}{\eta}\right)^{p-1}$. The Jacobian is

$$\eta^{p-1} = J[(l_1^2, \cdots, l_{p-1}^2, \eta) \rightarrow (y_1, \cdots, y_{p-1}, \eta)]$$

Thus

$$dF(y_1, \cdots, y_{p-1}, \eta) \stackrel{\text{def}}{=} dF(Y)$$

$$= C \exp(-\frac{p}{\lambda^2}\eta) \left[\prod_{i=1}^{p-1} (\eta y_i)^{n-p}\right] \left[p\eta - \sum_{i=1}^{p-1} \eta y_i\right]^{n-p} \left[\prod_{i<j<p} (\eta y_i - \eta y_j)^2\right]$$

$$\times \left[\prod_{i=1}^{p-1} \left(\eta y_i - p\eta + \sum_{j=1}^{p-1} \eta y_j\right)^2\right] \eta^{p-1} p(dY)$$

Now factor the joint density into a form having a term that is a function of only η.

$$dF(Y) = C\left(-\frac{p}{\lambda^2}\eta\right) \eta^{(n-p)(p-1)} \eta^{n-p} \eta^{2(p-1)(p-2)/2} \eta^{2(p-1)} \eta^{p-1} p$$

$$\times \left[\prod_{i=1}^{p-1} y_i{}^{n-p}\right] \left[p - \sum_{i=1}^{p-1} y_i\right]^{n-p} \left[\prod_{i<j<p} (y_i - y_j)^2\right] \left[\prod_{i=1}^{p-1} \left(y_i - p + \sum_{j=1}^{p-1} y_j\right)^2\right] (dY)$$

Collecting powers of η gives us

$$dF(Y) = C(-\frac{p}{\lambda^2}\eta) \eta^{np-1} \left[\prod_{i=1}^{p} y_i{}^{n-p}\right] \left[\prod_{i<j}^{p} (y_i - y_j)^2\right] (dY)$$

where $y_p = p - \sum_{i=1}^{p-1} y_i$ is used for shorthand notation and is not part of the change of variables, and the exponent of η is computed from

$$(n-p)(p-1) + (n-p) + (p-1)(p-2) + 3(p-1)$$

$$= p(n-p) + (p-1)(p+1) = np - p^2 + p^2 + p - p - 1 = np - 1$$

By the Neyman-Fisher Factorization Theorem, we see that η is independent of $(y_1, \cdots, y_{p-1})$. y_p is a function only of $(y_1, \cdots, y_{p-1})$, and u is a function only of $(y_1, \cdots, y_p)$. The variable $v = \operatorname{tr} A$ is a function only of η. Therefore u and v are statistically independent, which proves the theorem.□

The statistic $u = \frac{\det A}{\left[\frac{1}{p} \operatorname{tr} A\right]^p}$ is used to test the hypothesis $H_0 : \Sigma = \lambda^2 I_p$ versus the alternative hypothesis $H_a : \Sigma \neq \lambda^2 I_p$ for some fixed (but not necessarily known) λ^2. When H_0 is true, the cumulative distribution function is given by $F_u(x)$ which depends also on the parameters p and n. When the sample eigenvalues are equal, then $u = 1$, which is the maximum value of u. We know $u \leq 1$ by Hardy (p. 17, Theorem 9) [102]. We know from Okamoto [197] that the sample eigenvalues will all be different with probability 1. So, the smaller u is, the more likely $\Sigma = \lambda^2 I_p$ is not true. We want to choose a value x so that when $u < x$ we can decide to reject $H_0 : \Sigma = \lambda^2 I_p$ with a probability of rejecting H_0 when H_0 is true being less than α. Thus, we choose x so that

$$\Pr(u \leq x \mid H_0 : \Sigma = \lambda^2 I_p) = \alpha = F_u(x)$$

This is a one-sided test.

To obtain the marginal density $dF(y_1, \cdots, y_{p-1})$, integrate η out. Examining only those terms that contain η, evaluate

$$\mathcal{I} = \int_0^\infty \exp(-\frac{p}{\lambda^2}\eta)\eta^{np-1}d\eta$$

In the integral $\int_0^\infty e^{-\alpha x}x^m dx$ given in corollary 48, let $\alpha = \frac{p}{\lambda^2}$ and $m = np-1$. Then

$$\mathcal{I} = \left(\frac{\lambda^2}{p}\right)^{np}(np-1)!$$

Thus for

$$C_1 = C\mathcal{I}p = \frac{(np-1)!\pi^{p(p-1)}}{p^{np-1}\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}$$

then

$$dF(y_1, \cdots, y_{p-1}) = C_1\left[\prod_{i=1}^p y_i{}^{n-p}\right]\left[\prod_{i<j}^p (y_i - y_j)^2\right](d(y_1, \cdots, y_{p-1}))$$

Because $(y_1, \cdots, y_{p-1})$ is independent of η, we find

$$dF(\eta) = \frac{dF(y_1, \cdots, y_{p-1}, \eta)}{dF(y_1, \cdots, y_{p-1})}$$

$$= \frac{Cp\exp(-\frac{p}{\lambda^2}\eta)\eta^{np-1}\left[\prod_{i=1}^p y_i{}^{n-p}\right]\left[\prod_{i<j}^p (y_i - y_j)^2\right]}{Cp\mathcal{I}\left[\prod_{i=1}^p y_i{}^{n-p}\right]\left[\prod_{i<j}^p (y_i - y_j)^2\right]}d\eta$$

$$dF(\eta) = \frac{\left(\frac{p}{\lambda^2}\right)^{np}}{(np-1)!}\exp(-\frac{p}{\lambda^2}\eta)\eta^{np-1}d\eta$$

This is the probability density function of the average of the sample eigenvalues of $A \sim \mathbf{C}W_p(n, \lambda^2 I_p)$. To find $dF(\operatorname{tr} A)$, let $x = \operatorname{tr} A = p\eta$ be a change of variables.

$$dF(\operatorname{tr} A) = \frac{\left(\frac{p}{\lambda^2}\right)^{np}}{(np-1)!}\exp\left(-\frac{1}{\lambda^2}\operatorname{tr} A\right)\left(\frac{1}{p}\right)^{np-1}(\operatorname{tr} A)^{np-1}\frac{1}{p}d(\operatorname{tr} A)$$

or

$$dF(\operatorname{tr} A) = \frac{1}{\lambda^{2np}(np-1)!} \exp\left(-\frac{1}{\lambda^2}\operatorname{tr} A\right) (\operatorname{tr} A)^{np-1} d(\operatorname{tr} A)$$

We get another identity by looking at theorem 88, since

$$\mathcal{E}\{\operatorname{tr} A\} = \int_0^\infty (\operatorname{tr} A) dF(\operatorname{tr} A) = n(\operatorname{tr} \lambda^2 I_p) = \lambda^2 np$$

Then

$$\int_0^\infty (\operatorname{tr} A)^{np} \exp(-\frac{1}{\lambda^2}\operatorname{tr} A) d(\operatorname{tr} A) = \lambda^{2(np+1)}(np)!$$

for $A \sim \mathbf{C}W_p(n, \lambda^2 I_p)$. This same result is more straight-forwardly evaluated using the definition of the gamma function, letting $x = \operatorname{tr} A$.

Sphericity Test Statistic Density Function

We would like to find the density function for

$$u = \frac{\det A}{\left[\frac{1}{p}\operatorname{tr} A\right]^p} = \sum_{i=1}^p \frac{l_i^2}{\eta} = \prod_{i=1}^p y_i \tag{6.15}$$

We know $\sum_{i=1}^p y_i = p$, and the joint density of $(y_1, \cdots, y_{p-1})$ is

$$dF(y_1, \cdots, y_{p-1}) = C_1 \left[\prod_{i=1}^p {y_i}^{n-p}\right] \left[\prod_{i<j}^p (y_i - y_j)^2\right] (d(y_1, \cdots, y_{p-1}))$$

We need to do a change of variables $u = \prod_{i=1}^p y_i$, and $z_i = y_i$ for $2 \le i \le p-1$, and then integrate out the z_i. The challenge is to handle the nonlinearity introduced by $y_p = p - \sum_{i=1}^{p-1} y_i$ when evaluating both the $(y_i - y_j)^2$ terms and the Jacobian. The issue arises in evaluating y_1. We compute the inverse mappings now.

$$y_i = z_i,\ 2 \le i \le p-1$$

$$y_p = p - y_2 - y_3 - \cdots - y_{p-1} - y_1$$

$$u = y_1 y_2 \cdots y_{p-1} y_p = y_1 y_2 \cdots y_{p-1}(p - y_2 - y_3 - \cdots - y_{p-1} - y_1)$$

$$= y_1 y_2 \cdots y_{p-1}(p - y_2 - y_3 - \cdots - y_{p-1}) - y_1^2 y_2 \cdots y_{p-1}$$

Let $v = y_2 y_3 \cdots y_{p-1}$ and $w = \sum_{i=2}^{p-1} y_i$ as a shorthand notation. Then $u = y_1 v(p - w) - y_1^2 v$. Note that $v = z_2 z_3 \cdots z_{p-1}$ and $w = z_2 + z_3 + \cdots + z_{p-1}$. Solve for y_1 in terms of the new variables using completion of squares.

$$y_1^2 v - y_1 v(p - w) = -u$$

$$y_1^2 - 2y_1 \frac{1}{2}(p - w) + \frac{1}{4}(p - w)^2 = -\frac{u}{v} + \frac{1}{4}(p - w)^2$$

$$(y_1 - \frac{1}{2}(p - w))^2 = \frac{1}{4}(p - w)^2 - \frac{u}{v}$$

$$y_1 - \frac{1}{2}(p - w) = \pm \left[\frac{1}{4}(p - w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}$$

$$y_1 = \frac{1}{2}(p - w) \pm \left[\frac{1}{4}(p - w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}$$

We have two different values of y_1 that, together with $(y_2, \cdots, y_{p-1})$, map into the same value of $(u, z_2, \cdots z_{p-1})$. Let $dF(u, z_2, \cdots z_{p-1})$ be the joint probability density function of the transformed variables. It will be the sum of two functions representing the transformations from sets A_1 and A_2 into the set B, where set A_1 corresponds to all values of y_1 obtained with the $(+)$ solution and A_2 corresponds to all the values of y_1 obtained with the $(-)$ solution. To write these functions, we need to evaluate the Jacobian belonging to each A_i. Let

$$\varphi(y_1, \cdots, y_{p-1}) = dF(y_1, \cdots, y_{p-1})$$

to make the following discussion unambiguous. Let g_i be the inverse mappings from B to A_i. Let $J_i(y \rightarrow z) = \frac{\partial g_i}{\partial Y}$ denote the Jacobian that transforms variables $(y_1, \cdots, y_{p-1})$ in A_i into $(u, z_2, \cdots z_{p-1})$ in B. Then

$$dF(u, z_2, \cdots z_{p-1}) \underset{=}{\text{def}} dF(Z) = \varphi[g_1(Y)]\,|J_1(Y \rightarrow Z)| + \varphi[g_2(Y)]\,|J_2(Y \rightarrow Z)|$$

The Jacobian will be tedious, but straight forward, to evaluate because $\frac{\partial y_i}{\partial z_j} = \delta_{ij}$, $(i, j) \geq 2$. The Jacobian will thus have the form

$$|J_i(Y \rightarrow Z)| = \left| \det \begin{bmatrix} \frac{\partial y_1}{\partial u} & \frac{\partial y_2}{\partial u} & \cdots & \frac{\partial y_{p-1}}{\partial u} \\ \frac{\partial y_1}{\partial z_2} & \frac{\partial y_2}{\partial z_2} & \cdots & \frac{\partial y_{p-1}}{\partial z_2} \\ \frac{\partial y_1}{\partial z_3} & \frac{\partial y_2}{\partial z_3} & \cdots & \frac{\partial y_{p-1}}{\partial z_3} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial z_{p-1}} & \frac{\partial y_2}{\partial z_{p-1}} & \cdots & \frac{\partial y_{p-1}}{\partial z_{p-1}} \end{bmatrix} \right|_i$$

$$= \left| \det \begin{bmatrix} \frac{\partial y_1}{\partial u} & 0 & \cdots & 0 \\ \frac{\partial y_1}{\partial z_2} & 1 & & 0 \\ \vdots & & \ddots & \\ \frac{\partial y_1}{\partial z_{p-1}} & 0 & & 1 \end{bmatrix} \right|_i = \left| \frac{\partial y_1}{\partial u} \right|_i$$

The other terms drop out of the expansion of the above determinant down the first column because the cofactor matrices of $\frac{\partial y_1}{\partial z_j}$ all contain a first row of all zeros. Therefore, the determinant of those cofactor matrices evaluate to zero. Therefore, we do not have to evaluate the messy terms $\left(\frac{\partial y_1}{\partial z_j}\right)$. Now, evaluate $\frac{\partial y_1}{\partial u}$ for A_1 and A_2.

In A_1, we use the (+) solution for y_1.

$$\frac{\partial y_1}{\partial u} = \frac{\partial}{\partial u}\left\{\frac{1}{2}(p-w) + \left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}\right\}$$

$$= \frac{1}{2}\left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{-\frac{1}{2}}\left(-\frac{1}{v}\right) = \left\{-2v\left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}\right\}^{-1}$$

$$= -\{(vp-vw)^2 - 4uv\}^{-\frac{1}{2}}$$

$$= -\{[z_2\cdots z_{p-1}(p-z_2-\cdots z_{p-1})]^2 - 4uz_2\cdots z_{p-1}\}^{-\frac{1}{2}}$$

$$= -\{v^2t^2 - 4uv\}^{-\frac{1}{2}}$$

for $t = p - w$.

In A_2, we use the (−) solution for y_1.

$$\frac{\partial y_1}{\partial u} = \frac{\partial}{\partial u}\left\{\frac{1}{2}(p-w) - \left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}\right\}$$

$$= -\frac{1}{2}\left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{-\frac{1}{2}}\left(-\frac{1}{v}\right) = \left\{2v\left[\frac{1}{4}(p-w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}\right\}^{-1}$$

$$= \{(vp-vw)^2 - 4uv\}^{-\frac{1}{2}}$$

$$= \{[z_2\cdots z_{p-1}(p-z_2-\cdots z_{p-1})]^2 - 4uz_2\cdots z_{p-1}\}^{-\frac{1}{2}}$$

$$= \{v^2t^2 - 4uv\}^{-\frac{1}{2}}$$

for $t = p - w$.

Note that $\left|\frac{\partial y_1}{\partial u}\right|_+ = \left|\frac{\partial y_1}{\partial u}\right|_-$. Now we know we can simplify our expression for $dF(Z)$.

$$dF(Z) = \varphi[g_1(Y)]\,|J_1(Y \to Z)| + \varphi[g_2(Y)]\,|J_2(Y \to Z)|$$

$$= \{\varphi[g_1(Y)] + \varphi[g_2(Y)]\}\,|J(Y \to Z)|$$

where

$$|J| = \{[z_2 \cdots z_{p-1}(p - z_2 - \cdots z_{p-1})]^2 - 4uz_2 \cdots z_{p-1}\}^{-\frac{1}{2}}$$

Concentrate on evaluating $\varphi[g_1(Y)]$. The challenge is to find a convenient expression for $\prod_{i<j}^{p}(y_i - y_j)^2$. For $2 \le i < j < p$, the easy terms to evaluate, we get $(z_i - z_j)^2$. Consider $(y_i - y_p)^2$ for $2 \le i < p$. This is

$$(y_i - p + y_1 + y_2 + \cdots + y_{p-1})^2 = (y_1 - p + z_i + w)^2$$

$$= \left\{\frac{1}{2}(p - w) + \left[\frac{1}{4}(p - w)^2 - \frac{u}{v}\right]^{\frac{1}{2}} - p + w + z_i\right\}^2$$

where $w = z_2 + z_3 + \cdots + z_{p-1}$ and $v = z_2 z_3 \cdots z_{p-1}$. Let $t = p - w$ to simplify slightly to get

$$= \left\{-\frac{1}{2}t + \left[\frac{1}{4}t^2 - \frac{u}{v}\right]^{\frac{1}{2}} + z_i\right\}^2 = (y_i - y_p)^2$$

This is as simplified as I have been able to get. Now consider $(y_1 - y_j)^2$ for $2 \le j < p - 1$.

$$(y_1 - y_j)^2 = \left\{\frac{1}{2}t + \left[\frac{1}{4}t^2 - \frac{u}{v}\right]^{\frac{1}{2}} - z_j\right\}^2$$

Finally, we evaluate $(y_1 - y_p)^2$. Here, we have

$$(y_1 - y_p)^2 = \left\{y_1 - \left(p - \sum_{j=1}^{p-1} y_i\right)\right\}^2 = \{y_1 - p + y_1 + y_2 + \cdots + y_{p-1}\}^2$$

$$= \{2y_1 - p + w\}^2 = \{2y_1 - t\}^2 = \left\{t + \left[t^2 - \frac{4u}{v}\right]^{\frac{1}{2}} - t\right\}^2 = t^2 - \frac{4u}{v}$$

Putting this together, we find

$$h_1 \underset{=}{\text{def}} \prod_{i<j}^{p}(y_i-y_j)^2 = \left\{\prod_{j=2}^{p}(y_1-y_j)^2\right\}(y_1-y_p)^2\left\{\prod_{2\le i<j}^{p-1}(y_i-y_j)^2\right\}\left\{\prod_{2\le i}^{p-1}(y_i-y_p)^2\right\}$$

$$= \left\{\prod_{j=2}^{p-1}\left(\frac{1}{2}t+\left[\frac{1}{4}t^2-\frac{u}{v}\right]^{\frac{1}{2}}-z_j\right)^2\right\}\left\{t^2-\frac{4u}{v}\right\}$$

$$\times\left\{\prod_{2\le i<j}^{p-1}(z_i-z_j)^2\right\}\left\{\prod_{2\le i}^{p-1}\left(-\frac{1}{2}t+\left[\frac{1}{4}t^2-\frac{u}{v}\right]^{\frac{1}{2}}+z_j\right)^2\right\}$$

where $t = p - w = p - z_2 - z_3 - \cdots - z_{p-1}$ and $v = z_2 z_3 \cdots z_{p-1}$. The term $\prod_{i=1}^{p} y_i^{n-p} = u^{n-p}$ is true for both the "plus" solution density g_1 and the "minus" solution density g_2. So,

$$\varphi[g_1(Y)] = C_1 u^{n-p} h_1$$

Now, evaluate $\varphi[g_2(Y)]$, which corresponds to the $(-)$ solution for y_1. We evaluate $\prod_{i<j}^{p}(y_i-y_j)^2$ once more. For $2 \le i < j < p$, we get $(y_i-y_j)^2 = (z_i-z_j)^2$ as before. Consider $(y_i - y_p)^2$ for $2 \le i < p$. Then

$$(y_1 - t + z_i)^2 = \left(\frac{1}{2}t - \left[\frac{1}{4}t^2-\frac{u}{v}\right]^{\frac{1}{2}} - t + z_i\right)^2 = \left(-\frac{1}{2}t - \left[\frac{1}{4}t^2-\frac{u}{v}\right]^{\frac{1}{2}} + z_i\right)^2$$

This differs from the previous evaluation by the sign change for the coefficient of $\left[\frac{1}{4}t^2 - \frac{u}{v}\right]^{\frac{1}{2}}$. Now consider $(y_1 - y_j)^2$ for $2 \le j < p-1$. This is

$$(y_1 - y_j)^2 = \left(\frac{1}{2}t - \left[\frac{1}{4}t^2-\frac{u}{v}\right]^{\frac{1}{2}} - z_j\right)^2$$

Finally,

$$(y_1 - y_p)^2 = (2y_1)^2 = (t - (t^2 - \frac{4u}{v})^{\frac{1}{2}} - t)^2 = t^2 - \frac{4u}{v}$$

Thus

$$h_2 \underset{=}{\text{def}} \left\{\prod_{j=2}^{p-1}\left(\frac{1}{2}t - \left[\frac{1}{4}t^2 - \frac{u}{v}\right]^{\frac{1}{2}} - z_j\right)^2\right\}\left\{t^2 - \frac{4u}{v}\right\}$$

$$\times\left\{\prod_{2\le i<j}^{p-1}(z_i - z_j)^2\right\}\left\{\prod_{2\le i}^{p-1}\left(-\frac{1}{2}t - \left[\frac{1}{4}t^2 - \frac{u}{v}\right]^{\frac{1}{2}} + z_i\right)^2\right\}$$

The probability density function for the new variables $(u, z_2, \cdots, z_{p-1})$ is

$$dF(u, z_2, \cdots, z_{p-1}) = dF(Z) = C_1 u^{n-p}\{h_1+h_2\}[v^2t^2-4uv]^{-\frac{1}{2}}(d(u, z_2, \cdots, z_{p-1})) \tag{6.16}$$

where

$$C_1 = \frac{(np-1)!}{p^{np-1}}\frac{\pi^{p(p-1)}}{\mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)}$$

To find the density of u, integrate out $(z_2, \cdots, z_{p-1})$. The limits of integration are

$$(0, z_{k-1}) \quad \text{for } z_k \quad \text{where } 3 \le k \le p-1$$

$$(0, \infty) \quad \text{for } z_2$$

The case of $p = 1$ is trivial. The case for $p = 2$ is tractable, and a closed form solution is presented for $F_u(x) = \Pr(u \le x)$. The case for $p = 3$ is tedious and should be evaluated by some automated means. When $p > 3$, the general approach above is the appropriate tactic. The required integration is tedious for the general case. Let us look at some small values of p.

There is no point in doing an evaluation for $p = 1$. When $p = 1$, then $u = 1$ since $l_1^2 = \eta$. Thus $\Pr(u = 1) = 1$.

We are still dealing with a special case when $p = 2$. Let us go back to

basics.

$$dF(l_1^2, l_2^2) = \frac{\pi^2}{\lambda^{4n}} \frac{1}{\mathbf{C}\Gamma_2(n)\mathbf{C}\Gamma_2(2)} \exp\left[-\frac{1}{\lambda^2}(l_1^2 + l_2^2)\right] (l_1^2 l_2^2)^{n-2} (l_1^2 - l_2^2)^2 (dD)$$

Let $\eta = \frac{1}{2}(l_1^2 + l_2^2)$. Note: $l_2^2 = 2\eta - l_1^2$. Then the joint density of (l_1^2, η) is

$$dF(l_1^2, \eta) = C \exp\left(-\frac{2}{\lambda^2}\eta\right) (l_1^2)^{n-2} [p\eta - l_1^2]^{n-2} (2l_1^2 - p\eta)^2 p(d(l_1^2, \eta))$$

where

$$C = \frac{\pi^2}{\lambda^{4n}\mathbf{C}\Gamma_2(n)\mathbf{C}\Gamma_2(2)}$$

We note that

$$\mathbf{C}\Gamma_2(n) = \pi\Gamma(n)\Gamma(n-1) = \pi(n-1)!(n-2)!$$

and

$$\mathbf{C}\Gamma_2(2) = \pi(1!)(0!) = \pi$$

Let $y_i = l_i^2/\eta$. Note: $y_1 + y_2 = p = 2$. Let $u = y_1 y_2 = y_1(2 - y_1)$. With this change of variables, the joint density of (y_1, η) is given by

$$dF(y_1, \eta) = C \exp\left(-\frac{2}{\lambda^2}\eta\right) (\eta y_1)^{n-2} [p\eta - \eta y_1]^{n-2} (2\eta y_1 - p\eta)^2 \eta p(d(y_1, \eta))$$

$$= 2C \exp\left(-\frac{2}{\lambda^2}\eta\right) \eta^{2n-1} y_1^{n-2} [2 - y_1]^{n-2} (2y_1 - 2)^2 (d(y_1, \eta))$$

We do some intermediate bookkeeping to help us find the density of y_1.

$$\int_0^\infty \exp\left(-\frac{2}{\lambda^2}\eta\right) \eta^{2n-1} d\eta = \left(\frac{\lambda^2}{2}\right)^{2n} (2n-1)!$$

Evaluating $dF(y_1)$, we get

$$dF(y_1) = \frac{2^3\pi^2(2n-1)!}{\lambda^{4n}\pi(n-1)!(n-2)!\pi} \frac{\lambda^{4n}}{2^{2n}} y_1^{n-2} (2 - y_1)^{n-2} (y_1 - 1)^2 dy_1$$

$$= \frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!} y_1^{n-2}(2-y_1)^{n-2}(y_1-1)^2 dy_1$$

Since y_1 and η are independent, we see that

$$dF(\eta) = \frac{dF(y_1,\eta)}{dF(y_1)} = \frac{\frac{2^3\pi^2}{\lambda^{4n}\pi(n-1)!(n-2)!\pi} \exp\left(-\frac{2}{\lambda^2}\eta\right) \eta^{2n-1} y_1^{n-2}[2-y_1]^{n-2}(y_1-1)^2}{\frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!} y_1^{n-2}(2-y_1)^{n-2}(y_1-1)^2}$$

$$= \frac{\left(\frac{2}{\lambda^2}\right)^{2n}}{(2n-1)!} \exp\left(-\frac{2}{\lambda^2}\eta\right) \eta^{2n-1} d\eta$$

Our assurance this is correct is that it integrates to 1.

Now to find $u = y_1(2-y_1)$ for the case of $p = 2$. We see that $u = 2y_1 - y_1^2$ implies

$$y_1^2 - 2y_1 + 1 = 1 - u = (y_1 - 1)^2$$

Solving for y_1, we find $y_1 = 1 \pm \sqrt{1-u}$. We let A_1 be the set of all

$Y_1 = 1 + (1-u)^{\frac{1}{2}}$, and let A_2 be the set of all $y_1 = 1 - (1-u)^{\frac{1}{2}}$. In A_1, the critical computation in the Jacobian is $\frac{\partial y_1}{\partial u} = -\frac{1}{2}(1-u)^{-\frac{1}{2}}$, and in A_2 it is $\frac{\partial y_1}{\partial u} = \frac{1}{2}(1-u)^{-\frac{1}{2}}$. Note that $\left|\frac{\partial y_1}{\partial u}\right|_+ = \left|\frac{\partial y_1}{\partial u}\right|_- = |J|$ is the absolute value of the determinant of the Jacobian matrix. Thus, for the special case of $p = 2$,

$$dF(u) = \left(\frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!}\right) \times$$

$$\times \left\{ \left[1 + (1-u)^{\frac{1}{2}}\right]^{n-2} \left[1 - (1-u)^{\frac{1}{2}}\right]^{n-2} \left[(1-u)^{\frac{1}{2}}\right]^2 + \right.$$

$$\left. \left[1 - (1-u)^{\frac{1}{2}}\right]^{n-2} \left[1 + (1-u)^{\frac{1}{2}}\right]^{n-2} \left[-(1-u)^{\frac{1}{2}}\right]^2 \right\} \left|\frac{1}{2}(1-u)^{-\frac{1}{2}}\right| du$$

$$= \left(\frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!}\right) (1-u)^{\frac{1}{2}} \left[1 + (1-u)^{\frac{1}{2}}\right]^{n-2} \left[1 - (1-u)^{\frac{1}{2}}\right]^{n-2} du$$

$$dF(u) = \left(\frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!}\right) (1-u)^{\frac{1}{2}} u^{n-2} du \qquad (6.17)$$

The cumulative distribution $F_u(x) = \Pr\{u \leq x\}$ is found by integrating $dF(u)$ over the interval $[0, x]$, where $x \in [0, 1]$. The integral is evaluated by successive application of the chain rule. Note that $dF(u)$, and hence $F_u(x)$, is independent of λ^2. We truly are testing sphericity without regard to multivariate diameter.

From our combinatoric identity, given in proposition 103 with $m = n - 2$ and $a = \frac{1}{2}$, we see that

$$F_u(x) = \int_0^{\infty} dF(u) = \left(\frac{(2n-1)!}{2^{2n-3}(n-1)!(n-2)!}\right) \int_0^x u^{n-2}(1-u)^{\frac{1}{2}} du$$

$$= -\frac{(2n-1)!(1-u)^{\frac{3}{2}}}{3 \times 2^{2(n-1)}(n-1)!(n-2)!} \sum_{k=0}^{n-2} \frac{\binom{n-2}{k}}{\binom{k+\frac{3}{2}}{k}} u^{n-k-2}(1-u)^k \Bigg|_0^x$$

$$F_u(x) = \frac{(2n-1)!}{3 \times 2^{2(n-1)}(n-1)!(n-2)!} \times \tag{6.18}$$

$$\times \left[\frac{(n-2)!}{(n-\frac{1}{2})(n-\frac{3}{2})\cdots\frac{5}{2}} - (1-x)^{\frac{3}{2}} \sum_{k=0}^{n-2} \frac{\binom{n-2}{k}}{\binom{k+\frac{3}{2}}{k}} x^{n-k-2}(1-x)^k\right]$$

for $p = 2$ and finite n. When $n = 3$, $\Pr\{u \leq \frac{1}{2}\} \cong 0.19$.

The number of terms increases explosively with the dimension of the random vector. Even though the case of $p = 3$ is still a simplified special case, the number of terms is unmanageable using manual methods. In this case we want to find $dF(u, z_2)$ and then

$$df(u) = \int_{z_2} dF(u, z_2)$$

Reviewing some notation, $v = z_2$ and $w = z_2$. Evaluate

$$\prod_{i<j}^{p=3}(y_i - y_j)^2 = (y_1 - y_2)^2(y_1 - y_3)^2(y_2 - y_3)^2$$

where $y_3 = 3 - y_1 - y_2$, $y_2 = z_2$, and

$$y_1 = \frac{1}{2}(p - w) \pm \left[\frac{1}{4}(p - w)^2 - \frac{u}{v}\right]^{\frac{1}{2}}$$

Note that $y_1 - y_3 = 2y_1 + z_2 - 3$ and $y_2 - y_3 = y_1 + 2y_2 - 3$. There is both a (+) and a (−) solution. The (+) solution is given by

$$h_1 = \prod_{i<j}^{p=3}(y_i - y_j)_+^2 = \left[\frac{1}{2}(3 - z_2) + \left(\frac{1}{4}(3 - z_2)^2 - \frac{u}{z_2}\right)^{\frac{1}{2}} - z_2\right]^2$$

$$\times \left[(3 - z_2) + \left((3 - z_2)^2 - \frac{4u}{z_2}\right)^{\frac{1}{2}} + z_2 - 3\right]^2$$

$$\times \left[\frac{1}{2}(3 - z_2) + \left(\frac{1}{4}(3 - z_2)^2 - \frac{u}{z_2}\right)^{\frac{1}{2}} + 2z_2 - 3\right]^2$$

The (−) solution is given by

$$h_2 = \prod_{i<j}^{p=3}(y_i - y_j)_-^2 = \left[\frac{1}{2}(3 - z_2) - \left(\frac{1}{4}(3 - z_2)^2 - \frac{u}{z_2}\right)^{\frac{1}{2}} - z_2\right]^2$$

$$\times \left[(3 - z_2) - \left((3 - z_2)^2 - \frac{4u}{z_2}\right)^{\frac{1}{2}} + z_2 - 3\right]^2$$

$$\times \left[\frac{1}{2}(3 - z_2) - \left(\frac{1}{4}(3 - z_2)^2 - \frac{u}{z_2}\right)^{\frac{1}{2}} + 2z_2 - 3\right]^2$$

Then

$$dF(u, z_2) = C_1 u^{n-3}\{h_1 + h_2\}\left[z_2^2(3 - z_2)^2 - 4uz_2\right]^{-\frac{1}{2}} d(u, z_2)$$

where

$$C_1 = \frac{(3n-1)!}{3^{3n-1}} \frac{\pi^6}{\mathbf{C}\Gamma_3(n)\mathbf{C}\Gamma_3(3)} = \frac{(3n-1)!}{2 \times 3^{3n-1}(n-3)!(n-2)!(n-1)!}$$

To get the density of the statistic u, integrate over z_2.

Although a symbolic mathematics processor could evaluate the required integral over for small dimensions in a reasonable time, I think the numerical accuracy resulting from its evaluation would be worse than obtained by beginning with numerical integration.

6.3 Tests Motivated by Krishnaiah

In this section I provide joint distributions of some desirable test statistics and associated nuisance variables, when the sample eigenvalues obey a special case distribution. The distributions represent the nearest approach in this thesis to solving the original thesis question. These tests were motivated by Krishnaiah's works. Derivations are independent of Krishnaiah's work..

Krishnaiah has been a central figure in the development of tests on eigenvalues, including those making use of concepts from James' work on zonal polynomials and complex variables. His work is reported primarily in reports for the United States Air Force Aerospace Research Laboratories, and may be obtained through the United States Department of Commerce National Technical Information Service. The ordering information (AD numbers) are

included in the bibliography of this thesis. These reports may be characterized by their insight, briefness, and use of a lemma for integration which I have not yet tried to prove for the context of this thesis. Krishnaiah's works are reported in integral form.

The material that follows was directly motivated by the problems which Krishnaiah solved. I have not worked out all the details of Krishnaiah's work, so I have not yet made the necessary connections between his work and the work which follows. That is an important effort in the context of order estimation to be pursued later.

In all the work to follow, let the sample eigenvalues $D = \text{diag}(l_1^2, \cdots, l_p^2)$ estimate the population eigenvalues $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$. I will assume the following special case that the sample eigenvalues have the joint density function of D given by

$$\begin{aligned} dF(D) &= \left[\frac{|\det D|^{n-p}\,\pi^{p(p-1)}}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\sum_{k=1}^{p} \frac{l_k^2}{\lambda_k^2}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD) \\ &= \left[\frac{|\det D|^{n-p}\,\pi^{p(p-1)}}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\Lambda^{-2} D\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD) \\ &= \left[\frac{|\det D|^{n-p}\,\pi^{p(p-1)}}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] {}_0F_0(-\Lambda^{-2}, D) \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD) \qquad (6.19) \end{aligned}$$

This is the case when $D \sim \mathbf{C}W_p(n, \Lambda^2)$ such that D is further restricted to be diagonal. Thus, the elements of D are independently distributed χ^2.

This originally was considered as a result of the observation that zonal polynomials have the property that $Z_m(U^H X U) = Z_m(X)$ for all $U \in \mathbf{U}(n)$.

This leads to saying $Z_m(A) = Z_m(D)$ and $Z_m(-\Sigma^{-1}) = Z_m(-\Lambda^{-2})$. Further, noting that ${}_0F_0(-\Sigma^{-1}, A) = \text{etr}(\Sigma^{-1}A)$ and

$$ {}_0F_0(-\Sigma^{-1}, A) = \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Sigma^{-1}) Z_m(A)}{Z_m(I)} $$

and substituting this into the density function

$$ dF(D) = \left[\frac{|\det D|^{n-p} \pi^{p(p-1)}}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \right] $$

$$ \times \left[\sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Sigma^{-1}) Z_m(A)}{Z_m(I)} \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD) $$

$$ = \left[\frac{|\det D|^{n-p} \pi^{p(p-1)}}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \right] \left[{}_0F_0(-\Sigma^{-1}, A) \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD) $$

$$ = \left[\frac{|\det D|^{n-p} \pi^{p(p-1)}}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \right] \left[\text{etr}(-\Sigma^{-1} A) \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD) $$

led me to consider

$$ dF(D) = \left[\frac{|\det D|^{n-p} \pi^{p(p-1)}}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \right] $$

$$ \times \left[\sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Lambda^{-2}) Z_m(D)}{Z_m(I)} \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD) $$

$$ = \left[\frac{|\det D|^{n-p} \pi^{p(p-1)}}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \right] \left[\text{etr}(-\Lambda^{-2} D \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD) $$

The problem with this approach is the implication that $\text{etr}(-\Sigma^{-1}A) = \text{etr}(-\Lambda^{-2}D)$, which is not true.

6.3.1 Joint Density of Ratio of Adjacent Sample Eigenvalues

Proposition 1 *Let* $D = \text{diag}(l_1^2, \cdots, l_p^2)$ *be sample eigenvalues corresponding to* $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ *have the density function given in equation 6.19. Then the joint density of*

$$\nu = (\theta_1, \cdots, \theta_{p-1}), \quad \theta_i = \frac{l_i^2}{l_{i+1}^2}$$

is given by

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(n+p^2)}{[\det \Lambda^2]^n \, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\frac{1}{\lambda_p^2} \bigwedge_{k=p-1}^{1} \left(1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\theta_k\right)\right]^{-(n+p^2)}$$

$$\times \left[\prod_{i=1}^{p-1} \theta_i^{n+p-1+2i}\right] \left[\prod_{i<j}^{p-1} (\theta_j\theta_{j+1}\cdots\theta_{p-1})^2 (\theta_i\theta_{i+1}\cdots\theta_{j-1} - 1)^2\right] (d\nu)$$

where $\bigwedge_{k=a}^{b} (1+\gamma_k)$ *is a nested sum. This was motivated by the suggested transformations used by Krishnaiah and Waikar [144] related to their equation 4.3.*

Proof. Starting with equation 6.19, change variables from $(l_1^2, \cdots, l_p^2)$ to $(\theta_1, \cdots \theta_p)$ where

$$\theta_i = f_{i,i+1} = \frac{l_i^2}{l_{i+1}^2}$$

$1 \leq i \leq p-1$ and $\theta_p = l_p^2$. The transformation matrix from $(l_1^2, \cdots, l_p^2)$ to $(\theta_1, \cdots \theta_p)$ is given by

$$T = \begin{pmatrix} \frac{1}{l_2^2} & & & & \\ & \frac{1}{l_3^2} & & & \\ & & \ddots & & \\ & & & \frac{1}{l_p^2} & \\ & & & & 1 \end{pmatrix}$$

The Jacobian, in terms of l_i^2, is given by $|\det T^{-1}| = \left|\frac{1}{l_1^2} \det D\right|$. In terms of θ_i, we find

$$\begin{array}{lll}
\theta_1 = \frac{l_1^2}{l_2^2} & l_1^2 = l_2^2\theta_1 & = \theta_p\theta_{p-1}\cdots\theta_2\theta_1 = w_1(\Theta) \\
\theta_2 = \frac{l_2^2}{l_3^2} & l_2^2 = l_3^2\theta_2 & = \theta_p\theta_{p-1}\cdots\theta_3\theta_2 = w_2(\Theta) \\
\vdots & \vdots & \vdots \\
\theta_{p-1} = \frac{l_{p-1}^2}{l_p^2} & l_{p-1}^2 = l_p^2\theta_{p-1} & = \theta_p\theta_{p-1} \\
\theta_p = l_p^2 & l_p^2 = \theta_p & = \theta_p
\end{array}$$

In this form, the Jacobian is given by

$$J = \left|\det\begin{bmatrix} \frac{\partial w_1}{\partial\theta_1} & \frac{\partial w_1}{\partial\theta_2} & \cdots & \frac{\partial w_1}{\partial\theta_p} \\ \frac{\partial w_2}{\partial\theta_1} & \frac{\partial w_2}{\partial\theta_2} & \cdots & \frac{\partial w_2}{\partial\theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial w_{p-1}}{\partial\theta_1} & \frac{\partial w_{p-1}}{\partial\theta_2} & \cdots & \frac{\partial w_{p-1}}{\partial\theta_p} \\ \frac{\partial w_p}{\partial\theta_1} & \frac{\partial w_p}{\partial\theta_2} & \cdots & \frac{\partial w_p}{\partial\theta_p} \end{bmatrix}\right|$$

$$= \left|\det\begin{bmatrix} \theta_2\theta_3\cdots\theta_p & \theta_1\theta_3\cdots\theta_p & \cdots & \theta_1\cdots\theta_{p-1} \\ 0 & \theta_3\cdots\theta_p & \cdots & \theta_2\cdots\theta_{p-1} \\ \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & & \theta_{p-1} \\ 0 & 0 & \cdots & 1 \end{bmatrix}\right|$$

$$J = (\theta_2\cdots\theta_p)(\theta_3\cdots\theta_p)(\theta_4\cdots\theta_p)\cdots(\theta_{p-1}\theta_p)\theta_p\cdot 1$$

$$= \theta_p^{p-1}\theta_{p-1}^{p-2}\theta_{p-2}^{p-3}\cdots\theta_3^2\theta_2 = \prod_{i=2}^{p}\theta_i^{i-1}$$

The joint probability density function of $\Theta = (\theta_1, \cdots, \theta_p)$ is given by

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\exp\left(-\sum_{i=1}^{p} \frac{1}{\lambda_i^2} \prod_{j=i}^{p} \theta_j\right)\right]$$

$$\times \left\{\prod_{i=1}^{p} \left[\prod_{j=i}^{p} \theta_j^{n-p}\right]\right\} \left\{\prod_{i<j}^{p} \left[\left(\prod_{k=i}^{p} \theta_k\right) - \left(\prod_{l=j}^{p} \theta_l\right)\right]^2\right\} \left\{\prod_{i=2}^{p} \theta_i^{i-1}\right\} (d\Theta)$$

Note that

$$\sum_{i=1}^{p} \frac{1}{\lambda_i^2} \prod_{j=i}^{p} \theta_j = \frac{1}{\lambda_1^2}\theta_1\theta_2\cdots\theta_p + \frac{1}{\lambda_2^2}\theta_2\theta_3\cdots\theta_p + \cdots + \frac{1}{\lambda_{p-1}^2}\theta_{p-1}\theta_p + \frac{1}{\lambda_p^2}\theta_p$$

$$= \theta_p[\frac{1}{\lambda_p^2} + \theta_{p-1}[\frac{1}{\lambda_{p-1}^2} + \theta_{p-2}[\frac{1}{\lambda_{p-2}^2} + \cdots + \theta_2[\frac{1}{\lambda_2^2} + \theta_1\frac{1}{\lambda_1^2}]\cdots]]]$$

$$= \frac{1}{\lambda_p^2}\theta_p[1 + \frac{\lambda_p^2}{\lambda_{p-1}^2}\theta_{p-1}[1 + \frac{\lambda_{p-1}^2}{\lambda_{p-2}^2}\theta_{p-2}[1 \cdots \frac{\lambda_3^2}{\lambda_2^2}\theta_2[1 + \frac{\lambda_2^2}{\lambda_1^2}\theta_1]\cdots]]]$$

$$\stackrel{\text{def}}{=} \frac{1}{\lambda_p^2}\theta_p \bigwedge_{k=p-1}^{1} [1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\theta_k]$$

Therefore

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\exp\left(-\frac{1}{\lambda_p^2}\theta_p \bigwedge_{k=p-1}^{1} \left[1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\theta_k\right]\right)\right]$$

$$\times \left\{\prod_{i=1}^{p} \theta_i^{n-p+i}\right\} \left\{\prod_{i=2}^{p} \theta_i^{i-1}\right\} \left\{\prod_{i<j}^{p} (\theta_j\theta_{j+1}\cdots\theta_p)^2 \, (\theta_i\theta_{i+1}\cdots\theta_{j-1} - 1)^2\right\} (d\Theta)$$

To obtain the joint density of

$$\left(\frac{l_1^2}{l_2^2}, \frac{l_2^2}{l_3^2}, \ldots, \frac{l_{p-1}^2}{l_p^2}\right) = (\theta_1, \cdots, \theta_{p-1})$$

we integrate $dF(\Theta)$ on $\theta_p \in (0, \infty)$. We temporarily simplify the notation to help identify the integration problem. Let

$$C = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)}$$

$$\beta = \frac{1}{\lambda_p^2} \bigwedge_{k=p-1}^{1} \left[1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\theta_k\right]$$

$$g_1 = \prod_{i=1}^{p-1} \theta_i^{n-p+i}$$

$$g_2 = \prod_{i=2}^{p-1} \theta_i^{i-1} = \prod_{i=1}^{p-1} \theta_i^{i-1}$$

$$g_1 g_2 = \prod_{i=1}^{p-1} \theta_i$$

$$g_3 = \left[\prod_{i<j}^{p-1} (\theta_j\theta_{j+1}\cdots\theta_{p-1})^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]\left[\prod_{i=1}^{p-1}(\theta_i\theta_{i+1}\cdots\theta_{p-1}-1)^2\right]$$

The idea is for these to be cofactors of θ_p. The justification of g_3 is not obvious, so we give a little more detail. Let

$$\Psi(\Theta) = \prod_{i=1}^{p-1}\prod_{j=i+1}^{p} (\theta_j\theta_{j+1}\cdots\theta_p)^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2$$

$$= \theta_p^2(\theta_{p-1}-1)^2 \prod_{i=1}^{p-2}\left[\prod_{j=i+1}^{p} (\theta_j\theta_{j+1}\cdots\theta_p)^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]$$

$$= \theta_p^2(\theta_{p-1}-1)^2 \prod_{i=1}^{p-2}\left\{\left[\prod_{j=i+1}^{p} (\theta_j\theta_{j+1}\cdots\theta_p)^2\right]\left[\prod_{j=i+1}^{p} (\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]\right\}$$

$$= \theta_p^2(\theta_{p-1}-1)^2 \prod_{i=1}^{p-2}\left\{\left[\theta_p^{2(p-i)}\prod_{j=i+1}^{p-1} (\theta_j\theta_{j+1}\cdots\theta_{p-1})^2\right]\left[\prod_{j=i+1}^{p} (\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]\right\}$$

At this point, we need another observation. Note that

$$\sum_{i=1}^{p-1}(p-i) = p(p-2) - \sum_{i=1}^{p-2} i = p(p-2) - \frac{(p-2)(p-1)}{2}$$

$$= (p-2)\left(\frac{2p-p+1}{2}\right) = \frac{(p+1)(p-2)}{2}$$

Now we see that

$$\Psi(\Theta) = \theta_p^2(\theta_{p-1}-1)^2\theta_p^{(p+1)(p-2)}$$

$$\times \prod_{i=1}^{p-2}\left\{\left[\prod_{j=i+1}^{p-1} (\theta_j\theta_{j+1}\cdots\theta_{p-1})^2\right]\left[\prod_{j=i+1}^{p} (\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]\right\}$$

The following minor bookkeeping will help us. $2+(p+1)(p-2) = 2+p^2-p-2 = p(p-1)$. We get

$$\Psi(\Theta) = \theta_p^{p(p-1)}(\theta_{p-1}-1)^2$$

$$\times \prod_{i=1}^{p-2}\left\{(\theta_i\theta_{i+1}\cdots\theta_{p-1}-1)^2\left[\prod_{j=i+1}^{p-1}(\theta_j\theta_{j+1}\cdots\theta_{p-1})^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]\right\}$$

$$= \theta_p^{p(p-1)}\left[\prod_{i=1}^{p-1}(\theta_i\theta_{i+1}\cdots\theta_{p-1}-1)^2\right]\left[\prod_{i<j}^{p-1}(\theta_j\theta_{j+1}\cdots\theta_{p-1})^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right]$$

Using c, β, g_1, g_2, g_3, we rewrite $dF(\Theta)$ as

$$dF(\Theta) = Ce^{-\beta\theta_p}\theta_p^n g_1\theta_p^{p-1}g_2\theta_p^{p(p-1)}g_3(d\Theta)$$

$$= Cg_1g_2g_3\theta_p^{n+(p-1)(p+1)}e^{-\beta\theta_p}d(\theta_1,\theta_2,\cdots,\theta_p)$$

We now integrate, using lemma 62.

$$\int_{\theta_p=0}^{\infty} dF(\Theta) = Cg_1g_2g_3d(\theta_1,\theta_2,\cdots,\theta_{p-1})\int_0^{\infty}\theta_p^{n+(p-1)(p+1)}e^{-\beta\theta_p}(d\theta_p)$$

At this point, some more bookkeeping helps us. $n+(p-1)(p+1)+1 = n+p^2-1+1 = n+p^2$. Thus

$$\int_{\theta_p=0}^{\infty} dF(\Theta) = Cg_1g_2g_3d(\theta_1,\theta_2,\cdots,\theta_{p-1})\beta^{-(n+p^2)}\Gamma(n+p^2)$$

To simplify notation, let $\nu = (\theta_1,\theta_2,\cdots,\theta_{p-1})$. Then

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(n+p^2)}{[\det\Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\left[\prod_{k=1}^{p-1}\left(1+\frac{\lambda_{k+1}^2}{\lambda_k^2}\theta_k\right)\right]^{-(n+p^2)}$$

$$\times\left[\prod_{i=1}^{p-1}\theta_i^{n+p-1+2i}\right]\left[\prod_{i<j}^{p-1}\left\{(\theta_j\theta_{j+1}\cdots\theta_{p-1})^2(\theta_i\theta_{i+1}\cdots\theta_{j-1}-1)^2\right\}\right](d\nu)$$

Substituting $\frac{l_i^2}{l_{i+1}^2} = \theta_i$, but *not* doing the change of variables, we can *compute* this as

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(n+p^2)}{[\det \Lambda^2]^n \, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\frac{1}{\lambda_p^2} \bigwedge_{k=p-1}^{1} \left(1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\frac{l_k^2}{l_{k+1}^2}\right)\right]^{-(n+p^2)}$$

$$\times \left[\prod_{i=1}^{p-1} \left(\frac{l_i^2}{l_{i+1}^2}\right)^{n+p-1+2i}\right] \left[\prod_{j=2}^{p-1} \left(\frac{l_j^2}{l_p^2}\right)^{2(j-1)}\right] \left[\prod_{i<j} \left\{\left(\frac{l_i^2}{l_j^2} - 1\right)\right\}\right] (d\nu)$$

We need a few more notes.

$$\prod_{i=1}^{p-1} \left(\frac{l_i^2}{l_{i+1}^2}\right)^{n+p-1+2i} = \frac{l_1^{2(n+p+1)}}{l_2^{2(n+p+1)}} \frac{l_2^{2(n+p+3)}}{l_3^{2(n+p+3)}} \cdots \frac{l_{p-1}^{2(n+3p-2)}}{l_p^{2(n+3p-2)}} = \frac{l_1^{2(n+p+1)} l_2^4 l_3^4 \cdots l_{p-1}^4}{l_p^{2(n+3p-2)}}$$

$$\frac{l_j^4}{l_{j+1}^4}\frac{l_{j+1}^4}{l_{j+2}^4} \cdots \frac{l_{p-1}^4}{l_p^4} = \frac{l_j^4}{l_p^4} \text{ implies } \prod_{i<j} \left(\frac{l_j^2}{l_p^2}\right)^2 = \prod_{j=2}^{p-1} \left(\frac{l_j^2}{l_p^2}\right)^{2(j-1)} = \frac{1}{l_p^{2(p^2-p-4)}} \prod_{j=2}^{p-1} l_j^{4(j-1)}$$

where

$$\sum_{j=2}^{p-1} 2(j-1) = 2[-1-1+\sum_{j=1}^{p-1} j] = 2[-2 + \frac{(p-1)p}{2}] = -4 + p^2 - p$$

and

$$\frac{l_i^2}{l_{l+1}^2}\frac{l_{i+1}^2}{l_{i+2}^2} \cdots \frac{l_{j-1}^2}{l_j^2} - 1 = \frac{l_i^2}{l_j^2} - 1$$

Just a little more bookkeeping, and we see

$$\frac{l_1^{2(n+p+1)} l_2^4 l_3^4 \cdots l_{p-1}^4}{l_p^{2(n+3p-2)}} \frac{1}{l_p^{2(p^2-p-4)}} l_2^4 l_3^8 l_4^{12} \cdots l_{p-1}^{2(2p-4)} = \frac{l_1^{2(n+p+1)} l_2^8 l_3^{12} l_4^{16} \cdots l_{p-1}^{4(p-1)}}{l_p^{2(n+p^2+2p-6)}}$$

Putting it all together, the joint probability density function for $\nu = (\theta_1, \cdots, \theta_{p-1})$ written in terms of $\theta_i = \frac{l_i^2}{l_{i+1}^2}$ is

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(n+p^2)}{[\det \Lambda^2]^n \, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\frac{1}{\lambda_p^2} \bigwedge_{k=p-1}^{1} \left(1 + \frac{\lambda_{k+1}^2}{\lambda_k^2}\frac{l_k^2}{l_{k+1}^2}\right)\right]^{-(n+p^2)}$$

$$\times \frac{l_1^{2(n+p-1)}}{l_p^{2(n+3p-6)}} \left[\prod_{k=1}^{p-1} \left(\frac{l_k^2}{l_p^2}\right)^{2k}\right] \left[\prod_{i<j}^{p-1} \left\{\left(\frac{l_i^2}{l_j^2} - 1\right)\right\}\right] (d\nu)$$

Note that when all the $\{\lambda_i^2\}$ are equal, then the second factor is

$$\lambda^{2(n+p^2)} \left[\bigwedge_{k=p-1}^{1} \left(1 + \frac{l_k^2}{l_{k+1}^2}\right)\right]^{-(n+p^2)}$$

Note that

$$\bigwedge_{k=p-1}^{1} \left(1 + \frac{l_k^2}{l_{k+1}^2}\right) = \left[1 + \frac{l_{p-1}^2}{l_p^2} \left[1 + \frac{l_{p-2}^2}{l_{p-1}^2} \left[\cdots \left[\underbrace{1 + \frac{l_3^2}{l_4^2} \underbrace{\left[1 + \frac{l_2^2}{l_3^2}\left[1 + \frac{l_1^2}{l_2^2}\right]\right]}_{\left[1+\frac{l_2^2}{l_3^2}+\frac{l_1^2}{l_3^2}\right]}}_{1+\frac{l_3^2}{l_4^2}+\frac{l_2^2}{l_4^2}+\frac{l_1^2}{l_4^2}} \right] \cdots \right] \right] \right]$$

$$= 1 + \sum_{k=p+1}^{1} \frac{l_k^2}{l_p^2} = 1 + \frac{1}{l_p^2} \sum_{k=1}^{p-1} l_k^2$$

So, the second factor collapses to

$$\lambda^{2(n+p^2)} \left[1 + \frac{1}{l_p^2} \sum_{k=1}^{p-1} l_k^2\right]^{-(n+p^2)} = \lambda^{2(n+p^2)} \left[1 + \frac{1}{l_p^2}\left(-l_p^2 + \operatorname{tr}(D)\right)\right]^{-(n+p^2)}$$

$$= \lambda^{2(n+p^2)\left[\frac{1}{l_p^2} \operatorname{tr} D\right]^{-(n+p^2)}} = (\lambda^2 l_p^2)^{n+p^2} [\operatorname{tr} D]^{-(n+p^2)}$$

So, under the null hypothesis that $\lambda^2 = \lambda_1^2 = \lambda_2^2 = \cdots = \lambda_p^2$, we get the density function of ν under the null hypothesis as

$$dF(\nu \mid \lambda_i^2 = \lambda^2) = \frac{\pi^{p(p-1)}\Gamma(n+p^2)}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)} \left(\frac{\lambda^2 l_p^2}{\operatorname{tr} D}\right)^{n+p^2}$$

$$\times \frac{l_1^{2(n+p+1)}}{l_p^{2(n+3p-6)}} \left[\prod_{k=1}^{p-1} \left(\frac{l_k^2}{l_p^2}\right)^{2k}\right] \left[\prod_{i<j}^{p-1} \left(\frac{l_i^2}{l_j^2} - 1\right)\right] (d\nu)$$

Krishnaiah and Waiker [144] consider simultaneously testing

$$H_{i,i+1} : \lambda_i^2 = \lambda_{i+1}^2 \quad \text{against}$$

$$A_{i,i+1} : \lambda_i^2 > \lambda_{i+1}^2$$

for $1 \leq i < p$. $H_{i,i+1}$ is accepted or rejected according to the comparison of $\frac{l_i^2}{l_{i+1}^2}$ to a suitably chosen critical value $c_{i\alpha}$ where

$$\Pr\left\{1 \leq \frac{l_i^2}{l_{i+1}^2} \leq c_{i\alpha}; 1 \leq i \leq p-1 \mid H\right\} = (1-\alpha)$$

The total hypothesis H is accepted if and only if all the component hypotheses $H_{i,i+1}$ are accepted. The power of the test is given by

$$1 - \Pr\left\{1 \leq \frac{l_i^2}{l_{i+1}^2} \leq c_{i\alpha}; 1 \leq i \leq p-1 \mid A\right\}$$

where $A = \bigcup_{i=1}^{p-1} A_{i,i+1}$.

The joint density $dF(\nu)$ is the appropriate function for computing the required critical values $\{c_{i\alpha}\}$. Notice that values for the $\{\lambda_i^2\}$ must be assumed. Krishnaiah and Waiker [144] provided the test distribution for the case of the real variable Wishart matrix. □

6.3.2 Joint Density of the Ratio of an Arbitrary Sample Eigenvalue to the Smallest Sample Eigenvalue

Proposition 2 *Let the sample eigenvalues* $D = \text{diag}(l_1^2, \cdots, l_p^2)$ *estimate the*

population eigenvalues $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ *and have the joint density function given by*

$$dF(D) = \left[\frac{|\det D|^{n-p}\,\pi^{p(p-1)}}{[\det \Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\sum_{k=1}^{p}\frac{l_k^2}{\lambda_k^2}\right]\left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dD)$$

Then the joint density of $\nu = (\theta_1, \cdots, \theta_{p-1})$, $\theta_i = \frac{l_i^2}{l_p^2}$ *is given by*

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np+2)}{[\det \Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\left[\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1}\frac{\theta_i}{\lambda_i^2}\right]^{-(np+2)}$$

$$\times\left[\prod_{i=1}^{p-1}\theta_i^{n-p}\right]\left[\prod_{i=1}^{p-2}\left[(\theta_i - 1)^2\prod_{j=i+1}^{p-1}(\theta_i - \theta_j)^2\right]\right](d\nu)$$

This was motivated by the transformations suggested by Krishnaiah and Waikar related to their equation 4.7 [144].

Proof. Change variables from $(l_1^2, \cdots, l_p^2)$ to $(\theta_1, \cdots, \theta_p)$ where $\theta_i = \frac{l_i^2}{l_p^2}$, $1 \le i \le p-1$ and $\theta_p = l_p^2$. To compute the Jacobian, we note that $l_i^2 = l_p^2\theta_i = \theta_p\theta_i = w_i(\Theta)$. Then

$$\left|\det\begin{pmatrix} \frac{\partial w_1}{\partial\theta_1} & \frac{\partial w_1}{\partial\theta_2} & \cdots & \frac{\partial w_1}{\partial\theta_p} \\ \frac{\partial w_2}{\partial\theta_1} & \frac{\partial w_2}{\partial\theta_2} & \cdots & \frac{\partial w_2}{\partial\theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_p}{\partial\theta_1} & \frac{\partial w_p}{\partial\theta_2} & \cdots & \frac{\partial w_p}{\partial\theta_p} \end{pmatrix}\right| = \left|\det\begin{pmatrix} \theta_p & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \theta_p & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}\right| = \theta_p^{p-1}$$

We also need to chase some messy subscripts and isolate θ_p since we want to integrate on θ_p. We tackle the messiest one first.

$$\prod_{i<j}^{p}(l_l^2 - l_j^2)^2 = \prod_{i=1}^{p-1}\prod_{j=i+1}^{p}(l_l^2 - l_j^2)^2 = (l_{p-1}^2 - l_p^2)^2\prod_{i=1}^{p-2}\prod_{j=i+1}^{p}(l_i^2 - l_j^2)^2$$

$$= (l_{p-1}^2 - l_p^2)^2 \prod_{i=1}^{p-2} (l_i^2 - l_p^2)^2 \prod_{j=i+1}^{p-1} (l_i^2 - l_j^2)^2$$

Now substitute in the new variables. We get

$$\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 = (\theta_{p-1}\theta_p - \theta_p)^2 \prod_{i=1}^{p-2} (\theta_i\theta_p - \theta_p)^2 \prod_{j=i+1}^{p-1} (\theta_i\theta_p - \theta_j\theta_p)^2$$

$$= \theta_p^2(\theta_{p-1} - 1)^2 \prod_{i=1}^{p-2} \theta_p^2(\theta_i - 1)^2 \prod_{j=i+1}^{p-1} \theta_p^2(\theta_i - \theta_j)^2$$

$$= \theta_p^2(\theta_{p-1} - 1)^2 \theta_p^{2(p-2)} \prod_{i=1}^{p-2} (\theta_i - 1)^2 \theta_p^{2(p-1-i)} \prod_{j=i+1}^{p-1} (\theta_i - \theta_j)^2$$

$$= \theta_p^2 \theta_p^{2(p-2)} \theta_p^{2(p-1)(p-2)} \theta_p^{-(p-2)(p-1)} \prod_{i=1}^{p-2} (\theta_i - 1)^2 \prod_{j=i+1}^{p-1} (\theta_i - \theta_j)^2$$

We engage is some bookkeeping. $2 + 2(p-2) + 2(p-1)(p-2) - (p-2)(p-1)$ $= 2(p-1) + (p-1)[2(p-2) - (p-2)] = 2(p-1) + (p-1)(p-2) =$ $(p-1)[2 + p - 2] = p(p-1)$. Therefore,

$$\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 = \theta_p^{p(p-1)} \prod_{i=1}^{p-2} (\theta_i - 1)^2 \prod_{j=i+1}^{p-1} (\theta_i - \theta_j)^2$$

Consider also

$$\prod_{i=1}^{p} l_i^{2(n-p)} = \theta_p^{n-p} \prod_{i=1}^{p-1} \theta_i^{n-p} \theta_p^{n-p} = \theta_p^{n-p} \theta_p^{(n-p)(p-1)} \prod_{i=1}^{p-1} \theta_i^{n-p} = \theta_p^{p(n-p)} \prod_{i=1}^{p-1} \theta_i^{n-p}$$

Within the exponential function,

$$\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2} = \frac{l_p^2}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{l_i^2}{\lambda_i^2} = \frac{\theta_p}{\lambda_p^2} + \theta_p \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2} = \theta_p \left[\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2} \right]$$

Putting it all together, with the Jacobian, we get

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \exp \left[-\theta_p \left(\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2} \right) \right]$$

$$\times \theta_p^{p(n-p)} \left[\prod_{i=1}^{p-1} \theta_i^{n-p}\right] \theta_p^{p(p-1)} \left[\prod_{i=1}^{p-2}(\theta_i - 1)^2 \prod_{j=i+1}^{p-1}(\theta_i - \theta_j)^2\right] \theta_p^{p-1}(d\Theta)$$

We need to collect the powers of θ_p.

$$p(n-p) + p(p-1) + (p-1) = p(n-1) + p + 1 = pn + 1$$

Then

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \exp\left[-\theta_p \left(\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2}\right)\right]$$

$$\times \theta_p^{np+1} \left[\prod_{i=1}^{p-1} \theta_i^{n-p}\right] \left[\prod_{i=1}^{p-2}(\theta_i - 1)^2 \prod_{j=i+1}^{p-1}(\theta_i - \theta_j)^2\right] (d\Theta)$$

We want to integrate out θ_p to obtain the joint density of ν. Using lemma 62 we see

$$\int_0^\infty \theta_p^{np+1} \exp(-\beta\theta_p) d\theta_p = \beta^{-(np+2)} \Gamma(np+2)$$

where $\mathrm{Re}(np+2) > 0$ and

$$\beta = \frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2}$$

The joint density of

$$\left(\frac{l_1^2}{l_p^2}, \frac{l_2^2}{l_p^2}, \cdots, \frac{l_{p-1}^2}{l_p^2}\right) = \nu = (\theta_1, \theta_2, \cdots, \theta_{p-1})$$

is given by

$$dF(\nu) = \frac{\pi^{p(p-1)} \Gamma(np+2)}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{\theta_i}{\lambda_i^2}\right]^{-(np+2)}$$

$$\times \left[\prod_{i=1}^{p-1} \theta_i^{n-p}\right] \left[\prod_{i=1}^{p-2}(\theta_i - 1)^2 \prod_{j=i+1}^{p-1}(\theta_i - \theta_j)^2\right] (d\nu)$$

I want to rewrite this in terms of l_i^2, but not do a change of variables. First, some details.

$$\frac{1}{\lambda_p^2} + \sum_{i=1}^{p-1} \frac{1}{\lambda_i^2} \frac{l_i^2}{l_p^2} = \frac{1}{\lambda_p^2} \frac{l_p^2}{l_p^2} + \sum_{i=1}^{p-1} \frac{1}{\lambda_i^2} \frac{l_i^2}{l_p^2} = \frac{1}{l_p^2} \sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}$$

$$\prod_{i=1}^{p-1} \theta_i^{n-p} = \prod_{i=1}^{p-1} \left(\frac{l_i^2}{l_p^2} \right)^{n-p} = l_p^{-2(p-1)(n-p)} \prod_{i=1}^{p-1} l_i^{2(n-p)}$$

$$\prod_{i=1}^{p-2} (\theta_i - 1)^2 = \prod_{i=1}^{p-2} \left(\frac{l_i^2}{l_p^2} - 1 \right)^2 = l_p^{-4(p-2)} \prod_{i=1}^{p-2} \left(l_i^2 - l_p^2 \right)^2$$

$$\prod_{i=1}^{p-2} \prod_{j=i+1}^{p-1} (\theta_i - \theta_j)^j = \prod_{i=1}^{p-2} \prod_{j=i+1}^{p-1} \left(\frac{l_i^2}{l_p^2} - \frac{l_j^2}{l_p^2} \right)^2 = \prod_{i=1}^{p-2} l_p^{-4(p-1-i)} \prod_{j=i+1}^{p-1} \left(l_i^2 - l_j^2 \right)^2$$

$$= \prod_{i=1}^{p-2} l_p^{-4(p-1)} l_p^{4i} \prod_{j=i+1}^{p-1} \left(l_i^2 - l_j^2 \right)^2 = l_p^{-4(p-1)(p-2)} \prod_{i=1}^{p-2} l_p^{4i} \prod_{j=i+1}^{p-1} \left(l_i^2 - l_j^2 \right)^2$$

$$= l_p^{-4(p-1)(p-2)} l_p^{2(p-2)(p-1)} \prod_{i<j}^{p-1} (l_i^2 - l_j^2)^2 = l_p^{-2(p-1)(p-2)} \prod_{i<j}^{p-1} (l_i^2 - l_j^2)^2$$

Substituting back into the joint density function, we get

$$dF(\nu) = \frac{\pi^{p(p-1)} \Gamma(np+2)}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n) \mathrm{C}\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2} \right]^{-(np+2)} l_p^{2(np+2)} l_p^{-2(p-1)(n-p)}$$

$$\times \left[\prod_{i=1}^{p-1} l_i^{2(n-p)} \right] l_p^{-4(p-2)} \left[\prod_{i=1}^{p-2} (l_i^2 - l_p^2)^2 \right] l_p^{-2(p-1)(p-2)} \left[\prod_{i<j}^{p-1} (l_i^2 - l_j^2)^2 \right] (d\nu)$$

Now to collect the powers of l_p^2.

$$np + 2 - (p-1)(n-p) - 2(p-2) - (p-1)(p-2)$$

$$= np + 2 - (p-1)(n-2) - 2(p-2)$$

$$= np + 2 - np + n + 2p - 2 - 2p + 4$$

$$= n + 4$$

The joint density of

$$\left(\frac{l_1^2}{l_p^2}, \ldots, \frac{l_{p-1}^2}{l_p^2} \right)$$

is

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np+2)l_p^{2(n+4)}}{[\det \Lambda^2]^n \, C\Gamma_p(n)C\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right]^{-(np+2)}$$
$$\times \left(\frac{|\det D|}{l_p^2}\right)^{n-p} \left[\prod_{i=1}^{p-2}(l_i^2 - l_p^2)^2\right] \left[\prod_{i<j}^{p-1}(l_i^2 - l_j^2)^2\right] (d\nu)$$

We can do just a little more collapsing of terms.

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np+2)l_p^{2(n+4)}}{[\det \Lambda^2]^n \, C\Gamma_p(n)C\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right]^{-(np+2)}$$
$$\cdot \left[|\det \cdot|^{n-p}\right] \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right] (d\nu)$$

When we select the null hypothesis $H_{i,p} : \lambda_i^2 = \lambda_p^2$ for all $i < p$, which is $\lambda^2 = \lambda_1^2 = \lambda_2^2 = \cdots = \lambda_p^2$, we get

$$dF(\nu \mid \lambda_i^2 = \lambda^2 \text{ for all } i) = \frac{\pi^{p(p-1)}\Gamma(np+2)\lambda^{np+2}l_p^{2(p+4)}}{[\det \Lambda^2]^n \, C\Gamma_p(n)C\Gamma_p(p)}$$
$$\times \frac{|\det D|^{n-p}}{[\mathrm{tr}\, D]^{np+2}} \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right] (d\nu)$$

The alternate hypothesis is $A_{i,p} : \lambda_i^2 > \lambda_p^2$ for all $i < p$.

We follow Krishnaiah and Waikar [144] in constructing the test. We test all $\{H_{i,p}\}$ against all alternatives $\{A_{i,p}\}$. We accept or reject $H_{i,p}$ for $1 \le i \le p-1$ according to the comparison of the test statistic $\frac{l_i^2}{l_p^2}$ to the critical value $C_{i\alpha}$ where

$$\Pr\{1 \le \frac{l_i^2}{l_p^2} \le C_{i\alpha},\ 1 \le i \le p-1 \mid H\} = (1-\alpha)$$

The total hypothesis H is accepted if each individual hypothesis $H_{i,p}$ is accepted. The power of the test is

$$1 - \Pr\{1 \le \frac{l_i^2}{l_p^2} \le C_{i\alpha},\ 1 \le i \le p-1 \mid A\}$$

where $A = \bigcup_{i=1}^{p-1} A_{i,p}$.

The joint density of $dF(\nu)$ is the appropriate function for computing the required critical values $\{C_{i\alpha}\}$. Notice that λ_i^2 must be assumed. Krishnaiah and Waikar provide the test distribution for the case of the real variable Wishart matrix. □

6.3.3 Joint Density of Ratio of Sample Eigenvalues to Largest Sample Eigenvalue

Proposition 3 *Let the sample eigenvalues* $D = \text{diag}(l_1^2, \cdots, l_p^2)$ *estimate the population eigenvalues* $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ *and have the joint density function given by*

$$dF(D) = \left[\frac{|\det D|^{n-p}\, \pi^{p(p-1)}}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\sum_{k=1}^{p} \frac{l_k^2}{\lambda_k^2}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD)$$

Then the joint density of

$$\nu = (\theta_2, \cdots, \theta_p), \theta_i = \frac{l_i^2}{l_1^2}$$

is given by

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2}\right]^{np}$$

$$\times \left[\prod_{j=2}^{p} (1-\theta_j)^2 \theta_j^{n-p}\right] \left[\prod_{2=i<j}^{p} (\theta_i - \theta_j)^2\right] (d\nu)$$

This was motivated by the transformations suggested by Krishnaiah and Waikar [145].

Proof. Change variables from $(l_1^2, \cdots, l_p^2)$ to $(\theta_1, \cdots, \theta_p)$ where $\theta_i = \frac{l_i^2}{l_1^2}$, $2 \leq i \leq p$ and $\theta_1 = l_1^2$. To compute the Jacobian, we note that

$$l_i^2 = l_1^2 \theta_i = \theta_1 \theta_i = w_i(\Theta)$$

Then

$$J = \left| \det \begin{pmatrix} \frac{\partial w_1}{\partial \theta_1} & \frac{\partial w_1}{\partial \theta_2} & \cdots & \frac{\partial w_1}{\partial \theta_p} \\ \frac{\partial w_2}{\partial \theta_1} & \frac{\partial w_2}{\partial \theta_2} & \cdots & \frac{\partial w_2}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_p}{\partial \theta_1} & \frac{\partial w_p}{\partial \theta_2} & \cdots & \frac{\partial w_p}{\partial \theta_p} \end{pmatrix} \right| = \left| \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \theta_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \theta_1 \end{pmatrix} \right| = \theta_1^{p-1}$$

We do some subscript chasing to prepare for cleanly expressing the joint density of the new variables Θ, and to ease the integration over all θ_1.

$$\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (l_i^2 - l_j^2)^2 = \left[\prod_{j=2}^{p} (l_1^2 - l_j^2)^2 \right] \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (l_i^2 - l_j^2)^2$$

Now substitute the new variables. We get

$$\left[\prod_{j=2}^{p} (\theta_1 - \theta_1 \theta_j)^2 \right] \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_1 \theta_i - \theta_1 \theta_j)^2 = \left[\prod_{j=2}^{p} \theta_1^2 (1 - \theta_j)^2 \right] \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} \theta_1^2 (\theta_i - \theta_j)^2$$

$$= \theta_1^{2(p-1)} \left[\prod_{j=2}^{p} (1 - \theta_j)^2 \right] \left[\prod_{i=2}^{p-1} \theta_1^{2(p-i)} \right] \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_i - \theta_j)^2 \right]$$

$$= \theta_1^{2(p-1)} \left[\prod_{j=2}^{p} (1 - \theta_j)^2 \right] \left[\prod_{i=2}^{p-1} \theta_1^{2p} \right] \left[\prod_{i=2}^{p-1} \theta_1^{-2i} \right] \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_i - \theta_j)^2 \right]$$

$$= \theta_1^{2(p-1)} \theta_1^{2p(p-2)} \theta_1^{-(p-1)p+2} \left[\prod_{j=2}^{p} (1 - \theta_j)^2 \right] \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_i - \theta_j)^2 \right]$$

$$= \theta_1^{p(p-1)} \left[\prod_{j=2}^{p} (1 - \theta_j)^2 \right] \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_i - \theta_j)^2 \right]$$

$$= \theta_1^{p(p-1)} \left[\prod_{j=2}^{p}(1-\theta_j)^2\right] \left[\prod_{2=i<j}^{p}(\theta_i-\theta_j)^2\right]$$

For completeness and consistency of mistakes, we look at

$$\prod_{i=1}^{p} l_i^{2(n-p)} = l_1^{2(n-p)} \prod_{i=2}^{p} l_i^{2(n-p)} = \theta_1^{n-p} \prod_{i=2}^{p} \theta_1^{n-p}\theta_i^{n-p}$$

$$= \theta_1^{n-p}\theta_1^{(n-p)(p-1)} \prod_{i=2}^{p} \theta_i^{n-p} = \theta_1^{p(n-p)} \prod_{i=2}^{p} \theta_i^{n-p}$$

$$\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2} = \left(\sum_{i=2}^{p} \theta_1 \frac{\theta_i}{\lambda_i^2}\right) + \frac{\theta_1}{\lambda_1^2} = \theta_1 \left[\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2}\right]$$

Putting this all together with the Jacobian gives us

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n} \frac{\exp\left[-\theta_1\left(\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2}\right)\right]}{\mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)} \theta_1^{p(n-p)} \left[\prod_{i=2}^{p} \theta_i^{n-p}\right]$$

$$\times \theta_1^{p(p-1)} \left[\prod_{j=2}^{p}(1-\theta_j)^2\right] \left[\prod_{2=i<j}^{p}(\theta_i-\theta_j)^2\right] \theta_1^{p-1}(d\Theta)$$

We do some more bookkeeping to gather the θ_1 terms.

$$p(n-p) + p(p-1) + (p-1) = p(n-1) + p - 1 = np - 1$$

Thus

$$dF(\Theta) = \frac{\pi^{p(p-1)}\theta_1^{p(p-1)}}{[\det \Lambda^2]^n\, \mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)} \exp\left[-\theta_1\left(\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2}\right)\right]$$

$$\times \left[\prod_{j=2}^{p} \theta_j^{n-p}\right] \left[\prod_{j=2}^{p}(1-\theta_j)^2\right] \left[\prod_{2=i<j}^{p}(\theta_i-\theta_j)^2\right] (d\Theta)$$

To get the joint density of a test statistic, we integrate out θ_1. Using lemma 62,

$$\int_0^\infty \theta_1^{np-1} e^{-\beta\theta_1} d\theta_1 = \beta^{np}\Gamma(np)$$

where $\mathrm{Re}(np) > 0$ and

$$\beta = \frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2}$$

Using this result, the joint density function of

$$\nu = (\theta_2, \cdots, \theta_p) = \left(\frac{l_2^2}{l_1^2}, \frac{l_3^2}{l_1^2}, \cdots, \frac{l_p^2}{l_1^2} \right)$$

is given by

$$dF(\nu) = \frac{\pi^{p(p-1)} \Gamma(np)}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2} \right]^{np}$$
$$\times \left[\prod_{j=2}^{p} (1-\theta_j)^2 \theta_j^{n-p} \right] \left[\prod_{2=i<j}^{p} (\theta_i - \theta_j)^2 \right] (d\nu)$$

I want to write this in terms of l_i^2, without doing a change of variables.

This is for obtaining a computation form in terms of the original variables.

$$\frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{\theta_i}{\lambda_i^2} = \frac{1}{\lambda_1^2} + \sum_{i=2}^{p} \frac{1}{\lambda_i^2} \frac{l_i^2}{l_1^2} = \frac{1}{\lambda_1^2} \frac{l_1^2}{l_1^2} + \frac{1}{l_1^2} \sum_{i=2}^{p} \frac{l_i^2}{\lambda_i^2} = \frac{1}{l_1^2} \sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}$$

$$\prod_{j=2}^{p} (1-\theta_j)^2 = \prod_{j=2}^{p} (1 - \frac{l_j^2}{l_1^2})^2 = \prod_{j=2}^{p} l_1^{-4} (l_1^2 - l_j^2)^2 = l_1^{-4(p-1)} \prod_{j=2}^{p} (l_1^2 - l_j^2)^2$$

$$\prod_{j=2}^{p} \theta_j^{n-p} = \prod_{j=2}^{p} \left(\frac{l_j^2}{l_1^2} \right)^{n-p} = l_1^{-2(n-p)(p-1)} \prod_{j=2}^{p} l_j^{2(n-p)}$$

$$\prod_{2=i<j}^{p} (\theta_i - \theta_j)^2 = \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} \left(\frac{l_i^2}{l_1^2} - \frac{l_j^2}{l_1^2} \right)^2 = \prod_{i=2}^{p-1} l_1^{-2(p-i)} \prod_{j=i+1}^{p} (l_i^2 - l_j^2)^2$$

$$= \left[\prod_{i=2}^{p-1} l_1^{-4p} \right] \left[\prod_{i=2}^{p-1} l_1^{4i} \right] \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (l_i^2 - l_j^2)^2 = l_1^{-4p(p-2)} l_1^{2[(p-1)p-2]} \prod_{2=i<j}^{p} (l_i^2 - l_j^2)^2$$

$$= l_1^{2(-p^2+3p-2)} \prod_{2=i<j}^{p} (l_i^2 - l_j^2)^2$$

Putting it all together, we get the density function for

$$\left(\frac{l_2^2}{l_1^2}, \frac{l_3^2}{l_1^2}, \cdots, \frac{l_p^2}{l_1^2} \right)$$

in terms of l_i^2 as

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np) l_1^{-2np}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right] l_1^{-4(p-1)} \left[\prod_{j=2}^{p} (l_1^2 - l_j^2)^2\right]$$

$$\times \, l_1^{-2(n-p)(p-1)} \left[\prod_{j=2}^{p} l_j^{2(n-p)}\right] l_1^{2(-p^2+3p-2)} \left[\prod_{2=i<j}^{p} (l_i^2 - l_j^2)^2\right] (d\nu)$$

Collect the exponents of l_1^2. We get

$$-np - 2(p-1) - (n-p)(p-1) - p^2 + 3p - 2$$

$$= -np - 2p + 2 - np + p^2 + n - p - p^2 + 3p - 2$$

$$= -2np + n = n(1 - 2p)$$

Consolidating the l_1^2 terms gives us

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np) l_1^{2n(1-2p)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right]^{np}$$

$$\times \left[\prod_{j=2}^{p} (l_1^2 - l_j^2)^2 l_j^{2(n-p)}\right] \left[\prod_{2=i<j}^{p} (l_i^2 - l_j^2)^2\right] (d\nu)$$

$$= \frac{\pi^{p(p-1)}\Gamma(np) l_1^{2n(1-2p)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right]^{np} \left[\prod_{j=2}^{p} l_j^{2(n-p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (d\nu)$$

We have one more slight opportunity to economize on notation. $\prod_{j=2}^{p} l_j^{2(n-p)} = \left(\frac{|\det D|}{l_1^2}\right)^{n-p}$ and $n - 2np - n + p = p(1 - 2n)$. This gives us

$$dF(\nu) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \frac{|\det D|^{n-p}}{l_1^{2p(2n-1)}} \left[\sum_{i=1}^{p} \frac{l_i^2}{\lambda_i^2}\right]^{np} \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (d\nu)$$

When we select the null hypothesis $H_{i,1} : \lambda_1^2 = \lambda_i^2$ for all i, which is the same as $\lambda^2 = \lambda_1^2 = \lambda_2^2 = \cdots = \lambda_p^2$, we get

$$dF(\nu \mid \lambda_1^2 = \lambda_i^2 \text{ for all } i)$$

$$= \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n \, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \frac{|\det D|^{n-p}\,[\mathrm{tr}\, D]^{np}}{l_1^{2p(2n-1)}\lambda^{2np}} \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](d\nu)$$

$$= \frac{\pi^{p(p-1)}\Gamma(np)}{\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \frac{|\det D|^{n-p}\,[\mathrm{tr}\, D]^{np}}{l_1^{2p(2n-1)}\lambda^{4np}} \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](d\nu)$$

The alternate hypothesis is $A_{i,1} : \lambda_i^2 < \lambda_1^2$ for all $2 \leq i \leq p$.

We test all $\{H_{i,1}\}$ simultaneously against all alternatives $\{A_{i,1}\}$. We accept or reject $H_{i,1}$ for $2 \leq i \leq p$ according to the comparison of the test statistic $\frac{l_i^2}{l_1^2}$ against the critical value $C_{i\alpha}$ which is appropriately chosen for the desired significance level α such that

$$\Pr\{C_{i\alpha} \leq \frac{l_i^2}{l_1^2} \leq 1, 2 \leq i \leq p \mid H\} = (1 - \alpha)$$

The total hypothesis H is accepted if each individual hypothesis $H_{i,1}$ is accepted. The power of the test is

$$1 - \Pr\{C_{i\alpha} \leq \frac{l_i^2}{l_1^2} \leq 1, 2 \leq i \leq p \mid A\}$$

where $A = \bigcup_{i=2}^{p} A_{i,1}$.

The joint density $dF(\nu)$ is the appropriate function for computing the required critical values $\{C_{i\alpha}\}$. Notice that $\{\lambda_i^2\}$ must be assumed. □

6.3.4 Joint Density of Ratio of Arbitrary Sample Eigenvalue to Trace of Sample Covariance Matrix

Proposition 4 *Let the sample eigenvalues* $D = \mathrm{diag}(l_1^2, \cdots, l_p^2)$ *estimate the population eigenvalues* $\Lambda^2 = \mathrm{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ *and have the joint density func-*

tion given by

$$dF(D) = \left[\frac{|\det D|^{n-p}\,\pi^{p(p-1)}}{[\det \Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\sum_{k=1}^{p}\frac{l_k^2}{\lambda_k^2}\right]\left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dD)$$

Then the joint density of $\Psi = (\theta_2, \cdots, \theta_p)$, $\theta_i = \frac{l_i^2}{\operatorname{tr} D}$ *is given by*

$$dF(\Psi) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)e}\left[\sum_{i=2}^{p}\theta_i\left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2}\right)\right]^{-np}$$

$$\times\,[(1 - \theta_2 - \cdots - \theta_p)\theta_2 \cdots \theta_p]^{n-p}$$

$$\times\left[\prod_{j=2}^{p}(1 - \theta_2 - \cdots - \theta_{j-1} - 2\theta_j - \theta_{j+1} - \cdots - \theta_p)^2\right]\left[\prod_{2=i<j}^{p}(\theta_i - \theta_j)^2\right](d\Psi)$$

This was motivated by the transformations suggested by Krishnaiah and Schurmann [151].

Proof. The results and proof in [151] are for a complex Wishart matrix distributed as $\mathbf{C}W_p\left(\frac{n+p-1}{2}, I_p\right)$. Change variables from $(l_1^2, \cdots, l_p^2)$ to $(\nu_1, \cdots, \nu_p)$ where $\nu_1 = \sum_{i=1}^{p} l_i^2$ and $\nu_k = l_k^2$ for $2 \le k \le p$.

$$\begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_p \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}\begin{pmatrix} l_1^2 \\ l_2^2 \\ \vdots \\ l_p^2 \end{pmatrix}$$

or $N = BL$. The Jacobian is $\det B^{-1} = 1$.

To form the density function of $N = (\nu_1, \cdots, \nu_p)$ more easily, let us do some bookkeeping first.

$$\sum_{i=1}^{p}\frac{l_i^2}{\lambda_i^2} = \frac{\nu_1 - \nu_2 - \cdots - \nu_p}{\lambda_1^2} + \sum_{i=2}^{p}\frac{\nu_i}{\lambda_i^2} = 1 + \nu_2\left(\frac{1}{\lambda_2^2} - \frac{1}{\lambda_1^2}\right) + \cdots + \nu_p\left(\frac{1}{\lambda_p^2} - \frac{1}{\lambda_1^2}\right)$$

$$= 1 + \sum_{i=2}^{p} \nu_i \left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2} \right)$$

$$\prod_{i=1}^{p} l_i^2 = (\nu_1 - \nu_2 - \cdots - \nu_p)(\nu_2 \nu_3 \cdots \nu_p)$$

$$\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 = \left[\prod_{j=2}^{p} (l_1^2 - l_j^2)^2 \right] \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (l_i^2 - l_j^2)^2$$

$$l_1^2 - l_j^2 = \nu_1 - \nu_2 - \cdots - \nu_{j-1} - 2\nu_j - \nu_{j+1} - \cdots - \nu_p, \text{ for } 2 \leq j \leq p$$

$$l_i^2 - l_j^2 = \nu_i - \nu_j, \text{ for } i \neq 1$$

Substituting the new variables and including the Jacobian, we get

$$dF(N) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} \left\{ \exp \left[-1 - \sum_{i=2}^{p} \nu_i \left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2} \right) \right] \right\}$$

$$\times \left[\nu_1 \nu_2 \cdots \nu_p - (\nu_2 + \cdots + \nu_p) \nu_2 \nu_3 \cdots \nu_p \right]^{n-p}$$

$$\times \left[\prod_{j=2}^{p} (\nu_1 - \nu_2 - \cdots - \nu_{j-1} - 2\nu_j - \nu_{j+1} - \cdots - \nu_p)^2 \right]$$

$$\times \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\nu_i - \nu_j)^2 \right] \cdot 1 \cdot (dN)$$

The next step is again a change of variables. Let $\theta_1 = \nu_1$ and let $\theta_k = \frac{\nu_k}{\nu_1}$ for $2 \leq k \leq p$. Then $\nu_k = \nu_1 \theta_k = w_k(\Theta)$. The Jacobian is found by

$$\left| \det \begin{pmatrix} \frac{\partial w_1}{\partial \theta_1} & \frac{\partial w_1}{\partial \theta_2} & \cdots & \frac{\partial w_1}{\partial \theta_p} \\ \frac{\partial w_2}{\partial \theta_1} & \frac{\partial w_2}{\partial \theta_2} & \cdots & \frac{\partial w_2}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial w_p}{\partial \theta_1} & \frac{\partial w_p}{\partial \theta_2} & \cdots & \frac{\partial w_p}{\partial \theta_p} \end{pmatrix} \right| = \left| \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \theta_2 & \theta_1 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ \theta_p & \vdots & 0 & \theta_1 \end{pmatrix} \right| = \theta_1^{p-1}$$

To make the work easier, we do some more bookkeeping.

$$\exp \left[-1 - \sum_{i=2}^{p} \nu_i \left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2} \right) \right] = \exp \left[-1 - \theta_1 \sum_{i=2}^{p} \theta_i \left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2} \right) \right]$$

$$= \frac{1}{e} \exp[-\alpha\theta_1]$$

where

$$\alpha = \sum_{i=2}^{p} \theta_i \left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2} \right)$$

Then

$$\nu_1\nu_2 \cdots \nu_p - (\nu_2 + \cdots + \nu_p)\nu_2\nu_3 \cdots \nu_p = \theta_1^p\theta_2 \cdots \theta_p - \theta_1^p(\theta_2 + \cdots + \theta_p)\theta_2 \cdots \theta_p$$

$$= \theta_1^p(1 - \theta_2 - \cdots - \theta_p)\theta_2 \cdots \theta_p$$

$$\nu_1 - \nu_2 - \cdots - \nu_{j-1} - 2\nu_j - \nu_{j+1} - \cdots - \nu_p = \theta_1[1 - \theta_2 - \cdots \theta_{j-1} - 2\theta_j - \theta_{j+1} - \cdots - \theta_p]$$

$$\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\nu_i - \nu_j) = \prod_{i=2}^{p-1} \prod_{j=i+1}^{p} \theta_1(\theta_i - \theta_j)$$

We have effectively isolated all terms of θ_1. Collect these terms with the powers they are raised to in the density function. We get

$$\theta_1^{p(n-p)}\theta_1^{2(p-1)}\theta_1^{2\cdot\frac{1}{2}(p-1)(p-2)}\theta_1^{p-1}$$

where the last factor is the Jacobian of the transformation. Then

$$p(n-p) + 2(p-1) + (p-1)(p-2) + (p-1) = p(n-p) + (p-1)(2 + p - 2 + 1)$$

$$= pn - p^2 + p^2 - 1 = np - 1$$

Collect all the terms to get the density function of $\Theta = (\theta_1, \cdots, \theta_p)$.

$$dF(\Theta) = \frac{\pi^{p(p-1)}}{[\det \Lambda^2]^n \, C\Gamma_p(n) C\Gamma_p(p)} [(1 - \theta_2 - \cdots - \theta_p)\theta_2 \cdots \theta_p]^{n-p}$$

$$\times \left[\prod_{j=2}^{p} (1 - \theta_2 - \cdots \theta_{j-1} - 2\theta_j - \theta_{j+1} - \cdots - \theta_p)^2 \right] \left[\prod_{i=2}^{p-1} \prod_{j=i+1}^{p} (\theta_i - \theta_j)^2 \right]$$

$$\times \frac{1}{e}\theta_1^{np-1}e^{-\alpha\theta_1}(d\Theta)$$

where

$$\alpha = \sum_{i=2}^{p}\theta_i\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right)$$

Our goal is achieved by integrating out θ_1 to get the marginal density of $\Psi = (\theta_2, \cdots, \theta_p)$. Using lemma 62, we see that

$$\int_0^\infty \theta_1^{np-1}e^{-\alpha\theta_1}d\theta_1 = \alpha^{-np}\Gamma(np) = \alpha^{-np}\left[(np-1)!\right]$$

Therefore, the density of Ψ is given by

$$dF(\Psi) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det\Lambda^2]^n\,\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\left[\sum_{i=2}^{p}\theta_i\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right)\right]^{-np}$$

$$\times\left[(1-\theta_2-\cdots-\theta_p)\theta_2\cdots\theta_p\right]^{n-p}$$

$$\times\left[\prod_{j=2}^{p}(1-\theta_2-\cdots\theta_{j-1}-2\theta_j-\theta_{j+1}-\cdots-\theta_p)^2\right]\left[\prod_{i=2}^{p-1}\prod_{j=i+1}^{p}(\theta_i-\theta_j)^2\right](d\Psi)$$

We want to know how to compute this in terms of our original variables $(l_1^2, \cdots, l_p^2)$. We do some more bookkeeping.

$$\sum_{i=2}^{p}\theta_i\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right) = \sum_{i=2}^{p}\frac{l_i^2}{\operatorname{tr}D}\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right) = \frac{1}{\operatorname{tr}D}\sum_{i=2}^{p}l_i^2\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right)$$

$$(1-\theta_2-\cdots-\theta_p)\theta_2\cdots\theta_p = \left(1-\frac{l_2^2}{\operatorname{tr}D}-\cdots-\frac{l_p^2}{\operatorname{tr}D}\right)\frac{l_2^2}{\operatorname{tr}D}\frac{l_3^2}{\operatorname{tr}D}\cdots\frac{l_p^2}{\operatorname{tr}D}$$

$$= \left(\frac{1}{\operatorname{tr}D}\right)\left(\operatorname{tr}D-l_2^2-\cdots-l_p^2\right)\left(\frac{1}{\operatorname{tr}D}\right)^{p-1}(l_2^2\cdots l_p^2) = \left(\frac{1}{\operatorname{tr}D}\right)^p l_2^2\cdots l_p^2$$

$$= \frac{\det D}{[\operatorname{tr}D]^p}$$

$$1-\theta_2-\cdots-\theta_{j-1}-2\theta_j-\theta_{j+1}-\cdots-\theta_p$$

$$= \frac{\operatorname{tr} D}{\operatorname{tr} D} - \frac{l_2^2}{\operatorname{tr} D} - \cdots - \frac{l_{j-1}^2}{\operatorname{tr} D} - 2\frac{l_j^2}{\operatorname{tr} D} - \frac{l_{j+1}^2}{\operatorname{tr} D} - \cdots - \frac{l_p^2}{\operatorname{tr} D}$$

$$= \frac{1}{\operatorname{tr} D}\left[\operatorname{tr} D - l_2^2 - \cdots - l_{j-1}^2 - 2l_j^2 - l_{j+1}^2 - \cdots - l_p^2\right]$$

$$= \frac{1}{\operatorname{tr} D}\left[\operatorname{tr} D - (l_j^2 + \operatorname{tr} D - l_1^2)\right] = \frac{1}{\operatorname{tr} D}\left[l_1^2 - l_j^2\right]$$

$$\theta_i - \theta_j = \frac{l_i^2}{\operatorname{tr} D} - \frac{l_j^2}{\operatorname{tr} D} = \frac{1}{\operatorname{tr} D}\left(l_1^2 - l_j^2\right)$$

We collect the powers of $(\operatorname{tr} D)$ as a final bookkeeping task. To simplify, let $x = \operatorname{tr} D$. Then we have

$$x^{np}(\operatorname{tr} D)^{n-p} x^{-p(n-p)} x^{-2(p-1)} \left[\prod_{j=2}^{p}\left(l_1^2 - l_j^2\right)^2\right] x^{-\frac{1}{2}2(p-2)(p-1)} \left[\prod_{i=2}^{p-1}\prod_{j=i+1}^{p}\left(l_i^2 - l_j^2\right)^2\right]$$

Then

$$np - p(n-p) - 2(p-1) - (p-2)(p-1) = np - p(n-p) - (p-1)(2+p-2)$$

$$= np - p(n-p) - p(p-1) = np - np + p^2 - p^2 + p = p$$

This gives us

$$dF(\Psi) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n \, \mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)e}\left[\sum_{i=2}^{p} l_i^2\left(\frac{1}{\lambda_i^2} - \frac{1}{\lambda_1^2}\right)\right]^{-np}$$

$$\times \; [\operatorname{tr} D]^p \, [\det D]^{n-p}\left[\prod_{1=i<j}^{p}(l_i^2 - l_j^2)^2\right](d\Psi)$$

The idea to seek the joint density of

$$\left(\frac{l_2^2}{\operatorname{tr} D}, \cdots, \frac{l_p^2}{\operatorname{tr} D}\right)$$

was motivated by Krishnaiah and Schuurmann's suggestion to perform the change of variables $u_i = \frac{l_i^2}{\operatorname{tr} D}$ for $1 \leq i \leq p$.

Under the hypothesis

$$\bigcup_{i=2}^{p} H_{i,1} : \lambda_i^2 = \lambda_1^2$$

we see that $dF(\Psi) = 0$. The alternative hypothesis is given by

$$\bigcap_{i=2}^{p} A_{i,1} : \lambda_i^2 < \lambda_1^2$$

□

Corollary 4 . *Let the sample eigenvalues* $D = \text{diag}(l_1^2, \cdots, l_p^2)$ *estimate the population eigenvalues* $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ *be nonsingular such that* $\lambda_1^2 \neq \lambda_k^2$ *for* $k \geq 2$. *Let* D *have the joint density function given by*

$$dF(D) = \left[\frac{|\det D|^{n-p}\, \pi^{p(p-1)}}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\right] \exp\left[-\sum_{k=1}^{p} \frac{l_k^2}{\lambda_k^2}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD)$$

Let $\theta_1 = \text{tr}\, D$ *and* $\theta_k = \frac{l_k^2}{\text{tr}\, D}$ *for* $k \geq 2$. *Let*

$$\alpha = \sum_{k=2}^{p} \theta_k \left(\frac{1}{\lambda_k^2} - \frac{1}{\lambda_1^2}\right)$$

Let $\Theta = (\theta_1, \cdots, \theta_p)$ *and* $\Psi = (\theta_2, \cdots, \theta_p)$. *Then the conditional density of* Θ *given* Ψ *is*

$$dF(\Theta \mid \Psi) = \frac{1}{(np-1)!} \theta_1^{np-1} e^{-\alpha\theta_1} \alpha^{np} = dF(\Psi)$$

which is the density function for tr D.

Proof. From the proof of proposition 4, we have

$$dF(\Theta) = \frac{\pi^{p(p-1)}\Gamma(np)}{[\det \Lambda^2]^n\, \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)}\ [(1 - \theta_2 - \cdots - \theta_p)\theta_2 \cdots \theta_p]^{n-p}$$

$$\times \left[\prod_{j=2}^{p}(1-\theta_2-\cdots-\theta_{j-1}-2\theta_j-\theta_{j+1}-\cdots-\theta_p)^2\right]$$

$$\times \left[\prod_{i=2}^{p-1}\prod_{j=i+1}^{p}(\theta_i-\theta_j)^2\right]\frac{1}{e}\theta_1^{np-1}e^{-\alpha\theta_1}(d\Theta)$$

where

$$\alpha=\sum_{i=2}^{p}\theta_i\left(\frac{1}{\lambda_i^2}-\frac{1}{\lambda_1^2}\right)$$

Then

$$dF(\Psi)=\frac{\pi^{p(p-1)}}{[\det\Lambda^2]^n\,\mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)}\;[(1-\theta_2-\cdots-\theta_p)\theta_2\cdots\theta_p]^{n-p}$$

$$\times \left[\prod_{j=2}^{p}(1-\theta_2-\cdots-\theta_{j-1}-2\theta_j-\theta_{j+1}-\cdots-\theta_p)^2\right]$$

$$\times \left[\prod_{i=2}^{p-1}\prod_{j=i+1}^{p}(\theta_i-\theta_j)^2\right]\frac{1}{e}\Gamma(np)\alpha^{-np}(d\Psi)$$

Therefore

$$dF(\Theta\mid\Psi)=\frac{dF(\Theta)}{dF(\Psi)}=\frac{\theta_1^{np-1}e^{-\alpha\theta_1}}{\Gamma(np)\alpha^{-np}}d\theta_1=\frac{1}{(np-1)!}\theta_1^{np-1}e^{-\alpha\theta_1}\alpha^{np}d\theta_1$$

Under the hypothesis $H_{i,1}:\lambda_i^2=\lambda_1^2$, the term α is zero, and thus

$$dF(\Theta\mid\Psi)=0$$

The alternative is $A_{i,1}:\lambda_i^2<\lambda_1^2$ for at least one $i\in[2,p]$. □

Chapter 7

SUMMARY AND CONCLUSIONS

There are several conclusions from this research that need to be stated. Of immediate interest are the technical results which apply to the spatial processor order determination problem. The second kind of results from this research are the mathematical and statistical tools which are needed by engineers and physicists, but which are usually of little interest to traditional mathematicians and statisticians.

7.1 Results Related Directly to Order Determination

The immediate objective of this research was to derive a test statistic and its distribution for determining the number of significant sources observed by an arbitrary array for a small number of samples using a hypothesis testing approach. This is the problem of examining if eigenvalues of a covariance matrix from a complex multivariate Gaussian distribution are significantly different. This is the small sample complex principal components problem. The form of the required test statistics has been known for a long time. The challenge is to produce the distributions of the desired test statistics.

The problem of finding efficiently computable cumulative distribution functions for appropriate test statistics is still a problem I have not solved. In this thesis, several distributions relevant to the small sample system order determination problem have been found. These are highlighted below.

An exact solution we know how to compute which makes inefficient use of the data was constructed as an F-distributed statistic. This is theorem 6. It requires partitioning the data into two independent sets yielding two independent complex Wishart matrices. Then

$$F = \frac{mc_1^H W_1 c_1 c_2^H \Sigma_2 c_2}{nc_2^H W_2 c_2 c_1^H \Sigma_1 c_1} \sim dncF(2n, 2m, \frac{2c_1^H \delta_1 c_1}{c_1^H \Sigma_1 c_1}, \frac{2c_2^H \delta_2 c_2}{c_2^H \Sigma_2 c_2})$$

The values assigned to Σ_1 and Σ_2 are those specified in the hypotheses of the test. The form of the distribution becomes simplified when $n = m$. The cumulative distribution function for the $F(2n, 2n)$ distribution was derived, presenting a closed form result. This result is documented in theorem 71.

A closely related statistic is for testing hypotheses in MUSIC. This F-statistic is developed in section 6.2.2. It is distributed according to the distribution $F(2n, 2n)$.

$$F = \frac{C_1^H W_Y C_1 C_2^H V^H \Sigma V C_2}{C_2^H W_Z C_2 C_1^H Q^H (\Sigma + ARA^H) Q C_1}$$

Although this may look like a major breakthrough, it really is not. Covariance matrices Σ and R must be established as hypotheses for the test. Often the noise covariance is taken to be $\Sigma = \sigma^2 I$, and the vectors C_1 and C_2 of unit

length. Further simplification occurs when the hypothesis is $q = 1$ and if A is a vector of unit length.

The joint density function of the eigenvalues of one complex Wishart matrix with respect to another complex Wishart matrix $\{\det(A-\lambda B) = 0\}$ was found, paralleling Anderson's result [26] known for the case of real variables. This is theorem 7. Under the null hypothesis of sphericity, this is a piece of the signal subspace method which is based on examining the eigenstructure of the *signal* covariance matrix. See section 5.1.

A complexification of another Anderson result provides the joint density of *ordered* eigenvalues of an Hermitian matrix when the density of the Hermitian matrix is a function of only its eigenvalues. This is theorem 68. This is a powerful result because it allows us to examine generalizations. I have rederived a result of James [120] and Khatri [137] through complexifying Anderson's joint density of the eigenvalues of a matrix distributed as $\mathbf{C}W_p(n, I)$. This is theorem 69. This distribution is fairly simple, and it corresponds to the important case of a pre-whitened filter. James' result (theorem 70) for the joint density of the eigenvalues from $\mathbf{C}W_p(n, \Sigma)$ is also derived. I am not aware of any derivation in the literature of this distribution done for the complex case without reference to the derivation for the real variables case. James [120] wrote his result down by inspection from the form of the real variables case. Takemura [265] referred to his derivation for the real variables case. The

complete derivation for the complex variables case was made possible by the work of Gross and Richards [96].

The paper on zonal polynomials of one and two matrix arguments for the combined cases of real, complex Hermitian, and quaternion variables by Gross and Richards [96] is a major key to the pursuit of an expression for the density function of a test statistic for the small sample order identification problem. In particular, it is this proof which justifies the splitting theorem (proposition 41) for zonal polynomials. It is this splitting property that frees us from the prison of a specific coordinate system by allowing us to integrate over all rotations, leaving us with functions of only sample and parameter eigenvalues.

It is the abstractness of the mathematics involved that allowed solution of the problem. It was on this point that the validity of James' unproven result [120] for the joint distribution of the eigenvalues of the sample covariance matrix for the *complex* case hinged and had not been established by other means. A contribution to the engineering community by this thesis is the narrative parallel in appendix G provided to Gross and Richards' very good paper. Their paper contains key ideas for understanding how to investigate invariance problems. As a side benefit, it was discovered that their induction method hinged on a group theoretic version of the LDU decomposition which engineers are familiar with. See equations G.16 through G.25. I also provided an alternate proof of their lemma 5.2 (given in this thesis as theorem 98)

which relaxed one of their conditions. See the discussion for equations G.5 through G.11. This result opens the possibility of expression of the required distribution using other sets of polynomials which might be easier to compute or which might converge faster.

Following Muirhead's work [187] for the case of real variables, the joint density of the random variables $(u, \cdots)$ has been derived (given in equation 6.16), where u is the statistic for testing sphericity and is given by $u = \frac{\det A}{\left[\frac{1}{p} tr(A)\right]^p}$. It was also shown that $v = \operatorname{tr} A$ and u are independent. See theorem 10. The density of u for the case of $p = 2$ is given in equation 6.17. The cumulative distribution function for $p = 2$ is given in equation 6.18. The density of u for the case of $p = 3$ was determined to be computable. Its detail makes it a suitable evaluation problem for a symbolic mathematics processor.

The density function for the ratio of averages of disjoint sums of sequential sample eigenvalues of a complex Wishart matrix

$$T_{14} = \frac{(p - b + 1)(l_1^2 + \cdots + l_a^2)}{a(l_b^2 + \cdots l_p^2)}$$

was examined in section 6.2.1. The density function given as corollary 3 was determined in terms of a partitioning of $S = (C^H U^H \Sigma U C)^{-1}$ where Σ is evaluated as specified by the hypothesis. The matrix C defining the linear combinations to be compared is constructed as shown in the example by equation 6.14. Similarly, an expression for the cumulative distribution function is determined, which is given as theorem 9.

A number of results motivated by (but not paralleling) Krishnaiah's works [144] [145] [151] were produced in section 6.3 for the case of

$$D = \text{diag}(l_1^2, \cdots, l_p^2) \sim \mathbf{C}W_p(n, \Lambda^2)$$

The results are various ways of using the sample eigenvalues to test if all the population eigenvalues are equal. The tests differ in the details of the specification of the alternative hypothesis. The first method presented (section 6.3.1) examines the joint density of the ratio of adjacent sample eigenvalues. The second method (section 6.3.2) examines the joint density of the ratio of the sample eigenvalues to the smallest sample eigenvalue. A third method (section 6.3.3) is similar in spirit; it looks at the joint density of the ratio of the sample eigenvalues to the largest sample eigenvalue. A last method (section 6.3.4) examines the ratio of sample eigenvalues to the trace of the matrix. Even with this simplified distribution for D, the marginal densities of individual test statistics are difficult to evaluate in general. The densities for the null hypothesis of equal population eigenvalues has been provided. The testing problem is viewed through the mechanism of Roy's union-intersection principle.

Another contribution is the discussion of the details of the generalized maximum likelihood estimator of Kiefer and Wolfowitz [140] (section 4.2) which are generally unknown in the engineering community. Engineers familiar with their work in stochastic approximation [291] will find the discussion of gener-

alized maximum likelihood estimators to be closely related concepts via the mechanism of convergence of sequences. The generalized maximum likelihood estimator involves an application of the Radon-Nikodym derivative which is usually studied in a first course in real and complex analysis. Their powerful concept was written as a side comment in an article devoted to the study of statistical consistency. This thesis provides an exposition of their concept as a generalization of the classical hypothesis testing approach and to the estimation approaches others have taken. It also is a fairly nice discussion on the philosophy of what is going on when an estimation problem is done. It is a conceptual springboard to much more powerful generalizations.

7.2 Complex Statistics Tools for Engineers and Physicists

Some results which are necessary to support the research of this thesis have broader application. These results are collected in a systematic development of the statistics of complex random variables. I could not have efficiently developed a comprehensive theory of the statistics of complex variables without the very good works by Arnold [31], Muirhead [187], and others in the real variables case. This is a natural evolution of ideas. At the same time, it is cautioned that the *extension of real variables results to the complex case*

requires some care. In particular, the complex multiplication operator imposes a structure on the algebra that goes beyond treating $\mathbf{C}^n$ as merely $\mathbf{R}^{2n}$. This shows up most clearly when dealing with changes of variables and derivatives. It can, however, also be seen just from examining the algebraic theory involved. These differences have been demonstrated.

The tremendous similarity of results between the real and complex cases has occasionally led some extremely talented people to write incorrect results down by inspection. Very few people have worked on the statistics of complex variables and the reported results of several respected workers are not in agreement. The study of multivariate statistics of complex variables is still young enough that all results should be reexamined for correctness when their use is anticipated. That caution applies explicitly to this thesis as well as to the literature in general. This issue is important to this thesis because I needed specific results. In particular, I needed the density function of the complex Wishart distribution. A contribution of this thesis is the rederivation of this distribution, following two methods used by others (sections E.1.2 and E.1.3), and a third by mathematical induction (section E.1.1) which Arnold [31] applied in the real variables case. The agreement of the results from three different approaches builds confidence that the result is correct. I am pleased to report that Goodman [92] correctly reported the density function (with derivation) for the complex Wishart distribution.

I do not know of any similar systematic development of the distributions and properties of the complex matrix normal distribution (section D.2) and the complex Wishart distribution (section D.3). Among the properties examined in this thesis are the response under linear transformation of variables, conditional distributions, and conditions for independence. I have provided a derivation for the matrix complex normal distribution. The density function for this distribution has been previously reported in the literature without derivation by two well known researchers, and their results were not the same. I am pleased to report that Brillinger [45] correctly reported the complex matrix normal distribution. I have complexified Arnold's results [31] for the distribution of the trace of a linear transformation of a matrix complex normal random variable and the distribution of twice the trace of the argument of the exponential in the density function of the matrix complex normal distribution. The distribution of $2\,\mathrm{tr}(\Sigma^{-1}W)$ is found to be a chi-square variable. Special functionals of the complex Wishart distribution were shown to have a chi-square distribution, and with this observation an F-distributed statistic was constructed from two such independent functionals. A complex version of Hotelling's T^2 statistic was also derived (section D.4).

Another contribution is the development of the properties of a characteristic function in the context of complex variables (section B.4). This was motivated by the definition of the characteristic function of a complex variable

given by C. R. Rao [217]. Included in this is the development of expected values of moments for the complex case. An important part of this contribution is the demonstration that the expected value of moments are *not* found by applying a derivative. Rather, they are found by the complex conjugate of the derivative with respect to the transform variable evaluated at zero. Important cases are worked out. Variations on $\frac{\partial}{\partial T}\det(A+cBT)$ are computed. These are important because of the application to finding the expected value of moments of something related to the complex Wishart distribution. For the complex Wishart distribution, the characteristic function used is of $2W - \Delta(W)$ where $\Delta(W)$ is the matrix of elements on the diagonal of W. This fact is used to demonstrate the power of using the characteristic function as a generating function for moments. Various other results are given, including the distribution of $\det(\Sigma^{-1}W)$, W^{-1}, $W_1 + W_2$, and $(AW^{-1}A^H)^{-1}$ and some useful expected values such as the expected values of $[\det(W)]^k$, $\det(W^{-1})$, W, W^{-1}, $\mathrm{tr}(W)$, $(\mathrm{tr}\,W)^2$, and $\mathrm{var}(\mathrm{tr}(W))$. I am in debt to various results (section F.4) by Tague [264] which demonstrate the usefulness of the theory presented. This includes the expected values of $(W_2W_1^{-1}W_2)$, $(W^{-1}AW^{-1})$, W^{-2}, and $\mathrm{tr}(W^{-2})$, and $\mathrm{var}[\mathrm{tr}(W^{-1})]$. The work by Tague concludes with an example of computing the signal-to-noise ratio at the output of a beamformer (section F.5).

7.3 Other Results

There are a number of ideas developed which are not central to the main research theme, but which had to be developed in order to obtain the tools needed for the main theme research. Some of those which do not fit neatly into the above categories are identified here.

Of greatest practical importance to engineers and physicists are the results concerning the complex case for scalar derivatives of vectors and matrices, vector derivatives, and matrix derivatives (Appendix B). This is based on the observations made regarding the existence of complex derivatives with application of the Cauchy-Riemann conditions. The caution is that many results reported in the literature incorrectly engage in maximization of Hermitian forms by attempting a derivative approach. *Final results* are often valid because the same result can often be obtained by a completion of squares or projection approach to the problem. However, not all such extrema results are fortunate enough to be valid. *Attempting to avoid the derivative existence issue by treating the real and imaginary parts as separate variables is invalid.*

Other contributions of importance to engineers and physicists are the related results concerning change of variables for the complex multivariate case (Appendix C). Several important observations are in order. The first and most important observation is that there is no such thing as "the general case". A matrix without discernible structure is *not* a general case. You cannot apply

the change of variables for an unstructured matrix to a structured matrix and usually get the right result. Any structure that exists in a matrix *must* be accounted for in a change of variables. Otherwise, your results are simply wrong.

For nonlinear change of variables I copied Muirhead's approach [187] of using the exterior (wedge) product to simplify the algebra. Mathematicians discover this operator in the study of differential geometry. The wedge product is a tool commonly used by physicists and nuclear engineers, but rarely used by most other engineers. With this very practical application to change of variables for the complex case, this tool should become part of the working set of knowledge of all engineers involved in acoustic signal processing. This was a necessary tool for computing the Jacobian for the change of variables involving matrix quadratics, for example the form $Y = TT^H$. Most of these results are complexifications of results by Muirhead [187], Arnold [31], and Deemer and Olkin [67]. Some of these confirm results by Goodman [92], or confirm results or make corrections of editorial problems in Khatri [137]. The development of the Jacobians was a necessary part of the derivation of the various density functions in this thesis.

I have provided many decompositions of complex matrices and related results (appendix M) based on complexifying results done for the real case by others. These include special results for complex triangular matrices, eigen-

value decompositions with related results, proof of the relationship between the eigenvalues, determinants, and the trace of a matrix, square-root decompositions, polar decompositions, Cholesky decomposition, singular value decomposition, and various relationships between eigenvalues of X, $(aI + bX)$, and $(aI + bX)^{-1}$. Many of these results can be found in Stewart [259].

The various constructions of a complex vector space (appendix J) drive home the fact that $\mathbf{C}^n$ *is not merely* $\mathbf{R}^{2n}$, although I have seen the construction of a vector in $\mathbf{R}^{2n}$ by others, and have seen the *scalar* $\begin{pmatrix} x & -iy \\ iy & x \end{pmatrix}$ by others. This work was motivated by examples from Nomizu [193].

A few integrals were computed (appendix P), some of which do not appear in Gradshteyn and Ryzhik [94]. These were done in support of evaluation of a cumulative distribution function. Evaluation of $\int \frac{X^k}{(aX+b)^p} e^{-cX}\, dX$ was tedious (theorem 147), but is one most sophomores can do. Generalized even and generalized odd functions were defined (definitions 84 and 85) and their elementary properties demonstrated (section P.2.6). The integral $\int u^m(1-u)^a du$ was computed and interpreted as an expansion in terms of the probability of k failures in m trials when $0 < u < 1$ (proposition103). The complexification of Muirhead's matrix Laplace transform [187] of $(\det A)^{a-m}$ with respect to $\Sigma = \Sigma^H$ where $A = A^H$ is developed as theorem 150. This is the integral

$$\int_{A>0} \operatorname{etr}(-\Sigma^{-1}A)(\det A)^{a-m}(dA) = (\det \Sigma)^a \mathbf{C}\Gamma_m(a)$$

This provides an alternative interpretation of the complex Wishart density as the entire integrand in the normalized transform. This is merely the complexification of a relationship known by specialists working in the real variables case, yet it is an important one.

Other miscellaneous contributions include the generalized definition of the nested operator (definition 87) and development of the trigonometry of complex matrices (appendix N). The definition of the nested operator $\bigwedge_{k=1}^{n}(a_k \square b_k)$ is a generalization of Tuma's nested operator (section 8.11) [268] which has application in one of the test distribution density functions (proposition 1), and has application in recursive solutions of problems. I used the definition of e^A to complexify results by Curtis [64], which includes work regarding the matrix logarithm (section N.1).

I generalized the discussion by James [120] to show that his set of three simultaneous mappings can be formalized into the setting of a topological group theory (section H.6). I defined a group G whose elements are pairs of matrices with a special operator (section H.6.1). I then defined a set A upon which this group acts, where A happens to be the set of all complex multivariate normal distributions (section H.6.3). It is this generalization that justifies the application of the machinery of group representation theory. It makes explicit that we really are operating on distributions and not just parameters of a distribution.

7.4 Linear Algebra Results Verified for the Complex Case

Some results for complex variables are true merely because the space under consideration is a linear space for any arbitrary field. I verified all results I had a need for, not knowing ahead of time whether the result depended only on the linear algebra for an arbitrary field, or whether there was some modification needed to specialize the results to the complex case. The following results do not differ between the real and the complex cases:

1. Partitioned matrix right and left inverses (section K.3.1).
2. Partitioned matrix determinants (section K.4.2).
3. Eaton's lemma 1.35 [73]: $\det(I_n + AB) = \det(I_m + BA)$ with variations (lemma K.4.3).

7.5 Other Simple Results

The results identified here are results which are mundane. I have not bothered to see if anyone else has produced them. They are useful, but not challenging.

1. Properties of skew-Hermitian matrix $A = -A^H$ (definition 77).
2. Proof that if A is positive definite then A is Hermitian (proposition 48).

3. Many explicit expansions of the trace of a product of various matrices (section K.2). These were developed to support evaluation of functions of the complex Wishart distribution via the method of differential functions of the related characteristic function.

4. Complex matrix inversion lemmas (Section K.3.2).

5. Expression of the (p,p) element of an inverse matrix in terms of the elements of the original $p \times p$ matrix (lemma 41).

6. Proofs that $[\det(A)]^{-1} = \det(A^{-1})$ (proposition 57) $\det(A^*) = (\det A)^*$ (lemma 42), and $\det(A^H) = (\det A)^* = (\det A)^H$ (proposition 58).

7. Proof that for unitary A that $\det A = e^{i\theta}$ for arbitrary $\theta \in \mathbf{R}$ (lemma 43).

8. Proof that for orthonormal complex matrix A that $\det A = \pm 1$ (lemma 44).

9. From proposition 61:

$$\det \begin{pmatrix} A & I \\ B & I \end{pmatrix} = \det(A - B)$$

10. From proposition 62:

$$\det \begin{pmatrix} A & B \\ I & I \end{pmatrix} = \det(A - B)$$

11. From proposition 63:

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = (\det A)(\det B)$$

12. From lemma 49:

$$\det \begin{bmatrix} A \otimes I_p & C \otimes I_p \\ B \otimes I_p & D \otimes I_p \end{bmatrix} = \det \left[\begin{pmatrix} A & C \\ B & D \end{pmatrix} \otimes I_p \right] = \left[\det \begin{pmatrix} A & C \\ B & D \end{pmatrix} \right]^p$$

13. From proposition 65:

$$\det(I + A^2) = |\det(I + iA)|^2$$

7.6 Proofs for Results Stated by Others

1. Littlewood p. 19 [167]. If $A = -A^H$ then $(I + A)(I - A)^{-1}$ is unitary (proposition 59).

2. Littlewood p. 19 [167]. If B is unitary and -1 is not a characteristic root of B, then there exists $A = -A^H$ such that $B = (I + A)(I - A)^{-1}$ (proposition 60).

7.7 More Results

1. Two examples of structures involving Hermitian matrices failing to form a group (section H.5).

2. Proof that $A(\text{adj}\, A) = \det(A) I_n$ for the complex case (proposition 56). The point is that even for the complex case

$$\text{adj}(A) = [(-1)^{i+j} \det(X_{ij})]^T$$

rather than

$$[(-1)^{i+j} \det(X_{ij})]^H$$

where X_{ij} is the minor of a_{ij} for matrix A. This is a good example that intuition and experience cannot be trusted to guide the conversion of methods from real to complex variables.

3. Demonstration that the orthonormal bases produced by the Gram-Schmidt orthonormalization process (appendix L) are not unique, but depend upon the bilinear operator used in the algorithm, and that this bilinear operator is not required to be an inner product. This demonstrates that the property of having an orthonormal basis for a vector space does not imply that the space is an inner product space.

7.8 Comparisons.

The theory I have pursued is not yet fully developed. For selecting methods for use in systems being designed today (1994), use a different approach.

When constrained to serial processors, the estimation approach conceptually should yield quicker results than my sequential testing approach. However,

it should be observed that other estimation approaches often are sequential also. It is common to require the construction of a family of estimates and to pick out the "best" estimator from that family. As technology provides us with economical and practical parallel processors, this source of difference of approaches will become less important.

The more knowledge you have, the better decision you can make and the better assessment of the decision quality. The information theoretic approaches used now do not require establishment of measures of effectiveness for the quality of the produced estimate, although conceptually it can be done via providing confidence intervals. The hypothesis testing approach requires explicit identification of allowable error.

Is one method better than the other? It depends on your purpose. For applications, it also depends on available technology.

Chapter 8

FURTHER RESEARCH

Suggestions for further work identified in this section are of two types. The first type is covered in the section titled *Extending This Research.* That focus is on work needed to continue progress on the topic of the small sample order identification problem. The second type of recommendations focus on the development of ideas and tools less directly related to those used in this thesis. I think identification of these less directly applicable ideas is vitally important for the broader advancement of engineering and science.

The small sample order identification problem subtends several areas that desperately need more work. The small sample statistics of complex matrix random variables is still an area that has received little attention compared to the real variables case. The basic question of what are the appropriate properties of a small sample test statistic need to be examined.

Mathematical tools for constructing the needed distributions need to be collected, cataloged, and more extensively developed. This includes the systematic collection or development of complex matrix algebra and calculus beyond what was done in this thesis. Much of this work already exists in abstraction or is scattered throughout the literature. Carefulness in reading the literature is strongly recommended. Results identified as applying to the complex case may assume complex symmetric matrices rather than complex Hermitian ma-

trices. This is particularly true in literature dealing with group theory or zonal polynomials. Conditions for existence of complex derivatives are often ignored, resulting in errors in the literature, particularly in the area of adaptive beamforming. It is not uncommon to see the erroneous application in the complex case of gradients used in optimization and search algorithms. Treating the real and imaginary parts of complex variables separately in a gradient is invalid as a method of avoiding the existence of the complex derivative upon which the optimization methods depend. Jacobians for complex change of variables reported in the literature are not reliable. Similarly, distributional results are not yet reliably reported. Caveat emptor. Progress in related areas in the last few years further increases the urgency for completing a theory of complex multivariate analysis of stochastic variables. This effort will be of great benefit to people studying acoustics, signal processing, and other areas as well. It needs to be made accessible to the level of the engineering undergraduate senior. Specific areas of mathematical knowledge must be further developed. In particular, progress in extending our knowledge of zonal polynomials of several complex matrices is urgently needed. There is plenty of work yet to be done.

8.1 Extending This Research

This thesis has provided the tools for development of application theory, but is not mature enough yet for easy experimental or simulation use because of the difficulty of evaluating cumulative probability distribution functions. Research is still in the embryonic stage. The continuing work by Gross and Richards in theory and the continued use in applications by Tague are providing the theory, interest, and pressure that will eventually produce practical results.

8.1.1 Some References to Consider

There are a number of references that would be worthwhile to thoughtfully consider with respect to the problems of statistics related to the order estimation problem. Saw's 1977 paper [233] proposed a method of computing zonal polynomials which motivated further work on the problem which once was thought to be intractable. Farrell [80] reported on the calculation of complex zonal polynomials in 1980, which includes some tables. Understanding his paper requires understanding group characters, bisymmetric matrices, Young's diagrams and Haar measure. Saw's 1984 paper [234] establishes the connection between the ultraspherical polynomials and distributions on the m-sphere. Kushner and Meisner [159] reported in 1984 on integral and differential formulas for zonal polynomials. Watson's 1986 paper [277] discusses estimation theory on the sphere in a Bayesian setting. Yu's 1991 paper [295]

on recursive updating of the eigenvalue decomposition of a covariance matrix may represent a significant contribution. It would be of interest to determine how this affects the rank determination problem and the question of independent samples now assumed in forming the test statistic. Shenoy's 1991 Ph.D. thesis [240] on group representations and optimal recovery in signal modeling deals with concepts that have repeatedly surfaced in the background reading required to understand the work of Gross and Richards [96], thus indicating that it deserves careful attention.

8.1.2 Connecting Gross and Richards' Work to Stein and Weiss' Work

The bridge between abstract theory and engineering application will be narrowed when the connection between the work by Stein and Weiss [258] and Gross and Richards [96] are related in terms understandable to a well trained engineer. *This is the most urgent and productive next step.* Krantz' presentation of Stein and Weiss included some wonderful geometric interpretations of spherical harmonics and zonal polynomials. Stein and Weiss did their work for the case of real variables which stopped short of the splitting theorem used by James [120]. The work by Gross and Richards was done in a more general setting, yet it did not include generalizations of some key insights developed by Stein and Weiss. Making these connections must be done by someone that

understands both the traditional development of special functions and also the background material supporting Gross and Richards. It would be nice to complexify Stein and Weiss, and include the splitting theorem and other developments. The abstraction of Stein and Weiss' axis of spherical rotation needs to be determined in the Gross and Richards' framework.

The splitting theorem needs to be examined to determine if it truly establishes an equality (i.e., $\Leftarrow$ and $\Rightarrow$ both hold in the derivation) or whether the correct statement of the theorem is only $\Rightarrow$. The practical consequence is that we know from Gross and Richards work [96] that a complex zonal polynomial of Hermitian matrix argument $Z_m(W)$ has the same value as when the argument is the matrix of the eigenvalues of W. When this is connected with the splitting theorem, and if it is bidirectional, then via definitions of generalized hypergeometric functions of one and two matrix arguments we determine that $\exp(\Sigma^{-1}W) = \exp(\Lambda^{-2}L^2)$ where Λ^2 is the diagonal matrix of eigenvalues of Σ, and L^2 is the diagonal matrix of eigenvalues of W. I have generated a numerical counterexample to this.

Although Gross and Richards [96] defined a differential operator (p. 788, Section 2.2) as part of the development of zonal polynomials, it appears to not have been necessary to achieve the results they were after. In particular, their equation 2.2.8 is not required for proving their equation 2.2.5. If you define a polynomial as a vector in the way done by Broida and Williamson

[47], then the inner product between two polynomials defined by Gross and Richards which used the differential operator can be replaced by the usual vector inner product on the vector form of two polynomials. The application of differentiation used later to establish a set of coefficients to guarantee series convergence provides a nice but not necessary motivation. This is good in the sense that the ever present troubling issue of differentiability for the complex case can be avoided. However, it appears that proof of differentiability is a desirable achievement. That is part of the link back to the work by Stein and Weiss. I still have not yet determined if the set of functions I derived forms a complete set. Alternately, I have not yet determined the set for which the derived set of functions is complete.

It would be useful to examine the family of generalized functions resulting from the relaxed definition of the inner product. It is known that Fourier expansions converge more slowly than other expansions. It would be interesting to see if a careful choice of inner product could produce an oblique set of functions that converge faster or are easier to compute than zonal polynomials.

8.1.3 Computation of Zonal Polynomials

Takemura [265] remarked about the relationship between real and complex zonal polynomials. This linkage needs to be translated into the language of engineering. This link has already been made by other researchers. Separate

papers have been published on the computation of complex zonal polynomials. It is said that computing complex zonal polynomials is easier than for real zonal polynomials. I think the relationship can be obtained from a careful reading of Gross and Richards with that goal in mind. Care still needs to be exercised to determine if the "easy" computations apply to symmetric complex Wishart matrices or to Hermitian Wishart matrices.

Similarly, the works of Constantine need to be folded into this study. See references [56][57][58][59]. Constantine relied on zonal polynomials defined as arguments of positive definite complex *symmetric* (not Hermitian) matrices. We know that symmetry and Hermitian symmetry endow a matrix with different properties. Constantine remarks [57] (p.1272) that since zonal polynomials are polynomials in the characteristic roots, then the definition of zonal polynomials can be extended to arbitrary complex symmetric matrices. Consistently, he defines hypergeometric functions of a complex symmetric matrix variable in terms of zonal polynomials of complex symmetric matrix argument. Since Constantine [57] and James make use of the works of Herz [106], it makes sense to reexamine Herz' work more closely to determine the restrictions Herz uses. For example, Herz studies $m \times m$ complex symmetric matrices with a positive definite real part. Does it naturally follow that the same zonal polynomial is defined for Hermitian symmetric matrices? I can produce an Hermitian matrix and a different complex symmetric matrix which have the same set of eigen-

values. The work by Gross and Richards [96] shows that zonal polynomials are definable for Hermitian matrices, but are these the same as those treated by James and Constantine?

The importance of the question is because Constantine's work is the basis for the evaluation of zonal polynomials in a few specialized and very important cases, such as the noncentral real Wishart matrix. It may be that the basic properties carry over, but with different constants. This raises the issue of having to pay close attention to any derivative works based on James or Constantine when working with Hermitian matrices, to ensure that cited results of previous workers apply. It might not be enough to merely say that a result applies to complex matrices. It would be logical to define a complex Wishart distribution that applies to a complex symmetric matrix. I have not seen this issue made a point of. Noting how important matrix structure has been in modifying results from the real to the complex Hermitian case, the question needs to be asked. The goal is to validate or achieve similar results consistent with the work by Gross and Richards.

As an aside, another consideration is that zonal polynomials are symmetric functions of its arguments. A function $f(x, y)$ is called "symmetric" if it is invariant under permutation of its arguments. This means if you change the order of the arguments then the value of the function does not change. The trace and the determinant of a matrix are symmetric functions of the eigenval-

ues of a matrix. Thus, a zonal polynomial as developed by Gross and Richards [96] is a symmetric function of its Hermitian matrix argument. The value of the zonal polynomial depends only on the value of the unordered eigenvalues of its Hermitian matrix argument. What may be happening is that authors are unintentionally injecting ambiguity into the research by not differentiating between symmetry of the *function* and symmetry of the *function's argument.* When interest was restricted to real symmetric matrices there was no need to be careful to distinguish between the two because the answer was "yes" in both cases. When working with fields having more structure, more care is needed.

The comments of the preceding paragraphs are not criticisms of the work of the mentioned authors. Rather, the comments point out that the work they did is related to the present application, and because terminology is so similar the unwarned researcher may inadvertently apply results directly without first answering the question if the same set of assumptions are used.

8.1.4 Calculus of Zonal Polynomials

An immediate next step that will lead to useful results hinges on the ability to perform integration on zonal polynomials. This integration is necessary to evaluate marginal distributions of test statistics. The work in this thesis leads up to the point where changes of variables can be used to construct test

statistics much in the same way Krishnaiah has done for the real variables cases. Krishnaiah has stated the results for several integrals.

8.1.5 Distributional Theory and Tools

There is still much work to be done to support the development of statistics of complex variables for application to problems of engineering and physics. Critical to this process are two areas. The first is the development and systematization of Jacobians for changes of variables. A starting point is Roy's work of 1952 [228]. Distributional work relies very heavily on being able to perform well-selected changes of variables. There are plenty of results that have been done for the case of real variables that need to be adapted to the world of complex variables. Distributional work needs to begin with the Gaussian case, but must grow beyond it. Work such as Olkin and Roy's 1954 paper [198] is applicable and needs to be extended to the case of complex variables. For the physicist, this work needs to be done for quaternions as well as for complex variables. The second area that needs work is the study of invariance at an abstract level. This work has been partially addressed with the attention given to Gross and Richards.

8.1.6 Noncentral Distributions and Power Functions

Once we have the results for the central complex Wishart distribution based tests worked out, we must turn attention to the development of the noncentral complex Wishart distribution and the noncentral distribution of related test statistics. From these, we need to work out the power functions for the various tests. As we saw for the case of the central complex Wishart distribution, we cannot yet take for granted that previously published results are valid. Those results, where they exist, should be reexamined carefully, developing the related necessary tools.

8.2 Bridging Theory and Application

8.2.1 Important Authors to Consult

Krishnaiah is the author most prolific in examining specific tests based on eigenvalues of a real or complex Wishart matrix. Some effort may be worthwhile to understand the report by Krishnaiah and Shuurmann [151]. It appears that they have performed evaluations for special cases for which expressions might be available for zonal polynomials.

The works of C. R. Rao deserve much greater attention. There is a need

to examine his work thoroughly and apply it to this problem (as well as to other problems). One important work I rediscovered after the research phase of this thesis was finished is reference [214]. This includes work on matrix approximations specifically applied to complex matrices.

Another author deserving of attention is Steen Andersson. See references [27] [28] [29]. The first paper considers distributions of maximal invariants using quotient measures. This work should be studied in the context of Gross and Richards [96]. The second paper considers testing various real matrices to determine if they have complex or quaternion structures. The joint density of the eigenvalues are derived up to an unspecified norming constant, and the exact values of all norming constants are derived simultaneously using a method involving recursion formulae. The moments of the likelihood ratio statistics used in testing are obtained from these norming constants.

Likewise, anything written by Muirhead should receive attention. Muirhead's work is particularly useful in developing the machinery to obtain noncentral distributions. These are required for determining the power of tests. It is his application of areas of mathematics that are nontraditional for statistical (and engineering) work that enables his computation of some otherwise very difficult noncentral distributions.

An idea from Kshirsagar [154] is to look at the distribution of moments of the test statistic. This concept has not been attempted in this thesis, yet

it deserves consideration by future workers. In particular, he applies this concept to sphericity tests. Recall that one method for completely defining a distribution is to know all of the moments of the distribution. Kshirsagar is a recognized authority in multivariate analysis, and in principal components and test distribution theory in particular.

8.2.2 Small Sample Test Theory

The fundamental question of what constitutes a good estimator for small sample statistics deserves some study. The attention and controversy regarding consistency properties with respect to some information theoretic based methods (particularly AIC) implies that this question has yet to be authoritatively answered. For example, we know that consistency, as technically defined in statistics, is not a required nor necessarily desirable property [55]. However, lack of consistency has been referred to in the engineering literature [129] as a disqualifying property. That is appropriate for the large sample case, but not, in itself, appropriate for the small sample case. Refer to section 4.3.1 for a detailed discussion of this issue.

There are other questions as well. Will a biased minimum variance estimate do? Should you apply asymmetric confidence bounds based on some utility curve derived cost function? Must the test statistic for comparing estimators be invariant with respect to coordinate transformations? These and

other questions need to be collected and systematically examined. The idea of looking at parameter estimators for a distribution in terms of bias, etc., is not new and should be incorporated into the thinking about the order estimation problem. Bickel and Doksum [40] discuss this at length, and it is one of the finest texts on mathematical statistics that does not require the reader to have a background in measure theory.

8.2.3 Approximation Theory

The approximation techniques as discussed in Keener's well written text [131] need to be applied to the problem to obtain practical (easily computable) results once we understand what the correct exact forms are. Doing this does not require a radically nontraditional mathematical background for engineers once the basic form for zonal polynomials is understood.

8.2.4 Burnside's Theorem and Characteristic Functions

Newman's presentation [192] (p.166) of Burnside's theorem on irreducible sets of matrices raises an interesting question. The question is related to the subject content of this thesis in that the notion of eigenvalues is intimately wrapped in the theory of invariance and irreducibility. Burnside's theorem is stated as follows.

Theorem 11 *Let* $\mathbf{F}$ *be algebraically closed. Let* $\mathbf{G} = \{A\}$ *be a subgroup of* $\mathbf{GL}(n, \mathbf{F})$ *which is irreducible as a set of matrices. Then any relationship* $\sum_{p,q} \tau_{pq} a_{pq} = 0$ *for* $\tau \in \mathbf{F}$ *can hold for all* $A = (a_{pq})$ *of* $\mathbf{G}$ *if and only if* $\tau_{pq} = 0$ *with* $\tau_{pq} = 0$ *for all* p, q.

If you now consider $i\,\mathrm{Re}\{\sum_{p,q} \tau_{pq} a_{pq}\}$ as the argument of the exponential function and take its expectation, you have a characteristic function. When you generate moments using a characteristic function, you evaluate at $\Upsilon = 0$. So, the question is "What does Burnside's theorem say about characteristic functions?" Since characteristic functions present a sometimes easier way of achieving distributional results we need in the order estimation problem, answering this question for the complex case will give us insights to an important tool.

8.2.5 Sturm Separation Theorem and Parallel Processing

Another tool that deserves inquiry is the application of the Sturm Separation Theorem as found in C. R. Rao (p. 64, section 1f.2.13(vi))[213]. The advent of parallel processing makes use of this theorem to compute eigenvalues practical. The idea is that the eigenvalues of a principal minor of a matrix provide estimation limits for beginning a search for eigenvalues of the next size larger principal minor. Let A_k be the k^{th} principal minor of the square matrix A

formed by the upper left corner k rows and k columns. Let $\lambda_j^2(A_k)$ be the j^{th} eigenvalue of A_k. Number the eigenvalues so that $\lambda_1^2 > \lambda_2^2 > \cdots > \lambda_k^2$. Then the relationship between the eigenvalues is illustrated by the following lattice shown in figure 8.1.

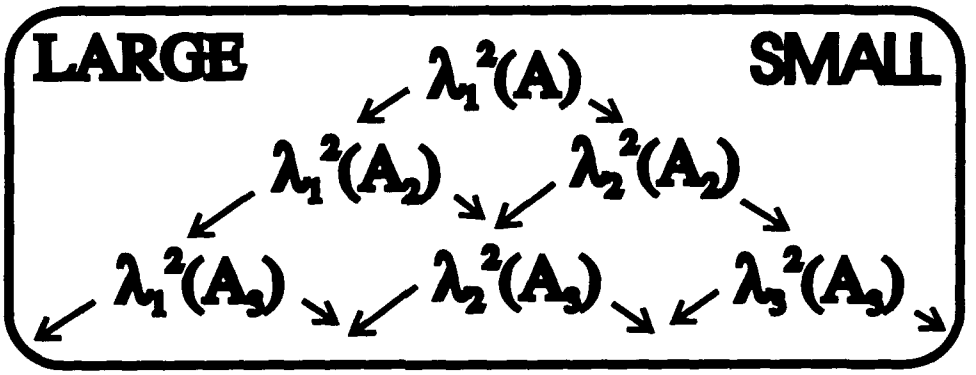


Figure 8.1. Sturm's Eigenvalue Lattice

Thus, eigenvalues on the same row of the lattice can be searched for independently on separate processors using search limits computed from the previous row of the lattice. Notice that the larger a matrix becomes, the more extreme the eigenvalues become. This approach might be good when accuracy is more important than speed. Note that at each row, you restart with the raw data. You do not lose precision merely because of accumulating finite word length errors from computations done with previous rows. At each step, we can still use the best of all the old techniques, including the iterative ones.

8.2.6 Empirical Characteristic Functions

Epps discussed characteristic functions from a geometric point of view in a 1993 tutorial [78]. In this paper, he proposed the use of empirical characteristic functions as a basic for hypothesis tests. Epps points out that such an approach makes handling of procedures for estimation of parameters of mixture distributions easier, which applies to the case of sonar.

8.3 Acoustics and Signal Processing

8.3.1 Processor Structure

As success is achieved in developing the necessary mathematical tools, some thought will need to be given to the implementation. Those tests based on a likelihood ratio test can be realized with an estimator-correlator (or estimator-subtractor) structure. This is the case for the sample eigenvalue ratios. The F-tests, however, were not derived from a likelihood ratio. Examining these tests and devising the relevant structures of associated processors is still needed.

8.3.2 Time Variation of Noise Field

There are two fundamental approaches supported by this thesis, yet further research results in the area of acoustic oceanography are needed to help make the decision regarding which approach is appropriate at the moment. In one

approach you only consider the sample covariance of signal plus noise, $\hat{S} + \hat{N}$. In the other approach, you obtain an independent estimate $\hat{N}^*$ of the noise and then look at $(\hat{S} + \hat{N}) - \hat{N}^*$. If your estimate of the noise covariance is good, you want to choose the second method. If your estimate of the noise covariance is bad, then you want to choose the first method. The trade-off point is a fundamental statistical question that needs to be answered.

Then you need to consider the acoustic oceanography aspects. Over what period of time is an estimate likely to be good enough to use? How do you best propagate noise estimates in time? The Kalman filter approach is one way. Use data such as sea state, precipitation rate, and noise in frequency bands different than the band of interest as concomitant variables. Also, geography, array geometry, and array platform orientation can be included in the sample noise covariance matrix prediction method. The paper by Scharf and Lytle [235] is related to these questions. So also is the paper by Scharf and Tufts [236].

8.3.3 Patterned Arrays

Patterned arrays impose a structure on the covariance matrix of data passing through the beamformer. This additional structure modifies the Jacobians for changes of variables applied in the derivation of probability distributions that ultimately show up in the sampling distribution of covariance matrix eigenval-

ues. The approach taken in this thesis is an important first step, but it is not sufficient for cases of special importance such as the line array with equally spaced elements. Idealized line arrays have been extensively studied, and the covariance matrix often modeled as a Toeplitz matrix. Random distortions in a towed line array invalidate assumptions that make an idealized line array efficient and simple to work with.

The inclusion of considering the matrix complex normal distribution begs the question of how this can be applied to rectangular receiving arrays. It has nicely built-in parameter matrices that can be thought of as a row covariance matrix and a column covariance matrix.

8.3.4 Multipath Detection

This work is related to the problem of detection in a multipath environment. To increase the probability of detection, you would like to consider several paths simultaneously rather than treating each path independently. Mirkin's thesis [183] on use of stochastic maximum likelihood estimators considers not only several paths simultaneously, but also the whole acoustic field. To formulate the problem, he requires that the number of sources be known. Techniques in this thesis form a piece of the problem of estimating the number of sources, but is not sufficient by itself to solve the whole problem.

The general unstructured array described at the beginning of this thesis

is a three dimensional array. Thus any discussion relating to determining the number or direction of signal arrival paths applies, regardless of the direction from which the signal arrives.

The clustering of arrival paths and signals into sources is another problem with another set of competing theories for attacking it, yet the statistical concepts of this research are a piece of that larger problem. Other related theories include cluster analysis, discriminant analysis, factor analysis, probabilistic neural networks, expert systems theory, etc.

Another way of viewing the problem, addressed by Buckley [48], is to look backwards at the problem. He does so for a general array and applies a Karhunen-Loève decomposition in his treatment of the problem. He uses a norm to determine if he has reconstructed his signal "well enough" as a way of selecting the number of significant singular values. It would be good to revisit his work with the view of making a determination from a statistical point of view.

8.3.5 Analogy of Temporal and Spatial Domain Signal Processing

It is routine to remark that there is a mapping between signal processing in the time domain and array processing in the spatial domain. It is filter theory applied in different domains. The usual mapping noted is the relation between

linear arrays and sampling in the time domain. It would be of interest to take theory of array processing and map it back to time domain processing to see what can be learned.

8.3.6 State Space Processing

The paper by Prasad and Chandna [208] proposed use of a state space approach to bearing determination for a uniform line array with a canonical correlation approach for solving for coefficients. Using these two ideas the concept might be generalizable to an arbitrary array. The techniques of this thesis come to bear in choosing the number of significant eigenvalues in the canonical correlation.

8.3.7 Application to Intensity Measurement

There might be some application of the statistical work developed in this thesis to the estimation of noise when setting up an intensity measurement experiment, and later accounting for noise in the analysis of data. Since pressure is treated as a complex quantity it is natural to apply statistics of complex variables when examining sources of variation. For example, one could estimate the noise field at different locations before turning the source on. During the experiment, you then have the ability to perform a hypothesis test to determine if a nodal line in the field has been found. You test to see if

the observed pressure is statistically significantly different from the previously measured noise-only case. The same test, interpreted in another way, tells you the chance that the data you are getting is something other than noise. For another example, you could use the covariance matrix decomposition of data from 4 microphones to determine directions of arrival in three dimensions and estimate the variance from that direction.

REFERENCES

[1] Milton Abramowitz and Irene A. Stegun (editors), *Handbook of Mathematical Functions* (AMS-55), with corrections, National Bureau of Standards (December 1972).

[2] Hirotugu Akaike, "On the use of an index of bias in the estimation of power spectra", *Annals of the Institute of Statistical Mathematics* **20** 55-69 (1968).

[3] Hirotugu Akaike, "Fitting autoregressive models for prediction", *Annals of the Institute of Statistical Mathematics* **21** 243-247 (1969).

[4] Hirotugu Akaike, "Statistical predictor identification", *Annals of the Institute of Statistical Mathematics* **22** 203-217 (1970).

[5] Hirotugu Akaike, "On a semiautomatic power spectrum estimation procedure", *Proceedings of the* 3^{rd} *Hawaii International Conference on System Sciences*, pp. 974-977 (1970).

[6] Hirotugu Akaike, "On a decision procedure for system identification", in pre-prints for: *IFAC Kyoto Symposium on System Engineering Approach to Computer Control*, pp. 485-490 (1970).

[7] Hirotugu Akaike, "Autoregressive model fitting for control", *Annals of the Institute of Statistical Mathematics* **23** 163-180 (1971).

[8] Hirotugu Akaike, *Determination of the Number of Factors by an Extended Maximum Likelihood Principle* (Research Memorandum 44) Institute of Statistical Mathematics (March 1971).

[9] Hirotugu Akaike, "Use of an information theoretic quantity for statistical model identification", *Proceedings of the* 5^{th} *Hawaii International Conference on System Sciences*, (Western Periodicals, North Hollywood, California) 249-250 (1972).

[10] Hirotugu Akaike, "Automatic data structure search by the maximum likelihood", *Proceedings of the* 5^{th} *Hawaii International Conference on System Sciences, Computers in Biomedicine Supplement* (Western Periodicals, North Hollywood, California) 99-101 (1972).

[11] Hirotugu Akaike and T. Nakagawa, *Statistical Analysis and Control of Dynamic Systems*, Saiensu-sha (Tokyo, 1972). (In Japanese, with a list of a computer program package TIMSAC for time series analysis and control written in a Fortran IV type language and with English comments.)

[12] Hirotutu Akaike, "Information theory and an extension of the maximum likelihood principle", *Proceedings of the 2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR, September 2-8, 1971, Supplement to Problems of Control and Information Theory*, edited by B. N. Petrov and F. Csaki, Akademiai Kiado (Budapest, 1973), pp. 267-281.

[13] Hirotugu Akaike, "Block toeplitz matrix inversion", *SIAM Journal of Applied Mathematics* **24** 234-241 (1973).

[14] Hirotugu Akaike, "Maximum likelihood identification of Gaussian autoregressive moving average models", *Biometrika* **60** (2) 255-265 (1973).

[15] Hirotugu Akaike, *Stochastic Theory of Minimal Realizations* (Technical Report 40), University of Manchester Institute of Science and Technology, Manchester, England: Department of Mathematics (1973), and *IEEE Transactions on Automatic Control* **AC-19** (6) 667-674 (December 1974).

[16] Hirotugu Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control* **AC-19** (6) 716-723 (December 1974).

[17] Hirotugu Akaike, "Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes", *Annals of the Institute of Statistical Mathematics* **26** 363-387 (1974).

[18] Hirotugu Akaike, "Markovian representation of stochastic processes by canonical variables", *SIAM Journal of Control* **13** 162-173 (1975).

[19] Hirotugu Akaike, E. Arahata, and T. Ozaki, *TIMSAC-74 - A Time Series Analysis and Control Program Package - (1)*, Computer Science Monographs **5**, Institute of Statistical Mathematics (Tokyo, March 1975).

[20] Hirotugu Akaike, "Canonical correlation analysis of time series and the use of an information criterion", in *System Identification*, edited by Demitrios G. Lainiotis (Academic Press, 1976), pp. 27-96.

[21] Hirotugu Akaike, "On entropy maximization principle", in *Applications of Statistics*, edited by P. R. Krishnaiah (North-Holland, 1977), pp. 27-41.

[22] Hirotugu Akaike, "On the likelihood of a time series model", *Institute Statisticians 1978 Conference on Time Series analysis and Forecasting*, Cambridge University (1978).

[23] Hirotugu Akaike, "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting", *Biometrika* **66** (2) 237-242 (1979).

[24] *Theodore Wilbur Anderson*, "Asymptotic theory for principal component analysis", *Annals of Mathematical Statistics* **34** 122-148 (1963).

[25] Theodore Wilbur Anderson, "Samuel Stanley Wilks, 1906-1964", *Annals of Mathematical Statistics* **36** (1) 1-27 (February 1965).

[26] *Theodore Wilbur Anderson, An Introduction to Multivariate Statistical Analysis,* 2^{nd} Edition, Wiley (1984).

[27] Steen Andersson, "Distributions of maximal invariants using quotient measures", *Annals of Statistics* **10** (3) 955-961 (1982).

[28] Steen A. Andersson, Hans K. Bröns and Soren Tolver Jensen, "Distribution of eigenvalues in multivariate statistical analysis", *Annals of Statistics* **11** (2) 392-415 (1983).

[29] Steen A. Andersson and Michael D. Perlman, "Two testing problems relating the real and complex multivariate normal distributions", *Journal of Multivariate Analysis* **15** 21-51 (1984).

[30] R. Arens, "Complex processes for envelopes of normal noise", *Institute of Radio Engineers (IRE) Transactions on Information Theory* **IT-3** 204-207 (1957).

[31] *Steven F. Arnold, The Theory of Linear Models and Multivariate Analysis*, Wiley (1981).

[32] ASTM Standard E 610-82.

[33] ASTM Standard E 1067-85.

[34] *M. Athans and F. C. Schweppe, Gradient Matrices and Matrix Calculations*, (Technical Note 1965-53), MIT Lincoln Laboratory (17 November 1965).

[35] Frank Ayres, Jr., *Matrices,* Schaum's Outline Series, McGraw-Hill (1962).

[36] Zhi-Dong Bai, Paruchuri R. Krishnaiah, and Lin-Cheng Zhao, "On rates of convergence of Efficient Detection Criteria in signal processing with white noise", *IEEE Transactions on Information Theory* **35** (2) 380-388 (March 1989).

[37] T. P. Barnett and R. W. Preisendorfer, "Multifield analog prediction of short-term climatic fluctuations using a climate state vector", *Journal of the Atmospheric Sciences* **35** 1771-1787 (1978).

[38] Julius S. Bendat and Allan G. Piersol, *Random Data: Analysis and Measurement Procedures,* Wiley (1971).

[39] Julius S. Bendat and Allan G. Piersol, *Engineering Applications of Correlation and Spectral Analysis,* Wiley (1980).

[40] Peter J. Bickel and Kjell A. Doksum, *Mathematical Statistics, Basic Ideas and Selected Topics,* Holden-Day (1977).

[41] Michel Bouvet, "On Bayesian quickest decision", 24^{th} *C D C 85* (1985).

[42] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control,* Holden-Day (1970).

[43] Glen E. Bredon, *Introduction to Compact Transformation Groups,* Academic Press (1972).

[44] David R. Brillinger, *Time Series: Data Analysis and Theory,* Holt, Rinehart and Winston, Inc. (1975).

[45] David R. Brillinger, *Time Series, Data Analysis and Theory,* Expanded Edition, Holden-Day (1981).

[46] Piet M. T. Broersen, "Selecting the order of autoregressive models from small samples", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-33** (4) 874-879 (August 1985).

[47] *Joel G. Broida and S. Gill Williamson, A Comprehensive Introduction to Linear Algebra,* Addison-Wesley (1989).

[48] Kevin M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-35** (3) 249-266 (March 1987).

[49] William S. Burdic, *Underwater Acoustic Systems Analysis,* Prentice-Hall (1984).

[50] George C. Canavos, *Applied Probability and Statistical Methods,* Little, Brown and Co. (1984).

[51] J. Capon, "High-resolution frequency-wavenumber spectrum analysis", *Proceedings of the IEEE* **57** (8) 1408-1418 (1969).

[52] J. Capon, "Maximum-likelihood spectral estimation", in *Nonlinear Methods of Spectral Analysis*, edited by S. Haykin, Springer (New York, 1979) pp. 155-179.

[53] M. Carmeli, *Statistical Theory of Energy Levels and Random Matrices in Physics* (ARL 73-0044), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (1973).

[54] W. G. Chen, J. P. Reilly, and K. M. Wong, *On the Detection of Number of Signals in High Resolution Array Processing,* (Research Memoranda), McMaster University, Hamilton, Ontario, Canada: Communications Research Laboratory (April 1988).

[55] William G. Cochran, *Sampling Techniques*, 3rd Edition, Wiley (1977).

[56] A. G. Constantine and A. T. James, "On the general canonical correlation distribution", *Annals of Mathematical Statistics* **29** 1146-1166 (1958).

[57] A. G. Constantine, "Some non-central distribution problems in multivariate analysis", *Annals of Mathematical Statistics* **34** (4) 1270-1285 (December 1963).

[58] A. G. Constantine, "The distribution of Hotelling's generalised T_0^2", *Annals of Mathematical Statistics* **37** (1) 215-225 (February 1966).

[59] A. G. Constantine and Robb J. Muirhead, "Asymptotic Expansions for distributions of latent roots in multivariate analysis", *Journal of Multivariate Analysis* **6** 369-391 (1976)

[60] D. B. Cooper and P. W. Cooper, "Nonsupervised adaptive signal detection and pattern recognition", *Information and Control* **7** 416-444 (1964).

[61] Henry Cox, "Interrelated problems in estimation and detection, I", presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics held at Enschede, The Netherlands (12-23 August 1968).

[62] H. Cramèr and M. R. Leadbetter, *Stationary and Related Stochastic Processes*, Wiley (1967).

[63] Christian F. Cremona and John A. Brandon, "Model order testing in discrete time domain identification methods", *Mechanical Systems and Signal Processing* **6** (3) 229-236 (May 1992).

[64] *Morton Landers Curtis, Matrix Groups*, 2^{nd} Edition, Springer-Verlag (1984).

[65] A. W. Davis, "On the ratios of individual latent roots to the trace of a Wishart matrix", *Journal of Multivariate Analysis* **2** 440-443 (1972).

[66] L. D. Davisson, "The prediction error of stationary Gaussian time series of unknown covariance", *IEEE Transactions on Information Theory* **IT-11** 527-532 (October 1965).

[67] *Walter L. Deemer and Ingram Olkin*, "The Jacobians of certain matrix transformations useful in multivariate analysis", *Biometrika* **38** (3 and 4) 345-367 (December 1951).

[68] David J. DeFatta, Joseph G. Lucas, and William S. Hodgkiss, *Digital Signal Processing: A System Design Approach*, Wiley (1988).

[69] J. G. de Gooijer, B. Abraham, A. Gould, and L. Robinson, "Methods for determining the order of an autoregressive-moving average process: a survey", *International Statistical Review* **55** (3) 301-329 (1985).

[70] *J. Dugundji*, "Envelopes and pre-envelopes of real waveforms", *IRE Transactions on Information Theory* **IT-4** (1) 53-57 (March 1958).

[71] J. J. Durbin, "The fitting of time series models", *Rev. L'Institut Intl. de Statisque* **28** (3) 233-244 (1960).

[72] *Paul S. Dwyer*, "Some applications of matrix derivatives in multivariate analysis", *American Statistical Association Journal 607-625 (June 1967).*

[73] Morris L. Eaton and M. D. Perlman, "The non-singularity of generalized sample covariance matrices", *Annals of Statistics* 710-717 (1970).

[74] *Morris L. Eaton, Multivariate Statistics, A Vector Space Approach*, Wiley (1983).

[75] Morris L. Eaton, *Group Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics **1**, Institute of Mathematical Statistics (Hayward, California, 1989).

[76] D. J. Edelblute, J. M. Fisk, and G. L. Kinnison, "Criteria for optimum-signal-detection theory for arrays", *Journal of the Acoustical Society of America* **41** (1) 199-205 (1967).

[77] Davis M. Egle, "Diffuse wave fields in solid media", *Journal of the Acoustical Society of America* **70** (2) 470-480 (August 1981).

[78] T. W. Epps, "Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference", *The American Statistician* **47** (1) 33-38 (February 1993).

[79] C. Fang and P. R. Krishnaiah, "Asymptotic distributions of functions of eigenvalues of the real and complex noncentral Wishart matrices", in *Statistics and Related Topics*, edited by M. Csörgö, D. A. Dawson, J. N. K. Rao, and A. K. Md. E. Saleh, North-Holland (1981) pp. 89-108.

[80] R. H. Farrell, "Calculation of complex zonal polynomials", in *Multivariate Analysis - V*, edited by P. R. Krishnaiah (North Holland, 1980), pp. 301-320.

[81] Meir Feder, "Maximum Entropy as a special case of the Minimum Description Length criterion", *IEEE Transactions on Information Theory* **IT-32** (6) 847-849 (November 1986).

[82] *Robert Ferlez*, informal discussions (August - December 1992).

[83] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population", *Biometrika* **10** 507 (1915).

[84] R. A. Fisher, "The mathematical distributions used in the common tests of significance", *Econometrica* **3** (4) 353-365 (October 1935).

[85] Gerald B. Folland, *Real Analysis, Modern Techniques and Their Applications*, Wiley (1984).

[86] Mark L. Fowler, "Signal detection using time-frequency and time-scale methods", Ph.D. thesis, The Pennsylvania State University, University Microfilms Inc. No. 92-14160 (1991).

[87] David Freedman and David Lane, "The empirical distribution of Fourier coefficients", *Annals of Statistics* **8** (6) 1244-1251 (1980).

[88] Benjamin Friedlander, "A signal subspace method for adaptive interference cancellation", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-36** (12) 1835-1845 (December 1988).

[89] Jean-Jacques Fuchs, "Estimating the number of sinusoids in additive white noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-36** (12) 1846-1853 (December 1988).

[90] M. A. Girshick, "Principal components", *Journal of the American Statistical Association* **31** (September 1936).

[91] M. A. Girshick, "On the sampling theory of roots of determinantal equations", *Annals of Mathematical Statistics* **10** 203-224 (1939).

[92] *N. R. Goodman*, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)", *Annals of Mathematical Statistics* **34**, 152-177 (1963).

[93] *N. R. Goodman*, "The distribution of the determinant of a complex Wishart distributed matrix", *Annals of Mathematical Statistics,* **34** 178-180 (1963).

[94] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products,* Corrected and Enlarged Edition, Academic Press (1980).

[95] *Franklin A. Graybill, Matrices with Applications in Statistics,* 2nd Edition, Wadsworth (1983).

[96] *Kenneth I. Gross and Donald St. P. Richards*, "Special functions of matrix argument. I: Algebraic induction, zonal polynomials, and hypergeometric functions", *Transactions of the American Mathematical Society* **301** (2) 781-811 (June 1987).

[97] *Kenneth I. Gross and Donald St. P. Richards*, "Total positivity, spherical series, and hypergeometric functions of matrix argument", *Journal of Approximation Theory* **59** 224-246 (1989).

[98] R. P. Gupta, "Asymptotic theory for principal component analysis in the complex case", *Journal of the Indian Statistical Association* **3** 97-106 (1965).

[99] L. Guttman, "Some necessary conditions for common-factor analysis", *Psychometrika* **19** 149-161 (1954).

[100] V. Haggan and M. B. Priestley, *Dimensionality reduction in multivariable stochastic systems*, (Technical Report 44), University of Manchester Institute of Science and Technology, Manchester, England: Department of Mathematics (1974).

[101] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression", *Journal of the Royal Statistical Society, Series B (Methodological)* **41** 190-195 (1979).

[102] *G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities,* 2^{nd} Edition, Cambridge University Press (1952).

[103] *Sabih I. Hayek*, "Integration of non-symmetric or odd functions", Lecture handout (07 February 1985).

[104] *Michiel Hazewinkel* (editor), *Encyclopaedia of Mathematics,* Kluwer (1992) [an annotated translation from Russian of: I. M. Vinogradov (editor), *Soviet Mathematical Encyclopaedia*].

[105] Sigurdur Helgason, *Groups and Geometric Analysis,* Academic Press (1984).

[106] *Carl S. Herz*, "Bessel functions of matrix argument", *Annals of Mathematics* **61** 474-523 (1955).

[107] Keith W. Hipel, "Geophysical model discrimination using the Akaike Information Criterion", *IEEE Transactions on Automatic Control* **AC-26** (2) 358-378 (April 1981).

[108] Paul G. Hoel, "A significance test for component analysis", *Annals of Mathematical Statistics* **8** 149-158 (1937).

[109] *Robert V. Hogg and Allen T. Craig, Introduction to Mathematical Statistics,* 4^{th} Edition, Macmillan (1978).

[110] Michael L. Honig and David G. Messerschmitt, *Adaptive Filtering: Structures, Algorithms, and Applications*, Kluwer (1984).

[111] J. D. Horel, "Complex principal component analysis: theory and examples", *Journal of Climate and Applied Meterology* **23** (12) 1660-1673 (December 1984).

[112] *Roger A. Horn and Charles R. Johnson, Matrix Analysis,* Cambridge University Press (1985).

[113] Hotelling, "Analysis of a complex of statistical variables into principal components", *The Journal of Educational Psychology* (September and October 1933).

[114] R. E. Ingham, "An integral which occurs in statistics", *Proceedings of the Cambridge Philosophical Society* **29** 271-276 (1933).

[115] *Kiyosi Itô* (editor), *Encyclopedic Dictionary of Mathematics*, 2nd Edition, MIT Press (1987) [a translation from Japanese, published by the Mathematical Society of Japan].

[116] Nathan Jacobson, *Basic Algebra*, W. H. Freeman (1985).

[117] Alan T. James, "The non-central Wishart distribution", *Proceedings of the Royal Society, Series A, Mathematical and Physical Sciences* **229** (1178) 364-366 (10 May 1955).

[118] Alan T. James, "The distribution of the latent roots of the covariance matrix", *Annals of Mathematical Statistics* **31** 151-158 (1960).

[119] Alan T. James, "Zonal polynomials of the real positive definite symmetric matrices", *Annals of Mathematics* **74** (3) 456-469 (November 1961).

[120] *Alan T. James*, "Distributions of matrix variates and latent roots derived from normal samples", *Annals of Mathematical Statistics* **35** 475-501 (1964).

[121] Alan T. James, "Inference on latent roots by calculation of hypergeometric functions of matrix argument", in *Multivariate Analysis* edited by Parchuri R. Krishnaiah, (Academic Press, 1966) pp. 209-235.

[122] W. J. Jobst and S. L. Adams, "Statistical analysis of ambient noise", *Journal of the Acoustical Society of America* **62** (1) 63-71 (July 1977).

[123] *Soren Johansen*, "The product limit estimator as maximum likelihood estimator", *Scandinavian Journal of Statistics* **5** 195-199 (1978).

[124] Charles R. Johnson, "Positive definite matrices", *American Mathematical Monthly* **77** 259-264 (1970).

[125] Don H. Johnson, "Application of spectral estimation methods to bearing estimation problems", *Proceedings of the IEEE* **70** (9) 1018-1028 (September 1982).

[126] Pierre Jolicoeur, "Principal components, factor analysis, and multivariate allometry: a small-sample direction test", *Biometrics* **40** (3) 685-690 (September 1984).

[127] D. G. Kabe, "Complex analogues of some classical noncentral multivariate distributions", *Australian Journal of Statistics* **8** 99-103 (1966).

[128] T. Kailath, "RKHS approach to detection and estimation problems - Part I: Deterministic signals in Gaussian noise", *IEEE Transactions on Information Theory* **17** 530-549 (1971).

[129] R. L. Kashyap, "Inconsistency of the AIC rule for estimating the order of autoregressive models", *IEEE Transactions on Automatic Control* **AC-25** (5) 996-998 (October 1980).

[130] Mostafa Kaveh and Arthur J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-34** (2) 331-341 (April 1986).

[131] James P. Keener, *Principles of Applied Mathematics: Transformation and Approximation*, Addison-Wesley (1988).

[132] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise", *Journal of the Society of Industrial and Applied Mathematics* **8** 309-341 (1960).

[133] E. J. Kelly, I. S. Reed, and W. L. Root, "A covariance approach to spectral moment estimation", *Journal of the Society of Industrial and Applied Mathematics* **8** 481-507 (1960).

[134] James G. Kelly, Robert N. Carpenter, and John A. Tague, "Object classification and acoustic imaging with active sonar", *Journal of the Acoustical Society of America* **91** (4) Part 1, 2073-2081 (April 1992).

[135] C. G. Khatri, "Wishart distribution. Queries and answers section", *Journal of the Indian Statistical Association* **1** 52-54 (1963).

[136] C. G. Khatri, "Further contributions to Wishartness and independence of second degree polynomials in normal vectors", *Journal of the Indian Statistical Association* **1** 61-70 (1963).

[137] *C. G. Khatri*, "Classical statistical analysis based on a certain multivariate complex Gaussian distribution", *Annals of Mathematical Statistics* **36** (1) 98-114 (February 1965).

[138] C. G. Khatri, "A test for reality of a covariance matrix in a certain complex Gaussian distribution", *Annals of Mathematical Statistics* **36** (i) 115-119 (February 1965).

[139] C. G. Khatri, "On the exact finite series distribution of the smallest or the largest root of matrices in three situations", *Journal of Multivariate Analysis* **2** 201-207 (1972).

[140] *Jack Carl Kiefer and Jacob Wolfowitz*, "Consistency of the maximum likclihood estimator in the presence of infinitely many incidental parameters", *Annals of Mathematical Statistics* **27** 887-906 (1956).

[141] *A. N. Kolmogorov and S. V. Fomin, Introductory Real Analysis,* Prentice-Hall (1970).

[142] Konstantinos Konstantinides and Kung Yao, "Statistical analysis of effective singular values in matrix rank determination", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36** (5) 757-763 (May 1988).

[143] Steven George Krantz, Math 523 lecture notes (19 February 1986).

[144] *Paruchuri R. Krishnaiah and V. B. Waikar, Simultaneous Tests for Equality of Latent Roots Against Certain Alternatives* (ARL 69-0119), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (July 1969), (NTIS No. AD-697 996).

[145] *Paruchuri R. Krishnaiah and V. B. Waikar, Simultaneous Tests for Equality of Latent Roots Against Certain Alternatives - II* (ARL 69-0178), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (October 1969), (NTIS No. AD-700 770).

[146] *Paruchuri R. Krishnaiah and T. C. Chang, On the Exact Distribution of the Smallest Root of the Wishart Matrix Using Zonal Polynomials* (ARL 70-0242), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (October 1970), (NTIS No. AD-717 703).

[147] *Paruchuri R. Krishnaiah and V. B. Waikar, Exact Distributions of the Intermediate Roots of a Class of Random Matrices* (ARL 70-0280), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (December 1970), (NTIS No. AD-718 427).

[148] *Paruchuri R. Krishnaiah and V. B. Waikar, Exact Joint Distributions of Few Roots of a Class of Random Matrices* (ARL 70-0345), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (December 1970), (NTIS No. AD-721 190).

[149] *Paruchuri R. Krishnaiah and V. B. Waikar, Exact Joint Distribution of Any Few Ordered Roots of a Class of Random Matrices* (ARL 71-0251), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (September 1971), (NTIS No. AD-734 495), reprinted in *Journal of Multivariate Analysis* **1** (3) 308-315 (September 1971).

[150] *Paruchuri R. Krishnaiah, On the Exact Distributions of the Test Statistics Based on the Eigenvalues of Complex Random Matrices* (ARL 73-0178), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (December 1973), (NTIS No. AD-776 315).

[151] *Paruchuri R. Krishnaiah and F. J. Schuumann, On the Exact Distributions of the Ratios of the Individual Roots to the Trace of the Complex and Real Wishart Matrices* (ARL 74-0026), Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (April 1974), (NTIS No. AD-780 617).

[152] *Paruchuri R. Krishnaiah*, "Some recent developments on complex multivariate distributions", *Journal of Multivariate Statistics* **6** 1-30 (1976).

[153] *Paruchuri R. Krishnaiah and Jack C. Lee*, "Inference on the eigenvalues of the covariance matrices of real and complex multivariate normal populations", *Multivariate Analysis IV*, edited by Paruchuri R. Krishnaiah (North Holland, 1977), pp. 96-97.

[154] Anant M. Kshirsagar, *Multivariate Analysis*, Marcel Dekker, Inc. (1972).

[155] Solomon Kullback, *Information Theory and Statistics*, Wiley (1959).

[156] Ramdas Kumaresan and Donald W. Tufts, "Estimating the angles of arrival of multiple plane waves", *IEEE Transactions on Aerospace and Electronic Systems* **AES-19** (1) 134-139 (January 1983).

[157] *Debasis Kundu*, an informal derivation (Spring, 1989).

[158] *Debasis Kundu*, "Some results in estimating the parameters of exponential signals in noise", Ph.D. thesis, The Pennsylvania State University, University Microfilms Inc. No. 90-18241 (1989).

[159] H. B. Kushner and Morris Meisner, "Formulas for zonal polynomials", *Journal of Multivariate Analysis* **14** 336-347 (1984).

[160] *Serge Lang, Complex Analysis,* 2^{nd} Edition, Springer-Verlag (1985).

[161] D. N. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices", *Biometrika* **43** 128-136 (1956).

[162] D. N. Lawley, "On testing a set of correlation coefficients for equality", *Annals of Mathematical Statistics* **34** 149-151 (1963).

[163] Harry B. Lee and Michael S. Wengrovitz, "Resolution threshold of Beamspace MUSIC for two closely spaced emitters", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-38** (9) 1545-1559 (September 1990).

[164] M. Lentner, "On the exact calculation of cumulative F probabilities", *Communications in Statistics, Series B* **5** 149-154 (1976).

[165] M. D. Levine and J. D. Irish, "A statistical description of temperature fine structure in the presence of internal waves", *Journal of Physical Oceanography* **11** 676-691 (1981).

[166] W. S. Liggett, Jr., "Passive sonar: fitting models to multiple time series", presented at the NATO Advanced Study Institute on Signal Processing held at the University of Technology, Loughborough, U.K. (1972), and published in *Signal Processing,* edited by J. W. Griffith et al. (Academic Press, 1973) pp. 327-345.

[167] *Dudley Ernest Littlewood, The Theory of Group Characters and Matrix Representations of Groups*, Clarendon Press (1940).

[168] Arthur E. Lord, Jr., "Acoustic emission - an update", *Physical Acoustics, Principles and Methods*, Vol. XV, edited by Warren P. Mason (Academic Press, 1981), pp. 295-360.

[169] John Makhoul, "Linear prediction: a tutorial review", *Proceedings of the IEEE* **63** 561-580 (April 1975).

[170] C. L. Mallows, "Some comments on C_p", *Technometrics* **15** 661-675 (1973).

[171] *K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate Analysis,* Academic Press (1979).

[172] Haruyoshi Matsumoto, "Characteristics of Sea MARC II Phase Data", *IEEE Journal of Ocean Engineering* **15** (4) 350-360 (October 1990).

[173] Suzanne T. McDaniel, "High frequency forward scatter from Arctic ice: temporal response", *IEEE Journal of Ocean Engineering* **17** (2) 216-221 (April 1992).

[174] M. L. Mehta, *Nuclear Physics* **18** 395-419 (1960).

[175] Neri Merhav, Michael Gutman, and Jacob Ziv, "On the estimation of the order of a Markov chain and universal data compression", *IEEE Transactions on Information Theory* **35** (5) 1014-1019 (September 1989).

[176] Neri Merhav, "The estimation of the model order in exponential families", *IEEE Transactions on Information Theory* **35** (5) 1109-1114 (September 1989).

[177] Kenneth S. Miller, *Multidimensional Gaussian Distributions*, Wiley (1964).

[178] Kenneth S. Miller and M. M. Rochwarger, "On estimating spectral moments in the presence of colored noise", *IEEE Transactions on Information Theory* **IT-16** 303-309 (1970).

[179] Kenneth S. Miller and M. M. Rochwarger, "A covariance approach to spectral moment estimation", *IEEE Transactions on Information Theory* **IT-18** 588-596 (1972).

[180] Kenneth S. Miller, *Complex Stochastic Processes*, Addison-Wesley, Reading (1974).

[181] Kenneth S. Miller, *Hypothesis Testing With Complex Distributions*, Robert E. Krieger Publishing Company, Huntington (1980).

[182] Michael I. Miller and Donald L. Snyder, "The role of likelihood and entropy in incomplete-data problems: applications to estimating point-process intensities and Toeplitz constrained covariances", *Proceedings of the IEEE* **75** (7) 892-907 (July 1987).

[183] Adam N. Mirkin, "Maximum likelihood estimation of the locations of multiple sources in an acoustic waveguide", Ph.D. thesis, The Pennsylvania State University (August 1992).

[184] Alireza Moghaddamjoo, "Eigenstructure Variability of the Multiple-Source Multiple-Sensor Covariance Matrix with Contaminated Gaussian Data", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36** (2) 153-167 (February 1988).

[185] Robert A. Monzingo and Thomas W. Miller, *Introduction to Adaptive Arrays,* Wiley, New York (1980).

[186] *Donald F. Morrison, Multivariate Statistical Methods*, 2^{nd} Edition, McGraw-Hill (1976).

[187] *Robb John Muirhead, Aspects of Multivariate Statistical Theory*, Wiley (1982).

[188] Robb John Muirhead, "Zonal polynomials", in *Encyclopedia of Statistical Sciences* **9,** edited by Samuel Kotz and Norman Lloyd Johnson III, Wiley (1988), pp. 676-682.

[189] B. N. Nagarsenker, P. B. Nagarsenker, and D. W. Quinn, "Asymptotic expansions of the non-central distributions of Wilks' statistic in the complex Gaussian case", *Communications in Statistics, Theory and Methods* **13** (12) 1493-1505 (1984).

[190] John Neter and William Wasserman, *Applied Linear Statistical Models*, Richard D. Irwin, Inc. (1974).

[191] John Neter, William Wasserman, and Michael H. Kutner, *Applied Linear Statistical Models*, 3rd Edition, Richard D. Irwin, Inc. (1990).

[192] Morris Newman, *Integral Matrices*, Academic Press (1972).

[193] *Katsumi Nomizu, Fundamentals of Linear Algebra*, McGraw-Hill (1966).

[194] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, "Sampling errors in estimation of empirical orthogonal functions", *Monthly Weather Review* **110** 699-706 (1982).

[195] Stephan J. Norton and Melvin Linzer, "Backprojection reconstruction of random source distributions", *Journal of the Acoustical Society of America* **81** (4) 977-985 (April 1987).

[196] *Adrian Ocneanu*, class lecture (31 August 1987).

[197] *Masashi Okamoto*, "Distinctness of the eigenvalues of a quadratic form in a multivariate sample", *Annals of Statistics* **1** (4) 763-765 (1973).

[198] Ingram Olkin and S. N. Roy, "On multivariate distribution theory", *Annals of Statistics* **25** 329-339 (1954).

[199] Ingram Olkin and Herman Rubin, "A characterization of the Wishart distribution", *Annals of Mathematical Statistics* **33** 1272-1280 (1962).

[200] Ingram Olkin, Leon J. Gleser, and Cyrus Derman, *Probability Models and Applications*, Macmillan Publishing Company (1980)

[201] J. E. Overland and R. W. Preisendorfer, "A significance test for principal components applied to a cyclone climatology", *Monthly Weather Review* **110** 1-4 (1982).

[202] Kalyanapuram Rangachari Parthasarathy, *Introduction to Probability and Measure*, Springer-Verlag (1977).

[203] Emanuel Parzen, "Some recent advances in time series modeling", *IEEE Transactions on Automatic Control* **AC-19** (6) 723-730 (December 1974).

[204] *G. P. Patil, M. T. Boswell, and M. V. Ratnaparkhi, Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, Vol. 2: Continuous Univariate Models,* International Co-operative Publishing House (Fairland, Maryland, 1984).

[205] *G. P. Patil, M. T. Boswell, M. V. Ratnaparkhi, and J. J. J. Roux, Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work, Volume 3: Multivariate Models*, International Co-operative Publishing House (Fairland, Maryland, 1984).

[206] Lionel Pelkowitz and Stuart C. Schwartz, "Asymptotically optimum sample size for quickest detection", *IEEE Transactions on Aerospace and Electronic Systems* **AES-23** (2) 263-272 (March 1987).

[207] *E. R. Phillips, An Introduction to Analysis and Integration Theory,* Dover (1984).

[208] Surendra Prasad and Bindu Chandna, "Direction of arrival estimation using stochastic model order reduction via state space modeling", *Signal Processing* **23** 157-177 (1991).

[209] M. B. Priestley, T. Subba Rao, and Howell Tong, "Identification of the structure of multivariable stochastic systems", *Multivariate Analysis III*, edited by P. R. Krishnaiah (Academic Press, 1973).

[210] M. B. Priestley, T. Subba Rao, and Howell Tong, "Applications of principal component analysis and factor analysis in the identification of multivariable systems", *IEEE Transactions on Automatic Control* **AC-19** (6) 730-734 (December 1974).

[211] Tarmo M. Pukkila and Paruchuri R. Krishnaiah, "On the use of autoregressive order determination criteria in univariate white noise tests", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-36** (5) 764-774 (May 1988).

[212] *Calyampudi Radhakrishna Rao*, "The use and interpretation of principal component analysis in applied research", *Sankhyā: The Indian Journal of Statistics, Series A* **26** 329-358 (1964).

[213] *Calyampudi Radhakrishna Rao, Linear Statistical Inference and Its Applications*, 2^{nd} Edition, Wiley ('973).

[214] Calyampudi Radhakrishna Rao, "Matrix approximations and reduction of dimensionality in multivariate statistical analysis", in *Multivariate Analysis* - **V**, edited by Paruchuri R. Krishnaiah, North-Holland (1980), pp. 3-22.

[215] *Calyampudi Radhakrishna Rao*, lecture notes (9 February 1989).

[216] *Calyampudi Radhakrishna Rao*, informal derivation (7 March 1989).

[217] *Calyampudi Radhakrishna Rao*, lecture notes (23 March 1989).

[218] *Calyampudi Radhakrishna Rao*, lecture notes (28 March 1989).

[219] Calyampudi Radhakrishna Rao and Yuehua Wu, "A strongly consistent procedure for model selection in a regression problem". *Biometrika* **76** (2) 369-374 (1989).

[220] S. O. Rice, "Mathematical analysis of random noise", *Bell System Technical Journal* **23** 282-332 (1944), and in *Selected Papers on Noise and Stochastic Processes*, edited by Nelson Wax, Dover Publications (New York) pp. 133-183.

[221] S. O. Rice, "Mathematical analysis of random noise", *Bell System Technical Journal* **24** 46-156 (1945).

[222] M. B. Richman, "Obliquely rotated principal components: an improved meteorological map typing technique", *Journal of Applied Meteorology* **20** 1145-1159 (1981).

[223] J. Riordan, *Combinatorial Identities*, Wiley (1968).

[224] Jorma Rissanen, "Modeling by shortest data description", *Automatica* **14** 465-471 (1978).

[225] Jorma Rissanen, "Estimation of structure by Minimum Description Length", *Circuits, Systems, and Signal Processing* **CSSP-1** (3-4) 395-406 (Birkhauser Boston, Inc., 1982).

[226] Jorma Rissanen, "Universal coding, information, prediction, and estimation", *IEEE Transactions on Information Theory* **IT-30** (4) 629-636 (July 1984).

[227] W. L. Root and T. S. Pitcher, "On the Fourier series expansion of random functions", *Annals of Mathematical Statistics* **26** 313-318 (1955).

[228] *S. N. Roy*, "Some useful results in Jacobians", *Calcutta Statistical Association Bulletin* **4** 117-122 (1952).

[229] *Walter Rudin*, *Principles of Mathematical Analysis*, 3rd Edition, McGraw-Hill (1976).

[230] *Walter Rudin*, *Real and Complex Analysis*, 3rd Edition, McGraw-Hill (1987).

[231] *Daniel H. Saracino*, *Abstract Algebra, A First Course*, Addison-Wesley (1980).

[232] Leonard J. Savage, *The Foundations of Statistics*, Wiley (1954).

[233] John G. Saw, "Zonal polynomials: an alternative approach", *Journal of Multivariate Analysis* **7** 461-467 (1977).

[234] John G. Saw, "Ultraspherical polynomials and statistics on the m-sphere", *Journal of Multivariate Analysis* **14** 105-113 (1984).

[235] Louis L. Scharf and Dean W. Lytle, "Stability of parameter estimates for a Gaussian process", *IEEE Transactions on Aerospace and Electronic Systems* **AES-9** (6) 847-851 (November 1973).

[236] Louis L. Scharf and Donald W. Tufts, "Rank reduction for modeling stationary signals", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-35** (3) 350-355 (March 1987).

[237] Louis L. Scharf and Barry D. Van Veen, "Low rank detectors for Gaussian random vectors", *IEEE Transactions on Acoustics, Speech, and Signal Processing,* **ASSP-35** (11) 1579-1582 (November 1987).

[238] Ralph O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation* **AP-34** (3) 276-280 (March 1986).

[239] G. Schwarz, "Estimating the dimension of a model", *Annals of Statistics* **6** 461-464 (1978).

[240] Ramachandra Ganesh Shenoy, "Group Representations and Optimal Recovery in Signal Modeling", Ph.D. thesis (August 1989).

[241] R. Shibata, "Selection of the order of an autoregressive model by Akaike's information criterion", *Biometrika* **63** 117-126 (1976).

[242] *G. E. Shilov and B. L. Gurevich, Integral, Measure and Derivative: A Unified Approach*, Prentice-Hall (1966).

[243] D. Short and J. Summerscales, "Amplitude distribution acoustic emission signatures of unidirectional fibre composite hybrid materials", *Composites* **15** (3), 200-206 (July 1984).

[244] M. Simaan, A. Hussain, and M. T. Hanna, "Nondestructive detection of flaws in material using transducer arrays", *Journal of Nondestructive Evaluation* **6** (1) 47-55 (March 1987).

[245] Anita Singh and K. C. S. Pillai, "On the exact non-null distribution of Wilks' L_{vc} criterion in the complex case", *Sankhyā: The Indian Journal of Statistics, Series B, Part 1* **45** 12-19 (1983).

[246] Naresh Kumar Sinha and B. Kuzta, *Modeling and Identification of Dynamic Systems*, Van Nostrand Reinhold (1983).

[247] V. Sisking, "Second moments of inverse Wishart-matrix elements", *Biometrika* **59** 691-692 (1972).

[248] Eugen J. Skudrzyk, *The Foundations of Acoustics, Basic Mathematics and Basic Acoustics*, Springer-Verlag (1971).

[249] V. I. Smirnov, *Linear Algebra and Group Theory*, McGraw-Hill (1960), p. 85.

[250] T. Söderström, "On model structure testing in system identification", *International Journal of Control* **26** (1) 1-18 (1977).

[251] E. D. Solomentsev, "Spherical harmonics", in *Encyclopaedia of Mathematics* **8** edited by Michiel Hazewinkel, Kluwer Academic Publ. (1992), p. 438.

[252] Daniel Solow, *How to Read and Do Proofs*, 2^{nd} Edition, Wiley (1990).

[253] Murray R. Spiegel, *Complex Variables*, Schaum's Outline Series in Mathematics, McGraw-Hill (1964).

[254] Michael Spivak, *Calculus on Manifolds, a Modern Approach to Classical Theorems of Advanced Calculus*, W. A. Benjamin (1965).

[255] Michael Spivak, *A Comprehensive Introduction to Differential Geometry*, 2^{nd} Edition, 5 volumes, Publish or Perish, Inc. (1979).

[256] *Muni Shanker Srivastava*, "On the complex Wishart distribution", *Annals of Mathematical Statistics* **36** (1) 313-315 (February 1965).

[257] *Muni Shanker Srivastava and C. G. Khatri, An Introduction to Multivariate Statistics*, North Holland (1979).

[258] *Elias M. Stein and Guido Weiss, Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Mathematical Series **32**, Princeton University Press (1971).

[259] *G. W. Stewart, Introduction to Matrix Computations*, Academic Press (1973).

[260] G. W. Stewart and Ji-guang Sun, *Matrix Perturbation Theory*, Academic Press (1990).

[261] Hans von Storch and Gerhard Hannoschöck, "Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes", *Journal of Climate and Applied Meteorology* **24** 716-724 (1985).

[262] *George P. H. Styan*, "Three useful expressions for expectations involving a Wishart matrix and its inverse", *Statistical Data Analysis and Inference,* edited by Y. Dodge (North-Holland, 1989), pp. 283-296.

[263] John A. Tague, "Estimation-correlation, modelling, and identification in adaptive array processors", Ph.D. thesis, The Pennsylvania State University (December 1987).

[264] *John A. Tague*, "Expectations of useful complex Wishart forms", (July 1992) accepted for publication in *Multidimensional Systems and Signal Processing.*

[265] *Akimichi Takemura, Zonal Polynomials*, Lecture Notes - Monograph Series **4**, Institute of Mathematical Statistics, Hayward, California (1984).

[266] W. Y. Tan, "Some distribution theory associated with complex Gaussian distribution", *Tamkang Journal of Mathematics* **7** 263-302 (Taipei, Taiwan 1968).

[267] *Derrick S. Tracy and Paul S. Dwyer*, "Multivariate maxima and minima with matrix derivatives", *American Statistical Association Journal,* 1576-1594 (December 1969).

[268] *Jan J. Tuma, Engineering Mathematics Handbook, Definitions, Theorems, Formulas, Tables,* 2^{nd} Edition, Enlarged, and Revised Edition, McGrall-Hill (1979).

[269] David E. Tyler, "The asymptotic distribution of principal component roots under local alternatives to multiple roots", *Annals of Statistics* **11** (4) 1232-1242 (1983).

[270] David E. Tyler, "A class of asymptotic tests for principal component vectors", *Annals of Statistics* **11** (4) 1243-1250 (1983).

[271] *Naum Iakolevich Vilenkin, Special Functions and the Theory of Group Representations,* translation of *Spetsial'nye funktsii i teoriia prestablenii grupp* (romanized form of Russian title), Translations of Mathematical Monographs **22**, American Mathematical Society (1968).

[272] *V. B. Waikar, T. C. Chang, and Paruchuri R. Krishnaiah, Exact Distributions of a Few Arbitrary Roots of Some Complex Random Matrices* (ARL 71-0035),Wright-Patterson Air Force Base, Ohio: Aerospace Research Laboratories (February 1971), (NTIS No. AD-726 594).

[273] V. B. Waikar, T. C. Chang, and Paruchuri R. Krishnaiah, *Australian Journal of Statistics* **14** 84-88 (1972).

[274] V. B. Waikar, "Distributions of characteristic roots of random matrices", in *Statistical Distributions in Scientific Work, Volume 1,* edited by G. P. Patil, S. Kotz, and J. Keith Ord, D. (Reidel Publishing Company, Cordrecht-Holland / Boston-USA, 1975), pp. 391-400.

[275] Hong Wang and Mostafa Kaveh, "On the performance of signal-subspace processing-Part I: narrow-band systems", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-34** (5) 1201-1209 (October 1986).

[276] Hong Wang and Mostafa Kaveh, "On the performance of signal-subspace processing-Part II: wide-band systems", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-35** (11) 1583-1591 (November 1987).

[277] G. S. Watson, "Some estimation theory on the sphere", *Annals of the Institute of Statistical Mathematics* **38 (2)** Part A 263-275 (1986)

[278] Mati Wax and Thomas Kailath, "Determining the number of signals by information theoretic criteria", in *Proceedings of the ASSP Spectrum Estimation Workshop II, Tampa, FL* (1983) pp. 192-196

[279] Mati Wax, Tie-Jun Shan, and Thomas Kailath, "Spatio-temporal spectral analysis by eigenstructure methods", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-32** (4) 817-827 (August 1984).

[280] Mati Wax and Thomas Kailath, "Detection of signals by information theoretic criteria", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-33** (2) 387-392 (April 1985).

[281] Mati Wax and Ilan Ziskind, "Detection of the number of coherent signals by the MDL principle", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-37** (8) 1190-1196 (August 1989).

[282] Mati Wax, "Detection and localization of multiple sources via the stochastic signals model", *IEEE Transactions on Signal Processing* **39** (11) 2450-2456 (November 1991).

[283] Anthony D. Whalen, *Detection of Signals in Noise,* Academic Press, pp. 61-83 (1971).

[284] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices", *Annals of Mathematics* **67** 325-326 (1958).

[285] E. P. Wigner, "Statistical properties of real symmetric matrices with many dimensions", *Proceedings of the Fourth Canadian Mathematical Congress* (University of Toronto Press, Toronto, Canada) 174-184 (1959).

[286] E. P. Wigner, "Distribution laws for the roots of a random Hermitian matrix", *Statistical Theories of Spectra: Fluctuations*, edited by C. E. Porter (Academic Press, 1965), pp. 446-461.

[287] E. P. Wigner, "Random matrices in physics", *SIAM Review* **9** 1-23 (1958).

[288] *Robert A. Wijsman, Invariant measures on Groups and Their Use in Statistics*, Lecture Notes-Monograph Series **14**, Institute of Mathematical Statistics (1990).

[289] Samuel S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses", *Annals of Mathematical Statistics* **9** 60-62 (1938).

[290] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population", *Biometrika* **20A** 32-52 (July 1928).

[291] Jacob Wolfowitz, "On stochastic approximation methods", *Annals of Mathematical Statistics* **27** 1151-1155 (December 1956).

[292] K. M. Wong, Q. T. Zhang, J. P. Reilly, and P. Yip, "A new criterion for the determination of the number of signals in high-resolution array processing", *Proceedings of ICASSP'88*, New York (April 1988) pp. 2769-2772.

[293] *R. A. Wooding*, "The multivariate distribution of complex normal variables", *Biometrika* **43** 212-215 (1956).

[294] *A. David Wunsch, Complex Variables With Applications*, Addison-Wesley (1983).

[295] Kai-Bor Yu, "Recursive updating the eigenvalue decomposition of a covariance matrix", *IEEE Transactions on Signal Processing* **39** (5) 1136-1145 (May 1991).

[296] Qi-tu Zhang, Kon Max Wong, Patrick C. Yip, and James P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37** (10) 1557-1567 (October 1989).

[297] L. C. Zhao, Paruchuri R. Krishnaiah, and Z. D. Bai, "On detection of number of signals in presence of white noise", *Journal of Multivariate Analysis* **20** (1) 1-25 (October 1986).

[298] L. C. Zhao, Paruchuri R. Krishnaiah, and Z. D. Bai, "On detection of number of signals when the noise covariance matrix is arbitrary", *Journal of Multivariate Analysis* **20** (1) 26-49 (October 1986).

[299] Lawrence J. Ziomek, *Underwater Acoustics: A Linear Systems Theory Approach*, Academic Press (1985).

[300] J. Ziv and A. Lempel, "Compression of individual sequences via variable rate coding", *IEEE Transactions on Information Theory* **IT-24** 530-536 (1978).

Appendix A

MATHEMATICAL BACKGROUND

A.1 Organization of Appendices

Much of the material presented in this and other appendices to follow is new in the sense that the forms presented here have not been expressed explicitly in the context of a development of the algebra of complex valued vectors and matrices. Likewise, most of it is old in the sense that the basic concepts are well understood, occasionally have been proven for a more general case which includes complex variables, or are so trivial that no one thought them important enough to write down for publication.

The appendices are arranged as follows. The first set (A-F) consists of material for this thesis which is necessary background and which I think readers would be most interested in referring to. This set includes the appendices on matrix differential operators, definitions and properties of distributions, and density functions of distributions. The material on characteristic functions is especially important, and also fun.

The second set of appendices (G-J) consists of very important abstract mathematical background which any researcher desiring to extend or critique this thesis must master, but which I think is of secondary interest to the reader who wants instead to be a user of the results of this thesis. This

second set includes the appendices on zonal polynomials, group theory, Hilbert space, and complex vector space. I expect this to be of primary value to those with engineering background who need to quickly understand the group theoretic concepts. A mathematician will find this section trite. I found the source material on zonal polynomials at times very difficult. Perhaps the most challenging and important contributions of this thesis are found in the appendix on zonal polynomials. My hope is that you will realize a time savings in developing an understanding by this translation from the language of mathematicians to the language of engineers.

The third set of appendices (K-P) consists of basic linear algebra of complex matrices. It is material I expect any technical senior undergraduate to be capable of producing, but which is not assembled elsewhere in texts or the research literature in an expository fashion. The real variables forms of most of these results are part of the routine working set of knowledge for people who have had one reasonable course in matrix algebra. It is included because the details of the results are used throughout this thesis and I have noticed that the similarity of forms with the real variables cases have occasionally led other researchers to make minor errors by a factor of 2, typically with constant multipliers or powers.

A.2 Mathematical Background References

Although this thesis is of primary importance to engineers, the type of mathematics used in this thesis is outside the usual background of either engineers or statisticians. There are a few good books that provide good preparation for the material in this thesis. Even though a thesis is supposed to be self-contained and stand alone, knowing which references are useful makes the process of reading and learning far more efficient. It may also help provide the background for specialized terminology which I may not have adequately defined for a serious reader.

A.2.1 Linear Algebra

There are some very good books on linear algebra. My favorite is the quality text by Broida and Williamson [47]. Beyond the regular fare of linear algebra texts an engineer is likely to study from, it introduces groups in a natural way early in order to build on the concept. It has a very nice treatment of determinants and polynomials. Its presentation of a polynomial as an n-tuple is the key to avoiding the differential operator which Gross and Richards used in their development of zonal polynomials. Broida and Williamson also discuss multilinear mappings, exterior products, and Hilbert spaces. Also of interest

to engineers and physicists, they have a very nice introduction to tensors.

Another linear algebra text with a solid development is by Nomizu [193]. This is an undergraduate text intended for mathematics majors. Chapter 8 of [193] is very important for distinguishing properties associated with Hermitian ($A = A^H$), unitary ($AA^H = A^HA = I$), normal ($AA^H = A^HA$), symmetric $\{[(Ax, y) = (x, Ay)]$ for all $x, y \in V\}$, and orthogonal transformations. Note that there exist complex orthogonal transformations ($A^TA = AA^T = I$) as well as unitary ($B^HB = BB^H = I$) transformations, yet $A \neq B$. I know of no other text that points out these distinctions. In extending work from the real field to the complex field, this means that one cannot take for granted that properties claimed as the important ingredients for a proof really are the ones being used. In working with the field of real numbers, often stronger conditions are hypothesized when weaker ones would do. This is because for symmetric positive definite matrices, the matrix having the various properties coincide.

The book on numerical linear algebra by G. W. Stewart [259] is a gentle, yet mathematically respectable, advanced undergraduate or first year graduate text that treats complex matrices when it can be done without much additional effort. Chapter 5 on eigenvalues and eigenvectors is developed in $\mathbf{C}^n$, which is the natural setting for discussing Gerschgorin disks. This is a classic example where working in $\mathbf{C}$ makes a topic real easy, and working in $\mathbf{R}$ makes the

discussion very complex. He has a nice treatment on norms and condition numbers. He recommends perturbation estimates as a quick, non-rigorous look at what might otherwise be a mathematically difficult problem. Stewart and Sun coauthored a fine sequel [260] devoted to perturbation analysis that is also worthy of use. Among other topics, this book examines the relationship between the singular values of a matrix and partitions of that matrix, and also singular values of linear transformations of that matrix.

The book by Horn and Johnson [112] is a major important text that deserves to be read as a prerequisite to Rao's text [213] on multivariate analysis. Among its various topics, this book discusses complex symmetric matrices and Gershgorin disks. It has an encyclopedic treatment of matrix algebra. It does not treat differentiation or integration of matrices.

A.2.2 Multivariate Statistics

C. R. Rao's book [213] is a wonderful treatment of multivariate statistics. He does not shy away from powerful generalizations where it can be done profitably. He uses matrix notation throughout. He shares with Skudrzyk [248] the wonderful habit of carefully laying down the mathematical tools before leaping into the subject material requiring it. Unlike this thesis, the mathematical background is not hidden in appendices. He introduces those points of measure theory necessary for later work. C. R. Rao's references are

extensive and reflect a respectful sense of history.

The text by Eaton [74] is suitable for preparation for working with complex statistics because he approaches the subject using vector space and invariance methods. Other than Miller's books, Eaton's book says more about statistics of complex variables than any other multivariate text I have found. He includes some discussion on complex statistics and the relationship to statistics of real variables. Eaton is a nice repository of clever insights that make derivations much easier. For example, he imposes the condition $T = T^H$ to take advantage of the Hermitian symmetry of the covariance matrix to generate a change of variables of the standard complex normal distribution. This is used to obtain a chi-square distribution which becomes the seed for growing the Wishart distribution. This is a worthy book to study after reading Broida and Williamson.

Another nice text on multivariate analysis is the one by Arnold [31]. This book is an important contribution to the literature. It is remarkable for its clear development of properties of multivariate distributions and testing. Arnold makes it natural to think of statistics from a multivariate point of view, and to view univariate statistics as special cases. He applies group theory sparingly, but does so where it is clearly advantageous. Arnold's proof of the real Wishart distribution density function by induction is a contribution to the conceptual development. His presentation of the real matrix normal

distribution is also important.

A.2.3 Abstractions

Group invariance in statistics is a topic in statistics that has generated much attention by leading researchers in statistics, and is slowly making its way into the training of graduate students in theoretical statistics. I know of no simple text that introduces the concepts of group invariance, yet this subject is at the core of knowledge needed to make progress on the order estimation problem. One monograph that has proven useful is the 1989 work by Eaton [75]. It begins with topological groups and discusses Haar measure by page 6. The second chapter on group actions covers very rapidly the material covered by Vilenkin. He covers the very important topic of maximal invariants.

A text that is referenced by most authors at some point in the development of theory regarding zonal polynomials is the book by Littlewood [167]. This book is not one that can be rushed through, but rather must be worked through. Be prepared for lots of subscripts and tensor style notation. The reader should also have a basic understanding of abstract algebra and group theory. Mastery of this text will build a background not available from other sources which is necessary to understand current literature. Of particular interest, he treats groups of unitary matrices.

A.3 A Real Complex Treatment

It may seem unreasonable to the educated reader why I have bothered to prove some apparently obvious theorems in matrix algebra. The answer is that not all properties taken for granted in the real variables case carry over to complex matrices. Lack of attention to such issues has led some very well respected researchers to write down erroneous results by inspection based on forms known from the real variables case.

A clue that a result might need to be reproved (in both senses of the word) is when it requires symmetry or uses a transpose. Symmetry and Hermitian symmetry can both apply to complex matrices, yet they impose different properties. Symmetry has group theoretic properties which were used by earlier workers in the development of zonal polynomials for the application to the real Wishart matrix. In the development of linear algebra for real variables, the properties of symmetry and adjointness often go together and thus are often not distinguished when a proof requiring a property is done. They are usually treated synonymously. In $\mathbf{C}^n$, you can not afford that luxury.

In the context of linear operators, the function of the complex vectors x and y defined by $< x, y >= x^H y$ defines an inner product in the n-dimensional complex space $\mathbf{C}^n$, whereas the function defined by $(x, y) = x^T y$ does not. In

both cases you can produce an orthonormal set of vectors using an appropriate different version of the Gram-Schmidt process, but the results will in general be different. I have defined the inner product to be linear in the second argument rather than according to the mathematician's preference for assigning that property to the first argument. This was done to make use of the Hermitian transpose notation which carries with it natural meanings within the context of acoustics, engineering, and physics.

The space $\mathbf{C}^n$ is *not* the same as $\mathbf{R}^{2n}$. The structure imposed by the multiplication operator for complex numbers changes the nature of the space. This appears to not be widely understood, and it is very important to this thesis. Because of this, a few examples will be given to illustrate the problems involved. Smirnov [249] provides the following example.

The vectors $u = (1+i, 2i)$ and $v = (1, 1+i)$ in $\mathbf{C}^2$ are linearly dependent over the field of complex numbers $\mathbf{C}$, but are linearly independent over the field of real numbers $\mathbf{R}$. In the complex field $\mathbf{C}$, we have

$$u - v = w = (i, -1+i)$$

This implies that

$$-iw = (1, i+1) = (1, 1+i) = v$$

We observe that

$$-i(u-v) = -i(i, -1+i) = (1, 1+i) = v$$

Therefore

$$-iu + iv - v = -iu - (1-i)v = iu + (1-i)v = 0$$

which proves that u and v are linearly dependent. Now consider the same case, but in the field of real numbers $\mathbf{R}$. We vectorize u and v, by putting the real parts in the first two elements and the imaginary parts in the next two elements to obtain $u = (1,0,1,2)$ and $v = (1,1,0,1)$. Then $u - v = (0,-1,1,1) = w$. Then $-w = (0,1,-1,-1) \neq v$. We see that u and v are linearly independent when considered in $\mathbf{R}$. To multiply by i in $\mathbf{R}$ where $x = a + ib = (a,b)$, you must compute $ix = ia - b = (-b,a)$. This implies $iw = (-1,-1,0,-1)$ or $-iw = (1,1,0,1) = v$. When you are restricted to $\mathbf{R}$, the representation of $\mathbf{C}^n$ in $\mathbf{R}$ is not merely $\mathbf{R}^{2n}$. You have to modify the definition of scalar multiplication to allow a corollary to $i = \sqrt{-1}$.

We can represent complex numbers in matrix form, but not every choice will do. There is even a problem with multiplication by scalar complex numbers to be considered. Suppose we let $u = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$ and $v = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Then

$$u - v = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = w$$

This is not even close. Suppose instead that we try

$$u = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & -2 \\ 2 & 0 \end{pmatrix} \qquad v = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Then

$$u - v = \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & -1 \\ 1 & -1 \end{pmatrix} = w$$

which is what we want. We see that $-i$ is represented by the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Matrix multiplication for the form $-iw$ is not defined because the matrices $-i$ and w are not conformable. That is trying to multiply a 2×2 matrix by a 4×2 matrix. Instead, we must compute

$$w(-i) = -wi = \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

This is what we want.

The understanding that there are problems with differentiation when working with quadratic forms in complex variables is not widespread. A common

error is to apply the method of Lagrange multipliers via using derivatives to solve optimization problems. For example, in an otherwise very nice text, one author attempted to solve the problem

$$\min_{w} w^H \Phi^T w \text{ subject to } c^H w = f$$

by Lagrange multipliers. The derivative $\frac{d}{dw} w^H \Phi^T w$ does not exist except in the case that Φ is a diagonal matrix. When Φ is diagonal, the derivative exists only at $w = 0$, at which point it is zero.

A.4 The Rest of the Story

In this appendix, I have attempted to justify the documentation of the development of topics in statistics, matrix algebra of complex variables, and group representation theory which will be presented in following appendices. The use of complex variables is perhaps "too natural" in that we presume we know how to properly make the transition based on our experience with univariate complex variables and with matrix algebra of real variables. This is a false oasis. To help ease the transition, some references which have been very helpful to me are recommended to you to help shorten your transition period. Perhaps these may also provide some enjoyment to you as well. The structure and repetition of this material can be appreciated much in the manner of a poem. It takes patience, knowledge of the language, appreciation of the cultural con-

text, some understanding of the object of study, and a good environment for reflection and introspection.

Appendix B

MATRIX DIFFERENTIAL OPERATORS

B.1 Complex Derivatives

A major difference between working with real variables and working with complex variables is the extra care required to ensure that the desired derivatives exists. This has a major impact on the allowable approaches used to solve optimization problems. The theory of complex differentiation and the Cauchy-Riemann equations are part of any good course on complex variables. The purpose of this section is to raise a caution flag that some often used relationships in the case of real variables do not work for complex variables. Frequent errors in the adaptive beamforming literature has demonstrated that discussion of this topic is necessary. We begin by considering some simple functions whose derivatives do not exist.

The derivative of a scalar function is defined by the following.

$$\frac{d}{dz}f(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

The derivative does not exist if your answer depends on the *path* though the complex plane taken as $\Delta z \to 0$.

B.1.1 Computation and Cauchy-Riemann Equations

Recall from the development of the Cauchy-Riemann equations that *when* the derivative exists, *then* there are two ways to compute it. These are given by Wunsch (pp. 52-55) [294]. Let $z = x + iy$, $u(z) = \mathrm{Re}\{f(z)\}$, and $v(z) = \mathrm{Im}\{f(z)\}$. Then

$$f'(z_0) = \frac{d}{dz}f(z)\bigg|_{z_0} = \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right)\bigg|_{x_0, y_0} = \frac{\partial}{\partial x}f(z)\bigg|_{z_0}$$

and

$$f'(z_0) = \frac{d}{dz}f(z)\bigg|_{z_0} = \left(-i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right)\bigg|_{x_0, y_0} = -i\frac{\partial}{\partial y}f(z)\bigg|_{z_0}$$

These generate the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{B.1}$$

Satisfying these equations is a *necessary, but not sufficient*, condition for the existence of the derivative at z_0. These conditions are sufficient when $u, v, \frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial u}{\partial y}$, and $\frac{\partial v}{\partial y}$ are all continuous functions in the neighborhood of z_0. See Wunsch [294] for an *excellent* tutorial on this subject.

By substituting the Cauchy-Riemann equations back into $f'(z)$, we find

$$\frac{d}{dz}f(z) = \frac{\partial}{\partial x}f(z) = \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)u(z) = i\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)v(z) = -i\frac{\partial}{\partial y}f(z)$$

when $\frac{d}{dz}f(z)$ exists. Note that $\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) f(z) = \frac{d}{dz}f(z) - \frac{d}{dz}f(z) = 0$.

Caution. When you face a situation in which a derivative does not exist, you can *not* assume the derivative to merely be zero and ignore it. When the derivative does not exist, that means that any subsequent work requiring that derivative does not apply and cannot legally be used. This means you need to determine if there is another way to solve your problem.

B.1.2 Derivative of the Conjugate of a Variable

The following material appears very elementary, yet some very respected authors in adaptive beamforming and complex statistics have not understood it. Therefore, inclusion of this material is mandatory. It is this particular derivative which is the root of most mistakes in the literature, and is the foundation for the lack of existence of many other derivatives discussed in this section.

Let $z = x + iy$ and $z^* = x - iy$. Then $\frac{d}{dz}z^*$ does not exist, *anywhere.* This proof comes from Spiegel (p. 71) [253]. By definition,

$$\frac{d}{dz}z^* = \lim_{\Delta z \to 0} \frac{(z + \Delta z)^* - (z)^*}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{(x + iy + \Delta x + i\Delta y)^* - (x + iy)^*}{\Delta x + i\Delta y}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{x - iy + \Delta x - i\Delta y - x + iy}{\Delta x + i\Delta y} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\Delta x - i\Delta y}{\Delta x + i\Delta y}$$

Suppose $\Delta x = 0$. Then $\frac{d}{dz}z^* = \lim_{\Delta y \to 0} \frac{-i\Delta y}{i\Delta y} = -1$. Now suppose $\Delta y = 0$. Then $\frac{d}{dz}z^* = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} = 1$. Thus the answer you get depends on the path you take to get to the point which you evaluate the derivative. Here it is seen that there

is *no* point z_0 at which $\frac{d}{dz}z^*$ exists. In particular, $\frac{d}{dz}z^*$ does not even exist at $z_0 = 0$.

Note that this implies $\frac{d}{dz}\,\mathrm{Re}\,(z)$ and $\frac{d}{dz}\,\mathrm{Im}(z)$ do not exist because $\mathrm{Re}\,(z) = \frac{1}{2}(z + z^*)$ and $\mathrm{Im}(z) = \frac{1}{2i}(z - z^*)$. Since the derivative is a linear operator and $\frac{d}{dz}z^*$ does not exist, then these other derivatives also do not exist.

B.1.3 Derivative of the Magnitude

This was motivated by the discussion by Wunsch [294]. Although I have not looked for a statement of this result elsewhere, it is so elementary that I presume this is not an original result.

Let $z = x + iy$ and thus $|z| = (x^2 + y^2)^{1/2}$. Then $\frac{d}{dz}|z|$ does not exist anywhere.

$$\frac{d}{dz}|z| = \lim_{\Delta z \to 0} \frac{|z + \Delta z| - |z|}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{|x + iy + \Delta x + i\Delta y| - |x + iy|}{\Delta x + i\Delta y}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{((x + \Delta x)^2 + (y + \Delta y)^2)^{1/2} - (x^2 + y^2)^{1/2}}{\Delta x + i\Delta y}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{(x^2 + y^2 + 2(x\Delta x + y\Delta y) + (\Delta x)^2 + (\Delta y)^2)^{1/2} - (x^2 + y^2)^{1/2}}{\Delta x + i\Delta y}$$

Suppose $\Delta x = 0$. Then

$$\frac{d}{dz}|z| = \lim_{\Delta y \to 0} \frac{(x^2 + y^2 + 2y(\Delta y) + (\Delta y)^2)^{1/2} - (x^2 + y^2)^{1/2}}{i\Delta y}$$

At this point, we note that

$$\sqrt{a + b} - \sqrt{a} = \sqrt{\left(\sqrt{a + b} - \sqrt{a}\right)^2} = \sqrt{2a + b - 2\sqrt{a(a + b)}}$$

which implies

$$\frac{d}{dz}|z| = \lim_{\Delta y \to 0}\left[\frac{2(x^2+y^2)}{-(\Delta y)^2} + \frac{2y}{-\Delta y} - 1 - 2\sqrt{\frac{(x^2+y^2)(x^2+y^2+2y(\Delta y)+(\Delta y)^2)}{(\Delta y)^4}}\right]^{\frac{1}{2}}$$

which is unbounded, and tends to $i\infty$ if $y \geq 0$. This limit does not exist. This is sufficient to show $\frac{d}{dz}|z|$ does not exist. Just for completeness' sake, suppose $\Delta y = 0$. Then

$$\frac{d}{dz}|z| = \lim_{\Delta x \to 0}\frac{(x^2+y^2+2x(\Delta x)+(\Delta x)^2)^{1/2} - (x^2+y^2)^{1/2}}{\Delta x}$$

This is unbounded, and tends to ∞ if $x \geq 0$. Suppose $z = 0$. If $\Delta x = 0$ then

$$\frac{d}{dz}|z| = \lim_{\Delta y \to 0}\left[-1 - 2\sqrt{\frac{1}{(\Delta y)^2}}\right]^{\frac{1}{2}} \to i\infty$$

Suppose $\Delta y = 0$. Then

$$\frac{d}{dz}|z| = \lim_{\Delta x \to 0}(1) = 1$$

Thus $\frac{d}{dz}|z|$ does not exist even at $z = 0$.

B.1.4 Derivative of the Magnitude Squared

This is essentially the work on pp. 56-57 of Wunsch [294].

Here, we find that $\frac{d}{dz}|z|^2$ exists only at $z = 0$, at which point the derivative is zero.

$$\frac{d}{dz}|z|^2 = \lim_{\Delta z \to 0}\frac{|z+\Delta z|^2 - |z|^2}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}}\frac{|(x+\Delta x)+i(y+\Delta y)|^2 - |x+iy|^2}{\Delta x + i\Delta y}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}}\frac{x^2+2x(\Delta x)+(\Delta x)^2+y^2+2y(\Delta y)+(\Delta y)^2-x^2-y^2}{\Delta x + i\Delta y}$$

$$= \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{2x(\Delta x) + 2y(\Delta y) + (\Delta x)^2 + (\Delta y)^2}{\Delta x + i\Delta y}$$

Suppose $z \neq 0$ and $\Delta x = 0$. Then

$$\frac{d}{dz}|z|^2 = \lim_{\Delta y \to 0} \frac{2y(\Delta y) + (\Delta y)^2}{i\Delta y} = \frac{2y}{i}$$

Suppose $z \neq 0$ and $\Delta y = 0$. Then

$$\frac{d}{dz}|z|^2 = \lim_{\Delta x \to 0} \frac{2x(\Delta x) + (\Delta x)^2}{\Delta x} = 2x$$

Thus $\frac{d}{dz}|z|^2$ does not exist for $z \neq 0$. Suppose $z = 0$. Then

$$\frac{d}{dz}|z|^2 = \lim_{\Delta z \to 0} \frac{|\Delta z|^2}{\Delta z} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{(\Delta x)^2 + (\Delta y)^2}{\Delta x + i\Delta y} = 0$$

Thus $\frac{d}{dz}|z|^2$ exists only at $z = 0$, at which point the derivative is zero.

B.1.5 Derivative with Respect to a Vector

The familiar rules for differentiating with respect to a vector or matrix holds as long as the function does not contain the conjugate of the variable you are differentiating with respect to. Several frequently used derivatives with respect to a vector are presented below. Real-variables versions of these have long been established. Complex versions have often been used, with corresponding evidence in the literature by well-respected authors that these results are well unknown. Systematic development of these results is an easy and major contribution of this thesis which could be done by any senior in engineering.

Let z be a complex vector. Then

$$\frac{d}{dz}f(z) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{d}{dz_1} \\ \vdots \\ \frac{d}{dz_n} \end{pmatrix} f(z) \tag{B.2}$$

For $a \in \mathbf{C}^n$, Let $f(z) = a^T z$. Then

$$\frac{d}{dz}a^T z = \frac{d}{dz}\sum_{i=1}^{n} a_i z_i = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a = \frac{d}{dz}z^T a \tag{B.3}$$

For $a \in \mathbf{C}^n$, Let $f(z) = a^H z$. Then

$$\frac{d}{dz}a^H z = \frac{d}{dz}\sum_{i=1}^{n} a_i^* z_i = a^* \tag{B.4}$$

Note that $\frac{d}{dz}z^H a$ *does not exist.*

For $A \in \mathbf{C}^{n\times n}$ and $f(z) = z^T A z$. Then

$$\frac{d}{dz}z^T A z = \frac{d}{dz}(z_1, \cdots, z_n)\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

where $a_i \in \mathbf{C}^n$. When expanded, this is

$$\frac{d}{dz}\left(\sum_{i=1}^{n} z_i a_i\right) z = \frac{d}{dz}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i a_{ij} z_j$$

To visualize this easier, I will write this out ad nauseam. It is

$$\frac{d}{dz}\begin{pmatrix} a_{11}z_1^2 + a_{12}z_1z_2 + \cdots + a_{1n}z_1z_n \\ +a_{21}z_2z_1 + a_{22}z_2^2 + \cdots + a_{2n}z_2z_n \\ +a_{31}z_3z_1 + a_{32}z_3z_2 + \cdots + a_{3n}z_3z_n \\ +\cdots \\ +a_{n1}z_nz_1 + a_{n2}z_nz_2 + \cdots + a_{nn}z_n^2 \end{pmatrix}$$

where the term in the brackets is a scalar and I have arranged it to give insight how the terms arose. Taking the derivatives, we get the column vector

$$\begin{pmatrix} 2a_{11}z_1 + (a_{12}+a_{21})z_2 + (a_{13}+a_{31})z_3 + \cdots + (a_{1n}+a_{n1})z_n \\ (a_{12}+a_{21})z_1 + 2a_{22}z_2 + (a_{23}+a_{32})z_3 + \cdots + (a_{2n}+a_{n2})z_n \\ (a_{13}+a_{31})z_1 + (a_{23}+a_{32})z_2 + 2a_{33}z_3 + \cdots + (a_{3n}+a_{n3})z_n \\ \vdots \\ (a_{1n}+a_{n1})z_1 + (a_{2n}+a_{n2})z_2 + (a_{3n}+a_{n3})z_n + \cdots + 2a_{nn}z_n \end{pmatrix}$$

Let

$$A = (a^1, a^2, \cdots, a^n) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Then

$$\frac{d}{dz}z^TAz = \begin{pmatrix} (a_1 + (a^1)^T)z \\ (a_2 + (a^2)^T)z \\ \vdots \\ (a_n + (a^n)^T)z \end{pmatrix} = (A + A^T)z \tag{B.5}$$

If $A = A^T$, then $\frac{d}{dz}z^TAz = 2Az$, which is the familiar result often cited in the real variables case. The only requirement here is that A must be square. Note that the matrix in the answer is always symmetric even though A is not required to be symmetric.

Some simplicity is achieved if $A = A^H$ because the (a^i) are merely transposed, and not conjugated. Thus $A = A^H$ implies

$$A + A^T = A + A^* = 2\,\mathrm{Re}(A)$$

Thus, if $A = A^H$, then

$$\frac{d}{dz}z^TAz = 2[\mathrm{Re}(A)]z \tag{B.6}$$

Now consider the case of $\frac{d}{dz}z^HAz$. This is the specific derivative that is most frequently abused in the literature when approaching a linear optimization problem in adaptive beamforming. Expanding, we get $\frac{d}{dz}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i^*a_{ij}z_j$. However, we recall that $\frac{d}{dz}z_i^*a_{ij}z_j$ does not exist when $i \neq j$. Therefore, $\frac{d}{dz}z^HAz$ does not exist when A is any matrix except a diagonal matrix. When A is a diagonal matrix, the derivative exists only when z is the zero vector. When this is true, the derivative is zero. When faced with the desire to try this derivative,

one should instead attempt other techniques like completion of squares or a projection based technique. Note that the method of Lagrange multipliers can be pursued without taking derivatives even though the usual approach in the real variables case almost always uses a derivative in the solution.

A similar form is $f(z) = z^T Ay$. Let

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Then

$$z^T Ay = (z_1, z_2, \cdots, z_n) \begin{pmatrix} a_1 y \\ a_2 y \\ \vdots \\ a_n y \end{pmatrix} = \sum_{i=1}^{n} z_i a_i y$$

Taking the derivatives, we get

$$\frac{d}{dz} \sum_{i=1}^{n} z_i a_i y = \begin{pmatrix} a_1 y \\ a_2 y \\ \vdots \\ a_n y \end{pmatrix} = Ay = \frac{d}{dz} z^T Ay \tag{B.7}$$

Since $z^T Ay$ is a scalar, it equals its transpose. Thus $\frac{d}{dz} y^T A^T z = Ay$. Similarly, since $y^H Az = z^T A^T y^*$, we see that $\frac{d}{dz} y^H Az = A^T y^*$. For the case of $z^H Ay$, we know the derivative $\frac{d}{dz} z^H Ay$ does not exist.

Let

$$\frac{d}{dz^T} f(z) = \left(\frac{d}{dz_1}, \frac{d}{dz_2}, \cdots, \frac{d}{dz_n} \right) f(z)$$

Then

$$\frac{d}{dz}\left(\frac{d}{dz^T}\right) f(z) = \frac{d^2}{dz dz^T} f(z) = \begin{pmatrix} \frac{d^2}{dz_1^2} & \frac{d^2}{dz_1 dz_2} & \cdots & \frac{d^2}{dz_1 dz_n} \\ \frac{d^2}{dz_2 dz_1} & \frac{d^2}{dz_2^2} & \cdots & \frac{d^2}{dz_2 dz_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2}{dz_n dz_1} & \frac{d^2}{dz_n dz_2} & \cdots & \frac{d^2}{dz_n^2} \end{pmatrix} f(z)$$

which is the Hessian matrix operator on f, which I will call $\mathcal{H}$. Note that $\frac{d^2}{dz^T dz} = \mathcal{H}^T$. If all of the derivatives are continuous, then $\mathcal{H} = \mathcal{H}^T$ since the continuity of derivatives allows the order of the derivatives to be exchanged.

Recall that

$$\frac{d}{dz} z^T A z = \begin{pmatrix} (a_1 + (a^1)^T) z \\ (a_2 + (a^2)^T) z \\ \vdots \\ (a_n + (a^n)^T) z \end{pmatrix} = (A + A^T) z$$

Then

$$\frac{d^2}{dz^T dz} z^T A z = A + A^T$$

Again, $\frac{d^2}{dz^T dz} z^H A z$ does not exist. Similarly, $\frac{d^2}{dz^H dz} z^H A z$ and $\frac{d^2}{dz^H dz} z^T A z$ do not exist. If $A = A^H$, then

$$\frac{d^2}{dz^T dz} z^T A z = 2\,\mathrm{Re}(A) \tag{B.8}$$

B.2 Derivative with Respect to a Matrix

Let Z be a complex matrix with elements $\{Z_{ij}\}$. Let $f(Z)$ be a scalar valued function of Z. The function f may take on complex values. When f is a differentiable complex function, then define

$$\frac{d}{dZ}f(Z) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial}{\partial Z_{11}} & \frac{\partial}{\partial Z_{12}} & \cdots & \frac{\partial}{\partial Z_{1n}} \\ \frac{\partial}{\partial Z_{21}} & \frac{\partial}{\partial Z_{22}} & \cdots & \frac{\partial}{\partial Z_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial Z_{m1}} & \frac{\partial}{\partial Z_{m2}} & \cdots & \frac{\partial}{\partial Z_{mn}} \end{pmatrix} f(Z) \tag{B.9}$$

Note that when this exists, then

$$\frac{d}{dZ}f(Z) = \frac{\partial}{\partial \operatorname{Re}\{Z\}}f(Z) \tag{B.10}$$

When $Z = X + iY$, then

$$\frac{d}{dZ}f(Z) = \frac{d}{dX}f(Z) = \begin{pmatrix} \frac{\partial}{\partial X_{11}} & \frac{\partial}{\partial X_{12}} & \cdots & \frac{\partial}{\partial X_{1n}} \\ \frac{\partial}{\partial X_{21}} & \frac{\partial}{\partial X_{22}} & \cdots & \frac{\partial}{\partial X_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial X_{m1}} & \frac{\partial}{\partial X_{m2}} & \cdots & \frac{\partial}{\partial X_{mn}} \end{pmatrix} f(Z)$$

Suppose $f(Z)$ is itself now a matrix,

$$f(Z) = \begin{pmatrix} f_{11}(Z) & f_{12}(Z) & \cdots & f_{1q}(Z) \\ f_{21}(Z) & f_{22}(Z) & \cdots & f_{2q}(Z) \\ \vdots & \vdots & \ddots & \vdots \\ f_{p1}(Z) & f_{p2}(Z) & \cdots & f_{pq}(Z) \end{pmatrix}$$

Then $\frac{d}{dZ}f(Z)$ is an $(mp \times nq)$ complex matrix where each $\frac{\partial}{\partial Z_{ij}}f(Z)$ is a $(p \times q)$ matrix. The desired form is the "right hand" direct product of $\frac{\partial}{\partial Z}$ and $f(Z)$. The matrix $\frac{d}{dZ}f(Z)$ has the form

$$\frac{d}{dZ}f(Z) = \begin{pmatrix} \frac{\partial f}{\partial Z_{11}} & \cdots & \frac{\partial f}{\partial Z_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial Z_{m1}} & \cdots & \frac{\partial f}{\partial Z_{mn}} \end{pmatrix} \tag{B.11}$$

where

$$\frac{\partial f}{\partial Z_{jk}} = \begin{pmatrix} \frac{\partial f_{11}}{\partial Z_{jk}} & \cdots & \frac{\partial f_{1q}}{\partial Z_{jk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{p1}}{\partial Z_{jk}} & \cdots & \frac{\partial f_{pq}}{\partial Z_{jk}} \end{pmatrix} \tag{B.12}$$

Caution. This is much different than doing a matrix multiplication of the operator matrix by the function matrix. The more correct analogy is the direct product $\left(\frac{\partial}{\partial Z}\right) \otimes (f(Z))$. Even with this interpretation there is danger of ambiguity. For example, let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. $A \otimes B$ has been interpreted differently by various authors. The interpretation that yields the form above is

$$A \otimes_R B = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \tag{B.13}$$

This is *not* the same as

$$A \otimes_L B = \begin{pmatrix} Ab_{11} & Ab_{12} \\ Ab_{21} & Ab_{22} \end{pmatrix} \tag{B.14}$$

which is a "left hand" direct product. Because of the ambiguity, it is good practice to write out a definition for your readers. If you choose to use the left hand direct product for $\left(\frac{\partial}{\partial Z}\right) \otimes_L (f(Z))$, make sure your further derivations are consistent.

A result for real variables that carries over to complex variables is the derivative of a determinant with respect to a matrix element. Let $A \in \mathbf{C}^{n\times n}$ be a complex matrix with elements a_{ij} and *minors* X_{ij} obtained from A by deleting row i and column j from A. Computing $\det(A)$ by cofactor expansion down column j gives us the following.

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(X_{ij}) \tag{B.15}$$

Note that a_{kj} appears only *once* in this expansion, when $i = k$. If we take the derivative of $\det(A)$ with respect to a_{kj}, we get just one term if all the a_{ij} are algebraically independent. Thus $\frac{\partial}{\partial a_{ij}} \det(A) = (-1)^{i+j} \det(X_{ij})$. When $A = A^T$, the elements off the major diagonal are algebraically dependent such that $a_{ij} = a_{ji}$. In this case,

$$\frac{\partial}{\partial a_{ij}} \det(A) = \begin{cases} (-1)^{i+j} \det(X_{ii}), & i = j \\ (-1)^{i+j} 2 \det(X_{ij}), & i \neq j \end{cases} \tag{B.16}$$

For the algebraically independent case, $\frac{\partial}{\partial A}\det(A) =$

$$\begin{pmatrix} (-1)^{1+1}\det(X_{11}) & (-1)^{1+2}\det(X_{12}) & \cdots & (-1)^{1+n}\det(X_{1n}) \\ (-1)^{2+1}\det(X_{21}) & (-1)^{2+2}\det(X_{22}) & \cdots & (-1)^{2+n}\det(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}\det(X_{n1}) & (-1)^{n+2}\det(X_{n2}) & \cdots & (-1)^{n+n}\det(X_{nn}) \end{pmatrix}$$

Thus we observe that

$$\frac{\partial}{\partial A}\det(A) = (\text{adj } A)^T \tag{B.17}$$

When A^{-1} exists, then

$$\frac{\partial}{\partial A}\det A = (A^{-1}\det A)^T = (\det A)(A^{-1})^T = (\det A)A^{-T} \tag{B.18}$$

Caution. Suppose $A = A^H$ is a 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{pmatrix}$. Then

$$\frac{\partial}{\partial a_{12}}\det A = \frac{\partial}{\partial a_{12}}[a_{11}a_{22} - |a_{12}|^2]$$

exists only at $a_{12} = 0$. Suppose $A = A^H$ is a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12}^* & a_{22} & a_{23} \\ a_{13}^* & a_{23}^* & a_{33} \end{pmatrix}$$

The determinant is

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{13}^*a_{23} + a_{12}^*a_{13}a_{23}^*$$

$$- |a_{13}|^2 a_{22} - a_{11}|a_{23}|^2 - |a_{12}|^2 a_{33}$$

From this, we see that for $i \neq j$ that $\frac{\partial}{\partial a_{ij}} \det A$ does not exist because $\frac{\partial}{\partial a_{ij}} a_{ij}^*$ does not exist *anywhere.*

Let us examine this one step further. Consider $T = T^H$ and B, both in $M_2(\mathbf{C})$, and $\frac{\partial}{\partial T} \det(BT)$.

$$BT = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} \\ t_{12}^* & t_{22} \end{pmatrix} = \begin{pmatrix} b_{11}t_{11} + b_{12}t_{12}^* & b_{11}t_{12} + b_{12}t_{22} \\ b_{21}t_{11} + b_{22}t_{12}^* & b_{21}t_{12} + b_{22}t_{22} \end{pmatrix}$$

and

$$\det(BT) = (b_{11}t_{11} + b_{12}t_{12}^*)(b_{21}t_{12} + b_{22}t_{22}) - (b_{21}t_{11} + b_{22}t_{12}^*)(b_{11}t_{12} + b_{12}t_{22})$$

From this we see that

$$\frac{\partial}{\partial t_{12}} \det(BT) = \left[\frac{\partial}{\partial t_{12}} (b_{11}t_{11} + b_{12}t_{12}^*)\right] (b_{21}t_{12} + b_{22}t_{22}) + (b_{11}t_{11} + b_{12}t_{12}^*)\, b_{21}$$
$$- \left[\frac{\partial}{\partial t_{12}} (b_{21}t_{11} + b_{22}t_{12}^*)\right] (b_{11}t_{12} + b_{12}t_{22}) - (b_{21}t_{11} + b_{22}t_{12}^*)\, b_{11}$$

which does not exist because $\frac{\partial}{\partial t_{12}} t_{12}^*$ does not exist, anywhere, even at zero. In fact, for $T = T^H$ and B, both in $M_n(\mathbf{C})$, only $\frac{\partial}{\partial t_{kk}} \det(BT)$ exists. Thus, $\frac{\partial}{\partial T} \det(BT)$ does *not* exist. Similarly, $\frac{\partial}{\partial T} \det(A - iBT)$ does not exist when $T = T^H$. This specific example will become very important in a moment (in both senses of the phrase).

The interested reader is encouraged to read papers by Tracy and Dwyer [267] and Dwyer [72] for the case of real variables.

One of the implications is that when dealing with complex variables, the usual maximization or minimization problems cannot, in general, be solved by

taking derivatives. This includes the usual implementation of the method of Lagrange multipliers for constrained optimization problems. (For a successful application of the method of Lagrange multipliers applied to complex variables for adaptive sonar beamforming, see the paper by Cox [61].)

B.2.1 Linear Transformation of a Vector

Let $z = Tx$ where $z, x \in \mathbf{C}^n$ and $T \in \mathbf{C}^{n \times n}$. Let f be a differentiable complex function. Then

$$\frac{\partial f}{\partial x} = T^T \frac{\partial f}{\partial z}$$

This is a complexification of an unnumbered lemma found in Muirhead (p. 240) [187], which is stated without proof.

Proof. By definition,

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Also,

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial z_j}{\partial x_i} \frac{\partial f}{\partial z_j}$$

Thus

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_n} & \cdots & \frac{\partial z_n}{\partial x_n} \end{pmatrix}}_{P} \begin{pmatrix} \frac{\partial f}{\partial z_1} \\ \vdots \\ \frac{\partial f}{\partial z_n} \end{pmatrix} \underset{=}{\text{def}} P \frac{\partial f}{\partial z}$$

Recall that

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} T_{11} & \cdots & T_{1n} \\ \vdots & \ddots & \vdots \\ T_{n1} & \cdots & T_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Tx = z$$

Then $z_j = \sum_{i=1}^{n} t_{ji} x_i$ and $\frac{\partial z_j}{\partial x_i} = t_{ji}$. From this we see $P = T^T$. Therefore

$$\frac{\partial f}{\partial x} = T^T \frac{\partial f}{\partial z}$$

B.2.2 Derivative of a Matrix with Respect to Itself

Lemma 1 *Let*

$$E_n = \begin{pmatrix} {}_n e_1 \\ \vdots \\ {}_n e_n \end{pmatrix}$$

where ${}_n e_j$ is the elementary vector of size n consisting of all zeros except for a 1 in the j^{th} position. Let $X \in \mathbf{C}^{n \times m}$. Then

$$\frac{dX}{dX} = E_n E_m^T$$

Proof.

$$\frac{dX}{dX} = \begin{pmatrix} \frac{dX}{dX_{11}} & \frac{dX}{dX_{12}} & \cdots & \frac{dX}{dX_{1m}} \\ \frac{dX}{dX_{21}} & \frac{dX}{dX_{22}} & \cdots & \frac{dX}{dX_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dX}{dX_{n1}} & \frac{dX}{dX_{n2}} & \cdots & \frac{dX}{dX_{nm}} \end{pmatrix}$$

Consider

$$\frac{dX}{dX_{ij}} = \begin{pmatrix} \frac{dX_{11}}{dX_{ij}} & \cdots & \frac{dX_{1j}}{dX_{ij}} & \cdots & \frac{dX_{1m}}{dX_{ij}} \\ \vdots & & \vdots & & \vdots \\ \frac{dX_{i1}}{dX_{ij}} & \cdots & \frac{dX_{ij}}{dX_{ij}} & \cdots & \frac{dX_{im}}{dX_{ij}} \\ \vdots & & \vdots & & \vdots \\ \frac{dX_{n1}}{dX_{ij}} & \cdots & \frac{dX_{nj}}{dX_{ij}} & \cdots & \frac{dX_{nm}}{dX_{ij}} \end{pmatrix} = {}_n e_i \, {}_m e_j^T$$

This is an $n \times m$ matrix of all zeros except a 1 in position (i, j). Thus

$$\frac{dX}{dX} = \begin{pmatrix} {}_n e_1 \, {}_m e_1^T & {}_n e_1 \, {}_m e_2^T & \cdots & {}_n e_1 \, {}_m e_m^T \\ {}_n e_2 \, {}_m e_1^T & {}_n e_2 \, {}_m e_2^T & \cdots & {}_n e_2 \, {}_m e_m^T \\ \vdots & \vdots & \ddots & \vdots \\ {}_n e_n \, {}_m e_1^T & {}_n e_n \, {}_m e_2^T & \cdots & {}_n e_n \, {}_m e_m^T \end{pmatrix} = E_n E_m^T$$

This is an $n^2 \times m^2$ sparse matrix with only nm non-zero cells. Each nonzero cell contains a 1. *Caution.* For $n \times n$ square matrix X,

$$X^2 = XX = \left(\left(\sum_{m=1}^{n} X_{im} X_{mj} \right)_{ij} \right)$$

$$\frac{d}{dX_{kl}} \left(\sum_{m=1}^{n} X_{im} X_{mj} \right)_{ij} = \begin{cases} 0, & i \neq k,\ j \neq l,\ i \neq j \\ X_{lj}, & i = k,\ j \neq l,\ i \neq j \\ X_{ik}, & i \neq k,\ j = l,\ i \neq j \\ X_{kk} + X_{ll}, & i = k,\ j = l,\ i \neq j \\ 2X_{kk}, & i = j = k = l \\ 0 & i = j = k,\ k \neq l \\ 0 & i = j = l,\ k \neq l \end{cases}$$

$$\frac{dX^2}{dX} = \begin{bmatrix} \begin{pmatrix} 2X_{11} & 0 \\ 0 & 0 \end{pmatrix} & \cdots \\ \begin{pmatrix} X_{12} & 0 & \cdots \\ X_{11}+X_{22} & X_{12} & \\ X_{32} & \vdots & \\ \vdots & & \\ X_{n2} & & \end{pmatrix} & \cdots \\ \vdots & \ddots \end{bmatrix}_{n^2 \times n^2} \underbrace{\neq \underbrace{2X}_{n \times n} \underbrace{\frac{dX}{dX}}_{n^2 \times n^2}}_{\text{meaningless}} \tag{B.19}$$

Thus, the usual differentiation scheme for polynomials does not apply. I.e., $\frac{dX^2}{dX} \neq 2X\frac{dX}{dX}$.

B.2.3 ~Athans and Schweppe Theorems

Athans and Schweppe [34] published a technical report with many matrix gradients for the case of real variables, complete with proofs. These have been quite useful in this and other works. This paper was brought to my attention by Ferlez [82] in a series of very interesting and helpful discussions. What follows is a complexification of this convenient paper. I have supplied the proofs and occasional related corollaries.

Proposition 5 *Let $Z \in \mathbf{C}^{n \times n}$. Then $\frac{\partial}{\partial Z}\operatorname{tr}(Z) = I_n$. This is a complexification of equation (1) in the appendix of [34].*

Proof. $\mathrm{tr}(Z) = \sum_{i=1}^{n} Z_{ii}$.

$$\frac{\partial}{\partial Z_{jk}} \mathrm{tr}(Z) = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$$

where the $\{Z_{jj}\}$ are all algebraically independent. In particular, $Z_{ii} \neq Z_{jj}^*$ for any $i \neq j$.

Proposition 6 *Let $A \in \mathbf{C}^{n\times m}$ and $Z \in \mathbf{C}^{m\times n}$. Then $\frac{\partial}{\partial Z} \mathrm{tr}(AZ) = A^T$. This is a complexification of equation (2) in the appendix of [34].*

Proof. By lemma 26, $\mathrm{tr}(AZ) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} Z_{ji}$. When $Z_{ij} \neq Z_{kl}^*$ for any $(i \neq k, j \neq l)$, then $\frac{\partial}{\partial Z_{kl}} \mathrm{tr}(AZ) = A_{lk}$. Thus $\frac{\partial}{\partial Z} \mathrm{tr}(AZ) = A^T$.

Proposition 7 *Let $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{n\times m}$. Then $\frac{\partial}{\partial Z} \mathrm{tr}(A^*Z) = A^H$.*

Proof. $\mathrm{tr}(A^*Z) = \sum_{k=1}^{n} \sum_{l=1}^{m} A_{kl}^* Z_{lk}$. $\frac{\partial_{ij}}{\partial Z} \mathrm{tr}(A^*Z) = A_{ji}^*$. Thus $\frac{\partial}{\partial Z} \mathrm{tr}(A^*Z) = A^H$.

Proposition 8 *Let $A^H = A \in \mathbf{C}^{n\times n}$ and $Z \in \mathbf{C}^{n\times n}$. Then $\frac{\partial}{\partial Z} \mathrm{tr}(A^*Z) = A$.*

Proof. By corollary 1, $\frac{\partial}{\partial Z} \mathrm{tr}(A^*Z) = A^H = A$.

Proposition 9 *Let $A \in \mathbf{C}^{n\times n}$ and $Z^H = Z \in \mathbf{C}^{n\times n}$. Then $\frac{\partial}{\partial Z} \mathrm{tr}(AZ)$ does not exist.*

Proof. $\frac{\partial}{\partial Z} \mathrm{tr}(AZ) = \frac{\partial}{\partial Z} \mathrm{tr}(AZ^H)$. $\mathrm{tr}(AZ^H)$ includes a term involving Z_{ij}^*. $\frac{\partial}{\partial Z_{ij}} Z_{ij}^*$ does not exist. Therefore $\frac{\partial}{\partial Z} \mathrm{tr}(AZ^H)$ does not exist. Note that this does not depend on the structure of A.

Proposition 10 *Let* $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{n\times m}$. *Then* $\mathrm{tr}(AZ^T) = A$, *and* $\frac{\partial}{\partial Z}\mathrm{tr}(A^TZ) = A$. *This is a complexification of equation (3) in the appendix of [34].*

Proof. By lemma 28,

$$\mathrm{tr}(AZ^T) = \mathrm{tr}(A^TZ) = \sum_{k=1}^{n}\sum_{l=1}^{m} A_{kl}Z_{kl}$$

Thus $\frac{\partial}{\partial Z_{ij}}\mathrm{tr}(AZ^T) = A_{ij}$ when all the Z_{ij} are algebraically independent. The full matrix is then $\mathrm{tr}(AZ^T) = A$.

Proposition 11 *Let* $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{n\times m}$. *Then* $\frac{\partial}{\partial Z}\mathrm{tr}(A^HZ) = A^*$.

Proof. By lemma 29, $\mathrm{tr}(A^HZ) = \sum_{k=1}^{m}\sum_{l=1}^{n} A^*_{lk}Z_{lk}$. $\frac{\partial}{\partial Z_{ij}}\mathrm{tr}(A^HZ) = A^*_{ij}$. The full matrix is $\frac{\partial}{\partial Z}\mathrm{tr}(A^HZ) = A^*$.

Proposition 12 *Let* $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{n\times m}$. *Then* $\frac{\partial}{\partial Z}\mathrm{tr}(AZ^H)$ *does not exist.*

Proof. $\mathrm{tr}(AZ^H)$ includes a term involving Z^*_{ij}. $\frac{\partial}{\partial Z_{ij}}Z^*_{ij}$ does not exist. Therefore $\frac{\partial}{\partial Z}\mathrm{tr}(AZ^H)$ does not exist.

Proposition 13 *Let* $A \in \mathbf{C}^{m\times p}$, $Z \in \mathbf{C}^{p\times q}$, $B \in \mathbf{C}^{q\times m}$. *Then* $\frac{\partial}{\partial Z}\mathrm{tr}(AZB) = A^TB^T$. *This is a complexification of equation (4) in the appendix of [34].*

Proof. By lemma 31,

$$\mathrm{tr}(AZB) = \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{q} A_{ij}Z_{jk}B_{ki}$$

$$\frac{\partial}{\partial Z_{rs}} \operatorname{tr}(AZB) = \sum_{i=1}^{m} A_{ir} B_{si} = (A_{1r}, \cdots, A_{mr}) \begin{pmatrix} B_{s1} \\ \vdots \\ B_{sm} \end{pmatrix} \stackrel{\text{def}}{=} A_r^T (B^s)^T$$

Then

$$\frac{\partial}{\partial Z} \operatorname{tr}(AZB) = \begin{pmatrix} A_1^T(B^1)^T & A_1^T(B^2)^T & \cdots & A_1^T(B^q)^T \\ A_2^T(B^1)^T & A_2^T(B^2)^T & \cdots & A_2^T(B^q)^T \\ \vdots & \vdots & \ddots & \vdots \\ A_p^T(B^1)^T & A_p^T(B^2)^T & \cdots & A_p^T(B^q)^T \end{pmatrix}$$

$$= \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_p^T \end{pmatrix} \begin{pmatrix} (B^1)^T & (B^2)^T & \cdots & (B^q)^T \end{pmatrix} = A^T B^T$$

where the elements of Z are algebraically independent. In particular, this does not exist if $Z_{ij} = Z_{kl}^*$ for any $(i \neq k, j \neq l)$.

Proposition 14 *Let $A \in \mathbf{C}^{m \times p}$, $Z \in \mathbf{C}^{q \times p}$, $B \in \mathbf{C}^{q \times m}$. Then $\frac{\partial}{\partial Z} \operatorname{tr}(AZ^T B) = BA$. This is a complexification of equation (5) in the appendix of [34].*

Proof. By lemma 32,

$$\operatorname{tr}(AZ^T B) = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{q} A_{ij} Z_{kj} B_{ki}$$

From this, we know

$$\frac{\partial}{\partial Z_{rs}} \operatorname{tr}(AZ^T B) = \sum_{i=1}^{m} A_{is} B_{ri} = A_s^T (B^r)^T = B^r A_s$$

since this is a scalar. A_s is a column vector, and B^r is a row vector. Then

$\frac{\partial}{\partial Z}\operatorname{tr}(AZ^TB) =$

$$\begin{pmatrix} B^1A_1 & B^1A_2 & \cdots & B^1A_p \\ B^2A_1 & B^2A_2 & \cdots & B^2A_p \\ \vdots & \vdots & \ddots & \vdots \\ B^qA_1 & B^qA_2 & \cdots & B^qA_p \end{pmatrix} = \begin{pmatrix} B^1 \\ B^2 \\ \vdots \\ B^q \end{pmatrix} \begin{pmatrix} A_1 & A_2 & \cdots & A_p \end{pmatrix} = BA$$

where all elements of Z are algebraically independent.

Proposition 15 *Let $A \in \mathbf{C}^{m\times p}$, $Z \in \mathbf{C}^{q\times p}$, $B \in \mathbf{C}^{q\times m}$. Then $\frac{\partial}{\partial Z}\operatorname{tr}(AZ^HB)$ does not exist.*

Proposition 16 *Let $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{m\times n}$. Then $\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ) = A$. This is a complexification of equation (6) in the appendix of [34].*

Proof. By lemma 26, $\operatorname{tr}(AZ) = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}Z_{ji}$. Then

$$\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ) = \begin{pmatrix} \frac{\partial}{\partial Z_{11}} & \cdots & \frac{\partial}{\partial Z_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial Z_{1n}} & \cdots & \frac{\partial}{\partial Z_{mn}} \end{pmatrix} \operatorname{tr}(AZ)$$

Looking at a particular element, we observe $\frac{\partial}{\partial Z_{kl}}\operatorname{tr}(AZ) = A_{lk}$. Therefore

$$\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{pmatrix} = A$$

Proposition 17 *Let* $A \in \mathbf{C}^{n\times m}$, $Z \in \mathbf{C}^{m\times n}$. *Then* $\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ^T) = A^T$. *This is a complexification of equation (7) in the appendix of [34].*

Proof. By lemma 27, $\operatorname{tr}(AZ^T) = \sum_{l=1}^{n}\sum_{k=1}^{m} A_{lk}Z_{lk}$. This implies $\frac{\partial}{\partial Z_{ij}}\operatorname{tr}(AZ^T) = A_{ij}$ which means $\frac{\partial}{\partial Z}\operatorname{tr}(AZ^T) = A$. This, in turn, implies

$$\left(\frac{\partial}{\partial Z}\right)^T \operatorname{tr}(AZ^T) = \frac{\partial}{\partial Z^T}\operatorname{tr}(AZ^T) = A^T$$

Proposition 18 *Let* $A, Z \in \mathbf{C}^{n\times m}$. *Then* $\frac{\partial}{\partial Z^H}\operatorname{tr}(AZ^T)$ *and* $\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ^H)$ *do not exist.*

Proof. $\frac{\partial}{\partial Z_{ij}}Z_{ij}^*$ does not exist. Also, $\frac{\partial}{\partial Z_{ij}^*}Z_{ij}$ does not exist.

Proposition 19 *Let* $A \in \mathbf{C}^{p\times m}$, $Z \in \mathbf{C}^{m\times n}$, $B \in \mathbf{C}^{n\times p}$. *Then* $\frac{\partial}{\partial Z^T}\operatorname{tr}(AZB) = BA$. *This is a complexification of equation (8) in the appendix of [34].*

Proof. By proposition 13, $\frac{\partial}{\partial Z}\operatorname{tr}(AZB) = A^TB^T$, which implies

$$\frac{\partial}{\partial Z^T}\operatorname{tr}(AZB) = [\frac{\partial}{\partial Z}\operatorname{tr}(AZB)]^T = (A^TB^T)^T = BA$$

Proposition 20 *Let* $A \in \mathbf{C}^{p\times m}$, $Z \in \mathbf{C}^{q\times m}$, $B \in \mathbf{C}^{q\times p}$. *Then* $\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ^TB) = A^TB^T$. *This is a complexification of equation (9) in the appendix of [34].*

Proof. By proposition 14, $\frac{\partial}{\partial Z}\operatorname{tr}(AZ^TB) = BA$. This implies

$$\frac{\partial}{\partial Z^T}\operatorname{tr}(AZ^TB) = [\frac{\partial}{\partial Z}\operatorname{tr}(AZ^TB)]^T = (BA)^T = A^TB^T$$

Proposition 21 *Let* $Z \in \mathbf{C}^{n\times n}$. *Then*

$$\frac{\partial}{\partial Z}\operatorname{tr}(ZZ) = \frac{\partial}{\partial Z}\operatorname{tr}(Z^2) = 2Z^T$$

and $\frac{\partial}{\partial Z^T}\operatorname{tr}(Z^2) = 2Z$. *This is a complexification of equation (10) in the appendix of [34].*

Proof. By lemma 30, $\operatorname{tr}(Z^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}Z_{ji}$. Therefore

$$\frac{\partial}{\partial Z_{kl}}\operatorname{tr}(Z^2) = Z_{lk} + Z_{lk} = 2Z_{lk}$$

which implies $\frac{\partial}{\partial Z}\operatorname{tr}(Z^2) = 2Z^T$. Similarly, $\frac{\partial}{\partial Z^T}\operatorname{tr}(Z^2) = 2Z$.

B.2.4 Derivative of Determinants

This topic addresses the computations that lead to the discovery that the function referred to as the characteristic function for the complex Wishart distribution was not the straight-forward function I had hoped it to be. The function of interest is $\Phi_{\hat{W}}(T) = [\det(I_p - i\Sigma T)]^{-n}$. I attempted to compute moments of the complex Wishart distribution and did not obtain some forms which I knew to be true via using other methods. Nevertheless, the following forms may be useful in another context. I have not diligently searched the literature for these results. Although I supplied the following results, they are simple enough to have been done by any senior in engineering after exposure to the simple results regarding differentiation of complex variables.

Theorem 12 *Let* $A \in \mathbf{C}^{n\times n}$, $B \in \mathbf{C}^{n\times m}$, *and* $T \in \mathbf{C}^{m\times n}$. *Then*

$$\frac{\partial}{\partial T}\det(A - iBT) = -i[\det(A - iBT)][(A - iBT)^{-1}B]^T$$

where $T_{ij} \neq T_{kl}^*$.

Proof. The $(ik)^{th}$ element of the matrix $(A - iBT)$ is $(A_{ik} - i\sum_{l=1}^{m} B_{il}T_{lk})_{ik}$. The operator $\frac{\partial}{\partial T}$ is a matrix of operators $\frac{\partial}{\partial T_{jk}}$. Note that every element of column k of $(A-iBT)$ contains T_{jk}. We expand the determinant $\det(A-iBT)$ down column k. Let $Q = A - iBT$ and let Q^{pk} be the minor of the element Q_{pk} obtained by removing row p and column k from Q. Then

$$\det(A - iBT) = \sum_{p=1}^{n}(-1)^{p+k}\left(A_{pk} - i\sum_{l=1}^{m} B_{pl}T_{lk}\right)_{pk}\det(Q^{pk})$$

The $(jk)^{th}$ element of $\frac{\partial}{\partial T}\det(A - iBT)$ is

$$\frac{\partial}{\partial T_{jk}}\det(A - iBT) = \sum_{p=1}^{n}(-1)^{p+k}\det(Q^{pk})(-iB_{pj})$$

The full matrix is

$$\frac{\partial}{\partial T}\det(A - iBT)$$

$$= \begin{pmatrix} \sum_{p=1}^{n}(-1)^{p+1}\det(Q^{p1})(-iB_{p1}) & \sum_{p=1}^{n}(-1)^{p+2}\det(Q^{p2})(-iB_{p1}) & \cdots \\ \sum_{p=1}^{n}(-1)^{p+1}\det(Q^{p1})(-iB_{p2}) & \sum_{p=1}^{n}(-1)^{p+2}\det(Q^{p2})(-iB_{p2}) & \cdots \\ \vdots & \vdots & \ddots \\ \sum_{p=1}^{n}(-1)^{p+1}\det(Q^{p1})(-iB_{pm}) & \sum_{p=1}^{n}(-1)^{p+2}\det(Q^{p2})(-iB_{pm}) & \cdots \end{pmatrix}$$

$$\begin{pmatrix} \cdots & \sum_{p=1}^{n}(-1)^{p+n}\det(Q^{pn})(-iB_{p1}) \\ \cdots & \sum_{p=1}^{n}(-1)^{p+n}\det(Q^{pn})(-iB_{p2}) \\ \ddots & \vdots \\ \cdots & \sum_{p=1}^{n}(-1)^{p+n}\det(Q^{pn})(-iB_{pm}) \end{pmatrix}$$

Let $B_p \underset{=}{\text{def}} (B_{p1}, B_{p2}, \cdots, B_{pm})$ be row p of matrix B. Then $\frac{\partial}{\partial T}\det(A - iBT)$

$$= -i\sum_{p=1}^{n} B_p^T \left[(-1)^{p+1}\det(Q^{p1}), (-1)^{p+2}\det(Q^{p2}), \cdots, (-1)^{p+n}\det(Q^{pn})\right]$$

Let

$$R^p \underset{=}{\text{def}} \left[(-1)^{p+1}\det(Q^{p1}), \cdots, (-1)^{p+n}\det(Q^{pn})\right]$$

Then $\frac{\partial}{\partial T}\det(A - iBT)$

$$= -i\left[B_1^T R^1 + B_2^T R^2 + \cdots + B_n^T R^n\right] = -i\sum_{p=1}^{n} B_p^T R^p$$

Then

$$\frac{\partial}{\partial T}\det(A - iBT) = -i(B_1^T, B_2^T, \cdots, B_n^T)\begin{pmatrix} R^1 \\ \vdots \\ R^n \end{pmatrix} = -iB^T\begin{pmatrix} R^1 \\ \vdots \\ R^n \end{pmatrix}$$

$$= -iB^T\begin{pmatrix} (-1)^{1+1}\det Q^{11} & (-1)^{1+2}\det Q^{12} & \cdots & (-1)^{1+n}\det Q^{1n} \\ (-1)^{2+1}\det Q^{21} & (-1)^{2+2}\det Q^{22} & \cdots & (-1)^{2+n}\det Q^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}\det Q^{n1} & (-1)^{n+1}\det Q^{n2} & \cdots & (-1)^{n+n}\det Q^{nn} \end{pmatrix}$$

$$\underset{=}{\text{def}} \; -iB^T R = C$$

We recognize $Q^{-1} = \frac{1}{\det Q} R^T$ when Q^{-1} exists. Thus when $(A - iBT)^{-1}$ exists, then $\frac{\partial}{\partial T} \det(A - iBT)$

$$= -i[\det(A - iBT)]B^T[(A - iBT)^{-1}]^T = -i[\det(A - iBT)][(A - iBT)^{-1}B]^T$$

Note that if $(A - iBT)$ does not have an inverse, then $\frac{\partial}{\partial T} \det(A - iBT) = C$ is still valid and will exist provided that each of the partial derivatives $\frac{\partial}{\partial T_{jk}}$ exists, which they do.□

Note from the discussion on complex derivatives that if $m = n$ and $T = T^T$, then $\frac{\partial}{\partial T} \det(A - iBT)$ does not exist. Regardless, the existence of the derivative does *not* depend on any lack of structure, or the presence of structure, on A or B.

Theorem 13 *Let* $A \in \mathbf{C}^{n \times n}$, $T \in \mathbf{C}^{n \times m}$, $B \in \mathbf{C}^{m \times n}$. *Then*

$$\frac{\partial}{\partial T} \det(A - iTB) = -i[\det(A - iTB)][B(A - iTB)^{-1}]^T$$

where $T_{ij} \neq T^*_{kl}$.

Proof. This is nearly the same as the previous identity. The $(jl)^{th}$ element of $A - iTB = Q$ is given by

$$\left(A_{jl} - i \sum_{p=1}^{m} T_{jp} B_{pl} \right)_{jl}$$

We expand across row j is evaluating $\det(A - iTB)$ to obtain

$$\det(A - iTB) = \sum_{l=1}^{n} \left(A_{jl} - i \sum_{p=1}^{m} T_{jp} B_{pl} \right) [\det(Q^{jl})](-1)^{j+l}$$

The $(jk)^{th}$ element of $\frac{\partial}{\partial T}\det(A-iTB)$ is $\sum_{l=1}^{n}(-i)(-i)^{j+l}B_{kl}\det(Q^{jl})$ where Q^{jl} is the minor of element (jl) of Q obtained by deleting row j and column l. The full evaluation is

$$\frac{\partial}{\partial T}\det(A-iTB) =$$

$$-i\sum_{l=1}^{n}(-1)^{l}\begin{pmatrix} (-1)^{1}[\det(Q^{1l})]B_{1l} & (-1)^{1}[\det(Q^{1l})]B_{2l} & \cdots & (-1)^{1}[\det(Q^{1l})]B_{ml} \\ (-1)^{2}[\det(Q^{2l})]B_{1l} & (-1)^{2}[\det(Q^{2l})]B_{2l} & \cdots & (-1)^{2}[\det(Q^{2l})]B_{ml} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n}[\det(Q^{nl})]B_{1l} & (-1)^{n}[\det(Q^{nl})]B_{2l} & \cdots & (-1)^{n}[\det(Q^{nl})]B_{ml} \end{pmatrix}$$

$$= -i\sum_{l=1}^{n}(-1)^{l}\begin{pmatrix} (-1)^{1}[\det(Q^{1l})][B_{1l},B_{2l},\cdots,B_{ml}] \\ (-1)^{2}[\det(Q^{2l})][B_{1l},B_{2l},\cdots,B_{ml}] \\ \vdots \\ (-1)^{n}[\det(Q^{nl})][B_{1l},B_{2l},\cdots,B_{ml}] \end{pmatrix}$$

Let B_l be column l of matrix B, and

$$Q^{l} = \begin{pmatrix} (-1)^{1+l}\det Q^{1l} \\ (-1)^{2+l}\det Q^{2l} \\ \vdots \\ (-1)^{n+l}\det Q^{nl} \end{pmatrix}$$

Then

$$\frac{\partial}{\partial T}\det(A - iTB) = -i\sum_{l=1}^{n} Q^l B_l^T = -i(Q^1, Q^2, \cdots, Q^n)\begin{pmatrix} B_1^T \\ \vdots \\ Bn^T \end{pmatrix}$$

$$= -i(Q^1, Q^2, \cdots, Q^n)B^T$$

We recognize $\frac{\partial}{\partial T}\det Q =$

$$-i\begin{pmatrix} (-1)^{1+1}\det Q^{11} & (-1)^{1+2}\det Q^{12} & \cdots & (-1)^{1+n}\det Q^{1n} \\ (-1)^{2+1}\det Q^{21} & (-1)^{2+2}\det Q^{22} & \cdots & (-1)^{2+n}\det Q^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}\det Q^{n1} & (-1)^{n+2}\det Q^{n2} & \cdots & (-1)^{n+n}\det Q^{nn} \end{pmatrix} B^T$$

$$= -i[\det(Q)]Q^{-T}B^T = -i[\det(Q)](BQ^{-1})^T$$

Therefore

$$\frac{\partial}{\partial T}\det(A - iTB) = -i[\det(A - iTB)][B(A - iTB)^{-1}]^T$$

As with the previous example, $\frac{\partial}{\partial T}\det(Q)$ can exist even when $\det Q = 0$. In that case, it is evaluated as above, before using the adjoint form for an inverse to simplify the notation.

Theorem 14 *Let* $A \in \mathbf{C}^{n\times n}$, $B \in \mathbf{C}^{n\times m}$, $T \in \mathbf{C}^{m\times n}$. *Then*

$$\frac{\partial}{\partial T}\det(A - i(TB)^T) = -i[\det(A - i(TB)^T)][A - i(TB)^T]^{-1}B^T$$

where $T_{ij} \neq T_{kl}^*$.

Proof. Element (jl) of matrix (TB) is $\sum_{p=1}^{m} T_{jp}B_{pl}$. The $(jl)^{th}$ element of (TB) is in the $(lj)^{th}$ element of $Q = A - i(TB)^T$. The $(lj)^{th}$ element of Q is given by

$$\left(A_{lj} - i\sum_{p=1}^{m} T_{jp}B_{pl}\right)_{lj}$$

Expand down column j to obtain

$$\det(A - i(TB)^T) = \sum_{l=1}^{n}(A_{lj} - i\sum_{p=1}^{m} T_{jp}B_{pl})[\det(Q^{lj})](-1)^{l+j}$$

The $(jk)^{th}$ element of $\frac{\partial}{\partial T}\det(A - i(TB)^T)$ is

$$\frac{\partial}{\partial T_{jk}}\sum_{l=1}^{n}\left(A_{lj} - i\sum_{p=1}^{m} T_{jp}B_{pl}\right)\left[\det Q^{lj}\right](-1)^{l+j} = (-i)\sum_{l=1}^{n}(-1)^{l+j}B_{kl}\det(Q^{lj})$$

where Q^{lj} is the $(lj)^{th}$ minor of Q obtained by deleting row l and column j from Q. Then $\frac{\partial}{\partial T}\det(A - i(TB)^T)$

$$= -i\sum_{l=1}^{n}(-1)^{l}\begin{pmatrix} (-1)^1[\det Q^{l1}]B_{1l} & (-1)^1[\det Q^{l1}]B_{2l} & \cdots & (-1)^1[\det Q^{l1}]B_{ml} \\ (-1)^2[\det Q^{l2}]B_{1l} & (-1)^2[\det Q^{l2}]B_{2l} & \cdots & (-1)^2[\det Q^{l2}]B_{ml} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^n[\det Q^{ln}]B_{1l} & (-1)^n[\det Q^{ln}]B_{2l} & \cdots & (-1)^n[\det Q^{ln}]B_{ml} \end{pmatrix}$$

$$= -i\sum_{l=1}^{n}(-1)^{l}\begin{pmatrix} (-1)^1\det Q^{l1} \\ (-1)^2\det Q^{l2} \\ \vdots \\ (-1)^n\det Q^{ln} \end{pmatrix}(B_{1l}, B_{2l}, \cdots, B_{ml})$$

Let

$$B_l = \begin{pmatrix} B_{1l} \\ \vdots \\ B_{ml} \end{pmatrix}$$

and let

$$Q^l = \left[(-1)^{1+l} \det Q^{l1}, (-1)^{2+l} \det Q^{l2}, \cdots, (-1)^{n+l} \det Q^{ln}\right]$$

Then

$$\frac{\partial}{\partial T} \det(A - i(TB)^T) = -i \sum_{l=1}^{n} (Q^l)^T B_l^T = -i[(Q^1)^T, (Q^2)^T, \cdots, (Q^n)^T] \begin{pmatrix} B_1^T \\ B_2^T \\ \vdots \\ B_n^T \end{pmatrix}$$

$$= -i \begin{pmatrix} (-1)^{1+1} \det Q^{11} & (-1)^{1+2} \det Q^{21} & \cdots & (-1)^{1+n} \det Q^{n1} \\ (-1)^{2+1} \det Q^{12} & (-1)^{2+2} \det Q^{22} & \cdots & (-1)^{2+n} \det Q^{n2} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1} \det Q^{1n} & (-1)^{n+2} \det Q^{2n} & \cdots & (-1)^{n+n} \det Q^{nn} \end{pmatrix} B^T$$

$$= -i[\det Q] Q^{-1} B^T$$

by Cramer's Rule, which implies

$$\frac{\partial}{\partial T} \det[A - i(TB)^T] = -i[\det(A - i(TB)^T)][A - i(TB)^T]^{-1} B^T$$

Theorem 15 *Let* $A \in \mathbf{C}^{n \times n}$, $B \in \mathbf{C}^{n \times m}$, $T \in \mathbf{C}^{m \times n}$, $c \in \mathbf{C}$. *Then*

$$\frac{\partial}{\partial T} \det(A - cBT) = -c[\det(A - cBT)][(A - cBT)^{-1} B]^T$$

where $T_{ij} \neq T^*_{kl}$.

Proof. Element (ik) of matrix $Z = A - cBT$ is

$$\left(A_{ik} - c\sum_{l=1}^{m} B_{il}T_{lk}\right)_{ik} = Z_{ik}$$

The operator $\frac{\partial}{\partial T}$ is a matrix of operators $\frac{\partial}{\partial T_{ik}}$. Note that every element of column k of Z contains T_{ik}. Expand $\det(Z)$ down column k. Let Z^{pk} be the minor of Z associated with the element Z_{pk}, where Z^{pk} is obtained from Z by removing row p and column k from Z. Then

$$\det Z = \sum_{p=1}^{n}(-1)^{p+k} Z_{pk} \det Z^{pk}$$

Element (jk) of $\frac{\partial}{\partial T} \det Z$ is

$$\frac{\partial}{\partial T_{jk}} \det Z = \sum_{p=1}^{n}(-1)^{p+k}[\det Z^{pk}](-cB_{pj})$$

The full matrix is $\frac{\partial}{\partial T} \det(A - cBT) =$

$$\begin{pmatrix} \sum_{p=1}^{n}(-1)^{p+1}[\det Z^{p1}](-cB_{p1}) & \cdots & \sum_{p=1}^{n}(-1)^{p+n}[\det Z^{pn}](-cB_{p1}) \\ \sum_{p=1}^{n}(-1)^{p+1}[\det Z^{p1}](-cB_{p2}) & \cdots & \sum_{p=1}^{n}(-1)^{p+n}[\det Z^{pn}](-cB_{p2}) \\ \vdots & & \vdots \\ \sum_{p=1}^{n}(-1)^{p+1}[\det Z^{p1}](-cB_{pm}) & \cdots & \sum_{p=1}^{n}(-1)^{p+n}[\det Z^{pn}](-cB_{pm}) \end{pmatrix}$$

Let row p of matrix B be $B_p = (B_{p1}, B_{p2}, \cdots, B_{pm})$. Then $\frac{\partial}{\partial T} \det(A - cBT) =$

$$-c\sum_{p=1}^{n} B_p^T \left[(-1)^{p+1} \det(Z^{p1}), (-1)^{p+2} \det(Z^{p2}), \cdots, (-1)^{p+n} \det(Z^{pn})\right]$$

$$= -c \sum_{p=1}^{n} B_p^T \det Z^p$$

where

$$\det Z^p \underset{\equiv}{\text{def}} \left[(-1)^{p+1} \det(Z^{p1}), (-1)^{p+2} \det(Z^{p2}), \cdots, (-1)^{p+n} \det(Z^{pn})\right]$$

Therefore $\frac{\partial}{\partial T} \det(A - cBT) =$

$$-c[B_1^T \det(Z^1) + B_2^T \det(Z^2) + \cdots + B_n^T \det(Z^n)]$$

$$-c(B_1^T, B_2^T, \cdots, B_n^T) \begin{pmatrix} \det Z^1 \\ \vdots \\ \det Z^n \end{pmatrix} = -cB^T \begin{pmatrix} \det Z^1 \\ \vdots \\ \det Z^n \end{pmatrix}$$

$$= -cB^T \begin{pmatrix} (-1)^{1+1} \det Z^{11} & (-1)^{1+2} \det Z^{12} & \cdots & (-1)^{1+n} \det Z^{1n} \\ (-1)^{2+1} \det Z^{21} & (-1)^{2+2} \det Z^{22} & \cdots & (-1)^{2+n} \det Z^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1} \det Z^{n1} & (-1)^{n+2} \det Z^{n2} & \cdots & (-1)^{n+n} \det Z^{nn} \end{pmatrix} = X$$

$$= -c[\det(A - cBT)]B^T[(A - cBT)^{-1}]^T$$

$$= -c[\det(A - cBT)][(A - cBT)^{-1}B]^T$$

since

$$Z^{-1} = \frac{1}{\det Z} \begin{pmatrix} (-1)^{1+1} \det Z^{11} & \cdots & (-1)^{1+n} \det Z^{1n} \\ \vdots & \ddots & \vdots \\ (-1)^{n+1} \det Z^{n1} & \cdots & (-1)^{n+n} \det Z^{nn} \end{pmatrix} = \frac{\text{adj } Z}{\det Z}$$

If Z^{-1} does not exist, $X = \frac{\partial}{\partial T} \det(A - cBT)$ is still valid and will exist provided that each $\frac{\partial}{\partial T_{jk}}$ exists.

Theorem 16 *Let* $A \in \mathbf{C}^{n\times n}$, *and* $B, T \in \mathbf{C}^{n\times m}$, $c \in \mathbf{C}$. *Then*

$$\frac{\partial}{\partial T}\det(A + cBT^T) = c[\det(A + cBT^T)](A + cBT^T)^{-1}B$$

where $T_{ij} \neq T_{kl}^*$.

Proof. Let $Z = A + cBT^T$. The $(pj)^{th}$ element of Z is given by

$$Z_{pj} = \left[A_{pj} + c\sum_{l=1}^{m} B_{pl}T_{jl}\right]_{pj}$$

Every element of column j of Z contains T_{jk}. Expand $\det Z$ down column j.

We see that

$$\det Z = \sum_{p=1}^{n}(-1)^{j+p}\left[A_{pj} + c\sum_{l=1}^{m} B_{pl}T_{jl}\right]\det Z^{pj}$$

where Z^{pj} is the minor of Z associated with Z_{pj}. Then

$$\frac{\partial}{\partial T_{jk}}\det Z = \sum_{p=1}^{n}(-1)^{j+p}(cB_{pk})\det Z^{pj}$$

$\frac{\partial}{\partial T}\det(Z) =$

$$\begin{pmatrix} \sum_{p=1}^{n}(-1)^{1+p}(cB_{p1})\det Z^{p1} & \sum_{p=1}^{n}(-1)^{1+p}(cB_{p2})\det Z^{p1} & \cdots & \sum_{p=1}^{n}(-1)^{1+p}(cB_{pm})\det Z^{p1} \\ \sum_{p=1}^{n}(-1)^{2+p}(cB_{p1})\det Z^{p2} & \sum_{p=1}^{n}(-1)^{2+p}(cB_{p2})\det Z^{p2} & \cdots & \sum_{p=1}^{n}(-1)^{2+p}(cB_{pm})\det Z^{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{p=1}^{n}(-1)^{n+p}(cB_{p1})\det Z^{pn} & \sum_{p=1}^{n}(-1)^{n+p}(cB_{p2})\det Z^{pn} & \cdots & \sum_{p=1}^{n}(-1)^{n+p}(cB_{pm})\det Z^{pn} \end{pmatrix}$$

$$= c\sum_{p=1}^{n} \begin{pmatrix} (-1)^{1+p} \det Z^{p1} \\ \vdots \\ (-1)^{n+p} \det Z^{pn} \end{pmatrix} (B_{p1}, B_{p2}, \cdots, B_{pm})$$

Let $B_p = (B_{p1}, B_{p2}, \cdots, B_{pm})$ and

$$\det Z^p = \begin{pmatrix} (-1)^{1+p} \det Z^{p1} \\ \vdots \\ (-1)^{n+p} \det Z^{pn} \end{pmatrix}$$

Then

$$\frac{\partial}{\partial T} \det(Z) = c\sum_{p=1}^{n} [\det Z^p] B_p = c[\det Z^1, \det Z^2, \cdots, \det Z^n] \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix}$$

$$= c \begin{pmatrix} (-1)^{1+1} \det Z^{11} & \cdots & (-1)^{1+n} \det Z^{n1} \\ \vdots & \ddots & \vdots \\ (-1)^{n+1} \det Z^{1n} & \cdots & (-1)^{n+n} \det Z^{nn} \end{pmatrix} B = cRB$$

$$= c[\operatorname{adj} Z]B = c[\det Z]Z^{-1}B$$

when Z^{-1} exists. Thus

$$\frac{\partial}{\partial T} \det(A + cBT^T) = c[\det(A + cBT^T)](A + cBT^T)^{-1}B$$

when $(A + cBT^T)^{-1}$ exists. Even when $(A + cBT^T)^{-1}$ does not exist,

$$\frac{\partial}{\partial T} \det(A + cBT^T) = cRB$$

is valid when each $\frac{\partial}{\partial T_{ij}}$ exists.

B.3 Differential Operator $D(Z) = \left(\frac{\partial}{\partial X} + i\frac{\partial}{\partial Y}\right)$

Work in this section was done by me. I have not searched the literature for similar results. They are simple enough to have been done by any senior in engineering after exposure to the simple results of differentiation of a complex variable.

Definition 4 *Define the differential operator $D(z)$ which operates on complex-valued functions by*

$$D(z)f(z) = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) f(x+iy)$$

where $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$. This form becomes useful when using characteristic functions to evaluate expected moments of a distribution. We want to learn some basic properties of $D(z)$.

First, look at the relationship between $D(z)f$ and $\frac{d}{dz}f$ for complex-valued function f. Suppose the derivative $\frac{df}{dz}$ exists. Then by the Cauchy-Riemann equations,

$$\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) f(z)\Bigg|_{z_0=0} = \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x}\right) f(z)\Bigg|_{z_0=0} = 0$$

This says that when $\frac{d}{dz}f(z)$ exists then $D(z)f(z) = 0$.

Let us consider a few simple cases.

$$D(z)z = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)(x+iy) = 1 - 1 = 0 \qquad \text{(B.20)}$$

$$D(z)z^* = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)(x - iy) = 1 + 1 = 2 \tag{B.21}$$

$$D(z)\,\mathrm{Re}(z) = D(z)\frac{1}{2}(Z + Z^*) = 1 \tag{B.22}$$

$$D(z)\,\mathrm{Im}(z) = D(z)\frac{1}{2i}(z - z^*) = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\frac{1}{2i}(2iy) = i \tag{B.23}$$

$$D(z^*)z = \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)(x + iy) = 1 + 1 = 2 \tag{B.24}$$

$$D(z)\,|z| = D(z)(x^2 + y^2)^{1/2} = \frac{1}{2}(x^2 + y^2)^{-1/2}D(z)(x^2 + y^2) \tag{B.25}$$

$$= \frac{1}{2}\frac{1}{|z|}(2x + i2y) = \frac{z}{|z|}$$

Thus $D(z)\,|z|$ produces a unit length vector pointing in the direction of z.

$$D(z^*)\,|z| = \frac{1}{2}\frac{1}{|z|}(2x - i2y) = \frac{z^*}{|z|} \tag{B.26}$$

$$D(z)\,|z|^2 = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)(x^2 + y^2) = 2x + i2y = 2z \tag{B.27}$$

$$D(z^*)\,|z|^2 = 2z^* \tag{B.28}$$

B.3.1 Vector Case

Let z now be a complex vector in $\mathbf{C}^n$. Then define

$$D(z)f(z) \stackrel{\text{def}}{=} \begin{pmatrix} D(z_1) \\ \vdots \\ D(z_n) \end{pmatrix} f(z) \tag{B.29}$$

For the following discussion, let $a \in \mathbf{C}^n$ and $A \in \mathbf{C}^{n\times n}$.

Let $f(z) = a^T z$. Then

$$D(z)a^T z = \begin{pmatrix} D(z_1) \\ \vdots \\ D(z_n) \end{pmatrix} \sum_{i=1}^{n} a_i z_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0 \tag{B.30}$$

the zero vector.

$$D(z)a^H z = 0 \tag{B.31}$$

$$D(z)z^T Az = D(z)\sum_{i=1}^{n}\sum_{j=1}^{n} z_i A_{ij} Z_j$$

$$= D(z)\begin{pmatrix} A_{11}z_1^2 + A_{12}z_1 z_2 + \cdots + A_{1n}z_1 z_n \\ +A_{21}z_2 z_1 + A_{22}z_2^2 + \cdots + A_{2n}z_2 z_n \\ +A_{31}z_3 z_1 + A_{32}z_3 z_2 + \cdots + A_{3n}z_3 z_n \\ +\cdots \\ +A_{n1}z_n z_1 + A_{n2}z_n z_2 + \cdots + A_{nn}z_n^2 \end{pmatrix} = 0 \tag{B.32}$$

$$D(z)z^H Az = D(z)\begin{pmatrix} A_{11}|z_1|^2 + A_{12}z_1^* z_2 + \cdots + A_{1n}z_1^* z_n \\ +A_{21}z_2^* z_1 + A_{22}|z_2|^2 + \cdots + A_{2n}z_2^* z_n \\ +\cdots \\ +A_{n1}z_n^* z_1 + A_{n2}a_n^* z_2 + \cdots + A_{nn}|z_n|^2 \end{pmatrix}$$

$$= \begin{pmatrix} 2A_{11}z_1 + 2A_{12}z_2 + \cdots + 2A_{1n}z_n \\ 2A_{21}z_1 + 2A_{22}z_2 + \cdots + 2A_{2n}z_n \\ \vdots \\ 2A_{n1}z_i + 2A_n z_2 + \cdots + 2A_{nn}z_n \end{pmatrix}$$

$$= 2 \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = 2Az \tag{B.33}$$

Note that this result does *not* depend on the symmetry of A. The structure of A does not influence the form of this result.

$$D(z)z^T Ay = D(z)\sum_{i=1}^{n} z_i A_i y = 0$$

for

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}$$

Note that this exists and is zero.

$$D(z)z^H Ay = D(z)\sum_{i=1}^{n} z_i^* A_i y = \begin{pmatrix} 2A_1 y \\ \vdots \\ 2A_n y \end{pmatrix} = 2Ay \tag{B.34}$$

$$D(z)y^T Az = D(z)\sum_{i=1}^{n} y^T A^i z_i = 0 \tag{B.35}$$

for $A = (A^1, A^2, \cdots, A^n)$. This exists and is zero.

$$D(z)y^T Az^* = D(z)\sum_{i=1}^{n} y^T A^i z_i^* = \begin{pmatrix} 2y^T A^1 \\ \vdots \\ 2y^T A^n \end{pmatrix}$$

$$= 2 \begin{pmatrix} (A^1)^T y \\ \vdots \\ (A^n)^T y \end{pmatrix} = 2A^T y \tag{B.36}$$

B.3.2 Matrix Case

Let z now be an unstructured matrix in $M_n(\mathbf{C})$.

Then

$$D(Z_{ij}) \det Z = D(Z_{ij}) \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) Z_{1\sigma(1)} Z_{2\sigma(2)} \cdots Z_{n\sigma(n)}$$

$$= D(Z_{ij}) \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma) \prod_{k=1}^{n} Z_{k\sigma(k)} = 0 \tag{B.37}$$

S_n is the permutation set on n letters, and $\sigma(k)$ is the k^{th} permutation in S_n.

Suppose $Z = Z^H$ is a 2×2 matrix $\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12}^* & Z_{22} \end{pmatrix}$. Then

$$D(Z_{12}) \det Z = D(Z_{12})(Z_{11} Z_{22} - |Z_{12}|^2) = -2Z_{12}$$

Now consider the 3×3 matrix $Z = Z^H$. Then $D(Z_{12}) \det Z$

$$= D(Z_{12})[Z_{11} Z_{22} Z_{33} + Z_{12} Z_{13}^* Z_{23} + Z_{12}^* Z_{13} Z_{23}^*$$

$$- |Z_{13}|^2 Z_{22} - Z_{11} |Z_{23}|^2 - |Z_{12}|^2 Z_{33}]$$

$$= 2Z_{13} Z_{23}^* - 2Z_{12} Z_{33} = 2(Z_{13} Z_{23}^* - Z_{12} Z_{33}) = -2 \det(\tilde{Z}^{12})$$

where $\tilde{Z}^{12}$ is the minor of $Z_{12}^* = Z_{21}$. Similarly,

$$D(Z_{13}) \det Z = 2Z_{12} Z_{23} - 2Z_{13} Z_{22} = 2 \det(\tilde{Z}^{13})$$

and

$$D(Z_{23}) \det Z = 2Z_{12}^* Z_{13} - 2Z_{11} Z_{23} = -2 \det \bar{Z}^{23}$$

Finally,

$$D(Z_{11}) \det Z = D(Z_{ii}) \det Z = 0$$

Putting this all together, we get

$$D(Z) \det Z = 2 \begin{pmatrix} 0 & -\det \bar{Z}^{12} & \det \bar{Z}^{13} \\ -\det Z^{12} & 0 & -\det \bar{Z}^{23} \\ \det Z^{13} & -\det Z^{23} & 0 \end{pmatrix}$$

$$= 2 \begin{pmatrix} 0 & (-1)^{1+2} \det \bar{Z}^{12} & (-1)^{1+3} \det \bar{Z}^{13} \\ (-1)^{1+2} \det Z^{12} & 0 & (-1)^{2+3} \det \bar{Z}^{23} \\ (-1)^{1+3} \det Z^{13} & (-1)^{2+3} \det Z^{23} & 0 \end{pmatrix}$$

$$= 2 \begin{pmatrix} 0 & (-1)^{1+2} \det Z^{12} & (-1)^{1+3} \det Z^{13} \\ (-1)^{1+2} \det \bar{Z}^{12} & 0 & (-1)^{2+3} \det Z^{23} \\ (-1)^{1+3} \det \bar{Z}^{13} & (-1)^{2+3} \det \bar{Z}^{23} & 0 \end{pmatrix}^T$$

Recall that

$$Z^{-1} = \frac{1}{\det Z} \begin{pmatrix} (-1)^{1+1} \det Z^{11} & (-1)^{1+2} \det Z^{12} & (-1)^{1+3} \det Z^{13} \\ (-1)^{2+1} \det Z^{21} & (-1)^{2+2} \det Z^{22} & (-1)^{2+3} \det Z^{23} \\ (-1)^{3+1} \det Z^{31} & (-1)^{3+2} \det Z^{32} & (-1)^{3+3} \det Z^{33} \end{pmatrix}^T$$

$$D(Z)\det Z =$$

$$2Z^{-1}[\det Z] - 2\begin{pmatrix} (-1)^{1+1}\det Z^{11} & 0 & 0 \\ 0 & (-1)^{2+2}\det Z^{22} & 0 \\ 0 & 0 & (-1)^{3+3}\det Z^{33} \end{pmatrix}$$

Thus

$$= 2\left[Z^{-1}[\det Z] - \mathrm{diag}\left(\det Z^{11}, \det Z^{22}, \det Z^{33}\right)\right]$$

In general, for $Z = Z^H \in M_n(\mathbf{C})$, we have

$$D(Z)\det(Z) = 2[Z^{-1}\det(Z) - \mathrm{diag}(\det Z^{11}, \cdots, \det Z^{nn})] \qquad \text{(B.38)}$$

B.3.3 Differential Functions of Determinants

These results were supplied by me. I have not diligently searched the literature for them. They are simple enough to have been done by any senior in engineering after brief exposure to the principles of differentiation of a complex variable.

Proposition 22 *Let* $A \in \mathbf{C}^{n\times n}$, $B \in \mathbf{C}^{n\times m}$, $T \in \mathbf{C}^{m\times n}$. *Then*

$$D(T)\det(A - iBT) = 0$$

Proof. The $(ik)^{th}$ element of the matrix $(A - iBT)$ is $(A_{ik} - i\sum_{l=1}^{m} B_{il}T_{lk})_{ik}$. Every element of column k of $(A - iBT)$ contains T_{jk}. Expand $\det(A - iBT)$ down column k. Let $Z = A - iBT$ and let Z^{pk} be the minor of element Z_{pk}

obtained by removing row p and column k from Z. Then

$$\det(Z) = \sum_{p=1}^{n}(-1)^{p+k}(A_{pk} - i\sum_{l=1}^{m} B_{pl}T_{lk}) \det Z^{pk}$$

The $(jk)^{th}$ element of $D(T)\det(A - iBT)$ is $D(T_{jk})\det Z = 0$. Thus

$$D(T)\det(A - iBT) = 0$$

Theorem 17 *Let $A, B, T^H = T \in \mathbf{C}^{n\times n}$. Then*

$$D(T)\det(A - iBT) = i[\Delta(Z^{-T}B) - 2Z^{-T}B]\det(Z)$$

where $Z = A - iBT$ and $\Delta(A)$ is a diagonal matrix of the diagonal elements of A. Further,

$$D(T)\det(I - iBT)\Big|_{T=0} = i[\Delta(B) - 2B]$$

Proof. We know from previous examples that

$$\det(Z) = \det(A - iBT) = \sum_{p=1}^{n}(-1)^{p+k}(A_{pk} - i\sum_{l=1}^{m} B_{pl}T_{lk}) \det(Z^{pk})$$

where Z^{pk} is the minor of $Z = A - iBT$ formed by deleting row p and column k. We now have the additional relationship that $T_{lk} = T_{kl}^*$. We know $D(T_{jk})T_{jk} = 0$ and $D(T_{jk})T_{jk}^* = 2$ for $j \neq k$. Also,

$$D(T_{jj})T_{jj} = D(T_{jj})\,\mathrm{Re}(T_{jj}) = 1$$

since $T_{jj} \in \mathbf{R}$. To locate nonzero entries of $D(T)\det(Z)$ we rewrite $\det(Z)$ as

$$\det(Z) = \sum_{p=1}^{n}(-1)^{p+k}(A_{pk} - i\sum_{l=1}^{m} B_{pl}T_{kl}^{*})\det(Z^{pk})$$

Then

$$D(T_{kj})\det(Z) = \sum_{p=1}^{n}(-1)^{p+k}(-i2B_{pj})\det(Z^{pk})$$

for $k \neq j$, and

$$D(T_{kk})\det(Z) = \sum_{p=1}^{n}(-1)^{p+k}(-iB_{pk})\det(Z^{pk})$$

Then alle zusammen, $D(T)\det(Z) =$

$$\begin{bmatrix} -i\sum_{p=1}^{n}(-1)^{p+1}B_{p1}\det(Z^{p1}) & \cdots & -i\sum_{p=1}^{n}(-1)^{p+1}2B_{pn}\det(Z^{p1}) \\ -i\sum_{p=1}^{n}(-1)^{p+2}2B_{p1}\det(Z^{p2}) & \cdots & -i\sum_{p=1}^{n}(-1)^{p+2}2B_{pn}\det(Z^{p2}) \\ \vdots & & \vdots \\ -i\sum_{p=1}^{n}(-1)^{p+n}2B_{p1}\det(Z^{pn}) & \cdots & -i\sum_{p=1}^{n}(-1)^{p+n}B_{pn}\det(Z^{pn}) \end{bmatrix}$$

Let $B_p = (B_{p1}, B_{p2}, \cdots, B_{pm})$ be row p of matrix B. Note that

$$F = \sum_{p=1}^{n}\begin{bmatrix} (-1)^{p+1}B_{p1}\det(Z^{p1}) & \cdots & (-1)^{p+1}B_{pn}\det(Z^{p1}) \\ (-1)^{p+2}B_{p1}\det(Z^{p2}) & \cdots & (-1)^{p+2}B_{pn}\det(Z^{p2}) \\ \vdots & & \vdots \\ (-1)^{p+n}B_{p1}\det(Z^{pn}) & \cdots & (-1)^{p+n}B_{pn}\det(Z^{pn}) \end{bmatrix}$$

$$= \sum_{p=1}^{n}\begin{bmatrix} (-1)^{p+1}\det(Z^{p1}) \\ \vdots \\ (-1)^{p+n}\det(Z^{pn}) \end{bmatrix} B_p$$

$$= \begin{bmatrix} (-1)^{1+1}\det(Z^{11}) & \cdots & (-1)^{n+1}\det(Z^{n1}) \\ \vdots & \ddots & \vdots \\ (-1)^{1+n}\det(Z^{1n}) & \cdots & (-1)^{n+n}\det(Z^{nn}) \end{bmatrix} \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} = [\det(Z)]Z^{-T}B$$

Substituting into our problem, we obtain

$$D(T)[\det(Z)] = -i2Z^{-T}B\det(Z) + i\Delta[Z^{-T}B\det(Z)]$$

where $\Delta(A)$ is the diagonal matrix whose elements are on the main diagonal of matrix A. Scalars commute, so we have

$$D(T)[\det(Z)] = i[\Delta(Z^{-T}B) - 2Z^{-T}B]\det(Z)$$

Expanded, we obtain

$$D(T)\det(A - iBT) = i[\Delta\{(A - iBT)^{-T}B\} - 2(A - iBT)^{-T}B]$$

When this is evaluated at $T = 0$, we obtain

$$D(T)[\det(Z)]|_{T=0} = i[\Delta(A^{-T}B) - 2A^{-T}B]$$

When we further simplify to $A = I$, then

$$D(T)\det(I - iBT)\Big|_{T=0} = i[\Delta(B) - 2B]$$

B.4 Complex Characteristic Functions

The discussion of characteristic functions of complex variables is almost non-existent in the literature. C. R. Rao [218] provided a definition which is the

starting basis for the following study. The remainder of the results in this section were developed with Ferlez [82]. Acknowledgment of his contribution does not constitute his certification that he has reviewed and approved the enclosed material. Although we developed different results, the work contained here is greatly extended beyond its original bounds and more thoroughly thought out because of the wonderful semester of discussions with Ferlez. Thus, even though I am responsible for these results, they would not have been developed without his active insights. If these results stand the test of close examination, then he should receive much of the positive credit.

In the case of real variables, the theory of characteristic functions has been cast as an application of the Fourier transform. Properties of the Fourier transform have been extensively developed and widely used, particularly by scientists and engineers working with time series data. A major attraction of characteristic functions to statistics is that they provide a conceptually simple way to evaluate the expected value of some linear combinations of random variables. In distribution theory, the Fourier transform provides a useful tool for obtaining nice results via its properties. A first exposure to this application in statistics often is in the proof of the Central Limit Theorem, which yields results more general than obtainable when restricting attention to moment generating functions.

Characteristic functions are important to the development of the proper-

ties of the distributions needed as background for the eventual development of the joint density of the sample eigenvalues of the complex Wishart matrix. The motivation for researching various forms of a characteristic function had its genesis in examining the existence of the derivative when applied to the characteristic function for the complex Wishart distribution. In particular, this examination included applying principles used in deriving the Cauchy-Riemann equations. The case will arise when examining moments of the complex Wishart distribution that blind application of the formula referred to as the "characteristic function" does not yield (at first blush) what is desired. There are two reasons for this. One reason is that the usually cited formula is not really the characteristic function of the complex Wishart distribution, but rather the formula is the characteristic function of another complex matrix variable that is algebraically related to the complex Wishart random variable. The second reason is that the derivative with respect to the transform variable matrix of the function does not exist. It was this discovery that lead to understanding the need for developing the theory of the characteristic function of complex matrix variables.

In the case of complex variables, the choices of functions to call a "characteristic function" widens. The unwary engineer (or theoretician) may apply one concept of a characteristic function to a result derived by another worker using a different concept. These different concepts appear similar, but they

have different properties. After worrying for a while over which version was the "correct" version, it became apparent that merely searching for a definition from an authority was not as instructive as attempting to construct families of functions and observing the resulting properties. After all, it is the set of properties of the characteristic function that make characteristic functions important, not the fact that someone dreamed up a formula with a fancy name that also has applications in other disciplines. The material that follows is the record of investigations. By reading it, you should be able to develop an idea of the different kinds of issues that need to be considered.

Some properties I prefer to have in a characteristic function for complex variables follow:

1. When reduced to the real variables case, the complex variable characteristic function should behave as the real variable characteristic function.

2. The complex characteristic function should be useful for computing expected values of moments.

3. The complex characteristic function should be useful in deriving characteristic functions of linear and affine functions of complex random variables.

4. It should be fairly easy to determine the existence of the desired operations, such as a derivative (if used) applied to characteristic functions to

compute moments

The following forms are the starting points for the discussion to follow.

$$\begin{aligned}
\Phi_z(T) &= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z)])\} \\
\Psi_z(T) &= \mathcal{E}\{\exp(i\,\mathrm{tr}(T^T Z))\} \\
\Omega_z(T) &= \mathcal{E}\{\exp(i\,\mathrm{tr}(T^H Z))\} \\
\nu_z(T) &= \mathcal{E}\{\exp(ig(< T, Z >))\}
\end{aligned}$$

B.4.1 Definition of Characteristic Function of a Complex Random Matrix

This study begins with a derivation of the characteristic function of a complex random variable, as presented by C. R. Rao.

Definition 5 *Let z be a complex vector random variable, and let t be any complex vector of the same dimension as z. Then the characteristic function is given by*

$$\Phi_z(t) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^H z)])\}$$

To see this, let $t = t_1 + it_2$ and $z = x + iy$. Then

$$t^H z = (t_1^T - it_2^T)(x + iy) = t_1^T x + t_2^T y + i(t_1^T y - t_2^T x)$$

Thus $\mathrm{Re}[t^H z] = t_1^T x + t_2^T y$. Consider $s^T = (t_1^T, t_2^T)$, and $w^T = (x^T, y^T)$ where $s, w \in \mathbf{R}^2$. Then $s^T w = t_1^T x + t_2^T y$ and

$$\Phi_w(s) = \mathcal{E}\{exp[i(s^T w)]\} = \mathcal{E}\{exp[i(t_1^T x + t_2^T y)]\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^H z)])\} = \Phi_z(t)$$

Note that $\mathrm{Re}(t^H z)$ satisfies the properties of an inner product.

B.4.2 Basic Properties of Characteristic Functions

Some of the work in this section was motivated by Arnold's discussion [31] of moment generating functions for the case of multivariate distributions of real-valued matrices. His presentation is in his equations 17.4 through 17.9. Except as otherwise noted, I have supplied all the work in this section.

Let the characteristic function of a matrix complex random variable be

$$\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z)])\} \tag{B.39}$$

where $T = (t_{ij})$ is a matrix that has the same dimensions as $Z_{m\times n}$. Then

$$\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\sum_{k=1}^{n}\sum_{j=1}^{m} t_{jk}^* z_{jk}])\} \tag{B.40}$$

where

$$\mathrm{tr}\left[\begin{pmatrix} t_{11}^* & t_{21}^* & \cdots & t_{m1}^* \\ t_{12}^* & t_{22}^* & \cdots & t_{m2}^* \\ \vdots & \vdots & \ddots & \vdots \\ t_{1n}^* & t_{2n}^* & \cdots & t_{mn}^* \end{pmatrix}\begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{pmatrix}\right]$$

$$= \sum_{j=1}^{m} t_{j1}^* z_{j1} + \sum_{j=1}^{m} t_{j2}^* z_{j2} + \cdots + \sum_{j=1}^{m} t_{jn}^* z_{jn} = \sum_{k=1}^{n}\sum_{j=1}^{m} t_{jk}^* z_{jk} \tag{B.41}$$

Proposition 23 *Let $a \in \mathbf{C}$, and let $W = aZ$. Then*

$$\Phi_{aZ}(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H aZ)])\} = \Phi_Z(a^* T) \tag{B.42}$$

Theorem 18 *Let* $A \in \mathbf{C}^{m\times n}$, $B \in \mathbf{C}^{p\times r}$, $C \in \mathbf{C}^{m\times r}$, $Y_{m\times r} = AZB + C$, *and* $Z \in \mathbf{C}^{n\times p}$. *Then*

$$\Phi_Y(T) = \Phi_{AZB+C}(T) = \exp(i\,\mathrm{Re}[\mathrm{tr}(T^H C)])\Phi_Z(A^H T B^H)$$

Proof. The proof consists merely of following the definition and applying the algebra.

$$\Phi_Y(T) = \Phi_{AZB+C}(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H\{AZB + C\})])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H AZB) + \mathrm{tr}(T^H C)])\}$$

$$= \exp(i\,\mathrm{Re}[\mathrm{tr}(T^H C)])\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H AZB)])\}$$

$$= \exp(i\,\mathrm{Re}[\mathrm{tr}(T^H C)])\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(BT^H AZ)])\}$$

$$= \exp(i\,\mathrm{Re}[\mathrm{tr}(T^H C)])\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}((A^H T B^H)^H Z)])\}$$

$$\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H C)])\Phi_Z(A^H T B^H) = \Phi_{AZB+C}(T)$$

□

A useful special application of this property is important enough to separately identify.

Corollary 5 *Let* $Z \in \mathbf{C}^{n\times m}$ *have characteristic function* $\Phi_Z(T)$. *Suppose you want the characteristic function of a weighted sum* y *of elements of* Z. *Let* $A^* \in \mathbf{C}^{n\times m}$ *specify the desired weights. Then* $\Phi_y(t) = \Phi_Z(At)$ *where* t *is a scalar.*

Proof. An arbitrary sum of weighted elements of Z is given by $y = \sum_{j=1}^{m}\sum_{k=1}^{n} A_{kj}^{*} Z_{kj}$. By lemma 29, $y = \mathrm{tr}(A^H Z)$. Since y is a scalar, its characteristic function is given by

$$\Phi_y(t) = \mathcal{E}\{\exp(i\,\mathrm{Re}[t^* y])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[t^*\,\mathrm{tr}(A^H Z)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^* A^H Z)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(At)^H Z])\} = \Phi_Z(At)$$

Note that matrix A is the complex conjugate of the desired set of weights.

This was motivated by Goodman's theorem (p.169) [92] for $\Phi_{\mathrm{tr}\,Z}(t)$. To get the characteristic function of Z_{ij}, just set element $A_{ij} = 1$ and all other elements of A to zero. To get the characteristic function of the sum of the k^{th} row of Z, set the k^{th} row of A to $(1, 1, \cdots, 1)$ and all other elements of A to zero.

Proposition 24 *Let $Z = (Z_1, Z_2)$ where Z_1 is $n \times p_1$. Similarly, let $T = (T_1, T_2)$ where T_1 is $n \times p_1$. Then*

$$\Phi_Z(T_1, 0) = \Phi_{Z_1}(T_1)$$

Proof.

$$\Phi_Z(T_1, 0) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T_1, 0)^H (Z_1, Z_2)])\} \tag{B.43}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\begin{pmatrix} T_1^H \\ 0 \end{pmatrix}(Z_1, Z_2)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\begin{pmatrix} T_1^H Z_1 & T_1^H Z_2 \\ 0 & 0 \end{pmatrix}])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_1^H Z_1\right)])\} = \Phi_{Z_1}(T_1) \tag{B.44}$$

□

Proposition 25 *Similarly, let* $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ *and* $S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$ *where* Y_1 *and* S_1 *are* $n_1 \times p$. *Then*

$$\Phi_Y \begin{pmatrix} S_1 \\ 0 \end{pmatrix} = \Phi_{Y_1}(S_1)$$

Proof.

$$\Phi_Y \begin{pmatrix} S_1 \\ 0 \end{pmatrix} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr} \begin{pmatrix} S_1 \\ 0 \end{pmatrix}^H \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}])\} \tag{B.45}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr} \begin{pmatrix} S_1^H & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(S_1^H Y_1)])\} = \Phi_{Y_1}(S_1) \tag{B.46}$$

□

Proposition 26 *Let* $Z = (Z_1, Z_2)$, Z_1 *and* Z_2 *be independent, and* $T = (T_1, T_2)$. *Then*

$$\Phi_Z(T) = \Phi_Z(T_1, 0)\Phi_Z(0, T_2)$$

Proof.

$$\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T_1, T_2)^H (Z_1, Z_2)])\} \tag{B.47}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr} \begin{pmatrix} T_1^H \\ T_2^H \end{pmatrix} (Z_1, Z_2)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\begin{pmatrix} T_1^H Z_1 & T_1^H Z_2 \\ T_2^H Z_1 & T_2^H Z_2 \end{pmatrix}])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_1^H Z_1\right) + \mathrm{tr}\left(T_2^H Z_2\right)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_1^H Z_1\right)])\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_2^H Z_2\right)])\}$$

If Z_1 and Z_2 are independent, then this equals

$$\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_1^H Z_1\right)])\}\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\left(T_2^H Z_2\right)])\}$$

$$= \Phi_{Z_1}(T_1)\Phi_{Z_2}(T_2) = \Phi_Z(T_1, 0)\Phi_Z(0, T_2) \tag{B.48}$$

□

Proposition 27 *Similarly, let* $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, Y_1 *and* Y_2 *are independent, and* $S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$. *Then*

$$\Phi_Y(S) = \Phi_{Y_1}\begin{pmatrix} S_1 \\ 0 \end{pmatrix}\Phi_{Y_2}\begin{pmatrix} 0 \\ S_2 \end{pmatrix}$$

Proof.

$$\Phi_Y(S) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\begin{pmatrix} S_1 \\ S_2 \end{pmatrix}^H \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}])\} \tag{B.49}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}\begin{pmatrix} S_1^H & S_2^H \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}])\}$$

$$= \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}\left(S_1^H Y_1 + S_2^H Y_2\right)])\}$$

$$= \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}\left(S_1^H Y_1\right)]) \exp(i \operatorname{Re}[\operatorname{tr}\left(S_2^H Y_2\right)])\}$$

If Y_1 and Y_2 are independent, then

$$= \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}\left(S_1^H Y_1\right)])\}\mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}\left(S_2^H Y_2\right)])\}$$

$$= \Phi_{Y_1}(S_1)\Phi_{Y_2}(S_2) = \Phi_{Y_1}\begin{pmatrix} S_1 \\ 0 \end{pmatrix}\Phi_{Y_2}\begin{pmatrix} 0 \\ S_2 \end{pmatrix} \tag{B.50}$$

□

Proposition 28 *The characteristic function of the transpose of the random variable is*

$$\Phi_{Z^T}(T) = \Phi_Z(T^T)$$

Proof.

$$\Phi_{Z^T}(T) = \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}(T^H Z^T)])\}$$

$$= \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}(ZT^*)^T])\} = \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}(ZT^*)])\}$$

since $\operatorname{tr} A^T = \operatorname{tr} A$, and this equals

$$\Phi_{Z^T}(T) = \mathcal{E}\{\exp(i \operatorname{Re}[\operatorname{tr}(T^* Z)])\} = \Phi_Z(T^T) \tag{B.51}$$

□

Proposition 29 *Similarly, the characteristic function of the Hermitian transpose of the random variable is*

$$\Phi_{Z^H}(T) = \Phi_{Z^*}(T^T)$$

Proof.

$$\Phi_{Z^H}(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z^H)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(Z^* T^*)^T])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(Z^* T^*)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^* Z^*)])\} = \Phi_{Z^*}(T^T) \qquad \text{(B.52)}$$

□

Discussion. Note that in general there is no simple connection between $\Phi_Z(T)$ and $\Phi_{Z^*}(T)$. Let $Z = A + iB$ in the following discussion.

$$\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}([A+iB]^H[C+iD])])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}([A-iB]^T[C+iD])])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^TC + B^TD - iB^TC + iA^TD)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^TC + B^TD)])\}$$

and

$$\Phi_{Z^*}(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z^*)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}([A+iB]^H[C+iD]^*)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}([A-iB]^T[C-iD])])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^TC - B^TD - iB^TC - iA^TD)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^TC - B^TD)])\}$$

As an additional curiosity,

$$\Phi_Z^*(T) = \left[\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^TC + B^TD)])\}\right]^* = \mathcal{E}\{\exp(-i\,\mathrm{Re}[\mathrm{tr}(A^TC + B^TD)])\}$$

If $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are independent, then

$$\Phi_Z(T) = \Phi_{\mathrm{Re}\,Z}(T)\Phi_{\mathrm{Im}\,Z}(T) \tag{B.53}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(A^T C)])\}\mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(B^T D)])\}$$

$$\Phi_{Z^*}(T) = \Phi_{\mathrm{Re}\,Z}(T)\Phi_{-\mathrm{Im}\,Z}(T) = \Phi_{\mathrm{Re}\,Z}(T)\Phi_{\mathrm{Im}(-Z)}(T)$$

$$= \mathcal{E}\left\{\exp\left(i\,\mathrm{Re}[\mathrm{tr}(A^T C)]\right)\right\}\mathcal{E}\left\{\exp\left(i\,\mathrm{Re}[\mathrm{tr}(-B^T D)]\right)\right\} = \Phi_{\mathrm{Re}\,Z}(T)\Phi^*_{\mathrm{Im}\,Z}(T)$$

The characteristic function of the *trace* of a random square matrix can be obtained from the characteristic function of the random square matrix by judicious choice of the transform variable.

Theorem 19 *Let $t \in \mathbf{C}$ and $Z \in \mathbf{C}^{n\times n}$. Then*

$$\Phi_{\mathrm{tr}\,Z}(t) = \Phi_Z(tI_n)$$

This is stated in Goodman (p. 169) [92] without proof.

Proof.

$$\Phi_{\mathrm{tr}\,Z}(t) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^*(\mathrm{tr}\,Z))])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^* Z)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^* I Z)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}((tI)^H Z)])\} = \Phi_Z(tI_n)$$

□

Theorem 20 *Likewise, let $a \in \mathbf{C}^n$, $t \in \mathbf{C}$, and $Z \in \mathbf{C}^{n\times n}$. Then*

$$\Phi_{a^H Z a}(t) = \Phi_Z(taa^H)$$

Proof.

$$\Phi_{a^H Z a}(t) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(t^* a^H Z a)])\} = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(at^* a^H Z)])\}$$

$$= \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}((ata^H)^H Z)])\} = \Phi_Z(ata^H) = \Phi_Z(taa^H)$$

since t is a scalar.

B.4.3 Just a Moment

Except for the statement of the distribution and characteristic function of the Cauchy distribution, this section was supplied by me. I have not searched the literature diligently for these results.

Here we examine the properties of the characteristic function as a moment generating function. This study is the beginning of a discovery of a few interesting surprises. We begin with a result that is standard when dealing with real variables to see where it will lead us.

Let $Z_{m\times n}$ be a complex matrix random variable with characteristic function $\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\mathrm{tr}(T^H Z)])\}$. We first want to find the $\mathcal{E}\{Z\}$ solution. Recall that

$$\Phi_Z(T) = \mathcal{E}\{\exp(i\,\mathrm{Re}[\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^* Z_{jk}])\} = \mathcal{E}\{\exp(i[\sum_{j=1}^{m}\sum_{k=1}^{n}(T_{Rjk}Z_{Rjk}+T_{Ijk}Z_{Ijk})])\}$$

where $T = T_R + iT_I$ and $Z = Z_R + iZ_I$. Then

$$\left.\frac{\partial}{\partial T_{Rpq}}\Phi_Z(T)\right|_{T=0}$$

$$= \mathcal{E}\{iZ_{Rpq}\exp(i[\sum_{j=1}^{m}\sum_{k=1}^{n}(T_{Rjk}Z_{Rjk} + T_{Ijk}Z_{Ijk})])\}\Big|_{T=0} = i\mathcal{E}\{Z_{Rpq}\} \quad \text{(B.54)}$$

since expectation is a linear operator. We similarly take the partial derivative with respect to T_{Ipq}. Then

$$\mathcal{E}\{Z_{pq}\} = \mathcal{E}\{Z_{Rpq} + iZ_{Ipq}\}$$

$$= \frac{1}{i}\left.\frac{\partial}{\partial T_{Rpq}}\Phi_Z(T)\right|_{T=0} + \left.\frac{\partial}{\partial T_{Ipq}}\Phi_Z(T)\right|_{T=0}$$

$$= (-i)\left[\frac{\partial}{\partial T_{Rpq}} + i\frac{\partial}{\partial T_{Ipq}}\right]\Phi_Z(T)\Big|_{T=0}$$

Note that

$$\frac{\partial}{\partial T_{pq}} = \left(\frac{\partial}{\partial T_{Rpq}} - i\frac{\partial}{\partial T_{Ipq}}\right) f(T_{pq})$$

by the conditions leading to the Cauchy-Riemann equations. Thus the operator we want to use to find $\mathcal{E}\{Z_{pq}\}$ is *not* the complex derivative $\frac{\partial}{\partial T_{pq}}\Phi_Z(T)$.

Let us use the operator we established by definition 4, given here by

$$D(T_{jk}) = \left(\frac{\partial}{\partial T_{Rjk}} + i \frac{\partial}{\partial T_{Ijk}} \right)$$

Then

$$\mathcal{E}\{Z_{jk}\} = (-i) D(T_{jk}) \Phi_Z(T) \Big|_{T=0}$$

$\mathcal{E}\{Z_{jk}\}$ exists when $D(T_{jk})\Phi_Z(T)$ exists. Define

$$D(T) = \begin{pmatrix} D(T_{11}) & D(T_{12}) & \cdots & D(T_{1n}) \\ D(T_{21}) & D(T_{22}) & \cdots & D(T_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ D(T_{m1}) & D(T_{m2}) & \cdots & D(T_{mn}) \end{pmatrix}$$

Then

$$\mathcal{E}\{Z\} = (-i) D(T) \Phi_Z(T) \Big|_{T=0}$$

This is valid for arbitrary m and n. Let $D^T(T)$ be the transpose of the operator matrix $D(T)$, and let $D^H(T)$ be the Hermitian transpose.

What about higher moments? It is instructive to examine this. My intuition failed me. The wrong solution is to work with forms like

$$D^T(T) D(T) \Phi_Z(T) \Big|_{T=0}$$

$$D^2(T)\Phi_Z(T)\Big|_{T=0}$$

$$D^H(T)D(T)\Phi_Z(T)\Big|_{T=0}$$

etc. The right solution it to work with forms like

$$D(T^TT)\Phi_Z(T)\Big|_{T=0}$$

$$D(T^HT)\Phi_Z(T)\Big|_{T=0}$$

etc.

Although $D(T)D(T)\Phi_Z(T)$ is not what we want for computing expected values of moments like $\mathcal{E}\{Z^2\}$, it might be something to use in another context. Do not throw it away; just put it into your tool box. Recall that $\Phi_Z(T)$ is a scalar valued function. $D(T)\Phi_Z(T)$ is a matrix with the dimensions of $T \in \mathbf{C}^{m\times n}$. If we now look at $D(T_{jk})$ of that matrix, we get a matrix of the same size. When we look at $D(T)D(T)\Phi_Z(T)$ we get a matrix that lives in

$\mathbf{C}^{m^2 \times n^2}$. Likewise, $D^T(T)D(T)\Phi_Z(T)$ lives in $\mathbf{C}^{mn \times mn}$. Elements of

$$D^2(T)\Phi_Z(T) = D(T)D(T)\Phi_Z(T)$$

look like the following.

$$\begin{pmatrix} D(T_{11})D(T) & D(T_{12})D(T) & \cdots & D(T_{1n})D(T) \\ D(T_{21})D(T) & D(T_{22})D(T) & \cdots & D(T_{2n})D(T) \\ \vdots & \vdots & \ddots & \vdots \\ D(T_{m1})D(T) & D(T_{m2})D(T) & \cdots & D(T_{mn})D(T) \end{pmatrix} \Phi_Z(T)$$

$$= \begin{pmatrix} D(T_{11})D(T_{11}) & \cdots & D(T_{11})D(T_{1n}) & \cdots & D(T_{1n})D(T_{1n}) \\ D(T_{11})D(T_{21}) & \cdots & D(T_{11})D(T_{2n}) & \cdots & D(T_{1n})D(T_{2n}) \\ \vdots & & \vdots & & \vdots \\ D(T_{11})D(T_{m1}) & \cdots & D(T_{11})D(T_{mn}) & \cdots & D(T_{1n})D(T_{mn}) \\ D(T_{21})D(T_{11}) & \cdots & D(T_{21})D(T_{1n}) & \cdots & D(T_{2n})D(T_{1n}) \\ \vdots & & \vdots & & \vdots \\ D(T_{21})D(T_{m1}) & \cdots & D(T_{21})D(T_{mn}) & \cdots & D(T_{2n})D(T_{mn}) \\ \vdots & & \vdots & & \vdots \\ D(T_{m1})D(T_{m1}) & \cdots & D(T_{m1})D(T_{mn}) & \cdots & D(T_{mn})D(T_{mn}) \end{pmatrix} \Phi_Z(T)$$

When this is evaluated at $T = 0$, we get:

$$-\begin{pmatrix} \mathcal{E}\{Z_{11}^2\} & \cdots & \mathcal{E}\{Z_{11}Z_{1n}\} & \cdots & \mathcal{E}\{Z_{1n}^2\} \\ \mathcal{E}\{Z_{11}Z_{21}\} & \cdots & \mathcal{E}\{Z_{11}Z_{2n}\} & \cdots & \mathcal{E}\{Z_{1n}Z_{2n}\} \\ \vdots & & \vdots & & \vdots \\ \mathcal{E}\{Z_{11}Z_{m1}\} & \cdots & \mathcal{E}\{Z_{11}Z_{mn}\} & \cdots & \mathcal{E}\{Z_{1n}Z_{mn}\} \\ \mathcal{E}\{Z_{21}Z_{11}\} & \cdots & \mathcal{E}\{Z_{21}Z_{1n}\} & \cdots & \mathcal{E}\{Z_{2n}Z_{1n}\} \\ \vdots & & \vdots & & \vdots \\ \mathcal{E}\{Z_{m1}^2\} & \cdots & \mathcal{E}\{Z_{m1}Z_{mn}\} & \cdots & \mathcal{E}\{Z_{mn}^2\} \end{pmatrix}$$

This might be more useful than the result we are seeking.

If you compute

$$D(T)D(T)\Phi_Z(T) - [D(T)\Phi_Z(T)] \otimes [D(T)\Phi_Z(T)]$$

you obtain something that looks like a simple generalization of a covariance matrix. It is

$$-\begin{pmatrix} \mathcal{E}\{Z_{11}^2\} - [\mathcal{E}\{Z_{11}\}]^2 & \cdots & \mathcal{E}\{Z_{1n}^2\} - [\mathcal{E}\{Z_{1n}\}]^2 \\ \mathcal{E}\{Z_{11}Z_{21}\} - \mathcal{E}\{Z_{11}\}\mathcal{E}\{Z_{21}\} & \cdots & \mathcal{E}\{Z_{1n}Z_{2n}\} - \mathcal{E}\{Z_{1n}\}\mathcal{E}\{Z_{2n}\} \\ \vdots & & \vdots \\ \mathcal{E}\{Z_{11}Z_{m1}\} - \mathcal{E}\{Z_{11}\}\mathcal{E}\{Z_{m1}\} & \cdots & \mathcal{E}\{Z_{1n}Z_{mn}\} - \mathcal{E}\{Z_{1n}\}\mathcal{E}\{Z_{mn}\} \\ \mathcal{E}\{Z_{21}Z_{11}\} - \mathcal{E}\{Z_{21}\}\mathcal{E}\{Z_{11}\} & \cdots & \mathcal{E}\{Z_{2n}Z_{1n}\} - \mathcal{E}\{Z_{2n}\}\mathcal{E}\{Z_{1n}\} \\ \vdots & & \vdots \\ \mathcal{E}\{Z_{m1}^2\} - [\mathcal{E}\{Z_{m1}\}]^2 & \cdots & \mathcal{E}\{Z_{mn}^2\} - [\mathcal{E}\{Z_{mn}\}]^2 \end{pmatrix}$$

$$= - \begin{pmatrix} \operatorname{cov}(Z_{11}, Z_{11}) & \cdots & \operatorname{cov}(Z_{1n}, Z_{1n}) \\ \operatorname{cov}(Z_{11}, Z_{21}) & \cdots & \operatorname{cov}(Z_{1n}, Z_{2n}) \\ \vdots & & \vdots \\ \operatorname{cov}(Z_{11}, Z_{m1}) & \cdots & \operatorname{cov}(Z_{1n}, Z_{mn}) \\ \operatorname{cov}(Z_{21}, Z_{11}) & \cdots & \operatorname{cov}(Z_{2n}, Z_{1n}) \\ \vdots & & \vdots \\ \operatorname{cov}(Z_{m1}, Z_{m1}) & \cdots & \operatorname{cov}(Z_{mn}, Z_{mn}) \end{pmatrix}$$

$$= -\operatorname{cov}(Z, Z) = -[\mathcal{E}\{Z \otimes Z\} - \mathcal{E}\{Z\} \otimes \mathcal{E}\{Z\}].$$

Perhaps this is what we *should* want, even though it is not the standard construct we usually seek. The statistician will object to my using $\operatorname{cov}(Z_{11}, Z_{11})$ rather than $\operatorname{var}(Z_{11})$, but I wanted to make the point of the pattern that developed.

What I set out to find was the expected value of functions like $Z^T Z$ or $Z^H Z$. Each element of the resulting matrix $Y = Z^T Z$ is a linear combination of products of elements of Z. This is more complicated than our previous situation. Note that

$$Z^T Z = \begin{pmatrix} Z_{11} & Z_{21} & \cdots & Z_{m1} \\ Z_{12} & Z_{22} & \cdots & Z_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1n} & Z_{2n} & \cdots & Z_{mn} \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{pmatrix}$$

Element (j, k) is given by $\left(\sum_{i=1}^{m} Z_{ij} Z_{ik} \right)_{jk}$. From the work on the first moment

of Z we recall

$$D(T)\Phi_Z(T) = i\mathcal{E}\{Z\exp(i[\sum_{j=1}^{m}\sum_{k=1}^{n}(T_{Rjk}Z_{Rjk} + T_{Ijk}Z_{Ijk})])\}$$

Element (p,q) is given by

$$D(T_{pq})\Phi_Z(T) = i\mathcal{E}\{Z_{pq}\exp(i[\sum_{j=1}^{m}\sum_{k=1}^{n}(T_{Rjk}Z_{Rjk} + T_{Ijk}Z_{Ijk})])\}$$

If we now apply $D(T_{rs})$ to this result, we get

$$D(T_{rs})D(T_{pq})\Phi_Z(T) = i^2\mathcal{E}\{Z_{rs}Z_{pq}\exp(i[\sum_{j=1}^{m}\sum_{k=1}^{n}(T_{Rjk}Z_{Rjk} + T_{Ijk}Z_{Ijk})])\}$$

Evaluating this at $T = 0$ gives us

$$D(T_{rs})D(T_{pq})\Phi_Z(T)\Big|_{T=0} = -\mathcal{E}\{Z_{rs}Z_{pq}\}$$

To obtain the expected value of element (s,q) of $Y = Z^TZ$, I must find

$$\mathcal{E}\{\sum_{i=1}^{m}Z_{is}Z_{iq}\} = \sum_{i=1}^{m}\mathcal{E}\{Z_{is}Z_{iq}\} = -\sum_{i=1}^{m}D(T_{is})D(T_{iq})\Phi_Z(T)\Big|_{T=0}$$

To make the next step easy, define

$$D_*(T_{rs}T_{pq}) \underline{\text{def}} D(T_{rs})D(T_{pq})$$

Note that this definition does not naturally follow from $D(T_{rs})$, but it will turn out to be a useful definition. The close critical reader would observe

$$D(T_{rs}T_{pq}) = D[\text{Re}(T_{rs}T_{pq}) + i\,\text{Im}(T_{rs}T_{pq})]$$

where

$$\mathrm{Re}(T_{rs}T_{pq}) = T_{Rrs}T_{Rpq} - T_{Irs}T_{Ipq}$$

and

$$\mathrm{Im}(T_{rs}T_{pq}) = T_{Irs}T_{Rpq} + T_{Rrs}T_{Ipq}$$

and thus

$$D(T_{rs}T_{pq}) = \left[\frac{\partial}{\partial\,\mathrm{Re}(T_{rs}T_{pq})} + i\frac{\partial}{\partial\,\mathrm{Im}(T_{rs}T_{pq})}\right]$$

This is *not* what I want. Using the proposed definition, then $\mathcal{E}\{Z^TZ\} =$

$$-\begin{pmatrix} \sum_{i=1}^{m} D(T_{i1})D(T_{i1}) & \sum_{i=1}^{m} D(T_{i1})D(T_{i2}) & \cdots & \sum_{i=1}^{m} D(T_{i1})D(T_{in}) \\ \sum_{i=1}^{m} D(T_{i2})D(T_{i1}) & \sum_{i=1}^{m} D(T_{i2})D(T_{i2}) & \cdots & \sum_{i=1}^{m} D(T_{i2})D(T_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} D(T_{in})D(T_{i1}) & \sum_{i=1}^{m} D(T_{in})D(T_{i2}) & \cdots & \sum_{i=1}^{m} D(T_{in})D(T_{in}) \end{pmatrix} \Phi_Z(T)\Big|_{T=0}$$

$$= \begin{pmatrix} D(T_{11}) & D(T_{21}) & \cdots & D(T_{m1}) \\ D(T_{12}) & D(T_{22}) & \cdots & D(T_{m2}) \\ \vdots & \vdots & \ddots & \vdots \\ D(T_{1n}) & D(T_{2n}) & \cdots & D(T_{mn}) \end{pmatrix} \times$$

$$\times \begin{pmatrix} D(T_{11}) & D(T_{12}) & \cdots & D(T_{1n}) \\ D(T_{21}) & D(T_{22}) & \cdots & D(T_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ D(T_{m1}) & D(T_{m2}) & \cdots & D(T_{mn}) \end{pmatrix} \Phi_Z(T)\Big|_{T=0}$$

We have already seen that this is *not*

$$-D^T(T)D(T)\Phi_Z(T)\Big|_{T=0}$$

This is why we need the suggested definition. When used, we get $\mathcal{E}\{Z^TZ\} =$

$$-\begin{pmatrix} \sum_{i=1}^{m} D_*(T_{i1}T_{i1}) & \sum_{i=1}^{m} D_*(T_{i1}T_{i2}) & \cdots & \sum_{i=1}^{m} D_*(T_{i1}T_{in}) \\ \sum_{i=1}^{m} D_*(T_{i2}T_{i1}) & \sum_{i=1}^{m} D_*(T_{i2}T_{i2}) & \cdots & \sum_{i=1}^{m} D_*(T_{i2}T_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} D_*(T_{in}T_{i1}) & \sum_{i=1}^{m} D_*(T_{in}T_{i2}) & \cdots & \sum_{i=1}^{m} D_*(T_{in}T_{in}) \end{pmatrix} \Phi_Z(T)\Big|_{T=0}$$

$$= -D_*(T^TT)\Phi_Z(T)\Big|_{T=0}$$

Extrapolating this concept,

$$\mathcal{E}\{Z^HZ\} = -D_*(T^HT)\Phi_Z(T)\Big|_{T=0}$$

when these exist. When Z is a square matrix, then

$$\mathcal{E}\{Z^2\} = -D_*(T^2)\Phi_Z(T)\Big|_{T=0}$$

Further,

$$\mathcal{E}\{Z^k\} = (-i)^k D_*(T^k)\Phi_Z(T)\Big|_{T=0}$$

Example 1 *Expectation of* $Z^T AZ$

Let $A = A^H$ be a matrix of complex constants, and let Z be a complex random matrix variable with characteristic function $\Phi_Z(T)$. Then $A = B^H B$ and $Z^H AZ = Y^H Y$ where $Y = BZ$.

$$\mathcal{E}\{Z^H AZ\} = \mathcal{E}\{Y^H Y\} = -D(T^H T)\Phi_Y(T)\Big|_{T=0}$$

$$= -D(T^H T)\Phi_{BZ}(T)\Big|_{T=0} \qquad = -D(T^H T)\Phi_Z(B^H T)\Big|_{T=0}$$

Similarly, $ZAZ^H = XX^H$ where $X = ZC$ and $A = CC^H$. Then

$$\mathcal{E}\{ZAZ^H\} = \mathcal{E}\{XX^H\} = -D(TT^H)\Phi_X(T)\Big|_{T=0}$$

$$= -D(TT^H)\Phi_{ZC}(T)\Big|_{T=0} \qquad = -D(TT^H)\Phi_Z(TC^H)\Big|_{T=0}$$

In a similar manner, for $A^T = A = B^T B > 0$ then

$$\mathcal{E}\{Z^T AZ\} = -D(T^T T)\Phi_Z(B^H T)\Big|_{T=0}$$

Notice that it is still the Hermitian transpose of the square root of A that appears as a factor in the transform variable matrix even though we are dealing with the characteristic function of a symmetric matrix variable. As review, this is a result of theorem 18.

Likewise, for $A^T = A = CC^T > 0$ then

$$\mathcal{E}\{ZAZ^T\} = -D(TT^T)\Phi_Z(TC^H)\Big|_{T=0}$$

Note that this general technique is not applicable for computing $\mathcal{E}\{ZAZ\}$ for $A > 0$ with no other restrictions on Z and A. We observe that $A > 0$ implies there is a "square root" decomposition $A = CC$. Then $ZAZ = ZCCZ = XY$ where $X = ZC$ and $Y = CZ$. There is not a simple relation on $\Phi_Z(T)$ that yields $\mathcal{E}\{XY\}$ with this approach.

If $Z = Z^H$ and $A = A^H$ then

$$\mathcal{E}\{ZAZ\} = \mathcal{E}\{ZAZ^H\} = \mathcal{E}\{Z^H AZ\}$$

which were given above.

Example 2 *Cauchy Distribution*

The Cauchy distribution provides a simple example of a distribution that does not have a "well-defined" mean. Here, it is an example that existence of a characteristic function does not imply existence of moments. The probability density function of the Cauchy distribution is given by

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+(x-\theta)^2}\right)dx,\ x \in \mathbf{R}$$

Its characteristic function is

$$\Phi_x(t) = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\pi}\left(\frac{1}{1+(x-\theta)^2}\right)dx,\ t \in \mathbf{R}$$

Perform the change of variables $y = x - \theta$. Then $x = y + \theta$ and $dx = dy$.

$$\Phi_x(t) = \int_{-\infty}^{\infty} e^{it(y+\theta)}\frac{1}{\pi}\left(\frac{1}{1+y^2}\right)dy = \frac{1}{\pi}e^{it\theta}\int_{-\infty}^{\infty} e^{ity}\left(\frac{1}{1+y^2}\right)dy$$

In order to apply Gradshteyn and Ryshik equation (3.354.5) [94], let $z = -y$. Then $y = -z$ and $dy = -dz$. Then

$$\Phi_x(t) = -\frac{1}{\pi}e^{it\theta}\int_{\infty}^{-\infty} e^{-itz}\left(\frac{1}{1+z^2}\right)dz = \frac{1}{\pi}e^{it\theta}\int_{-\infty}^{\infty} e^{-itz}\left(\frac{1}{1+z^2}\right)dz$$

By equation (3.354.5), this is

$$\Phi_x(t) = \frac{1}{\pi}e^{it\theta}\frac{\pi}{1}e^{-|1\cdot t|} = e^{it\theta-|t|}$$

When moments exist, they are found by differentiating the characteristic function and evaluating at $t = 0$.

$$\frac{d}{dt}\Phi_x(t) = \frac{d}{dt}e^{it\theta-|t|} = e^{it\theta-|t|}\frac{d}{dt}(it\theta - |t|)$$

Note that

$$\frac{d}{dt}|t| = \begin{cases} -1, & t < 0 \\ 1 & t > 0 \\ \text{undefined}, & t = 0 \end{cases}$$

Thus

$$\left.\frac{d}{dt}\Phi_x(t)\right|_{T=0}$$

is undefined, and $\mathcal{E}\{x\}$ does not exist.

B.4.4 Uncharacteristic Funcıions for a Moment

All of the results in this section were supplied by me. I have not diligently searched the literature for these results.

Uncharacteristic Function A

Define function $\Psi_Z(T)$ to be a function that maps $(Z,T) \mapsto \mathbf{C}$ where $Z,T \in \mathbf{C}^{m\times n}$ are matrices and all elements of T are algebraically independent. Let Z be a matrix complex random variable with distribution function $dF(Z)$. Finally, let

$$\Psi_Z(T) = \mathcal{E}\{\exp[i\,\mathrm{tr}(T^T Z)]\} = \int_Z \exp[i\,\mathrm{tr}(T^T Z)]dF(Z)$$

When we expand the definition, we get

$$\Psi_Z(T) = \mathcal{E}\{\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}Z_{jk}]\}$$

If we take the derivative we obtain

$$\frac{d}{dT_{pq}}\Psi_Z(T) = \frac{d}{dT_{pq}}\mathcal{E}\{\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}Z_{jk}]\}$$

$$= \mathcal{E}\{\frac{d}{dT_{pq}}\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}Z_{jk}]\}$$

where we assumed it is legal to interchange the derivative and the integral.

Applying the derivative we obtain

$$\frac{d}{dT_{pq}}\Psi_Z(T) = \mathcal{E}\{iZ_{pq}\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}Z_{jk}]\}$$

When we evaluate this expression at $T = 0$, then

$$\left.\frac{d}{dT_{pq}}\Psi_Z(T)\right|_{T=0} = i\mathcal{E}\{Z_{pq}\}$$

or solving for the first moment,

$$\mathcal{E}\{Z_{pq}\} = -i\left.\frac{d}{dT_{pq}}\Psi_Z(T)\right|_{T=0}$$

Extending this to the derivative with respect to a matrix, we obtain

$$\frac{\partial}{\partial T}\Psi_Z(T) = \Big|_{T=0} = i\mathcal{E}\{Z\}$$

or

$$\mathcal{E}\{Z\} = -i \frac{\partial}{\partial T} \Psi_Z(T) = \Big|_{T=0}$$

This is a property we seek in usual work with characteristic functions, in that here we are using a true derivative rather than a special differential operator.

Now, let $Z, T \in M_n(\mathbf{C})$. We need the property that

$$Z^k = Z \cdot Z \cdot \cdots \cdot Z = \prod_{i=1}^{k} Z$$

makes sense. Then we obtain the form similar to the real variables case.

$$\mathcal{E}\{Z^k\} = (-i)^k \left(\frac{d}{dT}\right)^k \Psi_Z(T) \Bigg|_{T=0}$$

where

$$\left(\frac{d}{dT}\right)^k = \underbrace{\left(\frac{d}{dT}\right)\left(\frac{d}{dT}\right)\cdots\left(\frac{d}{dT}\right)}_{k \text{ times}}$$

Note that $\mathrm{tr}(T^T Z)$ is *not* an inner product.

$$\mathrm{tr}(T^T Z) = \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk} Z_{jk} = \left(\sum_{j=1}^{m}\sum_{k=1}^{n} Z_{jk}^* T_{jk}^*\right)^* = [\mathrm{tr}(Z^H T^*)]^*$$

which violates the first property of an inner product which requires

$$< T, Z > = < Z, T >^*$$

Uncharacteristic Function B

Define $\Omega_Z(T) : (Z,T) \in \mathbf{C}^{m\times n} \rightarrow \mathbf{C}$ by $\Omega_Z(T) = \mathcal{E}\{\exp[i\,\mathrm{tr}(T^H Z)]\}$. Then we get

$$\Omega_Z(T) = \mathcal{E}\{\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^* Z_{jk}]\}$$

Recall that $\frac{d}{dT_{jk}} T_{jk}^*$ does not exist anywhere. To obtain our moments we must look at $\frac{d}{dT_{jk}^*}$ and its matrix extension $\left(\frac{\partial}{\partial T^*}\right)$. Then we get

$$\frac{\partial}{\partial T_{pq}^*}\Omega_Z(T) = \mathcal{E}\{iZ_{pq}\exp[i\sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^* Z_{jk}]\}$$

and

$$\left.\frac{\partial}{\partial T_{pq}^*}\Omega_Z(T)\right|_{T=0} = i\mathcal{E}\{Z_{pq}\}$$

For $X \in \mathbf{C}$, we note that $X = 0$ implies $X^* = 0$. We thus get the nearly familiar result

$$\mathcal{E}\{Z^k\} = (-i)^k \left(\frac{\partial}{\partial T^*}\right)^k \Omega_Z(T)\Bigg|_{T=0}$$

$\Omega_Z(T)$ may have some nice properties because $\mathrm{tr}(T^H Z)$ defines an inner product $< T, Z >$. We verify this.

1.

$$< T, Z >= \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^* Z_{jk} = \left(\sum_{j=1}^{m}\sum_{k=1}^{n} Z_{jk}^* T_{jk}\right)^* =< Z, T >^*$$

2.

$$< T, Z + X >= \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^{*}(Z_{jk} + X_{jk}) = \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^{*} Z_{jk} + \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^{*} X_{jk}$$

$$=< T, Z > + < T, X >$$

where also $X \in \mathbf{C}^{m \times n}$.

3.

$$< T, \alpha Z >= \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^{*} \alpha Z_{jk} = \alpha < T, Z >$$

4.

$$< T, T >= \sum_{j=1}^{m}\sum_{k=1}^{n} T_{jk}^{*} T_{jk} = \sum_{j=1}^{m}\sum_{k=1}^{n} |T_{jk}|^{2} \geq 0$$

for all $T \in \mathbf{C}^{m \times n}$.

5.

$$< T, T >= 0$$

if and only if $T = 0$.

Uncharacteristic Function C

Let $T, Y, Z \in \mathbf{C}^{m \times n}$ and $\alpha \in \mathbf{C}$. Let $g(Z)$ be a linear function. Thus

$$g(Y + \alpha A) = g(Y) + \alpha g(Z)$$

Let $< T, Z >$ be an inner product on the set of matrices in $\mathbf{C}^{m \times n}$. Define

$$\nu_{Z}(T) = \mathcal{E}\{\exp[ig(< T, Z >)]\}$$

Let us consider the properties of $\nu_Z(T)$.

$$\nu_{\alpha Z}(T) = \mathcal{E}\{\exp[ig(< T, \alpha Z >)]\} = \mathcal{E}\{\exp[ig(\alpha < T, Z >)]\}$$

$$= \mathcal{E}\{\exp[i\alpha g(< T, Z >)]\}$$

Also

$$\nu_{\alpha Z}(T) = \mathcal{E}\{\exp[ig(< \alpha^* T, Z >)]\} = \nu_Z(\alpha^* T)$$

Let $A \in \mathbf{C}^{m\times n}$, $B \in \mathbf{C}^{p\times r}$, $C \in \mathbf{C}^{m\times r}$, $Y_{m\times r} = AZB + C$, $T \in \mathbf{C}^{m\times r}$, and $Z \in \mathbf{C}^{n\times p}$. Then

$$\nu_Y(T) = \nu_{AZB+C}(T) = \mathcal{E}\{\exp[ig(< T, AZB + C >)]\}$$

$$= \mathcal{E}\{\exp[ig(< T, AZB > + < T, C >)]\}$$

$$= \mathcal{E}\{\exp[ig(< T, AZB >) + g(< T, C >)]\}$$

$$= \mathcal{E}\{\exp[ig(< T, AZB >)] \exp[ig(< T, C >)]\}$$

Since C is a matrix of constants, we can write this as

$$\mathcal{E}\{\exp[ig(< T, AZB >)]\} \exp[ig(< T, C >)]$$

When A^H is the adjoint of A we get

$$\mathcal{E}\{\exp[ig(< A^H T, ZB >)]\} \exp[ig(< T, C >)]$$

$$= \nu_{ZB}(A^H T) \exp[ig(< T, C >)]$$

When

$$g(< X, Y >) = \mathrm{tr}(X^H Y)$$

for conformable X^H and Y then

$$\nu_Y(T) = \mathcal{E}\{\exp[i\,\mathrm{tr}(BT^HAZ)]\}\exp[i\,\mathrm{tr}(T^HC)] = \nu_Z(A^HTB^H)\exp[i\,\mathrm{tr}(T^HC)]$$

which we established earlier in a similar form for $g(< X, Y >) = \mathrm{Re}[\mathrm{tr}(X^HY)]$.

Partition $Z = (Z_1, Z_2)$ where Z_1 is $n \times p_1$. Similarly, let $T = (T_1, T_2)$ where T_1 is $n \times p_1$. Then

$$\nu_Z(T_1, 0) = \mathcal{E}\{\exp[ig(< T, Z >)]\} = \mathcal{E}\{\exp[ig(< (T_1, 0), (Z_1, Z_2) >)]\}$$

Let $<X, Y >= h(X^HY)$. Then

$$\nu_Z(T_1, 0) = \mathcal{E}\{\exp[ig(h\{\begin{pmatrix} T_1^H \\ 0 \end{pmatrix}, (Z_1, Z_2)\})]\}$$

$$= \mathcal{E}\{\exp[ig(h\begin{Bmatrix} T_1^HZ_1 & T_1^HZ_2 \\ 0 & 0 \end{Bmatrix})]\}$$

If $h(X)$ is a function of only square submatrices on the main diagonal of X then we have

$$\nu_Z(T_1, 0) = \mathcal{E}\{\exp[ig(h\begin{Bmatrix} T_1^HZ_1 & 0 \\ 0 & 0 \end{Bmatrix})]\}$$

If $h(X) = \mathrm{tr}(X)$ then $\nu_Z(T_1, 0) =$

$$\mathcal{E}\{\exp[ig(\mathrm{tr}(T_1^HZ_1))]\} = \mathcal{E}\{\exp[ig(< T_1, Z_1 >)]\} = \nu_{Z_1}(T_1)$$

Now partition $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ and $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ where Z_1 and T_1 are $n_1 \times p$,

Then

$$\nu_Z \begin{pmatrix} T_1 \\ 0 \end{pmatrix} = \mathcal{E}\{\exp[ig(< \begin{pmatrix} T_1 \\ 0 \end{pmatrix}, \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} >)]\}$$

Let $< X, Y >= h(X^H Y)$. Then we have

$$\nu_Z \begin{pmatrix} T_1 \\ 0 \end{pmatrix} = \mathcal{E}\{\exp[ig(h[(T_1^H, 0), \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}])]\} = \mathcal{E}\{\exp[ig(h(T_1^H Z_1))]\}$$

$$= \mathcal{E}\{\exp[ig(< T_1, Z_1 >)]\} = \nu_{Z_1}(T_1)$$

Let $Z = (Z_1, Z_2)$ and $T = (T_1, T_2)$. Then

$$\nu_Z(T) = \mathcal{E}\{\exp[ig(< T, Z >)]\} = \mathcal{E}\{\exp[ig(h[(T_1, T_2)^H (Z_1, Z_2)])]\}$$

$$= \mathcal{E}\{\exp[ig(h \begin{pmatrix} T_1^H Z_1 & T_1^H Z_2 \\ T_2^H Z_1 & T_2^H Z_2 \end{pmatrix})]\}$$

where $< T, Z >= h(T^H Z)$. When h is a function of only square submatrices on the main diagonal, this is

$$\nu_Z(T) = \mathcal{E}\{\exp[ig(h \begin{pmatrix} T_1^H Z_1 & 0 \\ 0 & T_2^H Z_2 \end{pmatrix})]\}$$

When $h(X) = \mathrm{tr}(X)$ then

$$\nu_Z(T) = \mathcal{E}\{\exp[ig(\mathrm{tr}(T_1^H Z_1) + \mathrm{tr}(T_2^H Z_2))]\}$$

$$= \mathcal{E}\{\exp[ig(\mathrm{tr}(T_1^H Z_1)) + ig(\mathrm{tr}(T_2^H Z_2))]\}$$

$$= \mathcal{E}\{\exp[ig(\mathrm{tr}(T_1^H Z_1))] \exp[ig(\mathrm{tr}(T_2^H Z_2))]\}$$

When Z_1 and Z_2 are independent then

$$\nu_Z(T) = \mathcal{E}\{\exp[ig(\mathrm{tr}(T_1^H Z_1))]\}\mathcal{E}\{\exp[ig(\mathrm{tr}(T_2^H Z_2))]\}$$

$$= \nu_{Z_1}(T_1)\nu_{Z_2}(T_2) = \nu_Z(T_1, 0)\nu_Z(0, T_2)$$

Similarly, for $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$, and Z_1 independent of Z_2 we obtain

$$\nu_Z(T) = \nu_{Z_1}(T_1)\nu_{Z_2}(T_2) = \nu_Z \begin{pmatrix} T_1 \\ 0 \end{pmatrix} \nu_Z \begin{pmatrix} 0 \\ T_2 \end{pmatrix}$$

Consider Z^T and Z^H, where both are in $\mathbf{C}^{m\times n}$ and hence Z and Z^* are in $\mathbf{C}^{m\times n}$. Then

$$\nu_{Z^T}(T) = \mathcal{E}\{\exp[ig(< T, Z^T >)]\}$$

When $< X, Y >= h(\mathrm{tr}(X^H Y))$ then

$$\nu_{Z^T}(T) = \mathcal{E}\{\exp[ig(h[\mathrm{tr}(T^H Z^T)])]\} = \mathcal{E}\{\exp[ig(h[\mathrm{tr}(ZT^*)^T])]\}$$

$$= \mathcal{E}\{\exp[ig(h[\mathrm{tr}(ZT^*)])]\}$$

since $\mathrm{tr}\, X^T = \mathrm{tr}\, X$.

$$\nu_{Z^T}(T) = \mathcal{E}\{\exp[ig(h[\mathrm{tr}(T^* Z)])]\} = \nu_Z(T^T)$$

Similarly,

$$\nu_{Z^H}(T) = \mathcal{E}\{\exp[ig(h[\mathrm{tr}(T^H Z^H)])]\} = \mathcal{E}\{\exp[ig(h[\mathrm{tr}(Z^* T^*)^T])]\}$$

$$= \mathcal{E}\{\exp[ig(h[\mathrm{tr}(T^* Z^*)])]\} = \nu_{Z^*}(T^T)$$

Let $Z = X + iY$ and $T = R + iS$. Then

$$\nu_Z(T) = \mathcal{E}\{\exp[ig(< R + iS, X + iY >)]\}$$

$$= \mathcal{E}\{\exp[ig(< R + iS, X > + i < R + iS, Y >)]\}$$

$$= \mathcal{E}\{\exp[ig(< R, X > - i < S, X > + i < R, Y > + < S, Y >)]\}$$

$$= \mathcal{E}\{\exp[i\{g(< R, X >) + g(< S, Y >)\} + \{g(< S, X >) - g(< R, Y >)\}]\}$$

Similarly,

$$\nu_{Z^*}(T) = \mathcal{E}\{\exp[ig(< R + iS, X - iY >)]\}$$

$$= \mathcal{E}\{\exp[i\{g(< R, X >) - g(< S, Y >)\}$$

$$+\{g(< R, Y >) + g(< S, X >)\}]\}$$

and

$$\nu_Z^*(T) = \mathcal{E}\{\exp[-i\{g(< R, X >) + g(< S, Y >)\}$$

$$+\{g(< S, X >) - g(< R, Y >)\}]\}$$

If $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are independent, then

$$\nu_Z(T) = \nu_X(T)\nu_{iY}(T)$$

and

$$\nu_{Z^*}(T) = \nu_X(T)\nu_{(-i)Y}(T) = \nu_X(T)\nu_Y(iT)$$

Appendix C

COMPLEX CHANGE OF VARIABLES

C.1 Introduction to Changing Variables for the Complex Case

The reason for the existence of this chapter is to develop those Jacobians needed for changes of complex variables required for distributional results of this thesis. The theory for change of variables has long been worked out, but the specific forms required for application for multivariate statistics have not been systematically worked out for the complex variables case. Only isolated results appear in the literature, and I have not found some results needed for this thesis.

There are several issues that have arisen in this thesis related to this topic. The first is a need to recognize the difference between a mapping and a change of variables. We have unfortunately created confusion by the ambiguity of the American language by referring to both situations with the same terminology, whether we speak of transformations, mappings, or changing variables. At the abstract level, the basic difference is whether or not you are changing the measure involved. These must, in turn, be distinguished from mere renaming of variables, which is trivial and not discussed further. The picky reader can

consider renaming to be a trivial change of variables. The second issue is that changing variables in the case of the complex space $\mathbf{C}^n$ is like changing variables in the case of the real space $\mathbf{R}^{2n}$ with a need to pay special attention to the imposed algebraic structure. The third issue is the lack of results in the literature that apply to this thesis, or the statistics of complex variables in general.

There is a fourth issue which will not be dwelled on in this thesis, but it is important when reading the literature. When comparing results *in the literature, very close attention must be paid to what assumptions are being made.* It is not uncommon for $\mathbf{C}^n$ to be treated like $\mathbf{R}^{2n}$ with the results expressed in terms of $\mathbf{R}^{2n}$, but discussed as if they are expressed in terms of $\mathbf{C}^n$. This point is also made in the work on complex differentiation. It is important enough to be said twice.

The first issue is really not a trivial issue. If you apply a transformation to a variable, are you engaged in changing variables or are you merely picking a new point in space at which to evaluate your function? If you are interested in the invariance properties of a particular function or measure, you want to examine or describe the effect of that measure as you move about among your various measurable sets. You could properly want to know the relationship between different points or subsets in a space for which the measure is invariant. A rule used to choose one point in a space when you are given another point in

a space is a transformation or function that does a mapping, but it is not a change of variables if your object is to retain the same measure and observe its performance as the data changes.

A transformation is a change of variables situation when the intent is to change the measure being used. The usual situation is that you want to rescale your data to make it conceptually or mathematically easier to work with or explain, but you want resulting integrations done with the new scaling to provide the same answer as the integrations over the same set made with the previous scaling. A more generalized concept is to allow the outcome of the integration to change by a known function, not necessarily constant, but I have not seen that discussed in any texts.

For the second point, it is not a surprise, but it is important, that you must be diligent to consider both the imaginary and real parts of a complex variable. Where differences show up in applications is that of structure in multivariate data. You already know to take into account the effects of repeated block structure, bandedness, etc. There is one more component to the structure of the variable to consider, the effects of conjugation, summarized in table C.1. Structure is important. The theory for zonal polynomials and group representation theory as applied to complex variables has been done by others for the structure of the complex symmetric case but not the Hermitian case. Only Gross and Richards [96] have addressed the complex Hermitian case.

Table C.1. Structure in Complex Variables

Matrix Type	**Meaning**	**Diagonal Elements**	**Restrictions**
symmetric	$Z = Z^T$	$x + iy = x + iy$	No restrictions
Hermitian	$Z = Z^H$	$x + iy = x - iy$	$Z_{ii} \in \mathbf{R}$
skew-symmetric	$Z = -Z^T$	$x + iy = -x - iy$	$Z_{ii} = 0$
skew-Hermitian	$Z = -Z^H$	$x - iy = -x + iy$	$Z_{ii} \in \mathbf{C} \backslash \mathbf{R}$ (imaginary)

The third point is that the usual discussions about changes of variables do not address the implementation of the principles to complex variables. The mathematician would say that it is not necessary because the theory has been worked out for more general spaces. The engineer rarely studies those more general spaces, and he has a daily need to work in the complex field. This chapter archives a systematic development which includes results specific to this thesis and is useful for other future complex variable work. Perhaps this could be called the engineering of mathematics, if engineering is the development of methods and tools for translating theory into application.

A nearly novel feature of this chapter to engineers is the application of *exterior* products. Exterior products go by several names, depending on the discipline of the individual doing the discussion. Another popular name for them is *wedge* products. Wedge products greatly simplify the computation of

Jacobians for nonlinear changes of variables. The root of using wedge products for change of variables problems is found in differential geometry. Rudin (pp. 253-266) [229] provides a nice introduction. Muirhead [187] uses exterior products in developing results for the real variables case. I do not know of any reference that illustrates the procedure for application of exterior products to the complex variable case. However, physicists and nuclear engineers are likely to know of such a reference because they must deal with changes of variables for tensors.

C.1.1 Univariate Real Change of Variables

From basic calculus we recall the technique for change of variables. Let $h(x)$ be some function of the variable x. Let $y = g(x)$ be a one-to-one transformation of x to y which is valid over some set of $x \in A$ and $y \in B$. So, $g : A \rightarrow B$. Let the inverse transformation be given by $x = f(y)$. Let $\frac{\partial f}{\partial y}$ exist and be continuous over B and take on non-zero values somewhere in B. Then the function $\varphi(y)$ resulting from the change of variables is given by

$$\varphi(y) = h[f(y)]\,|J(x \rightarrow y)|$$

where $|J(x \rightarrow y)|$ is the absolute value of the determinant of the Jacobian matrix (univariate, in this case) given by

$$|J(x \rightarrow y)| = \left|\det\left(\frac{\partial f}{\partial y}\right)\right|$$

C.1.2 Traditional Multivariate Change of Variables

Bendat and Piersol (p. 59) [38] discuss the change of variables involving multi-valued functions of nice functions, like $\sin(x)$ and x^2. The probability of having arisen from one pre-image set is identical to the probability of having arisen from any other pre-image set. Their discussion is a restricted case of a more general treatment given here. One of the nicest treatments of the change of variables technique is given by Hogg and Craig (pp. 147-152) [109]. This introduction is taken from their pages 151-152 with only a few notational changes. It is important enough and short enough to be included here rather than merely just referenced.

Let x be an n-dimensional random variable, and let $\varphi(x)$ be the joint probability density function of x. Let A be the n-dimensional space where $\varphi(x) > 0$ and consider the transformation $y = u(x)$ which maps A onto B in the n-dimensional space of Y. To each point in A there corresponds only one point in B. However, a particular point in B may correspond to more than one point in A. Thus, the transformation might not be one-to-one.

Suppose that we can represent A as the union of a finite number of disjoint sets $\{A_i\}$ so that $y = u(x)$ defines a one-to-one transformation of each A_i onto B. Thus, to each point in B there will correspond exactly one point in each

of the A_i. Let $x = W_i(y)$ denote the inverse functions that map B to A_i. Let

$$\left(\frac{\partial W_i}{\partial y}\right) = \begin{pmatrix} \frac{\partial W_{1i}}{\partial y_1} & \frac{\partial W_{1i}}{\partial y_2} & \cdots & \frac{\partial W_{1i}}{\partial yn} \\ \frac{\partial W_{2i}}{\partial y_1} & \frac{\partial W_{2i}}{\partial y_2} & \cdots & \frac{\partial W_{2i}}{\partial yn} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial W_{ni}}{\partial y_1} & \frac{\partial W_{ni}}{\partial y2} & \cdots & \frac{\partial W_{ni}}{\partial yn} \end{pmatrix}$$

where each $\frac{\partial W_{ji}}{\partial y_l}$ is continuous on B and the Jacobian $J_i(x \to y) = \det\left(\frac{\partial W_i}{\partial y}\right)$ be nonzero somewhere on B. From a consideration of the probability of the union of k mutually exclusive events and applying the change of variable technique to the probability of each of these events, it can be seen that the joint probability density function of y is given by

$$\psi(y) = \begin{cases} \sum_{i=1}^{k} \varphi[W_i(y)]\,|J_i(x \to y)|\,, & y \in B \\ 0 & y \notin B \end{cases}$$

Example of $y=x^2$ on a Shaped Sample Space

Blame me for this example.

Consider the sample space $\Omega = (-2, -1, 0, 1, 2, 3, 4, 5, 6)$ and the transformation $y = x^2$. We want to determine the new density function $\psi(y)$ when we are given the original sample space and density function $\varphi(x)$. The details are given in table C.2. The point to observe here is that the points $(-3, -4, -5, -6)$ are not in the pre-image of $\psi(y)$. To solve the problem, the image and pre-image sets must be partitioned such that within a given partition it is possible to define a one-to-one transformation of variables. For the

Table C.2. Variables on a Shaped Sample Space

Ω	**−2**	**−1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
$y = x^2$	4	1	0	1	4	9	16	25	36
$+\sqrt{x^2}$	2	1	0	1	2	3	4	5	6
$-\sqrt{x^2}$	−2	−1	0	−1	−2	−3*	−4*	−5*	−6*

example just given, a set of partitions would look like figure C.1. Although a minimal set of partitions may exist, it is not necessary to use or even find it.

To generalize slightly, consider the continuous real random variable $x \in [-2, 6]$ having a probability density function of $\varphi(x)$. Change variables with the transformation $y = x^2$. We want to find the new density function $\psi(y)$. Then the inverse transformations are given by

$$W_1(y) = +\sqrt{y} \qquad 0 \leq y \leq 4$$

$$W_2(y) = -\sqrt{y} \qquad 0 \leq y \leq 4$$

$$W_3(y) = +\sqrt{y} \qquad 4 < y \leq 36$$

The magnitude of the Jacobian for this transformation is computed as follows.

$$|J(x \to y)| = \left|\det\left(\frac{\partial W_i}{\partial y}\right)\right| = \left|\det \frac{\partial \sqrt{y}}{\partial y}\right| = \frac{1}{2\sqrt{y}}$$

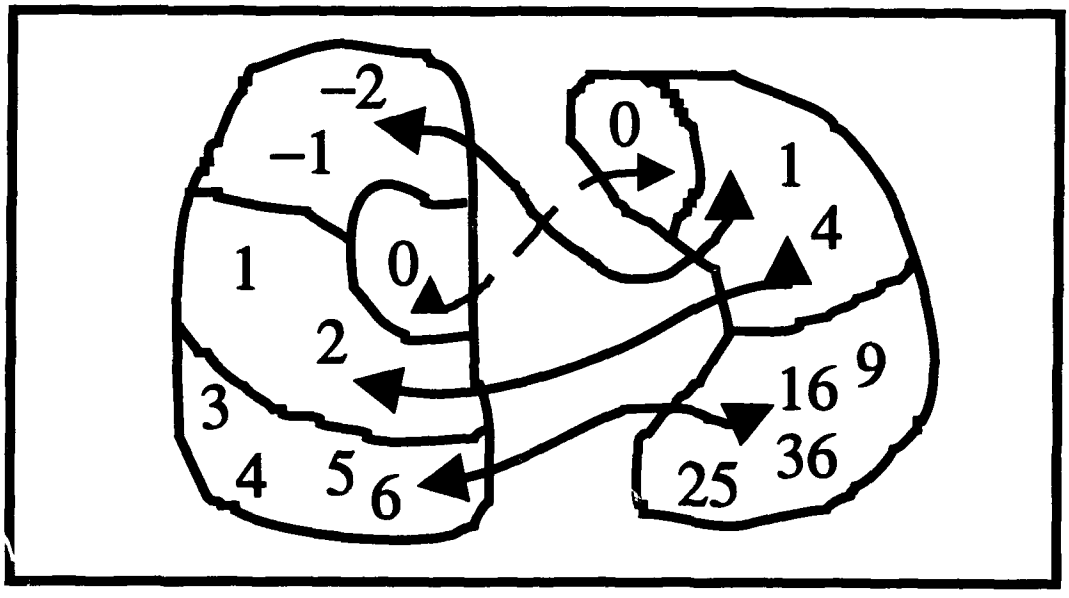


Figure C.1. Partitioning of Domain and Range for Multivalued Transformation

The new density function is given by

$$\psi(y) = \begin{cases} \varphi[W_1(y)]\,|J_1(x \to y)| + \varphi[W_2(y)]\,|J_2(x \to y)| & = 2\varphi[+\sqrt{y}]\left|\frac{1}{2\sqrt{y}}\right|, \\ & \quad \text{for } 0 \le y \le 4 \\ \varphi[W_3(y)]\,|J_3(x \to y)| & = \varphi[+\sqrt{y}]\left|\frac{1}{2\sqrt{y}}\right|, \\ & \quad \text{for } 4 < y \le 36 \end{cases}$$

C.2 Exterior (Wedge) Products

Exterior products are also known as wedge products because of the shape of the operator used to denote them, $\wedge$. Exterior products are most often studied by engineers and physicists when working with tensor calculus. The theory about

exterior products is grounded in k-forms and differential geometry. I will hide my ignorance of those areas by not explaining the theory of exterior products to you. Muirhead's text (pp. 50-57) [187] is a nice reference for those whose use of exterior products is limited to the need to compute complicated Jacobians. Another nice reference is Rudin's undergraduate text [229] on analysis. The more adventurous can consult Spivak's popular short book [254] on calculus on manifolds, and the truly bold can consult Spivak's more comprehensive work [255] on differential geometry. The reason to use exterior products is to make difficult Jacobians much easier to compute. While the use of exterior products is nice in the real variables case, their use for all but the simplest Jacobians is nearly mandatory in the complex variables case.

An exterior product $\bigwedge$ is an operator that maps a pair of differentials dx and dy into $\mathbf{R}$ and has the properties listed below:

$$dx \bigwedge dy = -dy \bigwedge dx$$

$$dx \bigwedge dx = 0$$

$$dx \bigwedge \alpha dy = \alpha dx \bigwedge dy, \text{ for constant } \alpha$$

$$dx \bigwedge (dy \bigwedge dz) = (dx \bigwedge dy) \bigwedge dy$$

$$dx \bigwedge (dy + dz) = (dx \bigwedge dy) + (dx \bigwedge dz)$$

The remainder of this section was supplied by me. The first property is useful for combining cross-product terms between real and imaginary parts of differentials when working with complex variables. The second property will

reduce our work on matrices that have some structure that repeats the use of any particular variable or its complex conjugate.

It is very useful to always consider a probability density function as a differential. As a reminder, it is useful to always write as a part of the density function the differential we are using. For example, instead of writing $\varphi(x)$, we are less likely to make conceptual mistakes by writing $\varphi(x)\,dx$. Let x be a complex variable. When viewed in $\mathbf{R}^2$, index the real part with R, and the imaginary part with I. Then we can write dx as $|dx_R\,dx_I|$ where we take the magnitude or absolute value since our interest is in scaling of differential volumes. In terms of the exterior product, we write $dx = |dx_R \bigwedge dx_I|$.

Let us examine the differential $dx \bigwedge dy = (dx \bigwedge dy)_R + i(dx \bigwedge dy)_I$.

$$dx \bigwedge dy = (dx_R + idx_I) \bigwedge (dy_R + idy_I)$$

$$= dx_R \bigwedge dy_R + dx_R \bigwedge (idy_I) + (idx_I) \bigwedge dy_R + (idx_I) \bigwedge (idy_I)$$

$$= dx_R \bigwedge dy_R + idx_R \bigwedge dy_I + idx_I \bigwedge dy_R + i^2 dx_I \bigwedge dy_I$$

$$= dx_R \bigwedge dy_R - dx_I \bigwedge dy_I + idx_R \bigwedge dy_I + idx_I \bigwedge dy_R$$

Therefore

$$(dx \bigwedge dy)_R = dx_R \bigwedge dy_R - dx_I \bigwedge dy_I$$

$$(dx \bigwedge dy)_I = dx_R \bigwedge dy_I + dx_I \bigwedge dy_R$$

Now, let the differential dy be the conjugate of dx. So, $dx = dx_R + idx_I$ and

$$dx^* = (dx_R + idx_I)^* = dx_R - idx_I$$

This leads to $dy_I = -dx_I$.

$$dx \bigwedge dx^* = (dx_R + idx_I) \bigwedge (dx_R - idx_I)$$

$$= \underbrace{dx_R \bigwedge dx_R}_{0} + dx_R \bigwedge (-idx_I) + (idx_I) \bigwedge dx_R + (idx_I) \bigwedge (-idx_I)$$

$$= -idx_R \bigwedge dx_I + idx_I \bigwedge dx_R - i^2 \underbrace{dx_I \bigwedge dx_I}_{0}$$

$$= -i2dx_R \bigwedge dx_I$$

Therefore $(dx \bigwedge dx^*)_R = 0$ and

$$(dx \bigwedge dx^*)_I = -2(dx_R \bigwedge dx_I)$$

Notice that

$$dx \bigwedge dx^* = -dx^* \bigwedge dx$$

C.2.1 Example of Wedge Products in Rectangular to Polar Coordinate Change of Variables

This was supplied by me. I expect that a senior in physics or nuclear engineering could easily produce this example.

Let $z = x + iy = re^{i\theta}$ and let $f(z) = f(x, y)$. We want to change variables from (x, y) to (r, θ). Find the Jacobian of the transformation. In other words, find $|J|$ where

$$dx \bigwedge dy = |J| \, dr \bigwedge d\theta$$

The result is the familiar $dx\,dy = r\,dr\,d\theta$. This result is often developed in undergraduate calculus courses without the benefit of exterior products. Notice how much simpler the development is here. With this approach there is no magic or serendipitous hindsight required.

We begin with $|z|^2 = x^2 + y^2 = r^2$, and from

$$x + iy = r(\cos\theta + i\sin\theta)$$

we see $x = r\cos\theta$ and $y = r\sin\theta$. Thus

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{y}{x}$$

We now take the differentials.

$$d\tan\theta = \left(\frac{1}{\cos^2\theta}\right)d\theta = \frac{1}{x}dy - \frac{y}{x^2}dx = -\frac{1}{x}\left(\frac{y}{x}dx - dy\right)$$

$$2rdr = 2xdx + 2ydy \Rightarrow dr = (x^2+y^2)^{-1/2}(xdx + ydy)$$

Taking the exterior (wedge) product, we see

$$2rdr \bigwedge \left(\frac{1}{\cos^2\theta}\right)d\theta = \frac{2r}{\cos^2\theta}dr \bigwedge d\theta$$

$$= (2xdx + 2ydy) \bigwedge \left[-\frac{1}{x}\left(\frac{y}{x}dx - dy\right)\right]$$

$$= -\frac{2}{x}(xdx + ydy) \bigwedge \left(\frac{y}{x}dx - dy\right)$$

$$= -\frac{2}{x}\left[y\underbrace{dx \bigwedge dx}_{0} - xdx \bigwedge dy + \frac{y^2}{x}\underbrace{dy \bigwedge dx}_{-dx \bigwedge dy} - y\underbrace{dy \bigwedge dy}_{0}\right]$$

$$= -\frac{2}{x}\left[-x - \frac{y^2}{x}\right]dx \bigwedge dy = \frac{2}{x^2}\left[x^2 + y^2\right]dx \bigwedge dy$$

Thus

$$dx \bigwedge dy = \frac{x^2}{2}\left[x^2+y^2\right]^{-1}\frac{2r}{\cos^2\theta}\left(dr \bigwedge d\theta\right)$$

Recall that $\cos^2\theta = \frac{x^2}{r^2}$, which means

$$dx \bigwedge dy = \frac{x^2}{2}\frac{1}{r^2}2r\left(\frac{r^2}{x^2}\right)dr \bigwedge d\theta = r\,dr\,d\theta.$$

Therefore $|J| = r$ and $dx\,dy = r\,dr\,d\theta$.

C.3 Jacobians for Complex Change of Variables

In the section we develop those Jacobians that do not require use of exterior products. In the next section we will treat those which do require exterior products. We begin with the most basic differential of a product of two matrices.

Theorem 21 *Let $X \in \mathbf{C}^{n\times m}$ and $Y \in \mathbf{C}^{m\times p}$. Then*

$$d(XY) = X[dY] + [dX]Y$$

and

$$(dX^n) = \sum_{k=0}^{n-1} X^k(dX)X^{n-1-k} = nX^{n-1}(dX)$$

when $n = m$. The first result is a complexification of the statement of Muirhead's problem 2.1 [187], given without proof. It is used by Srivastava [256]

in his derivation of the density for the complex Wishart distribution. I do not have a record of the pedigree of the second result.

Proof.

$$XY = \begin{pmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{pmatrix} \begin{pmatrix} Y_{11} & \cdots & Y_{1p} \\ \vdots & \ddots & \vdots \\ Y_{m1} & \cdots & Y_{mp} \end{pmatrix} = \left[\left(\sum_{k=1}^{m} X_{ik} Y_{kj} \right)_{ij} \right]$$

$$\left(d(XY)_{ij} \right) = \left[\left(\sum_{k=1}^{m} [(dX_{ik}) Y_{kj} + X_{ik} (dY_{kj})] \right)_{ij} \right]$$

$$= \left[\left(\sum_{k=1}^{m} (dX_{ik}) Y_{kj} \right) + \left(\sum_{k=1}^{m} X_{ik} (dY_{kj}) \right) \right] = (dX)dY + X(dY)$$

Thus

$$(dX^n) = \sum_{k=0}^{n-1} X^k (dX) X^{n-1-k} = nX^{n-1}(dX)$$

□

Note that (dX) is a scalar, whereas $\frac{dY}{dX}$ is an $(nq) \times (mp)$ matrix when Y is $q \times p$ and X is $n \times m$, and dX is a matrix of differentials (dX_{ij}).

Theorem 22 *Let x and y both be column vectors in $\mathbf{C}^n$, and let $B \in \mathbf{C}^{n \times n}$ such that $\mathrm{Re}(B^{-1})$ exists and let $B^{-1} = A \in \mathbf{C}^{n \times n}$ such that A is unstructured. Let $y = Ax$ be a complex linear transformation from x to y. Then $|J(x \to y)| = |\det(A^{-1})|^2 = |\det(B)|^2$. This is a complexification of Muirhead's theorem 2.1.1, stated in a slightly different form.*

Proof. Muirhead gave a proof for the real-variables case which used exterior products. The proof I have provided follows a more traditional approach. This

first proof of a Jacobian will dwell more on basics than future proofs. It is important to see the details once.

When forming the Jacobian of a change of variables in the complex case, recall that each part, the real part and the imaginary part, of the complex variable undergoes a change. Suppose we have some function

$$p_x(x_1, x_2, \cdots, x_n) = p_x(x_{R1}, x_{I1}, \cdots, x_{Rn}, x_{In})$$

where the subscripts R and I denote the real and imaginary parts of the associated complex variable. The goal is to find a function

$$p_y(y_1, y_2, \cdots, y_n) = p_y(y_{R1}, y_{I1}, \cdots, y_{Rn}, y_{In})$$

which is related to p_x by the mappings

$$y_{R1} = g_{R1}(x_{R1}, x_{I1}, \cdots, x_{Rn}, x_{In})$$

$$y_{I1} = g_{I1}(x_{R1}, x_{I1}, \cdots, x_{Rn}, x_{In})$$

$$\vdots$$

$$y_{Rn} = g_{Rn}(x_{R1}, x_{I1}, \cdots, x_{Rn}, x_{In})$$

$$y_{In} = g_{In}(x_{R1}, x_{I1}, \cdots, x_{Rn}, x_{In})$$

Let the inverse mappings be given by

$$x_{R1} = f_{R1}(y_{R1}, y_{I1}, \cdots, y_{Rn}, y_{In})$$

$$x_{I1} = f_{I1}(y_{R1}, y_{I1}, \cdots, y_{Rn}, y_{In})$$

$$\vdots$$

$$x_{Rn} = f_{Rn}(y_{R1}, y_{I1}, \cdots, y_{Rn}, y_{In})$$

$$x_{In} = f_{In}(y_{R1}, y_{I1}, \cdots, y_{Rn}, y_{In})$$

If the partial derivatives of f_{Rj} and f_{Ij} with respect to each of the y_{Rk} and y_{Ik} exist, then the Jacobian is the determinant of the matrix given by

$$\begin{pmatrix} \frac{\partial f_{R1}}{\partial y_{R1}} & \frac{\partial f_{I1}}{\partial y_{R1}} & \cdots & \frac{\partial f_{Rn}}{\partial y_{R1}} & \frac{\partial f_{In}}{\partial y_{R1}} \\ \frac{\partial f_{R1}}{\partial y_{I1}} & \frac{\partial f_{I1}}{\partial y_{I1}} & \cdots & \frac{\partial f_{Rn}}{\partial y_{I1}} & \frac{\partial f_{In}}{\partial y_{I1}} \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial f_{R1}}{\partial y_{Rn}} & \frac{\partial f_{I1}}{\partial y_{Rn}} & \cdots & \frac{\partial Rn}{\partial y_{Rn}} & \frac{\partial f_{In}}{\partial y_{Rn}} \\ \frac{\partial f_{R1}}{\partial y_{In}} & \frac{\partial f_{I1}}{\partial y_{In}} & \cdots & \frac{\partial f_{Rn}}{\partial y_{In}} & \frac{\partial f_{I1}}{\partial y_{In}} \end{pmatrix}$$

Consider the inverse transformation given by

$$\begin{pmatrix} x_{R1} + ix_{I1} \\ \vdots \\ x_{Rn} + ix_{In} \end{pmatrix} = \begin{pmatrix} B_{R11} + iB_{I11} & \cdots & B_{R1n} + iB_{I1n} \\ \vdots & \ddots & \vdots \\ B_{Rn1} + iB_{In1} & \cdots & B_{Rnn} + iB_{Inn} \end{pmatrix} \begin{pmatrix} y_{R1} + iy_{I1} \\ \vdots \\ y_{Rn} + iy_{In} \end{pmatrix}$$

Then

$$x_{Rk} + ix_{Ik} = \sum_{j=1}^{n} (B_{Rkj} + iB_{Ikj})(y_{Rj} + iy_{Ij})$$

$$= \sum_{j=1}^{n} [(B_{Rkj}y_{Rj} - B_{Ikj}y_{Ij}) + i(B_{Ikj}y_{Rj} + B_{Rkj}y_{Ij})]$$

By equating the real parts with each other, and doing likewise with the imaginary parts, we get

$$x_{Rk} = \sum_{j=1}^{n} (B_{Rkj} y_{Rj} - B_{Ikj} y_{Ij})$$

$$x_{Ik} = \sum_{j=1}^{n} (B_{Ikj} y_{Rj} + B_{Rkj} y_{Ij})$$

Thus the partial derivatives are given by

$$\begin{bmatrix} \frac{\partial x_{Rk}}{\partial y_{Rj}} & \frac{\partial x_{Rk}}{\partial y_{Ik}} \\ \frac{\partial x_{Ik}}{\partial y_{Rj}} & \frac{\partial x_{Ik}}{\partial y_{Ij}} \end{bmatrix} = \begin{pmatrix} B_{Rkj} & -B_{Ikj} \\ B_{Ikj} & B_{Rkj} \end{pmatrix}$$

In the transformation, it is the absolute value of the Jacobian which we need to evaluate. Knowing this relaxes the bookkeeping required while reforming the matrix of the inverse transformation. We seek the form of

$$|J| = \left| \det \left(\left(\frac{\partial x_k}{\partial y_j} \right) \right) \right|$$

$$= \left| \det \begin{pmatrix} B_{R11} & B_{I11} & B_{R21} & B_{I21} & \cdots & B_{Rn1} & B_{In1} \\ -B_{I11} & B_{R11} & -B_{I21} & B_{R21} & \cdots & -B_{In1} & B_{Rn1} \\ B_{R12} & B_{I12} & B_{R22} & B_{I22} & \cdots & B_{Rn2} & B_{In2} \\ -B_{I12} & B_{R12} & -B_{I22} & B_{R22} & \cdots & -B_{In2} & B_{Rn2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ B_{R1n} & B_{I1n} & B_{R2n} & B_{I2n} & \cdots & B_{Rnn} & B_{Inn} \\ -B_{I1n} & B_{R1n} & -B_{I2n} & B_{R2n} & \cdots & -B_{Inn} & B_{Rnn} \end{pmatrix} \right|$$

Notice that the indexing of the block matrices $B_{kj} = \begin{pmatrix} B_{Rkj} & B_{Ikj} \\ -B_{Ikj} & B_{Rkj} \end{pmatrix}$ in this equation for the Jacobian is the transpose of the matrix B when written

in its $\mathbf{R}^{2n}$ isometric form. By exchanging rows, we get

$$|J| = \left|\det \begin{pmatrix} B_{R11} & B_{I11} & B_{R21} & B_{I21} & \cdots & B_{Rn1} & B_{In1} \\ B_{R12} & B_{I12} & B_{R22} & B_{I22} & \cdots & B_{Rn2} & B_{In2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ B_{R1n} & B_{I1n} & B_{R2n} & B_{I2n} & \cdots & B_{Rnn} & B_{Inn} \\ -B_{I11} & B_{R11} & -B_{I21} & B_{R21} & \cdots & -B_{In1} & B_{Rn1} \\ -B_{I12} & B_{R12} & -B_{I22} & B_{R22} & \cdots & -B_{In2} & B_{Rn2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ -B_{I1n} & B_{R1n} & -B_{I2n} & B_{R2n} & \cdots & -B_{Inn} & B_{Rnn} \end{pmatrix}\right|$$

Notice that all the negative elements are now in the bottom half of the matrix.

By exchanging columns, we get

$$|J| = \left|\det \begin{pmatrix} B_{R11} & B_{R21} & \cdots & B_{Rn1} & B_{I11} & B_{I21} & \cdots & B_{In1} \\ B_{R12} & B_{R22} & \cdots & B_{Rn2} & B_{I12} & B_{I22} & \cdots & B_{In2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{R1n} & B_{R2n} & \cdots & B_{Rnn} & B_{I1n} & B_{I2n} & \cdots & B_{Inn} \\ -B_{I11} & -B_{I21} & \cdots & -B_{In1} & B_{R11} & B_{R21} & \cdots & B_{Rn1} \\ -B_{I12} & -B_{I22} & \cdots & -B_{In2} & B_{R12} & B_{R22} & \cdots & B_{Rn2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -B_{I1n} & -B_{I2n} & \cdots & -B_{Inn} & B_{R1n} & B_{R2n} & \cdots & B_{Rnn} \end{pmatrix}\right|$$

Now we see that all the negative elements are in the bottom left quadrant. We have

$$|J| = \left|\det\begin{pmatrix} B_R^T & B_I^T \\ -B_I^T & B_R^T \end{pmatrix}\right| = \left|\det\left[\begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix}^T\right]\right| = \left|\det\left[\begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix}\right]\right|$$

since the determinant of a matrix is equal to the determinant of its transpose.

Using lemma 45, this becomes $|J| = \left|\det[B_R]\det[B_R - B_I B_R^{-1}(-B_I)]\right|$ where we need B_R^{-1} to exist.

$$|J| = \left|\det[B_R]\det[B_R + B_I B_R^{-1} B_I]\right| = \left|\det[B_R]\det[(I + B_I B_R^{-1} B_I B_R^{-1})B_R]\right|$$

$$= \left|(\det[B_R])^2 \det[I + (B_I B_R^{-1})^2]\right|$$

$$= \left|(\det[B_R])^2 \det[\{I + iB_I B_R^{-1}\}\{I - iB_I B_R^{-1}\}]\right|$$

$$= \left|\{\det[I + iB_I B_R^{-1}]\det[B_R]\}\{\det[I - iB_I B_R^{-1}]\det[B_R]\}\right|$$

$$= |\{\det[B_R + iB_I]\}\{\det[B_R - iB_I]\}| = |\{\det[B]\}\{\det[B]\}^*| = |\det[B]|^2$$

We also know that

$$|\det[B]|^2 = |\{\det[B]\}\{\det[B]\}^*| = |\{\det[B]\}\{\det[B^*]\}|$$

by lemma 42, which becomes

$$|\{\det[BB^*]\}| = \{\det[|B|^2]\}$$

To summarize, given a complex linear transformation $y = Ax$ where each of y, A, and x are complex, with inverse transformation

$$x = A^{-1}y = By = (B_R + iB_I)y$$

then the Jacobian of the transformation is given by $|J(x \rightarrow y)| = |\det B|^2 = |\det A|^{-2}$. □

Proposition 30 *Let $x = By$ be a change of complex variables such that $x, y \in \mathbf{C}^n$ and $B \in \mathbf{C}^{n \times n}$ except that the first column of B is constrained to be a column of all ones. Then $|J(x \rightarrow y)| = |\det B|^2$ and $|J(y \rightarrow x)| = |\det B|^{-2}$.*

Proof. This issue arose in deriving a complex version of Anderson (pp. 522-530) [26] theorem 13.2.2. There was concern over the possibility of the Jacobian being anomalously zero due to all zero entries in the imaginary part of the first column of B. The important message of this proof is that the concern is unfounded.

Let

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & B_{12} & \cdots & B_{1n} \\ 1 & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & B_{n2} & \cdots & B_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Then

$$x_{Ri} + ix_{Ii} = y_{Ri} + \sum_{j=2}^{n}(B_{Rij}y_{Rj} - B_{Iij}y_{Ij}) + i\sum_{j=2}^{n}(B_{Iij}y_{Rj} + B_{Rij}y_{Ij})$$

where the subscripts R and I refer to the real and imaginary parts of their complex variables. The Jacobian matrix for the change of variables is given

by

$$\begin{pmatrix} \frac{\partial y_R}{\partial x_R} & \frac{\partial y_I}{\partial x_R} \\ \frac{\partial y_R}{\partial x_I} & \frac{\partial y_I}{\partial x_I} \end{pmatrix} = \begin{pmatrix} 1 & B_{R12} & \cdots & B_{R1n} & 0 & -B_{I12} & \cdots & -B_{I1n} \\ 1 & B_{R22} & \cdots & B_{R2n} & 0 & -B_{I22} & \cdots & -B_{I2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & B_{Rn2} & \cdots & B_{Rnn} & 0 & -B_{In2} & \cdots & \mathfrak{Z}_{Inn} \\ 0 & B_{I12} & \cdots & B_{I1n} & 1 & B_{R12} & \cdots & B_{R1n} \\ 0 & B_{I22} & \cdots & N_{I2n} & 1 & B_{R22} & \cdots & B_{R2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & B_{In2} & \cdots & B_{Inn} & 1 & B_{Rn2} & \cdots & B_{Rnn} \end{pmatrix}$$

$$= \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix}$$

As shown in theorem 22, the determinant of this matrix is found by the partitioned matrix determinant lemma 45, if the determinant exists. Thus

$$\det \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix} = \det(B_R)\det(B_R + B_I B_R^{-1} B_I)$$

Let $A_R = B_R^{-1}$. Then

$$B_I B_R^{-1} B_I = \begin{pmatrix} 0 & \sum_{j=2}^{n} B_{I1j} \sum_{i=1}^{n} A_{Rji} B_{Ii2} & \cdots & \sum_{j=2}^{n} B_{I1j} \sum_{i=1}^{n} A_{Rji} B_{Iin} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sum_{j=2}^{n} B_{Inj} \sum_{i=1}^{n} A_{Rji} B_{Ii2} & \cdots & \sum_{j=2}^{n} B_{Inj} \sum_{i=1}^{n} A_{Rji} B_{Iin} \end{pmatrix}$$

Even though column one is zero, when we consider $\det(B_R + B_I B_R^{-1} B_i)$ we observe that the determinant is not necessarily zero. Thus we can claim that

except for pathological cases, $\det \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix} = |\det B|^2$. □

Lemma 2 *Let $Y = TL$ be a change of complex variables where Y and L are in $\mathbf{C}^{n\times p}$ and T is upper triangular in $\in \mathbf{C}^{n\times n}$. Then the Jacobians of the transformations between Y and L are $|J(Y \to L)| = |\det T|^{2p} = \prod_{k=1}^{n} |T_{kk}|^{2p}$ and $|J(L \to Y)| = \prod_{k=1}^{n} |T_{kk}|^{-2p}$.*

Proof. We begin by treating the matrices as the sum of their real and imaginary parts.

$$Y_R + iY_I = (T_R + iT_I)(L_R + iL_I)$$

Then matching the real and imaginary parts we get

$$Y_R = T_R L_R - T_I L_I$$

and

$$Y_I = T_I L_R + T_R L_I$$

Then

$$\left(\frac{\partial Y_R}{\partial L_R}\right) = T_R \otimes I_p, \left(\frac{\partial Y_R}{\partial L_I}\right) = -T_I \otimes I_p, \left(\frac{\partial Y_I}{\partial L_R}\right) = T_I \otimes I_p$$

and

$$\left(\frac{\partial Y_I}{\partial L_I}\right) = T_R \otimes I_p$$

Note that this is merely the Cauchy-Riemann condition for the existence of the complex derivative at the points given by L. It is more apparent in the

form: $\left(\frac{\partial Y_R}{\partial L_R}\right) = \left(\frac{\partial Y_I}{\partial L_I}\right)$ and $\left(\frac{\partial Y_R}{\partial L_I}\right) = -\left(\frac{\partial Y_I}{\partial L_R}\right)$. To find the Jacobian, we will examine

$$\left|\det\begin{pmatrix} \frac{\partial Y_R}{\partial L_R} & \frac{\partial Y_R}{\partial L_I} \\ \frac{\partial Y_I}{\partial L_R} & \frac{\partial Y_I}{\partial L_I} \end{pmatrix}\right| = \left|\det\begin{pmatrix} T_R \otimes I_p & -T_I \otimes I_p \\ T_I \otimes I_p & T_R \otimes I_p \end{pmatrix}\right| = \left|\det\begin{pmatrix} T_R & -T_I \\ T_I & T_R \end{pmatrix}\right|^p$$

by lemma 49. We now apply the partitioned matrix determinant lemma 45 to get

$$\left|\det(T_R)\det(T_R + T_I T_R^{-1} T_I)\right|^p = \left|[\det(T_R)]^2 \det(I + T_R^{-1} T_I T_R^{-1} T_I)\right|^p$$

$$= \left|[\det(T_R)]^2 \det[I + (T_R^{-1} T_I)^2]\right|^p = \left|\det(T_R)\det(I + iT_R^{-1} T_I)\right|^{2p}$$

By proposition 65 this is

$$|\det(T_R + iT_I)|^{2p} = |\det T|^{2p}$$

since T_R is conformable with the matrices I and $T_R^{-1} T_I$. The last term is the magnitude of the determinant of complex triangular matrix T raised to the power $2p$. Thus the Jacobian of the change of variables $Y = TL$ is given by $|J(Y \to L)| = |\det T|^{2p}$ and $|J(L \to Y)| = |\det T|^{-2p}$. □

Lemma 3 *Let $Y = TA$ be a change of complex variables between Y and A where $Y \in \mathbf{C}^{n \times p}$, $A \in \mathbf{C}^{n \times p}$, and T is lower triangular in $\mathbf{C}^{n \times n}$ with positive real elements on the diagonal. Then $|J(Y \to A)| = \prod_{k=1}^{n} T_{kk}^{2p}$ and $|J(A \to Y)| = \prod_{k=1}^{n} T_{kk}^{-2p}$.*

Proof. By lemma 2, we know $|J(Y \to A)| = |\det T|^{2p}$. At this point in lemma 2, no use of the fact that T is triangular has been made. Since T is lower triangular with positive real diagonal elements, then $\det T = \prod_{k=1}^{n} T_{kk}$ and therefore $|J(Y \to A)| = \prod_{k=1}^{n} T_{kk}^{2p}$ and $|J(A \to Y)| = \prod_{k=1}^{n} T_{kk}^{-2p}$. □

Lemma 4 *Let $Y = AT$ be a change of complex variables between Y and A where $Y \in \mathbf{C}^{n \times p}$, $A \in \mathbf{C}^{n \times p}$, and T is upper triangular in $\mathbf{C}^{p \times p}$ with positive real elements on the diagonal. Then $|J(Y \to A)| = \prod_{k=1}^{p} T_{kk}^{2n}$ and $|J(A \to Y)| = \prod_{k=1}^{p} T_{kk}^{-2n}$.*

Proof. $Y = AT$ implies that the transpose is $Y^T = T^T A^T$. Note that $Y^T \in \mathbf{C}^{p \times n}$, $A^T \in \mathbf{C}^{p \times n}$, and $T^T \in \mathbf{C}^{p \times p}$. The matrix T^T is lower triangular. By lemma 3, $\left|J(Y^T \to A^T)\right| = \prod_{k=1}^{p} T_{kk}^{2n}$. Since the Jacobian determinant is scalar, it equals its transpose. Thus $\left|J(Y^T \to A^T)\right| = |J(Y \to A)|$. Therefore $|J(Y \to A)| = \prod_{k=1}^{n} T_{kk}^{2p}$ and $|J(A \to Y)| = \prod_{k=1}^{n} T_{kk}^{-2p}$. □

Lemma 5 *Let $Y = AT$ be a change of complex variables between Y and A where $Y \in \mathbf{C}^{n \times p}$, $A \in \mathbf{C}^{n \times p}$, and T is lower triangular in $\mathbf{C}^{p \times p}$ with positive real elements on the diagonal. Then $|J(Y \to A)| = \prod_{k=1}^{p} T_{kk}^{2n}$ and $|J(A \to Y)| = \prod_{k=1}^{p} T_{kk}^{-2n}$.*

Proof. $Y = AT$ implies $Y^T = T^T A^T$ where $Y^T \in \mathbf{C}^{p \times n}$, $A^T \in \mathbf{C}^{p \times n}$, and $T^T \in \mathbf{C}^{p \times p}$. T^T is upper triangular. By lemma 2, $\left|J(Y^T \to A^T)\right| = \prod_{k=1}^{p} T_{kk}^{2n}$ which implies $|J(Y \to A)| = \prod_{k=1}^{p} T_{kk}^{2n}$ and $|J(A \to Y)| = \prod_{k=1}^{p} T_{kk}^{-2n}$. □

Theorem 23 *Let $Y = TAT^H$ be a change of complex variables between Y and A where $Y \in \mathbf{C}^{p\times p}$, $A \in \mathbf{C}^{p\times p}$, and T is lower triangular in $\mathbf{C}^{p\times p}$ with positive real elements on the diagonal. Let $B = TT^H$. Then $|J(Y \to A)| = \prod_{k=1}^{p} T_{kk}^{4p} = |\det T|^{4p} = |\det B|^{2p} = (\det B)^{2p}$.*

Proof. Consider $Y = TAT^H$ as two transformations $Y_1 = AT^H$ and $Y = TY_1$. T^H is upper triangular with positive diagonal elements. Apply lemma 4. Then

$$|J(Y_1 \to A)| = \prod_{k=1}^{p} T_{kk}^{2p} = |\det T|^{2p}$$

Now apply lemma 3 to obtain

$$|J(Y \to Y_1)| = \prod_{k=1}^{p} T_{kk}^{2p} = |\det T|^{2p}$$

Thus

$$|J(Y \to A)| = |J(Y \to Y_1)| \cdot |J(Y_1 \to A)|$$

$$= |\det T|^{4p} = \left|\det TT^H\right|^{2p} = |\det B|^{2p} = \prod_{k=1}^{p} T_{kk}^{4p}$$

The magnitude symbols may be dropped since T_{kk} is real for all k. □

Example 3 *This quick example illustrates the minute details of the action in theorem 23.*

$$\begin{pmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{pmatrix} = \begin{pmatrix} T_{11} & & \\ T_{21} & T_{22} & \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} T_{11} & T_{21}^* & T_{31}^* \\ & T_{22} & T_{32}^* \\ & & T_{33} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11} & & \\ T_{21} & T_{22} & \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \begin{pmatrix} A_{11}T_{11} & A_{11}T_{21}^* + A_{12}T_{22} & A_{11}T_{31}^* + A_{12}T_{32}^* + A_{13}T_{33} \\ A_{21}T_{11} & A_{21}T_{21}^* + A_{22}T_{22} & A_{21}T_{31}^* + A_{22}T_{32}^* + A_{23}T_{33} \\ A_{31}T_{11} & A_{31}T_{21}^* + A_{32}T_{22} & A_{31}T_{31}^* + A_{32}T_{32}^* + T_{33}A_{33} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11}A_{11}T_{11} \\ T_{21}A_{11}T_{11} + T_{22}A_{21}T_{11} \\ T_{31}A_{11}T_{11} + T_{32}A_{21}T_{11} + T_{33}A_{31}T_{11} \end{pmatrix}$$

$$\begin{vmatrix} T_{11}A_{11}T_{21}^* + T_{11}A_{12}T_{22} \\ T_{21}A_{11}T_{21}^* + T_{21}A_{12}T_{22} + T_{22}A_{21}T_{21}^* + T_{22}A_{22}T_{22} \\ T_{31}A_{11}T_{21}^* + T_{31}A_{12}T_{22} + T_{32}A_{21}T_{21}^* + T_{32}A_{22}T_{22} + T_{33}A_{31}T_{21}^* + T_{33}A_{32}T_{22} \end{vmatrix}$$

$$\begin{pmatrix} T_{11}A_{11}T_{31}^* + T_{11}A_{12}T_{32}^* + T_{11}A_{13}T_{33} \\ T_{21}A_{11}T_{31}^* + T_{21}A_{12}T_{32}^* + T_{21}A_{13}T_{33} + T_{22}A_{21}T_{31}^* + T_{22}A_{22}T_{32}^* + T_{22}A_{23}T_{33} \\ \left[\begin{array}{c} T_{31}A_{11}T_{31}^* + T_{31}A_{12}T_{32}^* + T_{31}A_{13}T_{33} + T_{32}A_{21}T_{31}^* + T_{32}A_{22}T_{32}^* \\ +T_{32}A_{23}T_{33} + T_{33}A_{31}T_{31}^* + T_{33}A_{32}T_{32}^* + T_{33}A_{33}T_{33} \end{array} \right] \end{pmatrix}$$

The element for the third row in the last column has too many terms to fit onto one line, and it is therefore enclosed in square brackets. Recall that each

$A_{ij} = A_{Rij} + iA_{Iij}$. We get the following differentials.

$$
\begin{aligned}
dY_{R11} &= T_{11}^2 dA_{R11} & dY_{I11} &= T_{11}^2 dA_{I11} \\
dY_{R21} &= T_{11}T_{22} dA_{R21} + \cdots & dY_{I21} &= T_{11}T_{22} dA_{I21} + \cdots \\
dY_{R12} &= T_{11}T_{22} dA_{R12} + \cdots & dY_{I12} &= T_{11}T_{22} dA_{I12} + \cdots \\
dY_{R13} &= T_{11}T_{33} dA_{R13} + \cdots & dY_{I13} &= T_{11}T_{33} dA_{I13} + \cdots \\
dY_{R31} &= T_{11}T_{33} dA_{R31} + \cdots & dY_{I31} &= T_{11}T_{33} dA_{I31} + \cdots \\
dY_{R22} &= T_{22}^2 dA_{R22} + \cdots & dY_{I22} &= T_{22}^2 dA_{I22} + \cdots \\
dY_{R32} &= T_{22}T_{33} dA_{R32} + \cdots & dY_{I32} &= T_{22}T_{33} dA_{I32} + \cdots \\
dY_{R23} &= T_{22}T_{33} dA_{R23} + \cdots & dY_{I23} &= T_{22}T_{33} dA_{I23} + \cdots \\
dY_{R33} &= T_{33}^2 dA_{R33} + \cdots & dY_{I33} &= T_{33}^2 dA_{I33} + \cdots
\end{aligned}
$$

Therefore

$$dY = T_{11}^{12} T_{22}^{12} T_{33}^{12} dA = (\det T)^{12}$$

where $p = 3$. Let $B = TT^H$. Thus

$$dY = (\det T)^{4p} = (\det B)^{2p}$$

Theorem 24 *Let $Y = TAT^H$ be a change of complex variables between Y and A where $A^H = A > 0$ and Y are both in $\mathbf{C}^{p\times p}$, and $T \in \mathbf{C}^{p\times p}$ is lower triangular with positive real diagonal elements. Let $B = TT^H$. Then $|J(Y \to A)| = (\det T)^{2p} = (\det B)^{p}$ and $|J(A \to Y)| = (\det T)^{-2p} = (\det B)^{-p}$.*

Proof. By corollary 36 there exists a unique $p \times p$ lower triangular matrix L with positive real diagonal elements such that $A = LL^H$. Thus

$$Y = TLL^H T^H = CC^H$$

where $C = TL$. C is lower triangular in $\mathbf{C}^{p\times p}$ with positive real diagonal elements $C_{kk} = T_{kk}L_{kk}$. By theorem 26,

$$\left|J(A \to LL^H)\right| = 2^p \prod_{k=1}^{p} L_{kk}^{2(p-k)+1}$$

By lemma 6,

$$|J(C \to L)| = \prod_{k=1}^{p} T_{kk}^{2k-1}$$

By theorem 26,

$$\left|J(Y \to CC^H)\right| = 2^p \prod_{k=1}^{p} (T_{kk}L_{kk})^{2(p-k)+1}$$

Now we put the Jacobian all together where we take the inverse of the Jacobian $\left|J(A \to LL^H)\right|$.

$$|J(Y \to A)| = \left|J(Y \to CC^H)\right| \cdot |J(C \to L)| \cdot \left|J(LL^H \to A)\right|$$

$$= \left(2^p \prod_{k=1}^{p} (T_{kk}L_{kk})^{2(p-k)+1}\right)\left(\prod_{k=1}^{p} T_{kk}^{2k-1}\right)\left(2^{-p} \prod_{k=1}^{p} L_{kk}^{-2(p-k)-1}\right)$$

$$= \prod_{k=1}^{p} T_{kk}^{2p} = (\det T)^{2p} = (\det TT^H)^p = (\det B)^p$$

Therefore

$$|J(Y \to A)| = (\det T)^{2p} = (\det B)^p$$

and therefore

$$|J(A \rightarrow Y)| = (\det T)^{-2p} = (\det B)^{-p}$$

□

Proposition 31 *Let A and G both be lower triangular complex matrices where the elements on the diagonal are complex. Let $T = AG$ define a change of variables between T and A. Then*

$$|J(T \rightarrow A)| = \prod_{k=1}^{p} |G_{kk}|^{2(p-k+1)}$$

where $G_{kk} = G_{Rkk} + iG_{Ikk}$, and

$$|J(A \rightarrow T)| = \prod_{k=1}^{p} |G_{kk}|^{-2(p-k+1)}$$

This is the second equation Khatri section 2.5 [137], stated without proof, where he uses $J(T;A) = |J(T \rightarrow A)|$. This result differs from Khatri's result.

Proof.

$$T = AG = \begin{pmatrix} T_{11} & & & \\ T_{21} & T_{22} & & \\ \vdots & \vdots & \ddots & \\ T_{p1} & T_{p2} & \cdots & T_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \vdots & \vdots & \ddots & \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{pmatrix} \times$$

$$\times \begin{pmatrix} G_{11} & & & \\ G_{21} & G_{22} & & \\ \vdots & \vdots & \ddots & \\ G_{p1} & G_{p2} & \cdots & G_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}G_{11} & & & \\ A_{21}G_{11} + A_{22}G_{21} & A_{22}G_{22} & & \\ \vdots & \vdots & \ddots & \\ \sum_{k=1}^{p} A_{pk}G_{k1} & \sum_{k=2}^{p} A_{pk}G_{k2} & \cdots & A_{pp}G_{pp} \end{pmatrix}$$

where $T_{ij} = \sum_{k=j}^{i} A_{jk}G_{kj}$ for $i \geq j$ is a typical element. We expand this element.

$$T_{ij} = \sum_{k=j}^{i} \left[\left(A_{Rik}G_{Rkj} - A_{Iik}G_{Ikj} \right) + i \left(A_{Iik}G_{Rkj} + A_{Rik}G_{Ikj} \right) \right]$$

We now compute the Jacobian of this transformation, $|J(T \rightarrow A)|$. Consider the Jacobian matrix having the following rows.

$$\det \begin{bmatrix} \frac{\partial T_{R11}}{\partial A_{R11}} & \frac{\partial T_{R11}}{\partial A_{I11}} & \frac{\partial T_{R11}}{\partial A_{R21}} & \frac{\partial T_{R11}}{\partial A_{I21}} & \cdots & \frac{\partial T_{R11}}{\partial A_{Rpp}} & \frac{\partial T_{R11}}{\partial A_{Ipp}} \\ \frac{\partial T_{I11}}{\partial A_{R11}} & \frac{\partial T_{I11}}{\partial A_{I11}} & \frac{\partial T_{I11}}{\partial A_{R21}} & \frac{\partial T_{I11}}{\partial A_{I21}} & \cdots & \frac{\partial T_{I11}}{\partial A_{Rpp}} & \frac{\partial T_{I11}}{\partial A_{Ipp}} \\ \frac{\partial T_{R21}}{\partial A_{R11}} & \frac{\partial T_{R21}}{\partial A_{I11}} & \frac{\partial T_{R21}}{\partial A_{R21}} & \frac{\partial T_{R21}}{\partial A_{I21}} & \cdots & \frac{\partial T_{R21}}{\partial A_{Rpp}} & \frac{\partial T_{R21}}{\partial A_{Ipp}} \\ \frac{\partial T_{I21}}{\partial A_{R11}} & \frac{\partial T_{I21}}{\partial A_{I11}} & \frac{\partial T_{I21}}{\partial A_{R21}} & \frac{\partial T_{I21}}{\partial A_{I21}} & \cdots & \frac{\partial T_{I21}}{\partial A_{Rpp}} & \frac{\partial T_{I21}}{\partial A_{Ipp}} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial T_{Rpp}}{\partial A_{R11}} & \frac{\partial T_{Rpp}}{\partial A_{I11}} & \frac{\partial T_{Rpp}}{\partial A_{R21}} & \frac{\partial T_{Rpp}}{\partial A_{I21}} & \cdots & \frac{\partial T_{Rpp}}{\partial A_{Rpp}} & \frac{\partial T_{Rpp}}{\partial A_{Ipp}} \\ \frac{\partial T_{Ipp}}{\partial A_{R11}} & \frac{\partial T_{Ipp}}{\partial A_{I11}} & \frac{\partial T_{Ipp}}{\partial A_{R21}} & \frac{\partial T_{Ipp}}{\partial A_{I21}} & \cdots & \frac{\partial T_{Ipp}}{\partial A_{Rpp}} & \frac{\partial T_{Ipp}}{\partial A_{Ipp}} \end{bmatrix}$$

The partial derivatives yield a block structure as illustrated in the Jacobian matrix below. The left half of the matrix is given on top, and the right half of the matrix is given on bottom. What you should be looking for is the pattern.

$$\det \begin{bmatrix} G_{R11} & -G_{I11} & 0 & \cdots & \\ G_{I11} & G_{R11} & 0 & \cdots & \\ 0 & 0 & G_{R11} & -G_{I11} & G_{R21} \\ \vdots & \vdots & G_{I11} & G_{R11} & G_{I21} \\ & & 0 & 0 & G_{R22} \\ \vdots & & \vdots & \vdots & G_{I22} \\ \vdots & & & & 0 \\ \vdots & & & & \\ \vdots & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

$$
\begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & & & & & \vdots \\
-G_{I21} & 0 & \cdots & & & \vdots \\
G_{R21} & 0 & \cdots & & & \vdots \\
-G_{I22} & 0 & \cdots & & & \vdots \\
G_{R22} & 0 & \cdots & & & \\
0 & G_{R11} & -G_{I11} & G_{R21} & -G_{I21} & \cdots \\
 & G_{I11} & G_{R11} & G_{I21} & G_{R21} & \cdots \\
 & 0 & 0 & \vdots & \vdots & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}
$$

$$
= \det \begin{bmatrix}
\tilde{G}_{11} & 0 & \cdots & & & & & \\
0 & \tilde{G}_{11} & \tilde{G}_{21} & 0 & \cdots & & & \\
\vdots & 0 & \tilde{G}_{22} & 0 & \cdots & & & \\
 & \vdots & 0 & \tilde{G}_{11} & \tilde{G}_{21} & \tilde{G}_{31} & 0 & \cdots \\
 & & \vdots & 0 & \tilde{G}_{22} & \tilde{G}_{32} & 0 & \cdots \\
 & & & \vdots & 0 & \tilde{G}_{33} & 0 & \cdots \\
 & & & & \vdots & 0 & \ddots &
\end{bmatrix}
$$

The Jacobian is the magnitude of the determinant of this matrix.

$$
|J(T \to A)| = \left| \det \begin{bmatrix} \tilde{G}_{11} & {}_1H_{21} \\ {}_1H_{21} & {}_1H_{22} \end{bmatrix} \right|
$$

where the Jacobian matrix has been partitioned so that $\tilde{G}_{11} = \begin{pmatrix} G_{R11} & -G_{I11} \\ G_{I11} & G_{R11} \end{pmatrix}$, ${}_1H_{12} = 0$ and ${}_1H_{21} = 0$ where 0 is a matrix of zeros of appropriate dimensions. The lower left subscript on H identifies the block number. It is the number of row blocks that are excluded from the full Jacobian matrix. We continue with the sequential partitioning scheme to compute the Jacobian by using the partitioned matrix determinant lemma 45.

$$\left|\det \begin{bmatrix} \tilde{G}_{11} & {}_1H_{21} \\ {}_1H_{21} & {}_1H_{22} \end{bmatrix}\right| = \left|\det(\tilde{G}_{11})\right| \cdot \left|\det[{}_1H_{22} - {}_1H_{21}\tilde{G}_{11}^{-1}\,{}_1H_{12}]\right|$$

$$= \left|\det(\tilde{G}_{11})\right| \cdot |\det[{}_1H_{22}]| = \left|\det \tilde{G}_{11}\right|^2 |\det[{}_2H_{22}]|$$

$$= \left|\det \tilde{G}_{11}\right|^2 \left|\tilde{G}_{22}\right| |\det[{}_3H_{22}]| = \prod_{K=1}^{P} \left|\det \tilde{G}_{kk}\right|^{p-k+1}$$

Note that

$$\left|\det \tilde{G}_{kk}\right| = \left|\det \begin{pmatrix} G_{Rkk} & -G_{Ikk} \\ G_{Ikk} & G_{Rkk} \end{pmatrix}\right| = G_{Rkk}^2 + G_{Ikk}^2 = |G_{kk}|^2$$

where $G_{kk} = G_{Rkk} + iG_{Ikk}$. Therefore,

$$|J(T \to A)| = \prod_{k=1}^{p} |G_{kk}|^{2(p-k+1)}$$

By the inverse property of Jacobians,

$$|J(A \to T)| = \prod_{k=1}^{p} |G_{kk}|^{-2(p-k+1)}$$

□

Proposition 32 *Let A and G both be lower triangular complex $p \times p$ matrices with real diagonal elements. Let $T = AG$ define a change of variables between A and T. Then the Jacobians of the transformations are $|J(T \to A)| = \prod_{k=1}^{p} G_{kk}^{2(p-k)+1}$ and $|J(A \to T)| = \prod_{k=1}^{p} G_{kk}^{-2(p-k)-1}$. This is the fourth equation of Khatri section 2.5 [137], which is stated without proof. This result differs from Khatri's.*

Proof. Notice that since A and G are both lower triangular with real diagonal elements, then T also is lower diagonal with real diagonal elements. A typical element of T is given by $T_{Rjk} + iT_{Ijk} = T_{ij}$ where

$$T_{ij} = \sum_{k=j}^{i} [(A_{Rik}G_{Rkj} - A_{Iik}G_{Ikj}) + i(A_{Iik}G_{Rkj} + A_{Rik}G_{Ikj})]$$

$$= \sum_{q=k}^{j-1} [(A_{Rjq}G_{Rqk} - A_{Ijq}G_{Iqk}) + i(A_{Ijq}G_{Rqk} + A_{Rjq}G_{Iqk})] + A_{jj}G_{Rjk}$$

For example,

$$T_{11} = A_{11}G_{11}$$

$$T_{21} = A_{R21}G_{11} + i(A_{I21}G_{11}) + A_{22}G_{R21}$$

We separate the real part of the equation from the imaginary part. This implies

$$T_{R21} = A_{R21}G_{11} + A_{22}G_{R21}$$

$$T_{I21} = A_{I21}G_{11} + A_{22}G_{I21}$$

We repeat this procedure for each element that is not a pure real element or a pure imaginary element.

$$T_{22} = A_{22}G_{22}$$

$$T_{R31} = A_{R31}G_{11} + A_{R32}G_{R21} - A_{I32}G_{I21} + A_{33}G_{R31}$$

$$T_{I31} = A_{I31}G_{11} + A_{I32}G_{R21} + A_{R32}G_{I21} + A_{33}G_{I31}$$

$$T_{R32} = A_{R32}G_{22} + A_{33}G_{R32}$$

$$T_{I32} = A_{I32}G_{22} + A_{33}G_{I32}$$

$$T_{33} = A_{33}G_{33}$$

Then the Jacobian $|J(T \rightarrow A)|$ is computed as the determinant of

$$\begin{bmatrix}
G_{11} & & & & & & & & & \\
 & G_{11} & 0 & G_{R21} & & & & & & \\
 & 0 & G_{11} & G_{I21} & & & & & & \\
 & & & G_{22} & & & & & & \\
 & & & & G_{11} & 0 & G_{R21} & -G_{I21} & G_{R31} & \\
 & & & & & G_{11} & G_{I21} & G_{R21} & G_{I31} & \\
 & & & & & & G_{22} & 0 & G_{R32} & \\
 & & & & & & & G_{22} & G_{I32} & \\
 & & & & & & & & G_{33} & \\
 & & & & & & & & & \ddots
\end{bmatrix}$$

From this pattern we can see

$$|J(T \to A)| = \prod_{k=1}^{p} G_{kk}^{2(p-k)+1}$$

and

$$|J(A \to T)| = \prod_{k=1}^{p} G_{kk}^{-2(p-k)-1}$$

□

Proposition 33 *Let A and G both be lower triangular complex $p \times p$ matrices with complex diagonal elements. Let $T = AG$ define a change of variables between G and T. Then the Jacobians of the transformations are $|J(T \to G)| = \prod_{k=1}^{p} |A_{kk}|^{2k}$ and $|J(G \to T)| = \prod_{k=1}^{p} |A_{kk}|^{-2k}$. This is the first equation of Khatri section 2.5 [137], which was stated without proof. Closely compare this to lemma 31 to see that before we changed variables between A and G, the order of the constant and variable matrices have been changed. This result is different than Khatri's.*

Proof. The matrix $\frac{\partial T}{\partial G}$ is given by the following. The left half of the matrix is in the top display, and the right half of the matrix is in the bottom of the displayed pair. You should be looking at the pattern of the elements. The top

half is:

$$\begin{bmatrix}
A_{R11} & -A_{I11} & & & & \\
A_{I11} & A_{R11} & & & & \\
A_{R21} & -A_{I21} & A_{R22} & -A_{I22} & & \\
A_{I21} & A_{R21} & A_{I22} & A_{R22} & & \\
 & & & & A_{R22} & -A_{I22} \\
 & & & & A_{I22} & A_{R22} \\
A_{R31} & -A_{I31} & A_{R32} & -A_{I32} & & \\
A_{I31} & A_{R31} & A_{I32} & A_{R32} & & \\
 & & & & A_{R32} & -A_{I32} \\
 & & & & A_{I32} & A_{R32} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}$$

The bottom half is:

$$\begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & & & & & \\
\cdots & & & & & \\
\cdots & & & & & \\
\cdots & & & & & \\
\cdots & & & & & \\
A_{R33} & -A_{I33} & & & & \\
A_{I33} & A_{R33} & & & & \\
 & & A_{R33} & -A_{I33} & & \\
 & & A_{I33} & A_{R33} & & \\
 & & & & A_{R33} & -A_{I33} \\
 & & & & A_{I33} & A_{R33}
\end{bmatrix}$$

This patterned matrix gives a Jacobian determinant

$$|J(T \to G)| = \begin{vmatrix} \tilde{A}_{11} & & & & & & \\ \tilde{A}_{21} & \tilde{A}_{22} & & & & & \\ & & \tilde{A}_{22} & & & & \\ \tilde{A}_{31} & \tilde{A}_{32} & & \tilde{A}_{33} & & & \\ & & \tilde{A}_{32} & & \tilde{A}_{33} & & \\ & & & & & \tilde{A}_{33} & \\ & & & & & & \ddots \end{vmatrix} = \prod_{k=1}^{p} \left|\tilde{A}_{kk}\right|^{k} = \prod_{k=1}^{p} |A_{kk}|^{2k}$$

where $A_{kk} = A_{Rkk} + iA_{Ikk}$ and

$$\tilde{A}_{kk} = \begin{pmatrix} A_{Rkk} & -A_{Ikk} \\ A_{Ikk} & A_{Rkk} \end{pmatrix}$$

By the inversion property of Jacobians, we also have

$$|J(G \to T)| = \prod_{k=1}^{p} |A_{kk}|^{-2k}$$

□

Lemma 6 *Let A and G both be lower triangular complex $p \times p$ matrices with real-valued diagonal elements. Let $T = AG$ define a change of variables between G and T. Then the Jacobians of the transformations are $|J(T \to G)| = \prod_{k=1}^{p} A_{kk}^{2k-1}$ and $|J(G \to T)| = \prod_{k=1}^{p} A_{kk}^{-2k+1}$. This is the third equation of Khatri section 2.5 [137], which was stated without proof. This result differs from Khatri's.*

Proof. The Jacobian $|J(T \rightarrow G)|$ is computed as

$$\left| \det \begin{bmatrix} A_{11} & 0 & 0 & 0 & 0 & & & & & \\ A_{R21} & A_{22} & 0 & 0 & 0 & & & & & \\ A_{I21} & 0 & A_{22} & 0 & 0 & & & & & \\ 0 & 0 & 0 & A_{22} & 0 & & & & & \\ A_{R31} & A_{R32} & -A_{I32} & 0 & A_{33} & & & & & \\ A_{I31} & A_{I32} & A_{R32} & 0 & 0 & A_{33} & & & & \\ 0 & 0 & 0 & A_{R32} & 0 & 0 & A_{33} & & & \\ 0 & 0 & 0 & A_{I32} & 0 & \vdots & \vdots & A_{33} & & \\ 0 & \vdots & \vdots & \vdots & & & & \vdots & A_{33} & \\ \vdots & & & & & & & \vdots & \ddots & \end{bmatrix} \right| = \prod_{k=1}^{p} A_{kk}^{2k-1}$$

By the inverse property for Jacobians,

$$|J(G \rightarrow T)| = \prod_{k=1}^{p} A_{kk}^{-2k+1}$$

□

C.4 Jacobians Requiring Exterior Product Approach

It is true that only a few of the Jacobians to come will be derived using the exterior product. However, those to immediately follow are very important

to our handling matrix quadratic forms, such as the complex Wishart matrix. We begin slowly with a important case. This is worth following closely as an example of the power of using an exterior product approach for a nonlinear change of variables.

Theorem 25 *Let T be an upper triangular complex matrix of size $p \times p$ with positive real elements on the diagonal. Let $B = TT^H$. Then*

$$|J(B \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(k-1)+1}$$

and

$$|J(T \to B)| = 2^{-p} \prod_{k=1}^{p} T_{kk}^{-2(k-1)-1}$$

This is a complexification of a variation of Muirhead theorem 2.1.9 (p. 60) [187].

Proof. This is a complexification and slight expansion of Muirhead's proof. We begin by looking at the matrices.

$$B = TT^H = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{12}^* & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1p}^* & B_{2p}^* & \cdots & B_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11} & T_{12} & T_{13} & \cdots & T_{1p} \\ & T_{22} & T_{23} & \cdots & T_{2p} \\ & & T_{33} & \cdots & T_{3p} \\ & & & \ddots & \vdots \\ & & & & T_{pp} \end{pmatrix} \begin{pmatrix} T_{11} & & & & \\ T_{12}^* & T_{22} & & & \\ T_{13}^* & T_{23}^* & T_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ T_{1p}^* & T_{2p}^* & T_{3p}^* & \cdots & T_{pp} \end{pmatrix}$$

Note that the diagonal terms of T^H do not have the asterisk since $T_{ii}^* = T_{ii}$ because $T_{ii} \in \mathbf{R}$. Columns of the matrix TT^H are given below to make the pattern of entries clear. Columns 1 and 2 are:

$$\begin{matrix} T_{11}^2 + \sum_{j=2}^{p} |T_{1j}|^2 \\ T_{22}T_{12}^* + \sum_{j=3}^{p} T_{2j}T_{1j}^* \\ T_{33}T_{13}^* + \sum_{j=4}^{p} T_{3j}T_{1j}^* \\ T_{44}T_{14}^* + \sum_{j=5}^{p} T_{4j}T_{1j}^* \\ \vdots \\ T_{(p-1),(p-1)}T_{1,(p-1)}^* + T_{(p-1),p}T_{1,p}^* \\ T_{pp}T_{1p}^* \end{matrix} \qquad \begin{matrix} T_{12}T_{22} + \sum_{j=3}^{p} T_{1j}T_{2j}^* \\ T_{22}^2 + \sum_{j=3}^{p} |T_{2j}|^2 \\ T_{33}T_{23}^* + \sum_{j=4}^{p} T_{3j}T_{2j}^* \\ T_{44}T_{24}^* + \sum_{j=5}^{p} T_{4j}T_{2j}^* \\ \vdots \\ T_{(p-1),(p-1)}T_{2,(p-1)}^* + T_{(p-1),p}T_{2,p}^* \\ T_{pp}T_{2p}^* \end{matrix}$$

Columns 3 and 4 are:

$$\begin{array}{ll}
T_{13}T_{33} + \sum_{j=4}^{p} T_{1j}T_{3j}^{*} & T_{14}T_{44} + \sum_{j=5}^{p} T_{1j}T_{4j}^{*} \\
T_{23}T_{33} + \sum_{j=4}^{p} T_{2j}T_{3j}^{*} & T_{24}T_{44} + \sum_{j=5}^{p} T_{2j}T_{4j}^{*} \\
T_{33}^{2} + \sum_{j=4}^{p} |T_{3j}|^{2} & T_{34}T_{44} + \sum_{j=5}^{p} T_{3j}T_{4j}^{*} \\
T_{44}T_{34}^{*} + \sum_{j=5}^{p} T_{4j}T_{3j}^{*} & T_{44}^{2} + \sum_{j=5}^{p} |T_{4j}|^{2} \\
\vdots & \vdots \\
T_{(p-1),(p-1)}T_{3,(p-1)}^{*} + T_{(p-1),p}T_{3,p}^{*} & T_{(p-1),(p-1)}T_{4,(p-1)}^{*} + T_{(p-1),p}T_{4,p}^{*} \\
T_{pp}T_{3p}^{*} & T_{pp}T_{4p}^{*}
\end{array}$$

The last two columns are:

$$\begin{array}{ll}
T_{1,(p-1)}T_{(p-1),(p-1)} + T_{1,p}T_{(p-1),p}^{*} & T_{1p}T_{pp} \\
T_{2,(p-1)}T_{(p-1),(p-1)} + T_{2,p}T_{(p-1),p}^{*} & T_{2p}T_{pp} \\
T_{3,(p-1)}T_{(p-1),(p-1)} + T_{3,p}T_{(p-1),p}^{*} & T_{3p}T_{pp} \\
T_{4,(p-1)}T_{(p-1),(p-1)} + T_{4,p}T_{(p-1),p}^{*} & T_{4p}T_{pp} \\
\vdots & \vdots \\
T_{(p-1),(p-1)}^{2} + \left|T_{(p-1),p}\right|^{2} & T_{(p-1),p}T_{pp} \\
T_{pp}T_{(p-1),p}^{*} & T_{pp}^{2}
\end{array}$$

To find $|dB|$ we take the exterior product of the differentials. We begin with the term $T_{pp}^{2} = B_{pp}$ and work backwards through the array. This tactic simplifies the algebra in the following way. Once a term dT_{ij} is computed, it never needs to be computed again. In forming the overall exterior product, repeated differentials cause that product term to be zero; $dT_{ij} \bigwedge dT_{ij} = 0$. Our

next step is to form the differentials needed.

$$B_{pp} = T_{pp}^2 \qquad dB_{pp} = 2T_{pp}dT_{pp}$$

$$B_{(p-1),p} = T_{(p-1),p}T_{pp} = [T_{R(p-1,p)} + iT_{I(p-1,p)}]T_{pp}$$

$$B_{R(p-1,p)} = T_{R(p-1,p)}T_{pp} \qquad dB_{R(p-1.p)} = T_{pp}dT_{R(p-1,p)} + \cdots$$

$$B_{I(p-1,p)} = T_{I(p-1,p)}T_{pp} \qquad dB_{I(p-1.p)} = T_{pp}dT_{I(p-1,p)} + \cdots$$

$$\vdots \qquad\qquad \vdots$$

$$B_{R(1,p)} = T_{R(1,p)}T_{pp} \qquad dB_{R(1,p)} = T_{pp}dT_{R(1,p)} + \cdots$$

$$B_{I(1,p)} = T_{I(1,p)}T_{pp} \qquad dB_{I(1,p)} = T_{pp}dT_{I(1,p)} + \cdots$$

$$B_{(p-1,p-1)} = T_{(p-1,p-1)}^2 + T_{(p-1,p)}T_{(p-1,p)}^* \qquad dB_{(p-1,p-1)} = 2T_{(p-1,p-1)}dT_{(p-1,p-1)} + \cdots$$

$$B_{(1,p-1)} = T_{(1,p-1)}T_{(p-1,p-1)} + T_{1p}T_{(p-1,p)}$$

$$B_{R(1,p-1)} = T_{R(1,p-1)}T_{(p-1,p-1)} + \cdots \qquad dB_{R(1,p-1)} = T_{(p-1,p-1)}dT_{R(1,p-1)} + \cdots$$

$$B_{I(1,p-1)} = T_{I(1,p-1)}T_{(p-1,p-1)} + \cdots \qquad dB_{I(1,p-1)} = T_{(p-1,p-1)}dT_{I(1,p-1)} + \cdots$$

$$\vdots$$

$$B_{44} = T_{44}^2 + \sum_{j=5}^{p} T_{4j}T_{4j}^* \qquad dB_{44} = 2T_{44}dT_{44} + \cdots$$

$$B_{34} = T_{34}T_{44} + \cdots \qquad dB_{R34} = T_{44}dT_{R34} + \cdots$$

$$dB_{I34} = T_{44}dT_{I34} + \cdots$$

$$\vdots$$

$$B_{11} = T_{11}^2 + \cdots \qquad dB_{11} = 2T_{11}dT_{11} + \cdots$$

We examine $dB_{R(p-1,p)}$ as an example of the reduction in terms achieved due to recurrence of differentials in different terms.

$$dB_{R(p-1,p)} = (dT_{R(p-1,p)})T_{pp} + T_{R(p-1,p)}dT_{pp}$$

When the product $dB_{pp} \bigwedge dB_{R(p-1,p)}$ is formed we get

$$dB_{pp} \bigwedge dB_{R(p-1,p)} = 2T_{pp}dT_{pp} \bigwedge [T_{pp}dT_{R(p-1,p)} + T_{R(p-1,p)}dT_{pp}]$$

$$= 2T_{pp}^2 dT_{pp} \bigwedge dT_{R(p-1,p)} + 2T_{pp}T_{R(p-1,p)} \underbrace{dT_{pp} \bigwedge dT_{pp}}_{0} = 2T_{pp}^2 dT_{pp} \bigwedge dT_{R(p-1,p)}$$

Therefore, to simplify algebra, we need only to keep track of dB_{ij} terms which have not already appeared in our sequence of computations. Also note that if both terms in a product have appeared in an earlier computation, they do not need to be considered again. This is because their differentials will be multiplied by the same differentials from earlier computations, yielding zero. This is first seen in our computation of $dB_{R(1,p-1)}$.

$$B_{R(1,p-1)} = T_{R(1,p-1)}T_{(p-1,p-1)} + T_{R(1,p)}T_{R(p-1,p)} - T_{I(1,p)}T_{I(p-1,p)}$$

$dT_{R(1,p)}$ first appears in $dB_{R(1,p)}$

$dT_{R(p-1,p)}$ first appears in $dB_{R(p-1,p)}$

$dT_{I(1,p)}$ first appears in $dB_{I(1,p)}$

$dT_{I(p-1,p)}$ first appears in $dB_{I(p-1,p)}$

Note that for an Hermitian matrix $B = B^H$, we need only to look at the superdiagonal elements. This is because in doing so, we generate terms involving the differentials of all real and all imaginary components. For example, consider $B^*_{(p-1,p)}$.

$$B^*_{(p-1,p)} = T^*_{pp}T^*_{(p-1,p)} = T_{pp}[T_{R(p-1,p)} - iT_{I(p-1,p)}]$$

$$B^*_{R(p-1,p)} = T_{pp}T_{R(p-1,p)} \quad dB^*_{R(p-1,p)} = T_{pp}dT_{R(p-1,p)} + \cdots$$

$$B^*_{I(p-1,p)} = -T_{pp}T_{I(p-1,p)} \quad dB^*_{I(p-1,p)} = -T_{pp}dT_{I(p-1,p)} + \cdots$$

Recall that we already have

$$dB_{R(p-1,p)} = T_{pp}dT_{R(p-1,p)} + \cdots$$

$$dB_{I(p-1,p)} = T_{pp}dT_{I(p-1,p)} + \cdots$$

Thus, when the exterior product is taken, terms containing both dB_{ij} and dB^*_{ij} will be zero.

Observe that each dB_{kk} term contains a factor 2. Thus $\bigwedge_{k=1}^{p} dB_{kk} = 2^p T_{kk}dT_{kk}$. The off-diagonal terms need a bit more care because the differentials for both the real and imaginary parts must be considered. Note that each differential in a given column k has the factor T_{kk}. We therefore can think of taking the product of all the terms in the matrix where T^2_{kk} is the result of $dB_{R(j,k)} \bigwedge dB_{I(j,k)}$.

$$\begin{pmatrix} 2T_{11} & T^2_{22} & \cdots & T^2_{pp} \\ & 2T_{22} & \cdots & T^2_{pp} \\ & & \ddots & \vdots \\ & & & 2T_{pp} \end{pmatrix}$$

Thus $|J(B \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(k-1)+1}$.□

Theorem 26 *Let T be a lower triangular complex matrix of size $p \times p$ with positive real elements on the diagonal. Let $A = TT^H$. Then $|J(A \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(p-k)+1}$ and $|J(T \to A)| = 2^{-p} \prod_{k=1}^{p} T_{kk}^{-2(p-k)-1}$. This is a complexification of Deemer and Olkin theorem 4.1 [67]. This is used by Khatri [137] in his*

proof of theorem 2.8 just before equation 2.8.1. This is Srivastava and Khatri problem 1.29 [257].

Proof. This proof follows the model set for theorem 25 by Muirhead rather than complexifying Deemer and Olkin's proof.

$$A = TT^H = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{12}^* & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1p}^* & A_{2p}^* & \cdots & A_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11} & & & & & \\ T_{21} & T_{22} & & & & \\ T_{31} & T_{32} & T_{33} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ T_{p-1,p} & T_{p-1,2} & T_{p-1,3} & \cdots & T_{p-1,p-1} & \\ T_{p1} & T_{p2} & T_{p3} & \cdots & T_{p,p-1} & T_{pp} \end{pmatrix} \times$$

$$\times \begin{pmatrix} T_{11} & T_{21}^* & T_{31}^* & \cdots & T_{p-1,1}^* & T_{p1}^* \\ & T_{22} & T_{32}^* & \cdots & T_{p-1,2}^* & T_{p2}^* \\ & & T_{33} & \cdots & T_{p-1,3}^* & T_{p3}^* \\ & & & \ddots & \vdots & \vdots \\ & & & & T_{p-1,p-1} & T_{p,p-1}^* \\ & & & & & T_{pp} \end{pmatrix}$$

The columns of TT^H are given below. The first three columns are:

$$\begin{array}{ccc}
T_{11}^2 & T_{11}T_{21}^* & T_{11}T_{31}^* \\
T_{21}T_{11} & T_{21}T_{21}^* + T_{22}^2 & T_{21}T_{31}^* + T_{22}T_{32}^* \\
T_{31}T_{11} & T_{31}T_{21}^* + T_{32}T_{22} & T_{31}T_{31}^* + T_{32}T_{32}^* + T_{33}^2 \\
\vdots & \vdots & \vdots \\
T_{p-1,1}T_{11} & T_{p-1,1}T_{21}^* + T_{p-1,2}T_{22} & T_{p-1,1}T_{31}^* + T_{p-1,2}T_{32}^* + T_{p-1,3}T_{33} \\
T_{p1}T_{11} & T_{p1}T_{21}^* + T_{p2}T_{22} & T_{p,1}T_{31}^* + T_{p,2}T_{32}^* + T_{p,3}T_{33}
\end{array}$$

The next to the last column (p-1) is:

$$T_{11}T_{p-1,1}^*$$

$$T_{21}T_{p-1,1}^* + T_{22}T_{p-1,2}^*$$

$$T_{31}T_{p-1,1}^* + T_{32}T_{p-1,2}^* + T_{33}T_{p-1,3}^*$$

$$\vdots$$

$$T_{p-1,1}T_{p-1,1}^* + T_{p-1,2}T_{p-1,2}^* + T_{p-1,p-1}^2$$

$$T_{p,1}T_{p-1,1}^* + T_{p,2}T_{p-1,2}^* + \cdots + T_{p,p-1}T_{p-1,p-1}$$

The final column (p) is:

$$T_{11}T_{p1}^*$$

$$T_{21}T_{p1}^* + T_{22}T_{p2}^*$$

$$T_{31}T_{p1}^* + T_{32}T_{p2}^* + T_{33}T_{p3}^*$$

$$\vdots$$

$$T_{p-1,1}T_{p1}^* + T_{p-1,2}T_{p2}^* + \cdots + T_{p-1,p-1}T_{p,p-1}^*$$

$$T_{p1}T_{p1}^* + T_{p2}T_{p2}^* + \cdots + T_{pp}^2$$

From this matrix we compute our differentials.

$$\begin{aligned}
A_{11} &= T_{11}^2 & dA_{11} &= 2T_{11}dT_{11} \\
A_{21} &= T_{21}T_{11} & & \\
A_{R21} &= T_{R21}T_{11} & dA_{R21} &= T_{11}dT_{R21} + \cdots \\
A_{I21} &= T_{I21}T_{11} & dA_{I21} &= T_{11}dT_{I21} + \cdots \\
&\vdots & &\vdots \\
A_{Rp1} &= T_{Rp1}T_{11} & dA_{Rp1} &= T_{11}dT_{Rp1} + \cdots \\
A_{Ip1} &= T_{Ip1}T_{11} & dA_{Ip1} &= T_{11}dT_{Ip1} + \cdots \\
A_{22} &= T_{21}[T_{R21} - iT_{I21}] + T_{22}^2 & dA_{22} &= 2T_{22}dT_{22} + \cdots \\
&\vdots & &\vdots
\end{aligned}$$

Observing the pattern and recalling the patterns generated for the upper triangular case, we see that for the lower triangular case that we get

$$|J(A \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(p-k)+1}$$

$$|J(T \to A)| = 2^{-p} \prod_{k=1}^{p} T_{kk}^{-2(p-k)-1}$$

□

Theorem 27 *Let T be an upper triangular complex matrix of size $p \times p$ with positive real elements on the diagonal. Let $A = T^H T$. Then*

$$|J(A \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(p-k)+1}$$

and

$$|J(T \to A)| = 2^{-p} \prod_{k=1}^{p} T_{kk}^{-2(p-k)-1}$$

This is a complexification of Muirhead's theorem 2.1.9 [187]. This is also Goodman equation 5.25 [92].

Proof. This is a complexification of Muirhead's proof.

$$A = T^H T = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{12}^* & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1p}^* & A_{2p}^* & \cdots & A_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11} & & & \\ T_{12}^* & T_{22} & & \\ \vdots & \vdots & \ddots & \\ T_{1p}^* & T_{2p}^* & \cdots & T_{pp} \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1p} \\ & T_{22} & \cdots & T_{2p} \\ & & \ddots & \vdots \\ & & & T_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} T_{11}^* & T_{11}T_{12} & \cdots & T_{11}T_{1p} \\ T_{12}^* T_{11} & T_{12}^* T_{12} + T_{22}^2 & \cdots & T_{12}^* T_{1p} + T_{22} T_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ T_{1p}^* T_{11} & T_{1p}^* T_{12} + T_{2p}^* T_{22} & \cdots & T_{1p}^* T_{1p} + T_{2p}^* T_{2p} + T_{2p}^* T_{2p} + \cdots T_{pp}^2 \end{pmatrix}$$

Note the similarity of this with the A of lemma 26. Instead of forming the exterior product from the lower triangular terms of A, use the upper triangle

of A.

$$\begin{aligned}
A_{11} &= T_{11}^2 & dA_{11} &= 2T_{11}dT_{11} \\
A_{12} &= T_{11}T_{12} & & \\
A_{R12} &= T_{11}T_{R12} & dA_{R12} &= T_{11}dT_{R12} + \cdots \\
A_{I12} &= T_{11}T_{I12} & dA_{I12} &= T_{11}dT_{I12} + \cdots \\
&\vdots & &\vdots \\
A_{R1p} &= T_{11}T_{R1p} & dA_{R1p} &= T_{11}dT_{R1p} + \cdots \\
A_{Ip1} &= T_{11}T_{I1p} & dA_{I1p} &= T_{11}dT_{I1p} + \cdots \\
A_{22} &= T_{12}^*(T_{R12} + iT_{I12}) + T_{22}^2 & dA_{22} &= 2T_{22}dT_{22} + \cdots \\
&\vdots & &\vdots
\end{aligned}$$

Thus

$$|J(A \to T)| = 2^p \prod_{k=1}^{p} T_{kk}^{2(p-k)+1}$$

$$|J(T \to A)| = 2^{-p} \prod_{k=1}^{p} T_{kk}^{-2(p-k)-1}$$

Theorem 28 *Let $Y \in M_p(\mathbf{C})$ and $A = Y^H Y$. Then*

$$|J(Y \to A)| = \frac{\pi^{p^2}}{\mathbf{C}\Gamma_p(p)}$$

and

$$|J(A \to Y)| = \frac{\mathbf{C}\Gamma_p(p)}{\pi^{p^2}}$$

Proof. This lemma depends on integrals that are proven under the section of helpful integrals. From theorem 150 with $\Sigma = I$ we have

$$\int_{A>0} \operatorname{etr}(-A)(\det A)^{a-p}(dA) = \mathbf{C}\Gamma_p(a)$$

From proposition 105,

$$\int_{M_p(\mathbf{C})} \text{etr}(-Y^H Y) \det(Y^H Y)^{a-p} (dY) = \frac{\pi^{p^2} \mathbf{C}\Gamma_p(a)}{\mathbf{C}\Gamma_p(p)}$$

Let $A = Y^H Y$. Then

$$\frac{\mathbf{C}\Gamma_p(p)}{\pi^{p^2}} \int_{M_p(\mathbf{C})} \text{etr}(-Y^H Y)(\det(Y^H Y)^{a-p} (dY) = \mathbf{C}\Gamma_p(a)$$

Therefore $|J(A \rightarrow Y)| = \frac{\mathbf{C}\Gamma_p(p)}{\pi^{p^2}}$.

Theorem 29 *If the density of $Y \in \mathbf{C}^{p \times m}$ is a function of YY^H, $f(YY^H)(dY)$. then the density of $B = YY^H$ is given by*

$$g(B) = \frac{|\det B|^{mp} f(B)}{\pi^{-pm} \mathbf{C}\Gamma_p(m)} (dB)$$

The Jacobian of the transformation is

$$|J(Y \rightarrow B)| = \frac{\pi^{pm} |\det B|^{mp}}{\mathbf{C}\Gamma_p(m)}$$

This is theorem 1 of [256].

Proof. This is proven in theorem 67 which is a replication of Srivastava's derivation of the complex Wishart density. This theorem is stated here to keep it in context with similar theorems. This is a complexification of Anderson lemma 13.3.1 [26]. $\mathbf{C}\Gamma_p(m)$ is defined in the section on helpful integrals.

Theorem 30 *Let $x = By$ be a linear change of variables from complex vector $x \in \mathbf{C}^n$ to complex vector y where B is in $\mathbf{C}^{n \times n}$. Let $\text{Re}(B) > 0$. Then*

$$|J(x \rightarrow y)| = |\det B|^2$$

Proof. I do not have a record of the pedigree of this result or its proof. I presume that is known already to many people.

$$x = By = (x_R + ix_I) = (B_R + iB_I)(y_R + iy_I)$$

$$= (B_R y_R - B_I y_I) + i(B_I y_R + B_R y_I)$$

Thus, this is a change of variables of the form

$$\begin{pmatrix} x_R \\ x_I \end{pmatrix} = \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix} \begin{pmatrix} y_R \\ y_I \end{pmatrix}$$

The Jacobian is

$$|J(x \to y)| = \left| \det \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix} \right| = \left| \det(B_R) \det(B_R + B_I B_R^{-1} B_I) \right|$$

$$= \left| [\det(B_R)]^2 \det(I + B_R^{-1} B_I B_R^{-1} B_I) \right| = \left| [\det(B_R)]^2 \det[I + (B_R^{-1} B_I)^2] \right|$$

$$= \left| [\det(B_R)]^2 \det[(I + iB_R^{-1} B_I)(I - iB_R^{-1} B_I)] \right| = \left| [\det(B_R)]^2 \right| \cdot \left| \det(I + iB_R^{-1} B_I) \right|^2$$

$$= \left| \det(B_R) \det(I + iB_R^{-1} B_I) \right|^2 = |\det(B_R + iB_I)|^2 = |\det B|^2$$

By the inverse property, $|J(y \to x)| = |\det B|^{-2}$, when it exists. When $B = B^H$ then $|\det B| = \det B$ because the eigenvalues of B are real and $\det B = \prod_{i=1}^{n} \lambda_i^2$, by Goodman corollary 2.1 proof [92]. □

Theorem 31 *Let $X = BY$ be a complex linear transformation between the variables $X \in \mathbf{C}^{n \times m}$, $Y \in \mathbf{C}^{n \times m}$, and $B \in \mathbf{C}^{n \times n}$. Then $|J(X \to Y)| = |\det B|^{2m}$ and $|J(Y \to X)| = |\det B|^{-2m}$. This is a complexification of Muirhead [187] theorem 2.1.4.*

Proof. Muirhead provides a proof for the real case which uses exterior products. I have provided a more traditional proof. We know that

$$|J(X \to Y)| = |J(X_1 \to Y_1)J(X_2 \to Y_2)\cdots J(X_m \to Y_m)|$$

By theorem 30, $|J(X_k \to Y_k)| = |\det B|^2$. Thus $|J(X \to Y)| = |\det B|^{2m}$ and $|J(Y \to X)| = |\det B|^{-2m}$. □

Theorem 32 *Let $Z \in \mathbf{C}^{n\times m}$ where the rank of Z is m. Let $Z = H_1T$ where $H_1^H H_1 = I_m$ and T is an $m \times m$ upper triangular matrix with positive diagonal elements. Let H_2 (a function of H_1) be an $n \times (n-m)$ matrix such that $H = [H_1, H_2]$ is an Hermitian $n \times n$ matrix. In column vector notation, let*

$$H = [h_1, \cdots, h_m, h_{m+1}, \cdots, h_n]$$

where $\{h_i\}_1^m$ belong to H_1. Then

$$(dZ) = \left[\prod_{i=1}^{m} T_{ii}^{2n-m-i}\right](d\tau_{H_1})(H^H dH_1)$$

where

$$(H^H dH_1) = \bigwedge_{i=1}^{m}\bigwedge_{j=i+1}^{n}(h_j^H dh_i)$$

and

$$(d\tau_{H_1}) = \left\{\bigwedge_{j=1}^{m}\left[\left(h_j^H dh_j\right)T_{jj} + dT_{jj}\right]\right\}\bigwedge_{i<j}^{m} dT_{ij}$$

and $\bigwedge$ here is an exterior product operator. This is a complexification of Muirhead's theorem 2.1.13 [187].

Proof. The following is a complexification and expansion of Muirhead's proof. First, recall by C. R. Rao 1b.2(ix) [213] that Z can indeed be decomposed into H_1T where $H_1^H H_1 = I_m$, and T is upper triangular with positive real diagonal elements. Also, given a subunitary matrix such as H_1, we can always find a completion to H_1 to form unitary matrix H.

Our goal is to find (dZ) in terms of (dT) and (dH_1). What we will actually get is something almost what we want, and this thing we get turns out to meet our needs. We start by invoking theorem 21, $Z = H_1T$ implies that

$$dZ = (dH_1)T + H_1 dT$$

Consider $H^H dZ$, which involves our completed unitary matrix H.

$$H^H dZ = \begin{bmatrix} H_1^H \\ H_2^H \end{bmatrix} dZ = \begin{bmatrix} H_1^H(dH_1)T + H_1^H H_1 dT \\ H_2^H(dH_1)T + H_2^H H_1 dT \end{bmatrix} = \begin{bmatrix} H_1^H(dH_1)T + dT \\ H_2^H dH_1 T \end{bmatrix}$$

Note that dH_1 here is a matrix, not an exterior product. Also note by H being unitary that $H_1^H H_1 = I_m$ and $H_2^H H_1 = 0$, the zero matrix.

By theorem 31, the exterior product of $H^H dZ$ is

$$(H^H dZ) = \left|\det H^H\right|^{2m} (dZ) = (dZ)$$

So, if we find $(H^H dZ)$, we have also then found (dZ). Now, let us evaluate $(H^H dZ)$. To do this with the least effort, we will make use of the partition and a special property of the upper partition. The lower partition is simplest, so we begin there.

$$H_2^H(dH_1)T = (h_{m+1}, h_{m+2}, \cdots, h_n)^H (dh_1, dh_2, \cdots, dh_m)\, T$$

$$= \begin{pmatrix} h_{m+1}^H dh_1 & h_{m+1}^H dh_2 & \cdots & h_{m+1}^H dh_m \\ h_{m+2}^H dh_1 & h_{m+2}^H dh_2 & \cdots & h_{m+2}^H dh_m \\ \vdots & \vdots & \ddots & \vdots \\ h_n^H dh_1 & h_n^H dh_2 & \cdots & h_n^H dh_m \end{pmatrix}_{(n-m)\times m} \times T_{m\times m}$$

Recalling that for any matrix A, $\det A = \det A^T$, and applying theorem 30 to row j of $H_2^H(dH_1)T$ we obtain $|\det T|^2 \bigwedge_{i=1}^{m} h_j^H dh_i$. Thus the exterior product of all elements in $H_2^H(dH_1)T$ is given by

$$\bigwedge_{j=m+1}^{n} \left[|\det T|^2 \bigwedge_{i=1}^{m} h_j^H dh_i \right] = |\det T|^{2(n-m)} \bigwedge_{j=m+1}^{n} \bigwedge_{i=1}^{m} h_j^H dh_i$$

This also follows from lemma 4.

Now, consider the upper partition, $H_1^H(dH_1)T + dT$. T is upper triangular, and so is dT. Thus, the lower triangle (below the diagonal) consists only of the elements of $H_1^H(dH_1)T$. Recall that $H_1^H H_1 = I_m$. Thus

$$d(H_1^H H_1) = d(I_m) = 0 = [dH_1^H]H_1 + H_1^H[dH_1]$$

which is the zero matrix. This means that

$$H_1^H dH_1 = -[dH_1^H]H_1 = -[H_1^H dH_1]^H$$

Therefore $H_1^H(dH_1)$ is skew-Hermitian.

$$H_1^H dH_1 = \begin{pmatrix} i\,\mathrm{Im}(h_1^H dh_1) & -[h_2^H dh_1]^H & \cdots & -[h_m^H dh_1]^H \\ h_2^H dh_1 & i\,\mathrm{Im}(h_2^H dh_2) & \cdots & -[h_m^H dh_2]^H \\ h_3^H dh_1 & h_3^H dh_2 & \cdots & -[h_m^H dh_3]^H \\ \vdots & \vdots & & \vdots \\ h_m^H dh_1 & h_m^H dh_2 & \cdots & i\,\mathrm{Im}(h_m^H dh_m) \end{pmatrix}_{m\times m}$$

Note that for the case of real variables, the main diagonal consists of all zeros. Now evaluate $H_1^H(dH_1)T$, recalling that T is upper triangular. This matrix is given below by columns. The first two columns are:

$$\begin{array}{ll} i\,\mathrm{Im}(h_1^H dh_1)T_{11} & i\,\mathrm{Im}(h_1^H dh_1)T_{12} - [h_2^H dh_1]^H T_{22} \\ (h_2^H dh_1)T_{11} & (h_2^H dh_1)T_{12} + i\,\mathrm{Im}(h_2^H dh_2)T_{22} \\ (h_3^H dh_1)T_{11} & (h_3^H dh_1)T_{12} + (h_3^H dh_2)T_{22} \\ \vdots & \vdots \\ (h_m^H dh_1)T_{11} & (h_m^H dh_1)T_{12} + (h_m^H dh_2)T_{22} \end{array}$$

The last column is:

$$\begin{array}{l} i\,\mathrm{Im}(h_1^H dh_1)T_{1m} - \sum_{k=2}^{m} [h_k^H dh_1]^H T_{km} \\ (h_2^H dh_1)T_{1m} + i\,\mathrm{Im}(h_2^H dh_2)T_{2m} - \sum_{k=3}^{m} [h_k^H dh_2]^H T_{km} \\ \sum_{k=1}^{3} (h_3^H dh_k)T_{km} + i\,\mathrm{Im}(h_3^H dh_3)T_{3m} - \sum_{k=4}^{m} [h_k^H dh_3]^H T_{km} \\ \vdots \\ \sum_{k=1}^{m-1} (h_m^H dh_k)T_{km} + i\,\mathrm{Im}(h_m^H dh_m)T_{mm} \end{array}$$

In forming exterior products, once a term $(h_j^H dh_i)$ has been included, repeated terms with the same index will cause that particular product to be

zero. $(h_j^H dh_i) \bigwedge (h_j^H dh_i) = 0$. Also notice that $dZ \bigwedge (dZ)^* = 0$. The exterior product of elements *below* the main diagonal of $H_1^H (dH_1) T$ is also the exterior product of elements below the main diagonal of $H_1^H (dH_1) T + dT$. This is

$$\left[T_{11}^{m-1} \bigwedge_{j=2}^{m} (h_j^H dh_i) \right] \left[T_{22}^{m-2} \bigwedge_{j=3}^{m} (h_j^H dh_i) \right] \cdots \left[T_{m-1,m-1} \bigwedge (h_m^H dh_{m-1}) \right]$$

$$= T_{11}^{m-1} T_{22}^{m-2} \cdots T_{m-1,m-1} \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} (h_j^H dh_i)$$

Elements *along* the main diagonal of $H_1^H (dH_1) T + dT$ have an exterior product that is a bit more tedious. Since a typical element looks like

$$i \operatorname{Im}(h_j^H dh_j) T_{jj} + dT_{jj} = (h_j^H dh_j) T_{jj} + dT_{jj}$$

the exterior product is

$$\bigwedge_{j=1}^{m} \left[(h_j^H dh_j) T_{jj} + dT_{jj} \right]$$

When this is expanded, it has one term that is $\bigwedge_{j=1}^{m} dT_{jj}$, and another of the form $\left(\prod_{j=1}^{m} T_{jj} \right) \bigwedge_{j=1}^{m} (h_j^H dh_j)$. We do not get the simple form achieved when $h_j^T dh_j = 0$ in the case of real variables.

Elements *above* the main diagonal have an inner product of the form $\bigwedge_{i<j}^{m} dt_{ij} = \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dt_{ij}$. Putting this all together, we obtain

$$\left[|\det T|^{2(n-m)} \bigwedge_{i=m+1}^{n} \bigwedge_{i=1}^{m} (h_j^H dh_i) \right] \left[\prod_{k=1}^{m} T_{kk}^{m-k} \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} (h_j^H dh_i) \right] \times$$

$$\times \left\{ \bigwedge_{j=1}^{m} \left[(h_j^H dh_j) T_{jj} + dT_{jj} \right] \right\} \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dT_{ij}$$

$$= \left[\prod_{k=1}^{m} T_{kk}^{2n-m-k} \right] \left[\bigwedge_{i=1}^{m} \bigwedge_{j=i+1}^{n} (h_j^H dh_i) \right] \left\{ \bigwedge_{j=1}^{m} \left[(h_j^H dh_j) T_{jj} + dT_{jj} \right] \right\} \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dT_{ij}$$

where we ignore sign changes, and we recall $T_{kk} > 0$,

$$= \left[\prod_{k=1}^{m} T_{kk}^{2n-m-k}\right] (d\tau_{H_1})(H^H dH_1)$$

where

$$(d\tau_{H_1}) = \left\{\bigwedge_{j=1}^{m} \left[(h_j^H dh_j)T_{jj} + dT_{jj}\right]\right\} \bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dT_{ij}$$

and

$$(H^H dH_1) = \bigwedge_{i=1}^{m} \bigwedge_{j=i+1}^{n} (h_j^H dh_i)$$

□

Theorem 33 *Let $Z \in \mathbf{C}^{n\times m}$ where the rank of Z is m. Let $Z = H_1 T$ where $H_1^H H_1 = I_m$ and T is an $m \times m$ upper triangular matrix with positive diagonal elements. Let H_2 (a function of H_1) be an $n \times (n-m)$ matrix such that $H = [H_1, H_2]$ is an Hermitian $n \times n$ matrix. In column vector notation, let*

$$H = [h_1, h_2, \cdots, h_m, h_{m+1}, \cdots, h_n]$$

where $\{h_i\}_1^m$ belong to H_1. Let $A = Z^H Z$. Then

$$(dZ) = (\det A)^{(2n-3m)/2} \left\{\bigwedge_{j=1}^{m} [(h_j^H dh_j)T_{jj} + dT_{jj}]\right\} (dA_L)(H^H dH_1) \prod_{k=1}^{m} T_{kk}^k$$

where (dA_L) is the exterior product of elements of A below the main diagonal. This is a complexification of Muirhead theorem 2.1.14.

Proof. This is a complexification and expansion of Muirhead's proof. From theorem 32,

$$(dZ) = \left[\prod_{k=1}^{m} T_{kk}^{2n-m-k}\right] \left\{\bigwedge_{j=1}^{m} [(H^H dH_1)T_{jj} + dT_{jj}]\right\} \left[\bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dT_{ij}\right] (H^H dH_1)$$

Also,

$$A = Z^H Z = (H_1 T)^H (H_1 T) = T^H H_1^H H_1 T = T^H I_m T = T^H T$$

From theorem 27,

$$(dA) = 2^m \prod_{k=1}^{m} T_{kk}^{2(m-k)+1} (dT)$$

Note that $A = A^H$. Thus the exterior product of elements of A consist of the exterior product of the lower triangular submatrix of A. Partitioning that exterior product into diagonal and below diagonal elements, we get $(dA) = (dA_D)(dA_L)$

$$= \left[\underbrace{\left(2^m \prod_{k=1}^{m} T_{kk} \right) \left(\bigwedge_{k=1}^{m} dT_{kk} \right)}_{(dA_D)} \right] \left[\underbrace{\left(\prod_{k=1}^{m} T_{kk}^{2(m-k)} \right) \left(\bigwedge_{i=1}^{m-1} \bigwedge_{j=i+1}^{m} dT_{ij} \right)}_{(dA_L)} \right]$$

We will substitute (dA_D) into (dZ). We get $(dZ) =$

$$\left[\prod_{k=1}^{m} T_{kk}^{2n-3m+k} \right] \left\{ \bigwedge_{j=1}^{m} [(h_j^H dh_j) T_{jj} + dT_{jj}] \right\} \left[\prod_{k=1}^{m} T_{kk}^{2(m-k)} \right] \left(\bigwedge_{i<j}^{m} dT_{ij} \right) (H^H dH_1)$$

$$= \left[\prod_{k=1}^{m} T_{kk}^{2n-3m+k} \right] \left\{ \bigwedge_{j=1}^{m} [(h_j^H dh_j) T_{jj} + dT_{jj}] \right\} (dA_L)(H^H dH_1)$$

$$= (\det A)^{(2n-3m)/2} \left[\prod_{k=1}^{m} T_{kk}^{k} \right] \left\{ \bigwedge_{j=1}^{m} [(h_j^H dh_j) T_{jj} + dT_{jj}] \right\} (dA_L)(H^H dH_1)$$

We do not get the nice form given in Muirhead [187] for the case of real variables because the diagonal elements of $H_1^H (dH_1)$ are not zero. Elements of $h_i^H dh_i$ are purely imaginary. $H_1^H (dH_1)$ is skew-Hermitian. □

Theorem 34 *Let $X = BYC$ be a complex change of variables where $X \in \mathbf{C}^{n\times m}$, $Y \in \mathbf{C}^{n\times m}$, $B \in \mathbf{C}^{n\times n}$, and $C \in \mathbf{C}^{m\times m}$. Then*

$$|J(X \to Y)| = |\det B|^{2m} |\det C|^{2n}$$

This is a complexification of Muirhead theorem 2.1.5, Deemer and Olkin [67] theorem 3.6, and Arnold [31] Theorem A.16. This is also Khatri's theorem 2.3 [137].

Proof. This is a complexification of Deemer and Olkin's proof. Let $Z = BY$ and $X = ZC$. Then $|J(Z \to Y)| = |\det B|^{2m}$ and $|J(X \to Z)| = |\det C|^{2n}$ by theorem 31. Then

$$|J(X \to Y)| = |J(Z \to Y)| \cdot |J(X \to Z)|$$

This implies

$$|J(X \to Y)| = |\det B|^{2m} |\det C|^{2n}$$

and

$$|J(Y \to X)| = |\det B|^{-2m} |\det C|^{-2n}$$

□

Theorem 35 *Let $X = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Y \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}^T$ be a transformation of complex variables between X and Y. Let $X, Y \in \mathbf{C}^{n\times n}$, and let $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ be the*

identity matrix of rank r embedded in a null matrix of size $k \times n$ where $r < k$, $r < n$. Then $|J(X \rightarrow Y)| = 0$.

Proof. Let X equal

$$\begin{pmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(k-r)\times r} & 0_{(k-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} Y_{r\times r} & Y_{r\times(n-r)} \\ Y_{(n-r)\times r} & Y_{(n-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} I_r & 0_{r\times(k-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(k-r)} \end{pmatrix}$$

$$= \begin{pmatrix} Y_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}$$

In forming the exterior product, note that $dX_{nn} = 0dY_{nn}$. The zero factor causes the entire product to go to zero. Thus $|J(X \rightarrow Y)| = 0$. Alternatively, if $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ is $n \times n$ you could apply theorem 34. If $k = n = r$, then $|J(X \rightarrow Y)| = 1$. □

Lemma 7 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = Y^H$ and both X and Y in $\mathbf{C}^{n\times n}$. Let $B \in \mathbf{C}^{n\times n}$ be an elementary transformation matrix $B = \mathrm{diag}(1, \cdots, 1, a, 1, \cdots, 1)$, with the element a in the i^{th} position. Then $|J(X \rightarrow Y)| = |a|^{2n}$ and $|J(Y \rightarrow X)| = |a|^{-2n}$. This theorem was motivated by Deemer and Olkin's proof [67] of their theorem 3.7, and by Stewart p. 43, Example 4.26.1 [259] which addresses operator matrices for a real matrix.*

Proof.

$$\begin{pmatrix} X_{11} & X_{21}^* & \cdots & X_{i1}^* & \cdots & X_{n1}^* \\ X_{21} & X_{22} & \cdots & X_{i2}^* & \cdots & X_{n2}^* \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ii} & \cdots & X_{ni}^* \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{ni} & \cdots & X_{nn} \end{pmatrix}$$

$$= BYB^H = E_1 Y E_1^H =$$

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & a & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

$$\times \begin{pmatrix} Y_{11} & Y_{21}^* & \cdots & Y_{i1}^* & \cdots & Y_{n1}^* \\ Y_{21} & Y_{22} & \cdots & Y_{i2}^* & \cdots & Y_{n2}^* \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ Y_{i1} & Y_{i2} & \cdots & Y_{ii} & \cdots & Y_{ni}^* \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{ni} & \cdots & Y_{nn} \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & a^* & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} Y_{11} & Y_{21}^* & \cdots & a^*Y_{i1}^* & \cdots & Y_{n1}^* \\ Y_{21} & Y_{22} & \cdots & a^*Y_{i2}^* & \cdots & Y_{n2}^* \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ aY_{i1} & aY_{i2} & \cdots & |a|^2 Y_{ii} & \cdots & aY_{ni}^* \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & a^*Y_{ni} & \cdots & Y_{nn} \end{pmatrix}$$

In this matrix, there is one element (Y_{ii}) with a multiplier of $|a|^2$. There are $(i-1)$ elements in row i multiplied by a as coefficients of Y_{ik} in the lower triangle to the left of the main diagonal. There are $(n-i)$ elements in column i with a^* as coefficients of (Y_{ki}) below the main diagonal.

Use wedge products to compute the Jacobian of the transformation. To visualize the problem, first find the linear transformation for each element.

$$X_{ik} = X_{Rik} + iX_{Iik} = aY_{ik} = (a_R + ia_I)(Y_{Rik} + iY_{Iik})$$

$$= (\underbrace{a_R Y_{Rik} - a_I Y_{Iik}}_{X_{Rik}}) + i(\underbrace{a_I Y_{Rik} + a_R Y_{Iik}}_{X_{Iik}})$$

Thus

$$\begin{pmatrix} dX_{Rik} \\ dX_{Iik} \end{pmatrix} = \begin{pmatrix} a_R & -a_I \\ a_I & a_R \end{pmatrix} \begin{pmatrix} dY_{Rik} \\ dY_{Iik} \end{pmatrix}$$

and

$$dX_{Rik} \bigwedge dX_{Iik} = (a_R dY_{Rik} - a_I dY_{Iik}) \bigwedge (a_I dY_{Rik} + a_R dY_{Iik})$$

$$= a_R a_I dY_{Rik} \bigwedge dY_{Rik} + a_R^2 dY_{Rik} \bigwedge dY_{Iik} - a_I^2 dY_{Iik} \bigwedge dY_{Rik} - a_I a_R dY_{Iik} \bigwedge dY_{Iik}$$

$$= (a_R^2 + a_I^2) dY_{Rik} \bigwedge dY_{Iik} = |a|^2 \, dY_{Rik} \bigwedge dY_{Iik}$$

since by properties of the exterior product we know that $dZ \bigwedge dZ = 0$ and

$$dX \bigwedge dY = -dY \bigwedge dX$$

Observe that

$$X_{ik}^* = X_{Rik} - iX_{Iik} = (a_R - ia_I)(Y_{Rik} - iY_{Iik})$$

$$= (a_R Y_{Rik} - a_I Y_{Iik}) - i(a_I Y_{Rik} + a_R Y_{Iik})$$

Thus

$$dX_{Rik} = a_R d_{Rik} - a_I dy_{Iik}$$

$$dX_{Iik} = -a_I dY_{Rik} - a_R dY_{Iik}$$

When we take the wedge product $dX_{ik} \bigwedge dX_{ik}^*$ we observe that it goes to zero because we have repeated indices in our wedge product of

$$\left(dY_{Rik} \bigwedge dY_{Iik}\right) \bigwedge \left(dY_{Rik} \bigwedge dY_{Iik}\right)$$

Therefore we need only to consider the lower triangular and diagonal elements in evaluating the Jacobian of this transformation.

$$X_{ji} = a^* Y_{ji} = X_{Rji} + iX_{Iji} = (a_R + iA_I)^*(Y_{Rji} + iY_{Iji})$$

$$= (a_R - ia_I)(Y_{Rji} + iY_{Iji})$$

$$= (a_R Y_{Rji} + a_I Y_{Iji}) + i(a_R Y_{Iji} - a_I Y_{Rji})$$

$$dX_{Rji} = a_R dY_{Rji} + a_I dY_{Iji}$$

$$dX_{Iji} = a_R dY_{Iji} - a_I dY_{Rji}$$

$$dX_{Rji} \bigwedge dX_{Iji} = (a_R dY_{Rji} + a_I dY_{Iji}) \bigwedge (a_R dY_{Iji} - a_I dY_{Rji})$$

$$= a_R^2 dY_{Rji} \bigwedge dY_{Iji} - a_I^2 dY_{Iji} \bigwedge dY_{Rji} = (a_R^2 + a_I^2) dY_{Rji} \bigwedge dY_{Iji} = |a|^2 \, dY_{Rji} \bigwedge dY_{Iji}$$

$$X_{ii} = |a|^2 \, Y_{ii} = (X_{Rii} + iX_{Iii}) = |a|^2 \, (Y_{Rii} + iY_{Iii}) = |a|^2 \, Y_{Rii}$$

$$dX_{Rii} = |a|^2 \, dY_{Rii}$$

$Y_{ii} \in \mathbf{R}$ because $Y = Y^H$. Therefore $d\Lambda_{ii} = dX_{Rii}$. Finally,

$$dX_{Rjk} \bigwedge dX_{Ijk} = dY_{Rjk} \bigwedge dY_{Ijk}$$

Thus

$$(dX) = dX_{R11} \bigwedge dX_{R21} \bigwedge \cdots \bigwedge dX_{Rnn} = (|a|^2)^{1-1} (|a|^2)^{n-1} \, |a|^2 \, (dY) = |a|^{2n} \, (dY)$$

This leads to our conclusions, $|J(X \to Y)| = |a|^{2n}$ and $|J(Y \to X)| = |a|^{-2n}$.

□

Lemma 8 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = -Y^H$ and both X and Y be in $\mathbf{C}^{n\times n}$. Let $B \in \mathbf{C}^{n\times n}$ be an elementary transformation matrix $B = \mathrm{diag}(1,\cdots,1,a,1,\cdots,1)$, with the element $a \in \mathbf{C}$ in the i^{th} position. Then $|J(X \to Y)| = |a|^{2n}$ and $|J(Y \to X)| = |a|^{-2n}$.*

Proof. Follow the proof of lemma 7, taking into account that now $Y = -Y^H$ instead of $Y = Y^H$.

$$X = BYB^H = E_1 Y E_1^H =$$

$$\begin{pmatrix} Y_{11} & -Y_{21}^* & \cdots & -a^*Y_{i1}^* & \cdots & -Y_{n1}^* \\ Y_{21} & Y_{22} & \cdots & -a^*Y_{i2}^* & \cdots & -Y_{n2}^* \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ aY_{i1} & aY_{i2} & \cdots & |a|^2 Y_{ii} & \cdots & -aY_{ni}^* \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & a^*Y_{ni} & \cdots & Y_{nn} \end{pmatrix}$$

In this case, we still have $(i-1)$ elements in row i below the main diagonal that are multiplied by a. We have $(n-i)$ elements in column i below the main diagonal that are multiplied by a^*. We have one element on the main diagonal at position (i,i) that is multiplied by $|a|^2$. From lemma 7 we note that

$$dX_{Rji} \bigwedge dX_{Iji} = |a|^2 \, dY_{Rji} \bigwedge dY_{Iji}$$

the difference in cases is in the treatment of the diagonal. Y_{ii} is purely imaginary because $-Y = Y^H$. Thus $Y_{ii} \in \mathbf{C}\backslash\mathbf{R}$. So

$$X_{ii} = |a|^2 Y_{ii} = (X_{Rii} + iX_{Iii}) = |a|^2 (Y_{Rii} + iY_{Iii}) = i\,|a|^2 Y_{Iii}$$

Thus

$$dX_{Iii} = |a|^2 dY_{Iii} = dX_{ii}$$

since $Y_{Rii} = 0$. For unaffected elements,

$$dX_{Rjk} \bigwedge dX_{Ijk} = dY_{Rjk} \bigwedge dY_{Ijk}$$

Thus

$$(dX) = dX_{I11} \bigwedge dX \bigwedge \cdots \bigwedge dX_{Inn} = (|a|^2)^{i-1}(|a|^2)^{n-i}\,|a|^2\,(dY) = |a|^{2n}\,(dY)$$

Therefore

$$|J(X \to Y)| = |a|^{2n} \quad \text{and} \quad |J(Y \to X)| = |a|^{-2n}$$

□

Proposition 34 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = \mathrm{diag}(Y_1, \cdots, Y_n)$ where $Y_i \in \mathbf{R}$. Let $X \in \mathbf{C}^{n\times n}$. Let $B \in \mathbf{C}^{n\times n}$ be an elementary transformation matrix $B = \mathrm{diag}(1, \cdots, 1, a, 1, \cdots, 1)$, with the element a in the i^{th} position and $a \in \mathbf{C}$. Then $|J(X \to Y)| = |a|^2$ and $|J(Y \to X)| = |a|^{-2}$.*

Proof.

$$X = \begin{pmatrix} X_1 & & & \\ & \ddots & & \\ & & X_i & \\ & & & \ddots \\ & & & & X_n \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & a & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} \begin{pmatrix} Y_1 & & & \\ & \ddots & & \\ & & Y_i & \\ & & & \ddots \\ & & & & Y_n \end{pmatrix} \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & a^* & \\ & & & \ddots \\ & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} Y_1 & & & \\ & \ddots & & \\ & & |a|^2 Y_i & \\ & & & \ddots \\ & & & & Y_n \end{pmatrix}$$

Thus

$$(dX) = dY_1 \bigwedge \cdots \bigwedge dY_{i-1} \bigwedge |a|^2 dY_i \bigwedge dY_{i+1} \bigwedge \cdots \bigwedge dY_n = |a|^2 (dY)$$

So we see that $|J(X \to Y)| = |a|^2$ and $|J(Y \to X)| = |a|^{-2}$. □

Lemma 9 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = Y^H$ and both X and Y in $\mathbf{C}^{n\times n}$. Let $B \in \mathbf{C}^{n\times n}$ be an*

elementary transformation matrix

$$B = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 1 & & & & & \\ & & & & 1 & & & & \\ & & & \vdots & & \ddots & & & \\ & & & & & & 1 & & \\ & & & a & & \cdots & & 1 & \\ & & & & & & & & 1 \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{pmatrix} = E_2$$

The matrix B is shown here with the constant a in column i and row j. For some conformable matrix Z, BZ has the effect of multiplying row i of matrix Z by the constant a, and adding that result to row j. Then

$$|J(X \rightarrow Y)| = 1 = |J(Y \rightarrow X)|$$

This theorem was motivated by Deemer and Olkin's proof [67] of their theorem 3.7, and by Stewart p. 43, Example 4.26.3 [259] which addresses elementary operator matrices for a real matrix.

Proof. $X = BYB^H = E_2 Y E_2^H =$

$$\begin{pmatrix} X_{11} & X_{21}^* & \cdots & X_{i1}^* & \cdots & X_{n1}^* \\ X_{21} & X_{22} & \cdots & X_{i2}^* & \cdots & X_{n2}^* \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ii} & \cdots & X_{ni}^* \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{ni} & \cdots & X_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} Y_{11} & Y_{21}^* & \cdots & Y_{j1}^* + a^* Y_{i1} & \cdots & Y_{n1}^* \\ \vdots & \vdots & & \vdots & & \vdots \\ Y_{j1} + aY_{i1} & Y_{j2} + aY_{i2} & \cdots & Y_{jj} + aY_{ji}^* + a^* Y_{ji} + |a|^2 Y_{ii} & \cdots & Y_{nj}^* + aY_{ni}^* \\ \vdots & \vdots & & \vdots & & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nj} + a^* Y_{ni} & \cdots & Y_{nn} \end{pmatrix}$$

$$dX_{kl} = dX_{Rkl} + i dX_{Ikl}$$

Look separately at the real terms and the imaginary terms.

$$\left.\begin{aligned} dX_{Rkl} &= dY_{Rkl} \\ dX_{Ikl} &= dY_{Ikl} \end{aligned}\right\} l \leq k, k \neq j$$

Form the wedge product of the real terms and the imaginary terms.

$$dX_{Rkl} \bigwedge dX_{Ikl} = dY_{Rkl} \bigwedge dY_{Ikl}$$

We get the same results when we examine dX_{kl}^*, except for a sign change. Thus $dX_{kl} \bigwedge dX_{kl}^* = 0$. We therefore can restrict attention to the lower triangle plus

diagonal portion of X.

$$dX_{jl} = dX_{Rjl} + idX_{Ijl} = dY_{jl} + adY_{il}$$

$$= (dY_{Rjl} + idY_{Ijl}) + (a_R + ia_I)(dY_{Ril} + dY_{Iil})$$

$$= dY_{Rjl} + idY_{Ijl} + a_R dY_{Ril} - a_I dY_{Iil} + ia_I dY_{Ril} + ia_R dY_{Iil}$$

$$dX_{Rjl} = dY_{Rjl} + a_R dY_{Ril} - a_I dY_{Iil}$$

$$dX_{Ijl} = dY_{Ijl} + a_I dY_{Ril} + a_R dY_{Iil}$$

Note that

$$dX_{Ril} \bigwedge dX_{Iil} \bigwedge dX_{Rjl} = dY_{Ril} \bigwedge dY_{Iil} \bigwedge dY_{Rjl}$$

since the other terms of dX_{Rjl} of the exterior product go to zero. Similarly,

$$dX_{Ril} \bigwedge dX_{Iil} \bigwedge dX_{Rjl} \bigwedge dX_{Ijl} = dY_{Ril} \bigwedge dY_{Iil} \bigwedge dY_{Rjl} \bigwedge dY_{Ijl}$$

The terms with coefficients of a drop out. We thus can say

$$dX_{Rjl} \bigwedge dX_{Ijl} = dY_{Rjl} \bigwedge dY_{Ijl} + \cdots$$

without having to keep track of the other terms. Note that

$$X_{jj} = Y_{jj} + 2\,\mathrm{Re}[a^* Y_{ji}] + |a|^2 Y_{ii}$$

is real. Thus

$$dX_{jj} = dX_{Rjj} + \underbrace{dX_{Ijj}}_{0} = dY_{Rjj} + \underbrace{dY_{Ijj}}_{0} + \cdots,$$

and $dX_{Rjj} = dY_{Rjj}$ by the same reasoning.

Combining all the wedge products, we get $(dX) = (dY)$, and thus $|J(X \to Y)| = 1$ and $|J(Y \to X)| = 1$.

Proposition 35 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = -Y^H$ and let both X and Y be in $\mathbf{C}^{n\times n}$. Let $B \in \mathbf{C}^{n\times n}$ be an elementary transformation matrix $B = E_2 =$*

$$\begin{pmatrix}
1 & & & & & & & & \\
& \ddots & & & & & & & \\
& & 1 & & & & & & \\
& & & 1 & & & & & \\
& & & & 1 & & & & \\
& & & \vdots & & \ddots & & & \\
& & & & & & 1 & & \\
& & & a & & \cdots & & 1 & \\
& & & & & & & & 1 \\
& & & & & & & \ddots & \\
& & & & & & & & & 1
\end{pmatrix}$$

The matrix B is shown here with the constant a in column i and row j. For some conformable matrix Z, BZ has the effect of multiplying row i of matrix Z by the constant a, and adding that result to row j. Then

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

Proof. The proof is almost the same as for lemma 9. The matrix X looks slightly different because it, too, now is skew-Hermitian. What is different is that the sign of the coefficient of terms containing a may be different. For

terms having an a that are not on the diagonal, the wedge product including the associated variable will have been computed earlier. The noticeable change occurs on the diagonal. The cell of interest is of the form

$$X_{jj} = X_{Rjj} + iX_{Ijj} = Y_{jj} + a^* Y_{ji} - aY_{ji}^* + |a|^2 Y_{ii}$$

for $i < j$. We know Y_{jj} and Y_{ii} are imaginary numbers. Of interest is that

$$a^* Y_{ji} - aY_{ji}^* = 2i \operatorname{Im}(a^* Y_{ji})$$

is also imaginary. If $i > j$ then the sign will reverse. Since we are interested only in the absolute value of the Jacobian, we do not need to keep track of the signs in the wedge product. So, the term X_{ii} is imaginary.

Because all previous terms in dX_{jj} have been included in a wedge product, we can say $dX_{ii} = dX_{Iii} = dY_{Iii}$. Combining all the wedge products, we get $(dX) = (dY)$. Thus

$$|J(X \to Y)| = |J(Y \to X)| = 1$$

□

Example 4 *For the case of* $-Y^H = Y \in M_5(\mathbf{C})$ *and* $B = I + ae_4e_2^T$ *where* $a \in \mathbf{C}$, *we get* BYB^H

$$= \begin{pmatrix} Y_{11} & -Y_{21}^* & -Y_{31}^* & -Y_{41}^* - a^*Y_{21}^* & -Y_{51}^* \\ Y_{21} & Y_{22} & -Y_{32}^* & -Y_{42}^* + a^*Y_{22} & -Y_{52}^* \\ Y_{31} & Y_{32} & Y_{33} & -Y_{43}^* + a^*Y_{32} & -Y_{53}^* \\ Y_{41} + aY_{21} & Y_{42} + aY_{22} & Y_{43} - aY_{32}^* & Y_{44} - aY_{42}^* + a^*Y_{42} + |a|^2 Y_{22} & -Y_{54}^* \\ Y_{51} & Y_{52} & Y_{53} & Y_{54} + a^*Y_{52} & Y_{55} \end{pmatrix}$$

Compute (dX) *in the order of*

$$dX_{11}, dX_{21}, dX_{22}, dX_{31}, dX_{32},$$

$$dX_{33}, dX_{41}, dX_{42}, dX_{43}, dX_{44},$$

$$dX_{51}, dX_{52}, dX_{53}, dX_{54}, dX_{55}$$

Taking advantage of $dX_{kl} \wedge dX_{kl} = 0$, *and following this order, gives us a simple computation yielding* $(dX) = (dY)$.

Proposition 36 *Let* $X = BYB^H$ *be a transformation of complex variables between* X *and* Y. *Let* $Y = \mathrm{diag}(Y_1, \cdots, Y_n)$, *and let* X , $B \in M_n(\mathbf{C})$ *where* B *is the elementary transformation matrix* $B = I + ae_ie_j^T$. *The constant* $a \in \mathbf{C}$ *is a complex number and* $B_{ij} = a$, *where* i *and* j *are fixed. Then*

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

Proof.

$$\begin{pmatrix} X_1 & & & & & & & \\ & \ddots & & & & & & \\ & & X_j & 0 & 0 & X_{ji} & & \\ & & 0 & \ddots & & 0 & & \\ & & 0 & & \ddots & 0 & & \\ & & X_{ij} & 0 & 0 & X_i & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \\ & & & & & & & X_n \end{pmatrix}$$

$$= \begin{pmatrix} Y_1 & & & & & & & \\ & \ddots & & & & & & \\ & & Y_j & 0 & 0 & a^* Y_j & & \\ & & 0 & \ddots & & 0 & & \\ & & 0 & & \ddots & 0 & & \\ & & aY_j & 0 & 0 & Y_i + |a|^2 Y_j & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \\ & & & & & & & Y_n \end{pmatrix}$$

In the equation above, only two off-diagonal elements are non-zero. In evaluating the wedge products we have

$$dX_1 \bigwedge \cdots \bigwedge dX_{i-1} = dY_1 \bigwedge \cdots \bigwedge dY_{i-1}$$

where $1 \leq j \leq i-1$. Thus, when we get to including dX_i, we see that dX_j has already been accounted for. Therefore, $dX_i = dY_i$ and $|J(X \rightarrow Y)| = 1$. Thus $|J(Y \rightarrow X)| = 1$. □

Lemma 10 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = Y^H$, and let $X, Y, B \in \mathbf{C}^{n \times n}$. Let B be an elementary transformation matrix*

$$B = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & 1 & & & \\ & & & & 1 & & & & \\ & & & \vdots & \ddots & \vdots & & & \\ & & & & & & 1 & & \\ & & & 1 & \cdots & 0 & & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{pmatrix} = E_3$$

The diagonal of matrix B has a zero at positions i and j. Said differently, $B_{ii} = 0$ and $B_{jj} = 0$. B has off-diagonal ones in positions (i,j) and (j,i), so that $B_{ij} = 1$ and $B_{ji} = 1$. For some matrix Z, BZ has the effect of interchanging rows i and j. The Jacobian of this transformation is $|J(X \rightarrow Y)| = 1$ and $|J(Y \rightarrow X)| = 1$. This theorem was motivated by Deemer and Olkin's proof of their theorem 3.7 [67], and by Stewart (p. 43) example 4.26.2 [259] which addresses operator matrices for a real matrix.

Proof. $X = BYB^H =$

$$\begin{pmatrix} Y_{11} & \cdots & Y_{j1}^* & \cdots & Y_{i1}^* & \cdots & Y_{n1}^* \\ \vdots & & \vdots & & \vdots & & \vdots \\ Y_{j1} & \cdots & Y_{jj} & \cdots & Y_{ji} & \cdots & Y_{nj}^* \\ \vdots & & \vdots & & \vdots & & \vdots \\ Y_{i1} & \cdots & Y_{ji}^* & \cdots & Y_{ii} & \cdots & Y_{ni}^* \\ \vdots & & \vdots & & \vdots & & \vdots \\ Y_{n1} & \cdots & Y_{nj} & \cdots & Y_{ni} & \cdots & Y_{nn} \end{pmatrix}$$

Each element in X is an element in Y. $(dX) = (dY)$. Therefore $|J(X \rightarrow Y)| = 1$ and consequently $|J(Y \rightarrow X)| = 1$. □

Lemma 11 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = -Y^H$, and let $X, Y, B \in \mathbf{C}^{n \times n}$. Let B be an elementary transformation matrix that interchanges rows. Then*

$$|J(X \rightarrow Y)| = 1 = |J(Y \rightarrow X)|$$

Proof. B is merely a permutation of two of the vectors in $(e_1, e_2, \cdots, e_n)$. Thus BYB^H is merely a symmetric shuffle of the elements of Y. At worst, you only get sign changes in the wedge products. We are only interested in the absolute value of the Jacobian, so the sign changes are of no interest. □

Example 5 *Let* $-Y^H = Y \in M_5(\mathbf{C})$ *and let* $B = (e_1, e_4, e_3, e_2, e_5)$. *Then* $BYB^H =$

$$\begin{pmatrix} Y_{11} & -Y_{41}^* & -Y_{31}^* & -Y_{21}^* & -Y_{51}^* \\ Y_{41} & Y_{44} & Y_{43} & Y_{42} & -Y_{54}^* \\ Y_{31} & -Y_{43}^* & Y_{33} & Y_{32} & -Y_{53}^* \\ Y_{21} & -Y_{42}^* & -Y_{32}^* & Y_{22} & -Y_{52}^* \\ Y_{51} & Y_{54} & Y_{53} & Y_{52} & Y_{55} \end{pmatrix}$$

Proposition 37 *Let* $X = BYB^H$ *be a transformation of complex variables between* X *and* Y. *Let* $Y = \mathrm{diag}(Y_1, Y_2, \cdots, Y_n)$ *and let* $X, Y, B \in \mathbf{C}^{n \times n}$. *Let* B *be an elementary transformation matrix*

$$B = I - e_i e_i^T - e_j e_j^T + e_i e_j^T + e_j e_i^T$$

Then

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

Proof.

$$\begin{pmatrix} X_1 & & & & & \\ & \ddots & & & & \\ & & X_i & & & \\ & & & \ddots & & \\ & & & & X_j & \\ & & & & & \ddots \\ & & & & & X_n \end{pmatrix}$$

$$= \begin{pmatrix} Y_1 & & & & & \\ & \ddots & & & & \\ & & Y_j & & & \\ & & & \ddots & & \\ & & & & Y_i & \\ & & & & & \ddots \\ & & & & & Y_n \end{pmatrix}$$

Except for a sign change due to the permutation, the wedge products are identical. Thus the Jacobians are the same.

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

□

Theorem 36 *Inverses of Elementary Transformations.*

Table C.3. Elementary Operator Matrices

Lemma	Operation	Operator	Remark
7	*Scaling matrix.*	$E_1 = I + (a-1)\delta_{ii}$	*a is in position i*
9	$Row_j \leftarrow Row_j + aRow_i$	$E_2 = I + a\delta_{j,i}$	*a is in position* (j,i)
10	*Swap* Row_i *and* Row_j	$E_3 = \sigma_{ij}(e_1, \cdots, e_n)$	*Swap rows of I*

Let $\check{E}$ *denote the inverse of* E. *(This unusual notation is of temporary value in the conclusion. Once the conclusion is made, the bad notation can be forgotten.) Define elementary operator matrices* E_1, E_2, E_e *as in table C.3. The exhaustive formula for each is given in the lemma indicated in column 1.* $\sigma_{ij}(\cdot)$ *is the permutation that exchanges elements* i *and* j *of the argument. With these definitions, we note that the following relationships hold between the elementary matrix and its inverse.*

$\check{E}_1 = (E_1)^{-1}$ *is like* E_1 *with a reciprocal of the multiplication constant,* $\frac{1}{a}$

$\check{E}_2 = (E_2)^{-1}$ *is like* E_2 *with the sign changed on the multiplication constant,* $-a$

$\check{E}_3 = (E_3)^{-1} = E_3$

An explicit definition follows.

$$(E_1)^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \frac{1}{a} & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} = \check{E}_1$$

$$(E_2)^{-1} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ \vdots & \ddots & & & & \\ \vdots & & \ddots & & & \\ a & \cdots & \cdots & \ddots & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ \vdots & \ddots & & & & \\ \vdots & & \ddots & & & \\ -a & \cdots & \cdots & \ddots & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} = \check{E}_2$$

$$(E_3)^{-1} = \begin{pmatrix} 1 & & & & \\ & 0 & \cdots & 1 & \\ & & 1 & & \\ & \vdots & \ddots & \vdots & \\ & & & 1 & \\ & 1 & \cdots & 0 & \\ & & & & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 1 & & & & \\ & 0 & \cdots & 1 & \\ & & 1 & & \\ & \vdots & \ddots & \vdots & \\ & & & 1 & \\ & 1 & \cdots & 0 & \\ & & & & 1 \end{pmatrix} = \check{E}_3$$

Thus, each elementary transformation has an inverse. The proof is by simple matrix multiplication.

Proof.

$$E_1\check{E}_1 = I, \quad E_2\check{E}_2 = I, \quad E_3\check{E}_3 = I$$

□

Theorem 37 *Let $A \in \mathcal{F}^{m\times n}, m > n$, be a matrix over field $\mathcal{F}$ (real or complex), of rank $r < n$. Let E_1 be an elementary transformation matrix that multiplies a row by a constant when A is premultiplied by E_1. Let E_2 be an elementary transformation that adds to one row some constant multiple of another row, when A is premultiplied by E_2. Let E_3 interchange two rows of A when A is premultiplied by E_3. Let the notation of a superscript, such as $E_i^k A$, denote that matrix A is premultiplied by a set of k different elementary transformation matrices of type i. Then*

$$E_3^{n-r} E_1^r E_2^n A E_3^{n-r} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

for appropriate choices of $\{E_i^k\}$. The concept is simple and the proof is tedious and thus omitted.

Corollary 6 *If*

$$E_3^{n-r} E_1^r E_2^n A E_3^{n-r} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

as stated in theorem 37, then

$$A = \check{E}_2^n \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r}$$

Theorem 38 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = Y^H$, and let $B, X, Y \in \mathbf{C}^{n\times n}$ and let $\mathrm{rank}(B) = r$. Then $|J(X \to Y)| = |\det B|^{2n}$ if $r = n$, and is zero if $r < n$. Likewise, $|J(Y \to X)| = |\det B|^{-2n}$.*

Proof. From corollary 6, B can be written as

$$B = \check{E}_2^n \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r}$$

where the mark above the symbol indicates a matrix inverse, the superscript indicates the number of such matrices, and the subscript indicates the type of elementary transformation matrix. Then

$$X = \check{E}_2^n \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r} Y \left(\check{E}_3^{n-r}\right)^H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \left(\check{E}_3^{n-r}\right)^H \left(\check{E}_1^r\right)^H \left(\check{E}_2^n\right)^H$$

Now, apply the theorems that describe the Jacobians of individual elementary transformations, noting the pattern as suggested by Deemer and Olkin theorem 3.7 [67]. The subscript to the left of a matrix in the notation to follow is merely an index.

$$Y_3 = \underbrace{({}_{n-r}\check{E}_3) \cdots \underbrace{({}_2\check{E}_3) \underbrace{({}_1\check{E}_3) Y ({}_1\check{E}_3)^H}_{J_1=1} ({}_2\check{E}_3)^H}_{J_2=1} \cdots ({}_{n-r}\check{E}_3)^H}_{J_{n-r}=1}$$

$$|J(Y_3 \rightarrow Y)| = J_{n-r} \cdots J_2 J_1 = 1$$

$$Y_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Y_3 \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

By a previous theorem,

$$|J(Y_r \rightarrow Y_3)| = \begin{cases} 1, \text{ if } r = n \\ 0, \text{ otherwise} \end{cases}$$

If $|J(Y_r \to Y_3)| = 0$, then $|J(X \to Y)| = 0$ because $|J(X \to Y)|$ is the product of all the intermediate Jacobians. By lemma 7 we get

$$Y_1 = \underbrace{\underbrace{\underbrace{({}_r\check{E}_1)\cdots({}_2\check{E}_1)}_{J_r=|b_r|^{2n}}\underbrace{({}_1\check{E}_1)Y_r({}_1\check{E}_1)^H}_{J_1=|b_1|^{2n}}({}_2\check{E}_1)^H\cdots({}_r\check{E}_1)^H}_{J_2=|b_2|^{2n}}}_{J_r=|b_r|^{2n}}$$

Thus

$$|J(Y_1 \to Y_r)| = \prod_{i=1}^{r} J_i = \prod_{i=1}^{r} |b_i|^{2n} = \prod_{i=1}^{r} \left|\frac{1}{a_i}\right|^{2n}$$

where a_i is the complex constant multiplier appearing in the diagonal of the i^{th} elementary transformation of type E_1. Finally,

$$X = \underbrace{\underbrace{\underbrace{({}_n\check{E}_2)\cdots({}_2\check{E}_2)}_{J_n=1}\underbrace{({}_1\check{E}_2)Y_1({}_1\check{E}_2)^H}_{J_1=1}({}_2\check{E}_2)^H\cdots({}_n\check{E}_2)^H}_{J_2=1}}_{J_n=1}$$

Therefore

$$|J(X \to Y_1)| = 1$$

$$|J(X \to Y)| = |J(X \to Y_1)| \cdot |J(Y_1 \to Y_r)| \cdot |J(Y_r \to Y_3)| \cdot |J(Y_3 \to Y)|$$

$$= \begin{cases} \prod_{i=1}^{r} |b_i|^{2n}, & \text{if } r = n \\ 0, & \text{otherwise} \end{cases}$$

Now, consider the determinant of B.

$$\det B = \det[\check{E}_2'' \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r}]$$

$$= \det(\check{E}_2^n)\det(\check{E}_1^r)\det(\check{E}_3^{n-r})\det\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\det(\check{E}_3^{n-r})$$

$$= 1^n\left(\prod_{i=1}^{r} b_i\right)(-1)^{n-r}\left\{\begin{array}{l} 1, \text{ if } r = n \\ 0, \text{ otherwise} \end{array}\right\}(-1)^{n-r} = \left\{\begin{array}{l} \prod_{i=1}^{r} b_i, \text{ if } r = n \\ 0, \text{ otherwise} \end{array}\right.$$

$$(\det B)(\det B)^* = |\det B|^2 = \prod_{i=1}^{r} b_i b_i^* = \prod_{i=1}^{r} |b_i|^2$$

Let B be of full rank $r = n$. Then

$$|J(X \to Y)| = \prod_{i=1}^{n} |b_i|^{2n} = |\det B|^{2n}$$

Also, $|J(Y \to X)| = |\det B|^{-2n}$.

Note that this theorem is for matrices that are unstructured. When B has a special structure, such as being triangular, then that structure must be accounted for in determining the Jacobian.

Corollary 7 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = Y^H$, and let X, Y, and B be in $\mathbf{C}^{n\times n}$. Let B be unitary. Then*

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

Proof. From theorem 38, $|J(X \to Y)| = |\det B|^{2n}$. Recall that the determinant of any unitary matrix is 1. Therefore

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

□

In particular, when considering the eigenvalue decomposition $X = U\Lambda^2 U^H$, we see that $|J(X \to \Lambda^2)| = 1$ and $|J(\Lambda^2 \to X)| = 1$ where U is a fixed unitary matrix.

Theorem 39 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = -Y^H$, and let X, Y, and B be in $\mathbf{C}^{n\times n}$. Let* $\operatorname{rank}(B) = r$. *Then $|J(X \to Y)| = |\det B|^{2n}$ and $|J(Y \to X)| = |\det B|^{-2n}$ when $r = n$. When $r < n$ then $|J(X \to Y)| = 0$. This is a complexification of Muirhead's theorem 2.1.7 [187], which is stated without proof.*

Proof. If $Y = -Y^H$, then the diagonal of Y is pure imaginary. Thus, in following the proof of theorem 38, we note that

$$B = \check{E}_2^n \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r}$$

where the accent on top of the E indicates matrix inverse, the subscript indicates the type of the elementary transformation matrix, and the superscript indicates the number of matrices of that particular type. Then

$$X = \check{E}_2^n \check{E}_1^r \check{E}_3^{n-r} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \check{E}_3^{n-r} Y \left(\check{E}_3^{n-r}\right)^H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \left(\check{E}_3^{n-r}\right)^H \left(\check{E}_1^r\right)^H \left(\check{E}_2^n\right)^H$$

To help compute the Jacobian, expand the matrices, subscripting the specific matrices on the left. Let

$$Y_3 = \left({}_{n-r}\check{E}_3\right) \cdots \left({}_2\check{E}_3\right) \left({}_1\check{E}_3\right) Y \left({}_1\check{E}_3\right)^H \left({}_2\check{E}_3\right)^H \cdots \left({}_{n-r}\check{E}_3\right)^H$$

From lemma 11 we know that the Jabobian $|J(Y_3 \rightarrow Y)| = 1$. Now, let

$$Y_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Y_3 \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

We know

$$|J(Y_r \rightarrow Y_3)| = \begin{cases} 1, \text{ if } r = n \\ 0, \text{ otherwise} \end{cases}$$

Suppose $r = n$. Consider Y_1 next, with lemma 8. Let

$$Y_1 = \underbrace{\underbrace{\underbrace{\left({}_r\check{E}_1\right) \cdots \left({}_2\check{E}_1\right)}_{J_r = |b_r|^{2n}} \underbrace{\left({}_1\check{E}_1\right) Y_r \left({}_1\check{E}_1\right)^H}_{J_1 = |b_1|^{2n}} \left({}_2\check{E}_1\right)^H}_{J_2 = |b_2|^{2n}} \cdots \left({}_r\check{E}_1\right)^H}$$

Thus

$$|J(Y_1 \rightarrow J_r)| = \prod_{i=1}^{r} J_i = \prod_{i=1}^{r} |b_i|^{2n}$$

Finally,

$$X = \left({}_n\check{E}_2\right) \cdots \left({}_2\check{E}_2\right) \underbrace{\left({}_1\check{E}_2\right) Y_1 \left({}_1\check{E}_2\right)^H}_{J_1 = 1} \left({}_2\check{E}_2\right)^H \cdots \left({}_n\check{E}_2\right)^H$$

We see that $|J(X \rightarrow Y_1)| = 1$. Putting it all together,

$$|J(X \rightarrow Y)| = |J(X \rightarrow Y_1)| \cdot |J(Y_1 \rightarrow J_r)| \cdot |J(Y_r \rightarrow Y_3)| \cdot |J(Y_3 \rightarrow Y)|$$

$$= \begin{cases} \prod_{i=1}^{n} |b_i|^{2n}, \text{ for } r = n \\ 0, \text{ otherwise} \end{cases}$$

From the proof of theorem 38, if $\text{rank}(B) = n$, then $|\det B|^2 = \prod_{i=1}^{n} |b_i|^{2n}$. Thus $|J(X \rightarrow Y)| = |\det B|^{2n}$ and $|J(Y \rightarrow X)| = |\det B|^{-2}$. □

Discussion. In Muirhead's text (pp. 58-59) [187] attention is restricted to only real matrices. Thus the diagonal of Muirhead's $Y = -Y^T$ is zero. Compared to $Y = Y^T$, his skew-symmetric matrix has fewer algebraically independent variables. The effect is that the Jacobians for BYB^T for the two cases are different.

The complex case is simpler. The Hermitian Y has pure reals on the diagonal, whereas the skew-Hermitian Y has pure imaginary numbers on the diagonal. The Hermitian and the skew-Hermitian matrices have the same number of algebraically independent variables. The fact that the Jacobians turn out to be the same in the complex case is thus not inconsistent with this observation.

Corollary 8 *Let $X = BYB^H$ be a transformation of complex variables between X and Y. Let $Y = -Y^H$,and let X, Y, and B be in $\mathbf{C}^{n\times n}$. Let B be unitary ($B^H = B^{-1}$). Then $|J(X \to Y)| = 1 = |J(Y \to X)|$.*

Proof. From theorem 39, $|J(X \to Y)| = |\det B|^{2n}$. Since B is unitary, $\det B = 1$. Therefore

$$|J(X \to Y)| = 1 = |J(Y \to X)|$$

□

Theorem 40 *Let $X = Y^{-1}$ be a complex change of variables between X and Y. Let $X, Y \in \mathbf{C}^{n \times n}$. Then*

$$|J(X \to Y)| = |\det X|^{2n}$$

This is a complexification of Muirhead's theorem 2.1.8 [187].

Proof. This is a complexification of Muirhead's proof. $X = Y^{-1}$ implies $YX = I$. We compute the matrix differential of this according to theorem 21 to find that

$$(dY)X + Y(dX) = 0$$

This implies

$$(dY) = -Y(dX)X^{-1} = -Y(dX)X^{-H} = -X^{-1}(dX)X^{-H}$$

By theorem 38, $|J(Y \to X)| = |\det X^{-1}|^{2n}$. Therefore $|J(Y \to X)| = |\det X|^{-2n}$ and $|J(X \to Y)| = |\det X|^{2n}$. □

Appendix D

DISTRIBUTIONS, PART I

D.1 Complex Normal Distribution Introduction

A derivation of the probability density function for the vector complex distribution is given by Wooding [293] for the zero mean case. This is the form used by Goodman [92]. The most complete notationally consistent summary of basic results readily available are given by Anderson (problem 3.64) [26] and by Monzingo and Miller (appendix E.2) [185]. I strongly recommend that Goodman [92] be used as the source reference from which other results are constructed. He is a careful author.

Close reading of the literature is required if results of papers are to be compared on an equal basis. In particular, pay attention to the following issues. Are the results for the complex case being *presented* in terms of real or complex variables? Is the assumed covariance matrix of special form, or is it general? Answers to these questions will explain the variations in formulations of the characteristic function, as well as possibly other results.

The key to understanding the complex characteristic function is that each complex random variable can be thought of as two paired real random vari-

ables. Since a characteristic function is a special type of expected value, the integration of the associated density function must be carried out over all the space its random variable is defined for. Hence, the integration must be carried out over the entire complex plane. Equivalently, a double integration over $\mathbf{R}$ is required.

Not all vector complex normal distributions are the same. We are concerned about a very special complex normal distribution which is motivated by signal processing needs. I will follow the explanation provided by Goodman [92].

D.1.1 Definition of a Vector Complex Random Variable

A complex normal random variable is a complex variable whose real and imaginary parts are bivariate Gaussian distributed. Let $Z = X + iY$ be a p-variate column vector complex normal random variable. A p-variate complex normal random variable

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}$$

is a p-tuple of complex normal random variables such that the vector of real and imaginary parts

$$U = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ Y_1 \\ \vdots \\ Y_p \end{pmatrix}$$

is a $2p$-variate real normal random variable with a special covariance structure.

Wooding provided the following explanation. Consider

$$Z_n(t) = X_n(t) + iY_n(t)$$

This can be written in the form

$$Z_n(t) = \sum_k (C_{nk} - id_{nk}) \exp\{i\theta_k(t)\}$$

where the coefficients C_{nk} and d_{nk} are real. This complex Fourier series arises in numerous fields, particularly in theory related to time series. Expanding this into its real and imaginary parts yields

$$X_n(t) = \sum_k [C_{nk} \cos(\theta_k(t)) + d_{nk} \sin(\theta_k(t))]$$

$$Y_n(t) = \sum_k [C_{nk} \sin(\theta_k(t)) - d_{nk} \cos(\theta_k(t))]$$

The X_i and Y_i are in phase quadrature. (Note: Wooding's $Y_n(t)$ is the negative of what I am reporting here.) The covariance matrix satisfies the following

relations.

$$\mathcal{E}\{X_m X_n\} = \mathcal{E}\{Y_m Y_n\}$$

$$\mathcal{E}\{X_m Y_n\} = -\mathcal{E}\{Y_m X_n\}$$

When $\mathcal{E}\{X_m\} = \mathcal{E}\{Y_m\} = 0$, the definition of the complex variates will not involve Fourier series concepts.

When this restriction on the covariance matrix is made, reordering the elements of the real variable vector representation of the complex vector yields a covariance matrix with a special pattern. The vector $\begin{pmatrix} X \\ Y \end{pmatrix}$ has a normal distribution with mean vector $\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$ and covariance matrix $S = \begin{pmatrix} G & -F \\ F & G \end{pmatrix}$ where G is positive definite and $F = -F^T$ (skew symmetric). Then $Z = X + iY$ is said to have a complex normal distribution with mean $\mu = \mu_X + i\mu_Y$ and covariance matrix

$$\Sigma = \mathcal{E}\{(Z - \mu)(Z - \mu)^H\}$$

where H is the Hermitian (complex conjugate) transpose. Note that Σ is Hermitian and positive definite. We can express Σ also as the sum of real and imaginary parts, $\Sigma = Q + iR$.

D.1.2 Proof that Σ is Isomorphic to $2S$

This is Anderson problem 3.64(a) [26]. Let $\begin{pmatrix} X \\ Y \end{pmatrix}$ have a real multivariate normal distribution with mean vector $\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$ and covariance $S = \begin{pmatrix} G & -F \\ F & G \end{pmatrix}$ where G is positive definite and $F = -F^T$. Then, let $Z = X + iY$ have a special vector complex normal distribution with mean $\mu = \mu_X + i\mu_Y$ and covariance matrix

$$\Sigma = \mathcal{E}\{(Z-\mu)(Z-\mu)^H\}$$

where Σ is positive definite. Then

$$\Sigma = \mathcal{E}\{(Z-\mu)(Z-\mu)^H\} = \mathcal{E}\{(X+iY-\mu_X-i\mu_Y)(X+iY-\mu_X-i\mu_Y)^H\}$$

$$= \mathcal{E}\{(X+iY-\mu_X-i\mu_Y)(X^T-iY^T-\mu_X^T+i\mu_Y^T)\}$$

$$= \mathcal{E}\{(XX^T+iYX^T-\mu_X X^T-i\mu_Y X^T)-i(XY^T+iYY^T-\mu_X Y^T-i\mu_Y Y^T)$$

$$-(X\mu_X^T+iY\mu_X^T-\mu_X\mu_X^T-i\mu_Y\mu_X^T)+i(X\mu_Y^T+iY\mu_Y^T-\mu_X\mu_Y^T-i\mu_Y\mu_Y^T)\}$$

We expand and group all the real terms together and all the imaginary terms together.

$$= \mathcal{E}\{(XX^T+iYX^T-\mu_X X^T-i\mu_Y X^T)-(X\mu_X^T+iY\mu_X^T-\mu_X\mu_X^T-i\mu_Y\mu_X^T)$$

$$-i(XY^T+iYY^T-\mu_X Y^T-i\mu_Y Y^T)+i(X\mu_Y^T+iY\mu_Y^T-\mu_X\mu_Y^T-i\mu_Y\mu_Y^T)\}$$

$$= \mathcal{E}\{XX^T+iYX^T-\mu_X X^T-i\mu_Y X^T-X\mu_X^T-iY\mu_X^T+\mu_X\mu_X^T+i\mu_Y\mu_X^T$$

$$-iXY^T + YY^T + i\mu_X Y^T - \mu_Y Y^T + iX\mu_Y^T - Y\mu_Y^T - i\mu_X\mu_Y^T + \mu_Y\mu_Y^T\}$$

$$= \mathcal{E}\{XX^T - \mu_X X^T - X\mu_X^T + \mu_X\mu_X^T + YY^T - \mu_Y Y^T - Y\mu_Y^T + \mu_Y\mu_Y^T\}$$

$$+i\mathcal{E}\{YX^T - \mu_Y X^T - Y\mu_X^T + \mu_Y\mu_X^T - XY^T + \mu_X Y^T + X\mu_Y^T - \mu_X\mu_Y^T\}$$

$$= Q + iR$$

We look for a more compact expression for the real terms and the imaginary terms.

$$= \mathcal{E}\{(X - \mu_X)X^T - (X - \mu_X)\mu_X^T + (Y - \mu_Y)Y^T - (Y - \mu_Y)\mu_Y^T\}$$

$$+i\mathcal{E}\{(Y - \mu_Y)X^T - (Y - \mu_Y)\mu_X^T - (X - \mu_X)Y^T + (X - \mu_X)\mu_Y^T\}$$

$$= \mathcal{E}\{(X - \mu_X)(X^T - \mu_X^T) + (Y - \mu_Y)(Y^T - \mu_Y^T)\}$$

$$+i\mathcal{E}\{(Y - \mu_Y)(X^T - \mu_X^T) - (X - \mu_X)(Y^T - \mu_Y^T)\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} (X - \mu_X) & (Y - \mu_Y) \end{pmatrix}\begin{pmatrix} (X - \mu_X)^T \\ (Y - \mu_Y)^T \end{pmatrix}\right\}$$

$$+i\mathcal{E}\left\{\begin{pmatrix} (X - \mu_X) & (Y - \mu_Y) \end{pmatrix}\begin{pmatrix} -(Y - \mu_Y)^T \\ (X - \mu_X)^T \end{pmatrix}\right\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} (X - \mu_X) & (Y - \mu_Y) \end{pmatrix}\begin{pmatrix} (X - \mu_X) & (Y - \mu_Y) \end{pmatrix}^T\right\}$$

$$+i\mathcal{E}\left\{\begin{pmatrix} (X - \mu_X) & (Y - \mu_Y) \end{pmatrix}\begin{pmatrix} -(Y - \mu_Y) & (X - \mu_X) \end{pmatrix}^T\right\}$$

$$= Q + iR$$

We recombine terms to see if we can get further simplification in our notation.

$$= \mathcal{E}\left\{\left[\begin{pmatrix} (X-\mu_X) & (Y-\mu_Y) \end{pmatrix}\right]\right.$$

$$\left.\times\left[\begin{pmatrix} (X-\mu_X) & (Y-\mu_Y) \end{pmatrix}^T + i\begin{pmatrix} -(Y-\mu_Y) & (X-\mu_X) \end{pmatrix}^T\right]\right\}$$

$$= \mathcal{E}\left\{\left[\begin{pmatrix} (X-\mu_X) & (Y-\mu_Y) \end{pmatrix}\right]\right.$$

$$\left.\times\left[\begin{pmatrix} (X-\mu_X)-i(Y-\mu_Y) & (Y-\mu_Y)+i(X-\mu_X) \end{pmatrix}^T\right]\right\}$$

Now, express the complex vector as a partitioned real vector and compute the covariance matrix.

$$S = \begin{pmatrix} G & -F \\ F & G \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \mathcal{E}\left\{\begin{pmatrix} (X-\mu_X) \\ (Y-\mu_Y) \end{pmatrix}\begin{pmatrix} (X-\mu_X) \\ (Y-\mu_Y) \end{pmatrix}^T\right\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} (X-\mu_X) \\ (Y-\mu_Y) \end{pmatrix}\begin{pmatrix} (X-\mu_X)^T & (Y-\mu_Y)^T \end{pmatrix}\right\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} (X-\mu_X)(X-\mu_X)^T & (X-\mu_X)(Y-\mu_Y)^T \\ (Y-\mu_Y)(X-\mu_X)^T & (Y-\mu_Y)(Y-\mu_Y)^T \end{pmatrix}\right\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} (X-\mu_X)(X^T-\mu_X^T) & (X-\mu_X)(Y^T-\mu_Y^T) \\ (Y-\mu_Y)(X^T-\mu_X^T) & (Y-\mu_Y)(Y^T-\mu_Y^T) \end{pmatrix}\right\}$$

$$= \mathcal{E}\left\{\begin{pmatrix} XX^T-\mu_X X^T-X\mu_X^T+\mu_X\mu_X^T & XY^T-\mu_X Y^T-X\mu_Y^T+\mu_X\mu_Y^T \\ YX^T-\mu_Y X^T-Y\mu_X^T+\mu_Y\mu_X^T & YY^T-\mu_Y Y^T-Y\mu_Y^T+\mu_Y\mu_Y^T \end{pmatrix}\right\}$$

From the problem statement, the covariance matrix must have the special form such that $A = D$ and $C = -B$, Then $A = D$ implies

$$\mathcal{E}\{XX^T - \mu_X X^T - X\mu_X^T + \mu_X\mu_X^T\} = \mathcal{E}\{YY^T - \mu_Y Y^T - Y\mu_Y^T + \mu_Y\mu_Y^T\}$$

and $C = -B$ implies

$$\mathcal{E}\{YX^T - \mu_Y X^T - Y\mu_X^T + \mu_Y\mu_X^T\} = -\mathcal{E}\{(XY^T - \mu_X Y^T - X\mu_Y^T + \mu_X\mu_Y^T)\} =$$

$$= \mathcal{E}\{-XY^T + \mu_X Y^T + X\mu_Y^T - \mu_X\mu_Y^T\}$$

Observe that $C = B^T$, and by special requirements we impose $C = -B$. This is possible since B and C are square matrices of the same dimension. The special condition is possible in signal processing applications.

To demonstrate the required equalities, we know from the requirements

$$\begin{pmatrix} G & -F \\ F & G \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

that we need

$$2G = A + D$$

$$2F = C - B$$

Thus we compute

$$2G = A+D = \mathcal{E}\{XX^T - \mu_X X^T - X\mu_X^T + \mu_X\mu_X^T + YY^T - \mu_Y Y^T - Y\mu_Y^T + \mu_Y\mu_Y^T\}$$

$$Q = \mathcal{E}\{XX^T - \mu_X X^T - X\mu_X^T + \mu_X\mu_X^T + YY^T - \mu_Y Y^T - Y\mu_Y^T + \mu_Y\mu_Y^T\}$$

and

$$2F = C-B = \mathcal{E}\{YX^T - \mu_Y X^T - Y\mu_X^T + \mu_Y\mu_X^T - XY^T + \mu_X Y^T + X\mu_Y^T - \mu_X\mu_Y^T\}$$

$$R = \mathcal{E}\{YX^T - \mu_Y X^T - Y\mu_X^T + \mu_Y \mu_X^T - XY^T + \mu_X Y^T + X\mu_Y^T - \mu_X \mu_Y^T\}$$

We observe that $Q = 2G$ and $R = 2F$. From this, we know that any element in Σ will be double the value of the corresponding element of S. So, there is a one-to-one mapping between every element of S and Σ. We thus say that Σ is *isomorphic* to $2S$.

D.1.3 Density of Vector Complex Normal Distribution

Anderson problem 3.64(d) [26] defines the probability density function for the p-variate special vector complex normal distribution $\mathbf{C}N_p(\mu, \Sigma)$ as

$$f(z) = \frac{1}{\pi^p \det(\Sigma)} \exp[-(z-\mu)^H \Sigma^{-1}(z-\mu)]$$

The density for the special vector complex normal distribution is slightly different from the real variables case in three ways. First, the exponent uses the Hermitian transpose. Second, the exponent term is not divided in half. Third, the term preceding the exponent does not have a square root.

Recall that S is a $2p \times 2p$ matrix. Compared to S, each element of Σ is multiplied by 2. From the theory of determinants, we know

$$[\det(\Sigma)]^2 = \det(2S) = 2^{2p} \det(S)$$

which implies

$$\det(\Sigma) = 2^p [\det(S)]^{1/2}$$

or

$$[\det(S)]^{1/2} = 2^{-p}\det(\Sigma)$$

This result is different than theorem 2.5 of Goodman [92]. The 2π factor in the leading term of the density function for the real case $2p$-variate expression is raised to the $\frac{2p}{2} = p$ power. Thus, the complete leading term in the denominator is

$$(2\pi)^p 2^{-p}\det(\Sigma) = \pi^p \det(\Sigma)$$

D.1.4 Characteristic Function for the Standardized Vector Complex Normal Distribution

Derivation from the Standard Univariate Complex Normal Density Function

Let the random variable $z = x + iy$ have the univariate special complex normal distribution with zero mean and unit variance $\mathbf{C}N_1(0,1)$. The density function is given by

$$f(z) = \frac{1}{\pi}e^{-|z|^2} = \frac{1}{\pi}e^{-z^H z} = \frac{1}{\pi}e^{-(x-iy)(x+iy)} = \frac{1}{\pi}e^{-(x^2+y^2)} \qquad \text{(D.1)}$$

Let transform parameter be $t = \eta + i\tau$. The characteristic function is computed by

$$\Phi_z(t) = \mathcal{E}\{\exp[i\,\mathrm{Re}(t^H z)]\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\,\mathrm{Re}\{(\eta - i\tau)(x+iy)\}}\frac{1}{\pi}e^{-(x^2+y^2)}dxdy$$

$$= \frac{1}{\pi}\int_x\int_y e^{-(x^2+y^2)+i(\eta x+\tau y)}dxdy = \frac{1}{\pi}\left(\int_x e^{-x^2+i\eta x}dx\right)\left(\int_y e^{-y^2+ity}dy\right)$$

To solve the integral $\int_{-\infty}^{\infty} e^{i\eta x-x^2}dx$, note that the exponent can be placed into the form of a perfect square by observing the following standard trick.

$$-(x-\frac{1}{2}i\eta)^2 = -(x^2 - ix\eta - \frac{1}{4}\eta^2) = i\eta x - x^2 + \frac{1}{4}\eta^2$$

which implies that

$$i\eta x - x^2 = -(x-\frac{1}{2}i\eta)^2 - \frac{1}{4}\eta^2$$

Therefore

$$\int_{-\infty}^{\infty} e^{i\eta x-x^2}dx = \exp[-\frac{1}{4}\eta^2]\int_{-\infty}^{\infty}\exp[-(x-\frac{1}{2}i\eta)^2]dx$$

The integral is in the form of $\oint \exp[-z^2]dz$. *Note that the function* $\exp(-z^2)$ is analytic everywhere in the complex plane. Thus $\oint \exp[-z^2]dz = 0$. Consider the contour given in figure D.1.

The closed path of integration that begins on the real axis at $-K$, follows the real axis to $+K$, descends parallel to the imaginary axis in the negative direction to $K - \frac{1}{2}i\eta$, transits parallel to the real axis in the negative direction to $-K - \frac{1}{2}i\eta$, and the returns to the starting point by ascending parallel to the imaginary axis to the real axis again at $-K$. Evaluating the integral along this contour yields

$$\oint e^{-z^2}dz = 0 = \int_{-K}^{+K} e^{-z^2}dz + \int_{K}^{K-\frac{1}{2}i\eta} e^{-z^2}dz + \int_{K-\frac{1}{2}i\eta}^{-K-\frac{1}{2}i\eta} e^{-z^2}dz + \int_{-K-\frac{1}{2}i\eta}^{-K} e^{-z^2}dz$$

$$= \int_{-K}^{+K} e^{-x^2}dx + \int_0^{-\eta/2} e^{-(K+iu)^2}du + \int_K^{-K} e^{-(u-\frac{1}{2}i\eta)^2}du + \int_{-\eta/2}^{0} e^{-(-K+iu)^2}du$$

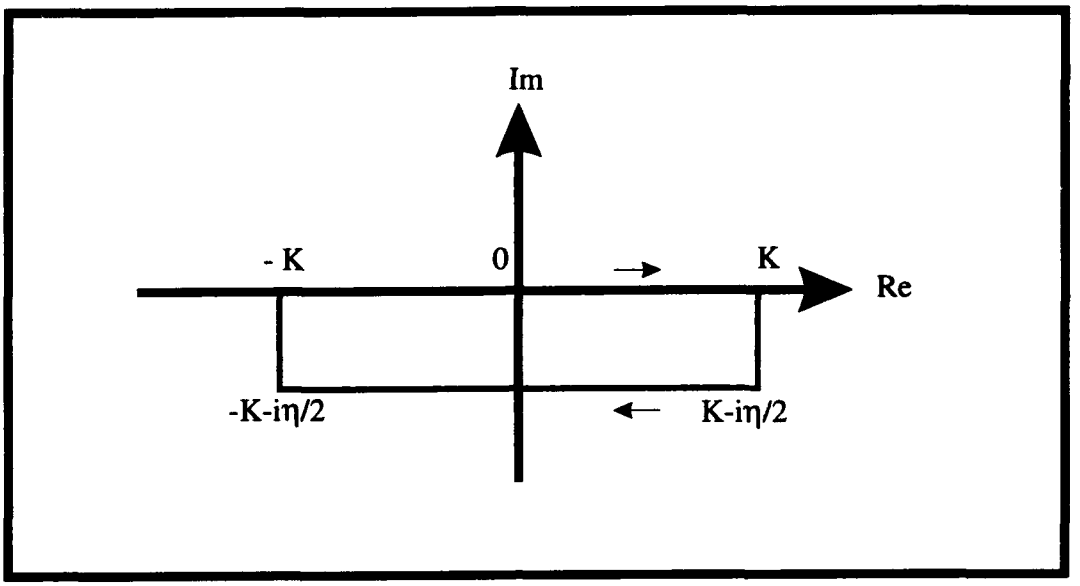


Figure D.1. Integration Contour to Get Characteristic Function

where a different appropriate change of variables has been made for each integral to simplify the limits. In the second integral, let $u = -i(z - K)$. In the third integral, let $u = z + \frac{1}{2}i\eta$. In the last integral, let $u = -i(z + K)$.

Examine what happens as K goes to infinity.

$$\lim_{K\to\infty}\left|\int_0^{-\eta/2} e^{-(K+iu)^2}du\right| = \lim_{K\to\infty}\left|\int_0^{-\eta/2} e^{-(K^2-u^2+2iKu)}du\right|$$

$$\leq \lim_{K\to\infty}\int_0^{-\eta/2}\left|e^{-(K^2-u^2)}\right|\cdot\left|e^{-2iKu}\right|du = 0$$

Similarly,

$$\lim_{K\to\infty}\left|\int_{-\eta/2}^0 e^{-(-K+iu)^2}du\right| = \lim_{K\to\infty}\left|\int_{-\eta/2}^0 e^{-(K-iu)^2}du\right|$$

$$\lim_{K\to\infty}\left|\int_{-\eta/2}^0 e^{-(K^2-u^2-2iKu)}du\right| \leq \lim_{K\to\infty}\int_{-\eta/2}^0\left|e^{-(K^2-u^2)}\right|\cdot\left|e^{-2iKu}\right|du = 0$$

The following integral is known by many. $\int_{-\infty}^{+\infty} e^{-x^2} dx = I$ is evaluated by looking at its square and performing a rectangular to polar coordinate conversion. I have not seen this trick used on any other integral.

$$I^2 = \left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{+\infty} e^{-w^2} dw\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+w^2)} dx\, dw$$

Let $x = r\cos\theta$, $w = r\sin\theta$, and $dx\ dw = r\, dr\, d\theta$. Then

$$I^2 = 4\int_0^{\infty}\int_0^{\pi/2} e^{-r^2} r\, dr\, d\theta = 4\left(\frac{\pi}{2}\right)\int_0^{\infty} e^{-r^2} r\, dr$$

$$= -\pi\int_0^{\infty} e^{-r^2}(-2r)\, dr = -\pi e^{r^2}\Big|_0^{\infty} = -\pi(0-1) = \pi$$

Therefore, $I = \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$.

Substitute these individual results back into the equation for $\oint e^{-z^2} dz$. Switch the limits on the remaining integral, thus changing its sign.

$$\oint e^{-z^2} dz = 0 = \sqrt{\pi} + 0 - \int_{-\infty}^{\infty} e^{-(u-\frac{1}{2}i\eta)^2} du + 0$$

Switching notation of dummy variables from u to x, we get

$$\int_{-\infty}^{\infty} e^{-(x-\frac{1}{2}i\eta)^2} dx = \sqrt{\pi}$$

Substituting this result yields

$$\int_{-\infty}^{\infty} \exp[i\eta x - x^2] dx = \sqrt{\pi}\exp[-\frac{1}{4}\eta^2]$$

Continuing the back-substitution, we get the characteristic function

$$\Phi_z(t) = \frac{1}{\pi}\left(\sqrt{\pi}\exp[-\frac{1}{4}\eta^2]\right)\left(\sqrt{\pi}\exp[-\frac{1}{4}\tau^2]\right) = \frac{1}{\pi}\pi\exp[-\frac{1}{4}(\eta^2+\tau^2)]$$

Finally, for the univariate standard special complex normal distribution, the characteristic function is

$$\Phi_z(t) = \exp[-\frac{1}{4}|t|^2] \tag{D.2}$$

Derivation from the Real Vector Normal Characteristic Function

Let $w = \begin{pmatrix} x \\ y \end{pmatrix}$ have a real normal distribution with mean vector $\mu_w = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ and covariance matrix $S = \begin{pmatrix} G & -F \\ F & G \end{pmatrix}$ where G is positive definite and $F = F^T$. Let $z = x + iy$ have a special complex normal distribution with mean $\mu = \mu_x + i\mu_y$ and covariance matrix

$$\Sigma = \mathcal{E}\{(z-\mu)(z-\mu)^H\} = Q + iR$$

where Σ is positive definite.

The characteristic function of w is known to be

$$\Phi_w(t) = \mathcal{E}\{\exp[it^T w]\} = \exp[it^T\mu_w - \frac{1}{2}t^T S t]$$

where $t = \begin{pmatrix} \eta \\ \tau \end{pmatrix}$ is a real vector. Then

$$\Phi_w(t) = \exp\left[i\begin{pmatrix} \eta^T & \tau^T \end{pmatrix}\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \eta^T & \tau^T \end{pmatrix}\begin{pmatrix} G & -F \\ F & G \end{pmatrix}\begin{pmatrix} \eta \\ \tau \end{pmatrix}\right]$$

$$= \exp\left[i\left(\eta^T\mu_x + \tau^T\mu_y\right) - \frac{1}{2}\left(\eta^T G\eta - \eta^T F\tau + \tau^T F\eta + \tau^T G\tau\right)\right]$$

Regardless of whether the distribution is expressed as paired real vectors or by complex vectors, the density function and the characteristic function should always have the same value for the corresponding identical parameters. With this in mind, let us examine similar forms with complex variables. Let $T = \eta + i\tau$. Then

$$T^H\mu = (\eta^T - i\tau^T)(\mu_x + i\mu_y) = \eta^T\mu_x + \tau^T\mu_y - i\tau^T\mu_x + i\eta^T\mu_y$$

Thus $\mathrm{Re}(T^H\mu) = t^T\mu_w$. Now examine the covariance term.

$$T^H\Sigma T = (\eta^T - i\tau^T)(Q + iR)(\eta + i\tau)$$

$$= \eta^TQ\eta + \tau^TQ\tau + \tau^TR\eta - \eta^TR\tau + i[\eta^TR\eta + \tau^TR\tau - \tau^TQ\eta + \eta^TQ\tau]$$

Recall that we proved $R = 2F$ and $Q = 2G$. Then $T^T\Sigma T =$

$$2\eta^TG\eta + 2\tau^TG\tau + 2\tau^TF\eta - 2\eta^TF\tau + i[2\eta^TF\eta + 2\tau^TF\tau - 2\tau^TG\eta + 2\eta^TG\tau]$$

Comparing this with $\frac{1}{2}t^TSt$, it is seen that

$$\frac{1}{4}\,\mathrm{Re}(T^H\Sigma T) = \frac{1}{2}t^TSt$$

Since Σ is Hermitian positive definite, by theorem 119 it can be factored as $\Sigma = CC^H$. Then

$$T^H\Sigma T = T^HCC^HT = (C^HT)^H(C^HT)$$

which is real. Thus $T^H\Sigma T = \mathrm{Re}[T^H\Sigma T]$, and therefore the characteristic function for the special vector complex normal distribution is

$$\Phi_z(T) = \exp\{i\,\mathrm{Re}[T^H\mu] - \frac{1}{4}T^H\Sigma T\} \qquad \text{(D.3)}$$

This differs from Anderson's result by the $\frac{1}{4}$ in the covariance term.

D.2 Matrix Complex Normal Distribution

The matrix complex normal distribution describes the random data whose quadratic form produces the complex Wishart distribution. In order to define and describe the properties of the complex Wishart distribution, the definition and properties of the complex Gaussian distribution must be understood. The material that follows is a complexification of Arnold's Section 17.2 [31]. I have also used characteristic functions where Arnold used moment generating functions.

D.2.1 Definition of the Matrix Complex Normal Distribution

Let $Z = (Z_{ij})$ be an $n \times p$ random matrix such that the $\{Z_{ij}\}$ are independent and each Z_{ij} is distributed according to the univariate complex normal distribution with zero mean and unit variance. Symbolically, we denote this as $Z_{ij} \sim CN_1(0,1)$. The characteristic function from equation D.2 is

$$\Phi_{Z_{ij}}(T_{ij}) = \exp(-\frac{1}{4}T_{ij}^* T_{ij}) \tag{D.4}$$

For $Z = (Z_{ij})_{n \times p}$ independent and identically distributed, then

$$\Phi_Z(T) = \prod_{j=1}^{p}\prod_{i=1}^{n} \exp(-\frac{1}{4}T_{ij}^* T_{ij}) = \exp[-\frac{1}{4}\operatorname{tr}(T^H T)] \tag{D.5}$$

Likewise, the density function (from equation D.1) is given by

$$f(Z) = \prod_{j=1}^{p} \prod_{i=1}^{n} \frac{1}{\pi} \exp(-Z_{ij}^{*} Z_{ij}) = \frac{1}{\pi^{pn}} \exp[-\operatorname{tr}(Z^H Z)] \tag{D.6}$$

Let $A \in \mathbf{C}^{n \times n}$, $B \in \mathbf{C}^{p \times r}$, $\mu \in \mathbf{C}^{m \times r}$, and $Y = AZB + \mu$. The symbol μ in statistics is often reserved to refer to the average or mean of a distribution, and it is often (but not necessarily) a parameter in the distribution functions. Then $Y \in \mathbf{C}^{m \times r}$ and

$$\begin{aligned} \Phi_Y(T) = \Phi_{AZB+\mu}(T) &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] \right\} \Phi_Z(A^H T B^H) \qquad \text{(D.7)} \\ &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] \right\} \exp\left\{ -\frac{1}{4} \operatorname{tr}\left[(A^H T B^H)^H (A^H T B^H) \right] \right\} \\ &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] - \frac{1}{4} \operatorname{tr}(B T^H A A^H T B^H) \right\} \\ &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] - \frac{1}{4} \operatorname{tr}(T^H A A^H T B^H B) \right\} \end{aligned}$$

Let

$$\Xi = AA^H \text{ and } \Sigma = B^H B \tag{D.8}$$

These symbols have special meanings in statistics. The symbol Σ is the more frequently used symbol, and it represents the covariance matrix of a distribution. When more than one covariance matrix is being discussed, the symbol Ξ is often the symbol of choice. In the case of the matrix complex normal distribution, Arnold [31] remarks that it helps to think of Ξ as representing the covariance between the *rows* of Y, and Σ as representing the covariance between the *columns* of Y. With these symbols defined, then

$$\Phi_Y(T) = \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] - \frac{1}{4} \operatorname{tr}(T^H \Xi T \Sigma) \right\} \tag{D.9}$$

Y has a special complex matrix normal distribution with parameters μ, Ξ, and Σ. We denote this by $Y \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. Note that this is not unique. For arbitrary scalar $a \in \mathbf{C}$, it is true that

$$Y \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma) = \mathbf{C}N_{m,r}(\mu, \frac{1}{a}\Xi, a\Sigma). \tag{D.10}$$

The parameter Ξ is often assumed known, and usually assumed to be the identity matrix I_m. Note the dimensions on the parameters: $\mu_{m\times r}$, $\Xi_{m\times m}$, and $\Sigma_{r\times r}$.

Proposition 38 *If* $z \sim \mathbf{C}N_1(0,1)$ *then* $z^* \sim \mathbf{C}N_1(0,1)$.

Proof. Let $z = x + iy$. Then $z \sim \mathbf{C}N_1(0,1)$ implies the expected value $\mathcal{E}(x+iy) = 0$ which implies and $\mathcal{E}(x) = 0$ and $\mathcal{E}(y) = 0$. We also see that the variance

$$\text{var}(z) = \mathcal{E}[(z-0)(z-0)^*] = \mathcal{E}(zz^*)$$

$$= \mathcal{E}[(x+iy)(x-iy)] = \mathcal{E}(x^2+y^2) = 1$$

$\mathcal{E}(x) = 0$ and $\mathcal{E}(y) = 0$ implies $\mathcal{E}(x-iy) = \mathcal{E}(z^*) = 0$. Then

$$1 = \mathcal{E}[(x+iy)(x-iy)] = \mathcal{E}(zz^*) = \mathcal{E}[z^*(z^*)^*]$$

$$= \mathcal{E}[(z^*-0)(z^*-0)^*] = \text{var}(z^*)$$

Therefore $\text{var}(z) = \text{var}(z^*)$. This completely defines the distribution. We conclude that

$$z \sim \mathbf{C}N_1(0,1) \Leftrightarrow z^* \sim \mathbf{C}N_1(0,1)$$

Proposition 39 *Let* $Y \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. *Let* $Y = (Y_{ij})$, $\mu = (\mu_{ij})$, $\Xi = (\Xi_{ij})$, *and* $\Sigma = (\Sigma_{ij})$. *Then* $\mathcal{E}(Y_{ij}) = \mu_{ij}$, $\text{var}(Y_{ij}) = \Xi_{ii}\Sigma_{jj}$, *and* $\text{cov}(Y_{ij}, Y_{i^*j^*}) = \Xi_{ii^*}\Sigma^*_{jj^*}$. *This is the complexification of Arnold's theorem 17.1 [31].*

Proof. This is a complexification of Arnold's proof. Let $Z_{ij} \sim CN(0,1)$. Then $\mathcal{E}(Z_{ij}) = 0$, $\text{var}(Z_{ij}) = 1$, and $\text{cov}(Z_{ij}, Z_{i^*j^*}) = 0$ unless $i = i^*$ and $j = j^*$. Define matrices A and B by $\Xi = AA^H$, $\Sigma = B^HB$. These factorizations exist for Hermitian positive definite Ξ and Σ as proven by theorem 119. Let $Y = AZB + \mu$. Then $Y \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$ by construction. Element Y_{ij} is given by

$$Y_{ij} = \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj} + \mu_{ij} \tag{D.11}$$

Then

$$\text{cov}(Y_{ij}, Y_{i^*j^*}) = \mathcal{E}\left\{\left[Y_{ij} - \mathcal{E}(Y_{ij})\right]\left[Y_{i^*j^*} - \mathcal{E}(Y_{i^*j^*})\right]^*\right\} \tag{D.12}$$

We consider the details of one of the arguments and note that the other argument has similar results.

$$\begin{aligned} Y_{ij} - \mathcal{E}(Y_{ij}) &= \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj} + \mu_{ij} - \mathcal{E}\left\{\sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj} + \mu_{ij}\right\} \\ &= \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj} + \mu_{ij} - \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}\mathcal{E}\left\{Z_{ks}\right\}B_{sj} + \mu_{ij} \end{aligned}$$

which implies

$$Y_{ij} - \mathcal{E}(Y_{ij}) = \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj} \tag{D.13}$$

We use this to evaluate the covariance term of D.12.

$$\text{cov}(Y_{ij}, Y_{i^*j^*}) = \mathcal{E}\left\{\left[\sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}Z_{ks}B_{sj}\right]\left[\sum_{k^*=1}^{n}\sum_{s^*=1}^{p} A_{i^*k^*}Z_{k^*s^*}B_{s^*j^*}\right]^*\right\}$$

$$= \sum_{k=1}^{n}\sum_{s=1}^{p}\sum_{k^*=1}^{n}\sum_{s^*=1}^{p} A_{ik}B_{sj}A^*_{i^*k^*}B^*_{s^*j^*} \times \underbrace{E\{Z_{ks}Z^*_{k^*s^*}\}}_{=0 \text{ except when } k^*=k,\ s^*=s}$$

$$= \sum_{k=1}^{n}\sum_{s=1}^{p} A_{ik}B_{sj}A^*_{i^*k}B^*_{sj^*} = \sum_{k=1}^{n} A_{ik}A^*_{i^*k}\sum_{s=1}^{p} B_{sj}B^*_{sj^*} = \Xi_{ii^*}\Sigma^*_{jj^*}$$

Therefore,

$$\mathrm{cov}(Y_{ij}, Y_{i^*j^*}) = \Xi_{ii^*}\Sigma^*_{jj^*} \tag{D.14}$$

The variance term is simpler to compute.

$$\mathrm{var}(Y_{ij}) = \mathrm{cov}(Y_{ij}, Y_{ij}) = \Xi_{ii}\Sigma^*_{jj} = \Xi_{ii}\Sigma_{jj} \tag{D.15}$$

since elements on the diagonal are in $\mathbf{R}$.

By this theorem, then, the covariance between two elements Y_{ij} and $Y_{i^*j^*}$ is just the covariance between the rows i and i^* multiplied by the conjugate of the covariance between the columns j and j^*.

D.2.2 Properties of the Matrix Complex Normal Distribution.

The properties studied here are the complexification of Arnold's theorems 17.2 and 17.3 [31], plus some corollaries motivated by these theorems.

Lemma 12 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. *If* $r = 1$ *and* $\Sigma = \sigma^2$ *(a scalar), then* $Z \sim \mathbf{C}N_m(\mu, \sigma^2\Xi)$.

Proof. The characteristic function of Z from equation D.9 is given by

$$\Phi_Z(T) = \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}(T^H\mu)\right) - \frac{1}{4}\mathrm{tr}(T^H\Xi T\Sigma)\right]$$

$$= \exp\left[i \operatorname{Re}\left(\operatorname{tr}(T^H \mu)\right) - \frac{1}{4}\operatorname{tr}(T^H [\sigma^2 \Xi] T)\right]$$

which is the characteristic function of $\mathbf{C}N_m(\mu, \sigma^2\Xi)$ by equation D.3. This is a complexification of Arnold's theorem 17.2(a), which was stated without proof.

Lemma 13 *Let $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. If α is a scalar, then*

$$\alpha Z \sim \mathbf{C}N_{m,r}(\alpha\mu, |\alpha|^2 \Xi, \Sigma) = \mathbf{C}N_{m,r}(\alpha\mu, \alpha^*\Xi, \alpha\Sigma)$$

$$= \mathbf{C}N_{m,r}(\alpha\mu, \alpha\Xi, \alpha^*\Sigma) = \mathbf{C}N_{m,r}(\alpha\mu, \Xi, |\alpha|^2 \Sigma) \qquad \text{(D.16)}$$

This is a complexification of Arnold's theorem 17.2(b), which was stated without proof.

Proof. From the characteristic function, we observe

$$\Phi_{\alpha Z}(T) = \mathcal{E}\left\{\exp\left[i \operatorname{Re}\left(\operatorname{tr}[T^H(\alpha Z)]\right)\right]\right\}$$

$$= \mathcal{E}\left\{\exp\left[i \operatorname{Re}\left(\operatorname{tr}[(\alpha^* T)^H Z]\right)\right]\right\} = \Phi_Z(\alpha^* T)$$

since α is a scalar. We continue by regrouping scalar α to obtain the results. Since α and α^* are scalars, they commute.

$$\Phi_Z(\alpha^* T) = \exp\left[i \operatorname{Re}\left(\operatorname{tr}[(\alpha^* T)^H \mu]\right) - \frac{1}{4}\operatorname{tr}\left((\alpha^* T)^H \Xi (\alpha^* T)\Sigma\right)\right]$$

$$= \exp\left[i \operatorname{Re}\left(\operatorname{tr}[T^H(\alpha\mu)]\right) - \frac{1}{4}\operatorname{tr}\left(T^H(\alpha\Xi)T(\alpha^*\Sigma)\right)\right]$$

These are the characteristic functions of the distributions cited in the results.

Corollary 9 *Let $Z \sim CN_{m,r}(\mu, \Xi, \Sigma)$. Then $Z^H \sim CN_{r,m}(\mu^H, \Sigma, \Xi)$. This is a complexification of Arnold's theorem 17.2(c), which was stated without proof.*

Proof. Let $Z = AXB + \mu$. Then $Z^H = B^H X^H A^H + \mu^H$ where $X \sim CN(0, I, I)$. The characteristic function of Z^H is given by

$$\Phi_{Z^H}(T) = \mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H(B^H X^H A^H + \mu^H)]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[A^H T^H B^H X^H]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\Phi_{X^H}(BTA)$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[(BTA)^H(BTA)]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[A^H T^H B^H BTA]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[T^H B^H BTAA^H]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^H]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[T^H\Sigma T\Xi]\right]$$

which is the characteristic function of the result, where $\Xi = AA^H$ and $\Sigma = B^H B$ as used earlier.

Theorem 41 *(*Very important*) Let $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$ where the matrix dimensions are $Z_{m\times r}$, $\mu_{m\times r}$, $\Xi_{m\times m}$, and $\Sigma_{r\times r}$. Let $Y = AZB + \nu$ where the dimensions are $Y_{n\times p}$, $A_{n\times m}$, $B_{r\times p}$, and $\nu_{n\times p}$. Then*

$$Y \sim \mathbf{C}N_{n,p}(\nu + A\mu B, A\Xi A^H, B^H\Sigma B)$$

This is a complexification of Arnold's theorem 17.2(d) [31], which was stated without proof.

Proof. From a practical standpoint for the future use of other people, this is one of the most important results in this thesis. The characteristic function is given by

$$\Phi_{AZB+\nu}(T) = \mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H(AZB+\nu)]\right)\right]\right\}$$

$$= \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H\nu]\right)\right]\mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[BT^HAZ]\right)\right]\right\}$$

$$= \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H\nu]\right)\right]\Phi_Z(A^HTB^H)$$

$$= \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H\nu]\right)\right]$$

$$\times \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[(A^HTB^H)^H\mu]\right) - \frac{1}{4}\mathrm{tr}[(A^HTB^H)^H\Xi(A^HTB^H)\Sigma]\right]$$

$$= \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H\nu] + \mathrm{tr}[BT^HA\mu]\right) - \frac{1}{4}\mathrm{tr}[BT^HA\Xi A^HTB^H\Sigma]\right]$$

$$= \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H(\nu+A\mu B)]\right) - \frac{1}{4}\mathrm{tr}[T^H(A\Xi A^H)T(B^H\Sigma B)]\right] = \Phi_Y(T) \tag{D.17}$$

which is the characteristic function of the result.

Theorem 42 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu,\Xi,\Sigma)$ *where the matrix dimensions are* $Z_{m\times r}$, $\mu_{m\times r}$, $\Xi_{m\times m}$, *and* $\Sigma_{r\times r}$. *Partition the random variable and parameters as follows. Let* $Z = (Z_1, Z_2)$, $\mu = (\mu_1,\mu_2)$, *and* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ *where* Z_1 *and* μ_1 *are* $m \times r_1$ *and* Σ_{11} *is* $r_1 \times r_1$. *Then* $Z_1 \sim \mathbf{C}N_{m,r_1}(\mu_1,\Xi,\Sigma_{11})$ *and* $Z_2 \sim \mathbf{C}N_{m,(r-r_1)}(\mu_2,\Xi,\Sigma_{22})$. *This is a complexification of Arnold's theorem 17.2(e) [31], which was stated without proof.*

Proof. Let $B = \begin{pmatrix} I_{r_1} \\ 0 \end{pmatrix}$. Then

$$Z_1 = ZB \sim \mathbf{C}N_{m,r_1}(\mu B, \Xi, B^H \Sigma B) = \mathbf{C}N_{m,r_1}(\mu_1, \Xi, \Sigma_{11})$$

by theorem 41. Similarly, let $D = \begin{pmatrix} 0 \\ I_{r-r_1} \end{pmatrix}$. Then

$$Z_2 = ZD \sim \mathbf{C}N_{m,(r-r_1)}(\mu D, \Xi, D^H \Sigma D) = \mathbf{C}N_{m,(r-r_1)}(\mu_2, \Xi, \Sigma_{22})$$

Theorem 43 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. *Partition the random variable and parameters as follows. Let* $Z = (Z_1, Z_2)$, $\mu = (\mu_1, \mu_2)$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, *and* $T = \begin{pmatrix} T_1 & T_2 \end{pmatrix}$. *Let* $\Xi \neq 0$. *Then* Z_1 *and* Z_2 *are independent if and only if* $\Sigma_{12} = 0$. *This is a complexification of Arnold's theorem 17.2(f) [31], which was stated without proof.*

Proof. Working with the characteristic function, equation D.9, we see that

$$\Phi_Z(T) = \exp\left[i \operatorname{Re}\left(\operatorname{tr}\left\{ \begin{pmatrix} T_1^H \\ T_2^H \end{pmatrix} (\mu_1, \mu_2) \right\} \right) \right.$$

$$\left. -\frac{1}{4} \operatorname{tr}\left\{ \begin{pmatrix} T_1^H \\ T_2^H \end{pmatrix} \Xi \begin{pmatrix} T_1 & T_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right\} \right]$$

$$= \exp\left[i \operatorname{Re}\left(\operatorname{tr}\left\{ \begin{pmatrix} T_1^H \mu_1 & T_1^H \mu_2 \\ T_2^H \mu_1 & T_2^H \mu_2 \end{pmatrix} \right\} \right) \right.$$

$$-\frac{1}{4}\operatorname{tr}\left\{\begin{pmatrix} T_1^H\Xi \\ T_2^H\Xi \end{pmatrix}\begin{pmatrix} T_1\Sigma_{11}+T_2\Sigma_{21} & T_1\Sigma_{12}+T_2\Sigma_{22} \end{pmatrix}\right\}\Bigg]$$

$$= \exp\Big[i\operatorname{Re}\Big(\operatorname{tr}[T_1^H\mu_1]+\operatorname{tr}[T_2^H\mu_2]\Big)$$

$$-\frac{1}{4}\operatorname{tr}[T_1^H\Xi T_1\Sigma_{11}+T_1^H\Xi T_2\Sigma_{21}]-\frac{1}{4}\operatorname{tr}[T_2^H\Xi T_1\Sigma_{12}+T_2^H\Xi T_2\Sigma_{22}]\Big]$$

$$= \exp\left[i\operatorname{Re}\Big(\operatorname{tr}[T_1^H\mu_1]\Big)-\frac{1}{4}\operatorname{tr}[T_1^H\Xi T_1\Sigma_{11}]\right]\times$$

$$\times\exp\left[i\operatorname{Re}\Big(\operatorname{tr}[T_2^H\mu_2]\Big)-\frac{1}{4}\operatorname{tr}[T_2^H\Xi T_2\Sigma_{22}]\right]\times$$

$$\times\exp\left[-\frac{1}{4}\operatorname{tr}[T_1^H\Xi T_2\Sigma_{21}]-\frac{1}{4}\operatorname{tr}[T_2^H\Xi T_1\Sigma_{12}]\right]$$

$$= \Phi_{Z_1}(T_1)\Phi_{Z_2}(T_2) \text{ if and only if } \Sigma_{12}=0 \qquad \text{(D.18)}$$

By the Neyman-Fisher factorization theorem, Z_1 and Z_2 are independent if and only if $\Sigma_{12}=0$.

Theorem 44 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu,\Xi,\Sigma)$. *Partition the random variable and parameters as follows. Let* $Z=(Z_1,Z_2)$, $\mu=(\mu_1,\mu_2)$, $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$. *Let* Σ_{22} *be nonsingular and define* $\Sigma_{11.2}\underline{\text{def}}\,\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. *Then the conditional distribution of* Z_1 *given* Z_2 *is given by*

$$(Z_1 \mid Z_2) \sim \mathbf{C}N_{m,r_1}\Big((\mu_1+(Z_2-\mu_2)\Sigma_{22}^{-1}\Sigma_{21}),\Xi,(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\Big)$$

$$= \mathbf{C}N_{m,r_1}\Big((\mu_1+(Z_2-\mu_2)\Sigma_{22}^{-1}\Sigma_{21}),\Xi,\Sigma_{11.2}\Big)$$

This is a complexification of Arnold's theorem 17.2(g) [31], which was stated without proof.

Proof. Let $B = \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix}$ and consider the transformation

$$Y = (Y_1, Y_2) = ZB = (Z_1, Z_2)B$$

Then

$$(Y_1, Y_2) = (Z_1 - Z_2\Sigma_{22}^{-1}\Sigma_{21}, Z_2)$$

Thus $\mu B = (\mu_1 - \mu_2\Sigma_{22}^{-1}\Sigma_{21}, \mu_2)$. The covariance is found by

$$B^H\Sigma B = \begin{pmatrix} I & -\Sigma_{21}^H\Sigma_{22}^{-H} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{21}^H\Sigma_{22}^{-H}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{21}^H\Sigma_{22}^{-H}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12} - \Sigma_{21}^H\Sigma_{22}^{-H}\Sigma_{22} \\ \Sigma_{21} - \Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \tag{D.19}$$

where $\Sigma = \Sigma^H$. By theorem 41, $(Y_1, Y_2) \sim \mathbf{C}N_{m,r}(\mu B, \Xi, B^H\Sigma B)$. By theorem 43, Y_1 and Y_2 are independent. Since Y_1 and Y_2 are independent, then the density factors as

$$f(Y_1 \mid Y_2) = \frac{f(Y_1, Y_2)}{f(Y_2)} = \frac{f(Y_1)f(Y_2)}{f(Y_2)} = f(Y_1) \tag{D.20}$$

Thus

$$(Y_1 \mid Y_2) \sim \mathbf{C}N_{m,r_1}((\mu_1 - \mu_2\Sigma_{22}^{-1}\Sigma_{21}), \Xi, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})) \tag{D.21}$$

Recall that $Y_2 = Z_2$ and $Y_1 = Z_1 - Z_2\Sigma_{22}^{-1}\Sigma_{21}$ which implies

$$Z_1 = Y_1 + Z_2\Sigma_{22}^{-1}\Sigma_{21}$$

In the conditional density $f(Y_1 \mid Y_2 = Z_2)$, Z_2 is a constant. Apply theorem 41 to find $f(Z_1 \mid Z_2)$. Here, $\nu = Z_2\Sigma_{22}^{-1}\Sigma_{21}$. The row and column covariance matrices remain the same, and the mean becomes

$$\mu_1 - \mu_2\Sigma_{22}^{-1}\Sigma_{21} + Z_2\Sigma_{22}^{-1}\Sigma_{21} = \mu_1 + (Z_2 - \mu_2)\Sigma_{22}^{-1}\Sigma_{21}$$

Thus

$$(Z_1 \mid Z_2) \sim \mathbf{C}N_{m,r_1}((\mu_1 + (Z_2 - \mu_2)\Sigma_{22}^{-1}\Sigma_{21}), \Xi, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}))$$

$$= \mathbf{C}N_{m,r_1}((\mu_1 + (Z_2 - \mu_2)\Sigma_{22}^{-1}\Sigma_{21}), \Xi, \Sigma_{11.2}) \qquad \text{(D.22)}$$

Corollary 10 *Let $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. Then $Z^* \sim \mathbf{C}N_{r,m}(\mu^*, \Xi^*, \Sigma^*)$. This is a variation on the complexification of Arnold's theorem 17.2(c) [31].*

Proof. Let $X \sim \mathbf{C}N(0, I, I)$ and $Z = AXB + \mu$ where $\Sigma = B^HB$ and $\Xi = AA^H$. Then $Z^* = A^*X^*B^* + \mu^*$. Using the characteristic function, we see

$$\Phi_{Z^*}(T) = \mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H(A^*X^*B^* + \mu^*)]\right)\right]\right\}$$

$$= \mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^HA^*X^*B^* + T^H\mu^*]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^HA^*X^*B^*]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[(B^*T^HA^*)X^*]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\mathcal{E}\left\{\exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[(A^TTB^T)^HX^*]\right)\right]\right\}$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\Phi_{Z^*}(A^TTB^T)$$

By proposition 38, $\Phi_Z(T) = \Phi_{Z^*}(T)$. Thus

$$\Phi_{Z^*}(T) = \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\Phi_Z(A^TTB^T)$$

By equation D.5 we find that

$$\Phi_{Z^*}(T) = \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[(A^TTB^T)^H(A^TTB^T)]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[B^*T^HA^*A^TTB^T]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[T^HA^*A^TTB^TB^*]\right]$$

$$= \exp\left[i\,Re\left(tr[T^H\mu^*]\right)\right]\exp\left[-\frac{1}{4}\,\mathrm{tr}[T^H\Xi^*T\Sigma^*]\right]$$

This is the characteristic function of a variable distributed as

$$Z^* \sim \mathbf{C}N_{r,m}(\mu^*, \Xi^*, \Sigma^*)$$

Theorem 45 *Let $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$ where the matrix dimensions are $Z_{m\times r}$, $\mu_{m\times r}$, $\Xi_{m\times m}$, and $\Sigma_{r\times r}$. Partition the random variable and parameters as follows. Let $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, and $\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix}$ where Z_1 and μ_1 are $m_1 \times r$ and Ξ_{11} is $m_1 \times m_1$. Then $Z_1 \sim \mathbf{C}N_{m_1,r}(\mu_1, \Xi_{11}, \Sigma)$ and*

$$Z_2 \sim \mathbf{C}N_{(m-m_1),r}(\mu_2, \Xi_{22}, \Sigma)$$

This is a variation on a complexification of Arnold's theorem 17.2(e).

Proof. Let $A = (I_{m_1}, 0)$. Then by theorem 41,

$$Z_1 = AZ \sim \mathbf{C}N_{m_1,r}(A\mu, A\Xi A^H, \Sigma) = \mathbf{C}N_{m_1,r}(\mu_1, \Xi_{11}, \Sigma) \tag{D.23}$$

Likewise, Let $B = (0, I_{m-m_1})$. Then again by theorem 41,

$$Z_2 = BZ \sim \mathbf{C}N_{(m-m_1),r}(B\mu, B\Xi B^H, \Sigma) = \mathbf{C}N_{(m-m_1),r}(\mu_2, \Xi_{22}, \Sigma) \tag{D.24}$$

Theorem 46 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. *Partition the random variable and parameters as follows. Let* $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix}$, *and* $T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$. *Let* $\Sigma \neq 0$. *Then* Z_1 *and* Z_2 *are independent if and only if* $\Xi_{12} = 0$. *This is a variation of a complexification of Arnold's theorem 17.2(f) [31].*

Proof. The characteristic function of Z is given by

$$\Phi_Z(T) = \exp\left[i\, Re\left(tr\left\{ (T_1^H, T_2^H) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right\} \right) \right.$$

$$\left. -\frac{1}{4} \operatorname{tr}\left\{ (T_1^H, T_2^H) \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \Sigma \right\} \right]$$

$$= \exp\left[i\, Re\left(tr[T_1^H \mu_1 + T_2^H \mu_2] \right) \right.$$

$$\left. -\frac{1}{4} \operatorname{tr}\left\{ \left(T_1^H \Xi_{11} T_1 + T_2^H \Xi_{21} T_1 + T_1^H \Xi_{12} T_2 + T_2^H \Xi_{22} T_2 \right) \Sigma \right\} \right]$$

$$= \Phi_{Z_1}(T_1) \Phi_{Z_2}(T_2) \exp\left[-\frac{1}{4} \operatorname{tr}[\left(T_2^H \Xi_{21} T_1 + T_1^H \Xi_{12} T_2 \right) \Sigma] \right]$$

$$= \Phi_{Z_1}(T_1)\Phi_{Z_2}(T_2) \text{ if and only if } \Xi_{12} = \Xi_{21}^H = 0 \tag{D.25}$$

Thus by the Neyman-Fisher factorization theorem, Z_1 and Z_2 are independent if and only if $\Xi_{12} = 0$.

Theorem 47 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. *Partition the random variable and parameters as follows. Let* $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix}$. *Let* Ξ_{22} *be nonsingular and define* $\Xi_{11.2} \underline{\underline{\text{def}}} \Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21}$. *Then the conditional distribution of* Z_1 *given* Z_2 *is given by*

$$(Z_1 \mid Z_2) \sim \mathbf{C}N_{m_1,r}\left((\mu_1 + \Xi_{21}\Xi_{22}^{-1}(Z_2 - \mu_2)), (\Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21}), \Sigma\right)$$

$$= \mathbf{C}N_{m_1,r}\left((\mu_1 + \Xi_{21}\Xi_{22}^{-1}(Z_2 - \mu_2)), \Xi_{11.2}, \Sigma\right)$$

This is a variation on a complexification of Arnold's theorem 17.2(g) [31].

Proof. Let $A = \begin{pmatrix} I & -\Xi_{12}\Xi_{22}^{-1} \\ 0 & I \end{pmatrix}$ and consider the transformation

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = AZ = A \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

Then

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} Z_1 - \Xi_{12}\Xi_{22}^{-1}Z_2 \\ Z_2 \end{pmatrix} \tag{D.26}$$

Thus the mean of the transformed random variable is

$$A\mu = \begin{pmatrix} \mu_1 - \Xi_{12}\Xi_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}$$

and its covariance is

$$A\Xi A^H = \begin{pmatrix} I & -\Xi_{12}\Xi_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Xi_{22}^{-H}\Xi_{12}^{H} & I \end{pmatrix}$$

$$\begin{pmatrix} \Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21} & \Xi_{12} - \Xi_{12}\Xi_{22}^{-1}\Xi_{22} \\ \Xi_{21} & \Xi_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Xi_{22}^{-H}\Xi_{12}^{H} & I \end{pmatrix}$$

$$= \begin{pmatrix} \Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21} - \Xi_{12}\Xi_{22}^{-H}\Xi_{12}^{H} + \Xi_{12}\Xi_{22}^{-1}\Xi_{22}\Xi_{22}^{-H}\Xi_{12}^{H} & \Xi_{12} - \Xi_{12}\Xi_{22}^{-1}\Xi_{22} \\ \Xi_{21} - \Xi_{22}\Xi_{22}^{-H}\Xi_{12}^{H} & \Xi_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21} & 0 \\ 0 & \Xi_{22} \end{pmatrix} \text{ where } \Xi = \Xi^H \qquad \text{(D.27)}$$

By theorem 41, $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathbf{C}N_{m,r}(A\mu, A\Xi A^H, \Sigma)$. By theorem 46, Y_1 and Y_2 are independent and therefore

$$f(Y_1 \mid Y_2) = \frac{f(Y_1, Y_2)}{f(Y_2)} = \frac{f(Y_1)f(Y_2)}{f(Y_2)} = f(Y_1) \qquad \text{(D.28)}$$

Thus

$$(Y_1 \mid Y_2) \sim \mathbf{C}N_{m_1,r}((\mu_1 - \Xi_{12}\Xi_{22}^{-1}\mu_2), (\Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21}), \Sigma) \qquad \text{(D.29)}$$

Recall that $Y_2 = Z_2$ and $Y_1 = Z_1 - \Xi_{12}\Xi_{22}^{-1}Z_2$ which implies $Z_1 = Y_1 + \Xi_{12}\Xi_{22}^{-1}Z_2$. In the conditional density $f(Y_1 \mid Y_2 = Z_2)$, Z_2 is a constant. Apply theorem 41 to find $f(Z_1 \mid Z_2)$. Here $\nu = \Xi_{12}\Xi_{22}^{-1}Z_2$. The distribution mean is

$$\mu_1 - \Xi_{12}\Xi_{22}^{-1}\mu_2 + \Xi_{12}\Xi_{22}^{-1}Z_2 = \mu_1 + \Xi_{12}\Xi_{22}^{-1}(Z_2 - \mu_2) \qquad \text{(D.30)}$$

Therefore the conditional distribution of Z_1 given Z_2 is

$$(Z_1 \mid Z_2) \sim \mathbf{C}N_{m_1,r}((\mu_1 + \Xi_{12}\Xi_{22}^{-1}(Z_2 - \mu_2)), (\Xi_{11} - \Xi_{12}\Xi_{22}^{-1}\Xi_{21}), \Sigma) \quad \text{(D.31)}$$

Theorem 48 *Let $X \sim \mathbf{C}N_{p,m}(\mu_1, \Xi_1, \Sigma_1)$ and $Y \sim \mathbf{C}N_{p,m}(\mu_2, \Xi_2, \Sigma_2)$ be independent matrix complex normal random variables of the same size matrices. Then the distribution of the sum $X + Y$ is given by*

$$X + Y \sim \mathbf{C}N_{p,m}(\mu_1 + \mu_2, \Xi_1 + \Xi_2, \Sigma_1 + \Sigma_2)$$

This common theorem was supplied by me.

Proof. Since X and Y are independent, their joint distribution is given by

$$Z = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \sim \mathbf{C}N_{2p,2m}\left[\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \begin{pmatrix} \Xi_1 & 0 \\ 0 & \Xi_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}\right]$$

Define $A = \begin{pmatrix} I_p & I_p \end{pmatrix}$ and $B = \begin{pmatrix} I_m \\ I_m \end{pmatrix}$. Then $AZB = X + Y$. By theorem 41, the distribution of $X + Y$ is

$$X + Y \sim \mathbf{C}N_{p,m}(\mu_1 + \mu_2, \Xi_1 + \Xi_2, \Sigma_1 + \Sigma_2)$$

where we observe

$$A\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} B = \begin{pmatrix} I_p & I_p \end{pmatrix}\begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}\begin{pmatrix} I_m \\ I_m \end{pmatrix} = \mu_1 + \mu_2$$

This is a $p \times m$ matrix. Also,

$$A\begin{pmatrix} \Xi_1 & 0 \\ 0 & \Xi_2 \end{pmatrix} A^H = \Xi_1 + \Xi_2$$

and

$$B^H \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} B = \Sigma_1 + \Sigma_2$$

□

D.2.3 Specialization to the Vector Complex Normal Distribution

We specialize a very few results to the vector complex normal distribution since this is the form most engineers finish there statistical preparation with. Let $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ be an n-dimensional random vector such that the z_j are independent and each element is distributed according to the standard univariate complex normal distribution $CN_1(0,1)$. The characteristic function of the individual elements is $\Phi_{z_j}(t) = \exp\left[\frac{1}{4}t_j^* t_j\right]$. For $z = (z_j)_n$ independent and identically distributed, then the characteristic function of the vector random variable is given by

$$\Phi_z(t) = \prod_{j=1}^{n} \exp\left[-\frac{1}{4}t_j^* t_j\right] = \exp\left[-\frac{1}{4}t^H t\right] \tag{D.32}$$

The density function for the vector random variable of independent and identically distributed univariate complex Gaussian elements is given by

$$f(z) = \prod_{i=1}^{n} \frac{1}{\pi} \exp\left[-z_j^* z_j\right] = \frac{1}{\pi^n} \exp\left[-z^H z\right] = \frac{1}{\pi^n} \operatorname{etr}\left[-z^H z\right] \tag{D.33}$$

We denote the standardized vector complex normal distribution by $\mathbf{C}N_n(0, I_n)$.

Corollary 11 *Let* $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \sim \mathbf{C}N_n(0, I_n)$, $A \in \mathbf{C}^{m\times n}$, $\mu \in \mathbf{C}^m$, $\Xi = AA^H$, *and* $y = Az + \mu$. *Then* $y \sim CN_m(\mu, \Xi)$.

Proof. By the transformation $y = Az + \mu$, we see that $y \in \mathbf{C}^m$. By theorem 18 from properties of a characteristic function, we see that

$$\Phi_y(t) = \Phi_{Az+\mu}(t) = \exp\left\{i \operatorname{Re}\left[t^H\mu\right]\right\} \Phi_z(A^H t) \tag{D.34}$$

$$= \exp\left\{i \operatorname{Re}\left[t^H\mu\right]\right\} \exp\left[-\frac{1}{4}(A^H t)^H (A^H t)\right] = \exp\left\{i \operatorname{Re}\left[t^H\mu\right] - \frac{1}{4} t^H AA^H t\right\}$$

$$= \exp\left\{i \operatorname{Re}\left[t^H\mu\right] - \frac{1}{4} t^H \Xi t\right\}$$

Thus $y \sim \mathbf{C}N_m(\mu, \Xi)$. The characteristic function presented here differs from that given by problem 2.66 of Anderson [26].

Corollary 12 *Let*

$$z = \begin{pmatrix} z_1 & \cdots & z_n \end{pmatrix} \sim \mathbf{C}N_n(0, I_n)$$

$B \in \mathbf{C}^{n\times p}$,

$$\nu = \begin{pmatrix} \nu_1 & \cdots & \nu_p \end{pmatrix} \in \mathbf{C}^p$$

$\Sigma = B^H B$ *and* $y = zB + \nu$. *Then* $y \sim \mathbf{C}N_p(\nu, \Sigma)$.

Proof. For independent and identically distributed $(z_j)_n$, we get the characteristic function

$$\Phi_z(t) = \prod_{i=1}^{n} \exp\left[-\frac{1}{4} t_j^* t_j\right] = \prod_{i=1}^{n} \exp\left[-\frac{1}{4} t_j t_j^*\right] \tag{D.35}$$

$$= \prod_{i=1}^{n} \exp\left[-\frac{1}{4} t t^H\right] = \prod_{i=1}^{n} \exp\left[-\frac{1}{4} \operatorname{tr}\left[t^H t\right]\right]$$

where $t = \begin{pmatrix} t_1 & \cdots & t_n \end{pmatrix}$. The density function is given by

$$f(z) = \prod_{i=1}^{n} \frac{1}{\pi} \exp\left[-z_j^* z_j\right] = \prod_{i=1}^{n} \frac{1}{\pi} \exp\left[-z_j z_j^*\right] \tag{D.36}$$

$$= \frac{1}{\pi^n} \exp\left[-z z^H\right] = \frac{1}{\pi^n} \exp\left[-\operatorname{tr}\left[z^H z\right]\right]$$

By the transformation $y = zB + \nu$, we see that $y \in \mathbf{C}^p$. From theorem 18, the characteristic is given by

$$\Phi_y(t) = \Phi_{zB+\nu}(t) = \exp\left\{i \operatorname{Re}\left[\operatorname{tr}(t^H \nu)\right]\right\} \Phi_z(tB^H) \tag{D.37}$$

where we retain the notation t but modify it so that it is now $t = \begin{pmatrix} t_1 & \cdots & t_p \end{pmatrix}$. Then

$$\Phi_y(t) = \exp\left\{i \operatorname{Re}\left[\operatorname{tr}(t^H \nu)\right]\right\} \exp\left[-\frac{1}{4} \operatorname{tr}(B t^H t B^H)\right] \tag{D.38}$$

$$= \exp\left\{i \operatorname{Re}\left[\operatorname{tr}(t^H \nu)\right]\right\} \exp\left[-\frac{1}{4} \operatorname{tr}(t B^H B t^H)\right]$$

$$= \exp\left\{i \operatorname{Re}\left[\operatorname{tr}(t^H \nu)\right] - \frac{1}{4} \operatorname{tr}(t \Sigma t^H)\right\} = \exp\left\{i \operatorname{Re}\left[\nu t^H\right] - \frac{1}{4} t \Sigma t^H\right\} \tag{D.39}$$

Therefore $y \sim \mathbf{C}N_p(\nu, \Sigma)$.

Theorem 49 *Let* $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ *where* Z_k *and* μ_k *are row vectors, and the* $(Z_k)_n$ *are independently distributed according to the p-variate vector complex normal distribution* $\mathbf{C}N_p(\mu_k, \Sigma)$. *Then* $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. *This is a complexification of the first part of Arnold's theorem 17.3 [31].*

Proof. This is a complexification of Arnold's proof, where I have also used characteristic functions rather than moment generating functions. By equation D.39,

$$\Phi_{Z_k}(T_k) = \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T_k^H \mu_k)\right] - \frac{1}{4}\operatorname{tr}(T_k \Sigma T_k^H)\right\} \tag{D.40}$$

Thus $\Phi_Z(T) = \prod_{k=1}^{n} \Phi_{Z_k}(T_k)$ by independence. Then

$$\Phi_Z(T) = \prod_{k=1}^{n} \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T_k^H \mu_k)\right] - \frac{1}{4}\operatorname{tr}(T_k \Sigma T_k^H)\right\}$$

$$= \prod_{k=1}^{n} \exp\left\{ i \operatorname{Re}\left[\mu_k T_k^H\right] - \frac{1}{4} T_k \Sigma T_k^H \right\} = \exp\left\{ i \operatorname{Re}\left[\sum_{k=1}^{n} \mu_k T_k^H\right] - \frac{1}{4}\sum_{k=1}^{n} T_k \Sigma T_k^H \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\begin{pmatrix} \mu_1 T_1^H & & \\ & \ddots & \\ & & \mu_n T_n^H \end{pmatrix}\right] - \frac{1}{4}\operatorname{tr}\begin{pmatrix} T_1 \Sigma T_1^H & & \\ & \ddots & \\ & & T_n \Sigma T_n^H \end{pmatrix} \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \begin{pmatrix} T_1^H & \cdots & T_n^H \end{pmatrix}\right] - \frac{1}{4}\operatorname{tr}\left[\begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix} \Sigma \begin{pmatrix} T_1^H & \cdots & T_n^H \end{pmatrix}\right] \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(\mu T^H)\right] - \frac{1}{4}\operatorname{tr}\left[T\Sigma T^H\right]\right\} \tag{D.41}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}(T^H \mu)\right] - \frac{1}{4}\operatorname{tr}\left[T^H I T\Sigma\right]\right\} \tag{D.42}$$

which is the characteristic function of $\mathbf{C}N_{n,p}(\mu, I, \Sigma)$.

Theorem 50 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$ *and partition* $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ *where* Z_k *and* μ_k *are row vectors. Then the* Z_k *are independently distributed according to* $Z_k^T \sim \mathbf{C}N_p(\mu_k^T, \Sigma)$. *This is a complexification of the second part of Arnold's theorem 17.3 [31].*

Proof. Note that $\Xi = I$ and thus $\Xi_{ij} = \delta_{ij}$. Then by theorem 46, each of the Z_k are independently distributed according to $Z_k^T \sim \mathbf{C}N_p(\mu_k^T, \Sigma)$.

D.2.4 Matrix Complex Normal Density Function

Theorem 51 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, \Xi, \Sigma)$ *where* Ξ *and* Σ *are Hermitian positive definite. Then* Z *has the joint density function*

$$f(Z) = \frac{1}{\pi^{pn}\left|\det \Xi\right|^p \left|\det \Sigma\right|^n} \operatorname{etr}\left(-\Xi^{-1}(Z-\mu)\Sigma^{-1}(Z-\mu)^H\right) \tag{D.43}$$

This is a complexification of Arnold's theorem 17.4 [31].

Proof. This is a complexification of Arnold's proof. Let $X \sim \mathbf{C}N_{n,p}(0, I_n, I_p)$, or equivalently let $X_{jk} \sim \mathbf{C}N_1(0,1)$. Then by equation D.6,

$$f(X) = \frac{1}{\pi^{pn}} \operatorname{etr}\left(-X^H X\right) \tag{D.44}$$

Let A and B be nonsingular matrices such that $\Xi = AA^H$ and $\Sigma = B^H B$, as in equation D.8. By theorem 119, this factorization is possible because Ξ and Σ are positive definite. Let $Z = AXB + \mu \sim \mathbf{C}N_{n,p}(\mu, \Xi, \Sigma)$ by theorem 41. Then $X = A^{-1}(Z-\mu)B^{-1}$. By theorem 34, the absolute value of the Jacobian of this transformation is given by

$$|J| = |\det A|^{-2p} \, |\det B|^{-2n} \tag{D.45}$$

where the result is modified by our previous regarding the Jacobian of a complex linear transformation. Thus

$$|J| = |\det \Xi|^{-p} \, |\det \Sigma|^{-n} \tag{D.46}$$

Therefore, Z has the density

$$f_Z(Z) = f_X(A^{-1}(Z-\mu)B^{-1}) \, |J| = f_X(A^{-1}(Z-\mu)B^{-1}) \, |\det \Xi|^{-p} \, |\det \Sigma|^{-n}$$

$$= \frac{1}{\pi^{pn} \, |\det \Xi|^p \, |\det \Sigma|^n} \operatorname{etr}\left(-[A^{-1}(Z-\mu)B^{-1}]^H [A^{-1}(Z-\mu)B^{-1}]\right)$$

$$= \frac{1}{\pi^{pn} \, |\det \Xi|^p \, |\det \Sigma|^n} \operatorname{etr}\left(-B^{-H}(Z-\mu)^H A^{-H} A^{-1}(Z-\mu)B^{-1}\right)$$

$$= \frac{1}{\pi^{pn} \, |\det \Xi|^p \, |\det \Sigma|^n} \operatorname{etr}\left(-A^{-H} A^{-1}(Z-\mu)B^{-1} B^{-H}(Z-\mu)^H\right)$$

$$= \frac{1}{\pi^{pn} \, |\det \Xi|^p \, |\det \Sigma|^n} \operatorname{etr}\left(-(AA^H)^{-1}(Z-\mu)(B^H B)^{-1}(Z-\mu)^H\right) \tag{D.47}$$

$$f_Z(Z) = \frac{1}{\pi^{pn} \, |\det \Xi|^p \, |\det \Sigma|^n} \operatorname{etr}\left(-\Xi^{-1}(Z-\mu)\Sigma^{-1}(Z-\mu)^H\right) \tag{D.48}$$

D.2.5 Specialization to the Vector Complex Normal Density Function

The special case of the vector complex normal distribution is widely used in applications, and thus deserves explicit attention. The following are corollaries to the matrix complex normal density which was just derived.

Corollary 13 *Let* $z \sim \mathbf{C}N_{1,p}(\mu, 1, \Sigma)$ *where* Σ *is Hermitian positive definite. Then* z *has the joint density*

$$f_z(z) = \frac{1}{\pi^p \left|\det \Sigma\right|} \operatorname{etr}\left(-(z-\mu)\Sigma^{-1}(z-\mu)^H\right)$$

Recall here that z *is a row vector.*

Corollary 14 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$ *where* Σ *is Hermitian positive definite. Then* Z *has the joint density*

$$f_Z(Z) = \frac{1}{\pi^{pn} \left|\det \Sigma\right|^n} \operatorname{etr}\left(-(Z-\mu)\Sigma^{-1}(Z-\mu)^H\right)$$

Recall here that Z *is a matrix whose rows are independent.*

Corollary 15 *Let* $z \sim \mathbf{C}N_{n,1}(\mu, \Xi, 1)$ *where* Ξ *is Hermitian positive definite. Then* z *has the joint density*

$$f_z(z) = \frac{1}{\pi^n \left|\det \Xi\right|} \operatorname{etr}\left(-(Z-\mu)^H\Xi^{-1}(Z-\mu)\right)$$

Here, z *is a column vector.*

Corollary 16 *Let $Z \sim \mathbf{C}N_{n,p}(\mu, \Xi, I)$ where Ξ is Hermitian positive definite. Then Z has the joint density*

$$f_Z(Z) = \frac{1}{\pi^{pn} |\det \Xi|^p} \operatorname{etr}\left(-(Z-\mu)^H \Xi^{-1}(Z-\mu)\right)$$

Here, Z is a matrix whose columns are independent. This is the form usually seen in the literature.

D.3 Complex Wishart Distribution

The object of this section is to develop the definition and properties of the complex Wishart distribution. The development that follows is primarily a complexification of Arnold's section 17.3 [31].

D.3.1 Introduction

Definition 6 *Let matrix* $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ *be distributed according to the matrix complex normal distribution $\mathbf{C}N_{n,p}(\mu, I_n, \Sigma)$. The row vectors $\{Z_i\}_{i=1}^n$ are independent, and $Z_i^T \sim \mathbf{C}N_p(\mu_i^T, \Sigma)$. Let*

$$W = Z^H Z = \sum_{i=1}^{n} Z_i^H Z_i$$

Then W is defined to have a complex Wishart distribution. W is a $p \times p$ complex matrix that is Hermitian nonnegative definite. We identify this dis-

tribution by notation $\mathbf{C}W_p(n, \Sigma, \mu^H\mu)$ *or* $\mathbf{C}W_p(n, \Sigma, \delta)$ *where* δ *is called the* noncentrality *parameter.*

Lemma 14 *Let* W *have a complex Wishart distribution derived from* $Z \sim \mathbf{C}N_{n,p}(\mu, I_n, \Sigma)$. *Then the dependence of the distribution of* W *on the matrix mean parameter* μ *is only through the noncentrality parameter* $\delta = \mu^H\mu$. *This is a complexification of Arnold's lemma 17.5 [31].*

Proof. This is a complexification of Arnold's proof. Use an invariance argument to show this result. Let Γ be an $n \times n$ unitary matrix. Then

$$Y = \Gamma Z \sim \mathbf{C}N_{n,p}(\Gamma\mu, \Gamma I \Gamma^H, \Sigma) = \mathbf{C}N_{n,p}(\Gamma\mu, I_n, \Sigma) \tag{D.49}$$

by theorem 41. Further,

$$Y^H Y = (\Gamma Z)^H \Gamma Z = Z^H \Gamma^H \Gamma Z = Z^H Z = W \tag{D.50}$$

The distribution of W is the same for any unitary transformation Γ of Z. If $F(W; \mu)$ is the distribution function of W for a particular μ, then $F(W; \mu) = F(W; \Gamma\mu)$. Hence, F is invariant under the group G of unitary transformations $g(\mu) = \Gamma\mu$. Note the following.

$$\delta(\Gamma\mu) \underset{=}{\text{def}} (\Gamma\mu)^H \Gamma\mu = \mu^H \Gamma^H \Gamma \mu = \mu^H\mu \underset{=}{\text{def}} \delta(\mu) \tag{D.51}$$

Thus δ is invariant under G. Also,

$$[\delta(\mu) = \delta(\nu)] \Rightarrow [\mu^H\mu = \nu^H\nu] \tag{D.52}$$

By theorem 123, there exists a unitary transformation Γ such that $\mu = \Gamma\nu$. Thus δ is a maximal invariant by Arnold's definition (p. 13) [31], where $T = \delta$.

So, we have a group G of invertible unitary functions $g(\mu) = \Gamma\mu$ that map space $\mathcal{C} = \{\mu\}$ onto itself, a maximal invariant $\delta(\mu)$ under G, and a function satisfying $F_1(g(\mu)) = F(W;\Gamma\mu) = F(W;\mu) = F_1(\mu)$. Thus by Arnold lemma 1.11 [31], there exists a function $k(\delta)$ such that $F_1(\mu) = k(\delta(\mu)) = k(\mu^H\mu)$. Thus the distribution of W depends on μ only through $\mu^H\mu$, as claimed. □

Similar to Arnold's observation for the real variables case, we note that the distribution of W defined by equations D.49 and D.50 is a p-dimensional complex Wishart distribution with n degrees of freedom, on the covariance matrix Σ, and with noncentrality matrix

$$\delta = \mu^H\mu \tag{D.53}$$

This distribution is symbolized by $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$. Note that both Σ and δ are Hermitian nonnegative definite, which is symbolized by $\Sigma \geq 0$ and $\delta \geq 0$. If $\delta = 0$, then W has a *central* complex Wishart distribution which we denote by $W \sim \mathbf{C}W_p(n, \Sigma)$. If $\delta \neq 0$, then W has a *noncentral* complex Wishart distribution.

Notation for this and other distributions are not standardized. I have adopted Arnold's notation. For the real Wishart distribution, I have also seen variations of $W(p, n; \Sigma, \delta)$. For this reason, it is always best to define your notation at least once in your work.

D.3.2 Properties of the Complex Wishart Distribution

Theorem 52 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$. *Then* $\mathcal{E}\{W\} = n\Sigma + \delta$. *This is a complexification of Arnold's theorem 17.6(a) [31].*

Proof. This is a complexification of Arnold's proof. Let $W = \sum_{i=1}^{n} Z_i^H Z_i$, where the $Z_i \sim \mathbf{C}N_p(\mu_i, \Sigma)$ are independent row vectors. Then

$$\Sigma = \mathcal{E}\{(Z_i - \mu_i)^H (Z_i - \mu_i)\} = \mathcal{E}\{Z_i^H Z_i - \mu_i^H Z_i - Z_i^H \mu_i + \mu_i^H \mu_i\}$$

$$= \mathcal{E}\{Z_i^H Z_i\} - \mu_i^H \mathcal{E}\{Z_i\} - \mathcal{E}\{Z_i^H\}\mu_i + \mu_i^H \mu_i = \mathcal{E}\{Z_i^H Z_i\} - \mu_i^H \mu_i$$

By rearranging the equation, we get $\mathcal{E}\{Z_i^H Z_i\} = \Sigma + \mu_i^H \mu_i$. Therefore,

$$\mathcal{E}\{W\} = \mathcal{E}\{\sum_{i=1}^{n} Z_i^H Z_i\} = \sum_{i=1}^{n} \mathcal{E}\{Z_i^H Z_i\} = \sum_{i=1}^{n} [\Sigma + \mu_i^H \mu_i]$$

$$= n\Sigma + \sum_{i=1}^{n} \mu_i^H \mu_i = n\Sigma + \mu^H \mu = n\Sigma + \delta$$

Therefor, $\mathcal{E}\{W\} = n\Sigma + \delta$. □

This proof is important because it was the independent information source that provided a clue that the function presented by Goodman [92] and Anderson [26] as the characteristic function of the complex Wishart distribution was the characteristic function of something slightly different. This became apparent when I tried to compute the first moment by differentiation without getting the above result. Goodman gives a correct statement of what set of variables he supplied the characteristic function of, but the importance of what he said was not obvious to me until I tried to use the function for some computational purpose.

Lemma 15 *Let*

$$V^T = (X_1, Y_1, X_2, Y_2, \cdots, X_n, Y_n) = (V_1, V_2, V_3, V_4, \cdots, V_{2n-1}, V_{2n})$$

be distributed according to the real vector normal distribution $N_{2n}(\nu, \frac{1}{2}\sigma^2 I)$ *where*

$$\nu^T = (\nu_1, \cdots, \nu_{2n}) = (\mu_{R1}, \mu_{I1}, \cdots, \mu_{Rn}, \mu_{In})$$

and $\sigma^2 > 0$ *is a real scalar. Then*

$$\frac{2}{\sigma^2} V^T V \sim \chi^2_{2n}\left(\frac{2\nu^T \nu}{\sigma^2}\right)$$

The variable names of X_k, Y_k, μ_{Rk}, μ_{Ik} *are defined here to suggest notation used in the proof of theorem 53. This is a slight variation of Arnold's lemma 3.8 [31].*

Proof. This is a slight modification of Arnold's proof, where I accounted for the variation of the theorem and also used characteristic functions rather than moment generating functions. Because $V \sim N_{2n}(\nu, \frac{1}{2}\sigma^2 I)$, the V_k are independent and $V_k \sim N_1(\nu_k, \frac{1}{2}\sigma^2)$. From this, we compute the characteristic function of the joint distribution as follows.

$$\Phi_V(t) = \prod_{k=1}^{2n} \Phi_{V_k}(t_k) = \prod_{k=1}^{2n} \exp\left[it_k\nu_k - \frac{1}{2}t_k^2\left(\frac{1}{2}\sigma^2\right)\right]$$

$$= \exp\left\{i\left(t_1\nu_1 + \cdots + t_{2n}\nu_{2n}\right) - \frac{1}{4}\sigma^2\left(t_1^2 + \cdots + t_{2n}^2\right)\right\} = \exp\left(it^T\nu - \frac{1}{4}\sigma^2 t^T t\right)$$

Recall that $V^T V = \sum_{k=1}^{2n} V_k^2$, $\nu^T\nu = \sum_{k=1}^{2n} \nu_k^2$, and $\frac{V_k\sqrt{2}}{\sigma} \sim N_1\left(\frac{\nu_k\sqrt{2}}{\sigma}, 1\right)$. Therefore

$$\left(\frac{V\sqrt{2}}{\sigma}\right)^T \left(\frac{V\sqrt{2}}{\sigma}\right) = \frac{2}{\sigma^2} V^T V \sim \chi^2_{2n}\left(\frac{2\nu^T\nu}{\sigma^2}\right)$$

by the definition of the noncentral χ^2 distribution given in Arnold section 1.4 [31]. □

Notation. It is common for the distributional notation to be slightly abused as follows to simplify discussion. The abusive notation $u \sim a\chi^2_n(\delta)$ is intended to mean $\frac{u}{a} \sim \chi^2_n(\delta)$. If you expand the shorthand notation for the distribution into its density function for both cases, it becomes obvious that the paired relationship is not strictly true. That would be a statement about how a change of variables is implemented. However, the abusive notation does allow simplification of other developments which involve ratios such that the abusive constants divide out, and no one is the wiser. Statisticians have also used this convention with other distributions in journal articles. Caveat emptor.

Theorem 53 *(*Important*) Let* $W \sim \mathbf{C}W_{p=1}(n, \Sigma = \sigma^2 > 0, \delta)$. *Then* $\frac{2}{\sigma^2}W \sim \chi^2_{2n}\left(\frac{2}{\sigma^2}\delta\right)$. *This is a complexification of Arnold theorem 17.6(b) [31], which was stated without proof.*

Proof. Let $Z \sim \mathbf{C}N_n(\mu, \sigma^2 I)$. This implies $Z_k \sim \mathbf{C}N_1(\mu_k, \sigma^2)$. Let $Z_k = X_k + iY_k$. Recall that

$$W = Z^H Z = \sum_{k=1}^{n} Z_k^H Z_k = \sum_{k=1}^{n} (X_k + iY_k)^H (X_k + iY_k) = \sum_{k=1}^{n} (X_k^2 + Y_k^2)$$

Examining Z_k, we see that $Z_k \sim \mathbf{C}N_1(\mu_k, \sigma^2)$ implies that the real and imaginary parts of Z_k are distributed as $X_k \sim N_1(\mu_{Rk}, \frac{1}{2}\sigma^2)$ and $Y_k \sim N_1(\mu_{Ik}, \frac{1}{2}\sigma^2)$.

Recall that $\mathbf{C}N_1(\mu,\sigma^2)$ is isomorphic to $N_2\left(\begin{pmatrix}\mu_R\\ \mu_I\end{pmatrix},\frac{1}{2}\begin{pmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{pmatrix}\right)$. Let $\sigma=\sigma_R+i\sigma_I$. Then

$$\sigma^2=(\sigma_R+i\sigma_I)(\sigma_R-i\sigma_I)=\sigma_R^2+\sigma_I^2$$

Written in matrix analogy,

$$\frac{1}{2}\begin{pmatrix}\sigma_R & -\sigma_I\\ \sigma_I & \sigma_R\end{pmatrix}\begin{pmatrix}\sigma_R & \sigma_I\\ -\sigma_I & \sigma_R\end{pmatrix}=\frac{1}{2}\begin{pmatrix}\sigma_R^2+\sigma_I^2 & 0\\ 0 & \sigma_R^2+\sigma_I^2\end{pmatrix}$$

is isomorphic to the above. Using the notation from lemma 15, we note that $W=V^TV=\sum_{k=1}^{n}(X_k^2+Y_k^2)$ and $\nu^T\nu=\mu^H\mu$. Thus when $W\sim\mathbf{C}W_1(n,\sigma^2,\delta)$, we have $\frac{2}{\sigma^2}W\sim\chi^2_{2n}\left(\frac{2}{\sigma^2}\mu^H\mu\right)=\chi^2_{2n}\left(\frac{2}{\sigma^2}\delta\right)$.□

Discussion. We will see this theorem used later in developing the form of the density function for the complex Wishart distribution by using a proof by the principle of finite induction. Tague [264] used this theorem in his development of the signal-to-noise ratio at the output of a beamformer.

Lemma 16 *Let $W\sim\mathbf{C}W_p(n,\Sigma,\delta)$. Let $a\geq 0$ be a real scalar, and let $a=b^*b$. Then*

$$aW\sim\mathbf{C}W_p(n,a\Sigma,a\delta)$$

This is a complexification of Arnold's theorem 17.6(c) [31], which was stated without proof.

Proof. Let $W = Z^H Z$ where $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Then

$$aW = aZ^H Z = (bZ)^H (bZ)$$

By lemma 13 we know

$$bZ \sim \mathbf{C}N_{n,p}(b\mu, I, |b|^2 \Sigma) = \mathbf{C}N_{n,p}(b\mu, I, a\Sigma)$$

We also know that

$$a\delta = (b\mu)^H (b\mu) = a\mu^H \mu$$

Therefore $aW \sim \mathbf{C}W_p(n, \Sigma, \delta)$. □

Theorem 54 *(*Important*) Let $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ and $A \in \mathbf{C}^{k \times p}$. Then*

$$AWA^H \sim \mathbf{C}W_k(n, A\Sigma A^H, A\delta A^H)$$

This is a complexification of Arnold's theorem 17.6(d) [31].

Proof. This is a complexification of Arnold's proof. Let $W = Z^H Z$ where $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Then

$$ZA^H \sim \mathbf{C}N_{n,k}(\mu A^H, I, A\Sigma A^H)$$

by theorem 41. Thus $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ and

$$(ZA^H)^H (ZA^H) = AZ^H ZA^H = AWA^H \sim \mathbf{C}W_k(n, A\Sigma A^H, A\delta A^H)$$

where

$$(\mu A^H)^H (\mu A^H) = A\mu^H \mu A^H = A\delta A^H$$

□

Theorem 55 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$, $A = (I, 0)$, *and* $B = (0, I)$. *Define the following partitions:* $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, *and* $\delta = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$ *where* W_{11}, Σ_{11}, *and* δ_{11} *are* $k \times k$ *matrices. Then* $AWA^H = W_{11} \sim \mathbf{C}W_k(n, \Sigma_{11}, \delta_{11})$ *and* $BWB^H = W_{22} \sim \mathbf{C}W_{p-k}(n, \Sigma_{22}, \delta_{22})$. *This is a complexification of Arnold's theorem 17.6(e) [31], which was stated without proof.*

Proof. The results for both AWA^H and BWB^H follow directly from theorem 54.

Theorem 56 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$. *Partition* W, Σ, *and* δ *into identical blocks of* $p_1, p_2, \cdots, p_q$ *rows and columns where* $p_1 + \cdots + p_q = p$.

$$W = \begin{pmatrix} W_{11} & \cdots & W_{1q} \\ \vdots & \ddots & \vdots \\ W_{q1} & \cdots & W_{qq} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & \ddots & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix} \quad \delta = \begin{pmatrix} \delta_{11} & \cdots & \delta_{1q} \\ \vdots & \ddots & \vdots \\ \delta_{q1} & \cdots & \delta_{qq} \end{pmatrix}$$

Let $\Sigma_{ij} = 0$ *for* $i \neq j$. *Then the* $\{W_{ii}\}$ *are independent and* $W_{ii} \sim \mathbf{C}W_{p_i}(n, \Sigma_{ii}, \delta_{ii})$. *This is a complexification and generalization of Anderson's theorem 7.3.5 [26] to the noncentral complex Wishart case.*

Proof. This is a complexification and generalization of Anderson's proof. W has the same distribution as $\sum_{k=1}^{n} Z_k Z_k^H$ where the Z_k are independent and

distributed as $Z_k \sim \mathbf{C}N_p(\mu, \Sigma)$. Partition Z_k and μ into blocks of $p_1, p_2, \cdots, p_q$ rows. We know by theorem 43 or theorem 46 that the $\{Z_k^{(i)}\}$ are independent because $\Sigma_{ij} = 0$. Because the Z_k are independent also, we know

$$Z_1^{(1)}, \cdots, Z_1^{(q)}, Z_2^{(2)}, \cdots, Z_2^{(q)}, \cdots, Z_n^{(1)}, \cdots, Z_n^{(q)}$$

are independent. Thus

$$W_{11} = \sum_{k=1}^{n} Z_k^{(1)}[Z_k^{(1)}]^H, \cdots, W_{qq} = \sum_{k=1}^{n} Z_k^{(q)}[Z_k^{(q)}]^H$$

are independent. Let $A_i \in \mathbf{C}^{p_i \times p}$ where $A = (0, I_{p_i}, 0)$. By theorem 54,

$$W_{ii} = AWA^H \sim \mathbf{C}W_{p_i}(n, \Sigma_{ii}, \delta_{ii})$$

□

Corollary 17 *As in theorem 56, let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ *and partition* W, Σ, *and* δ *into identical blocks of* $p_1, p_2, \cdots, p_q$ *rows and columns where* $p_1 + \cdots + p_q = p$ *and let* $\Sigma_{ij} = 0$ *for* $i \neq j$. *When* $\delta = 0$, *then* $W_{ii} \sim \mathbf{C}W_{p_i}(n, \Sigma_{ii})$. *This is a complexification of theorem 7.3.5 of Anderson [26].*

Proof. Substitute $\delta_{ii} = 0$ into theorem 56.

Theorem 57 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ *and let* c *be any nonzero* $p \times 1$ *vector. Then*

$$\frac{2c^H W c}{c^H \Sigma c} \sim \chi^2_{2n}\left(\frac{2c^H \delta c}{c^H \Sigma c}\right)$$

This is a complexification of Graybill theorem 9.3.2(4).

Proof. This proof differs from Graybill's in order to take advantage of other work already presented here. By theorem 54,

$$c^H W c \sim CW_1(n, c^H \Sigma c, c^H \delta c)$$

Applying theorem 53, where $c^H \Sigma c > 0$ to make sure the denominator does not go to zero, we obtain our final result that

$$\frac{2c^H W c}{c^H \Sigma c} \sim \chi^2_{2n}\left(\frac{2c^H \delta c}{c^H \Sigma c}\right)$$

□

Let $W = UL^2U^H$ be the eigenvalue decomposition of $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$. Consider a linear combination of the sample eigenvalues, given by $c^H L^2 c$, where c is a $p \times 1$ vector of known fixed constants. Then by theorem 54 we have

$$c^H L^2 c \sim \mathbf{C}W_1(n, c^H U^H \Sigma U c, c^H U^H \delta U c)$$

Note that $c^H L^2 c$ is a scalar, as are now all the parameters of the distribution. Then by theorem 53, when $c^H U^H \Sigma U c \neq 0$,

$$\frac{2c^H L^2 c}{c^H U^H \Sigma U c} \sim \chi^2_{2n}\left(\frac{2c^H U^H \delta U c}{c^H U^H \Sigma U c}\right)$$

We know that when U is unitary, then the similarity transformation $U^H W U$ has the same eigenvalues as W. Then, when we let c be a column vector of ones, we get

$$\frac{2\,\mathrm{tr}(W)}{c^H U^H \Sigma U c} \sim \chi^2_{2n}\left(\frac{2c^H U^H \delta U c}{c^H U^H \Sigma U c}\right)$$

Remark. Because $U^H \Sigma U$ is not generally a diagonal matrix, we conclude that the sample eigenvalues are generally not independent. Thus, disjoint linear combinations of sample eigenvalues from the same sample set are not independent. Therefore, the ratio of disjoint linear combinations of sample eigenvalues from the same sample set is generally not F-distributed.

Now, consider the case when $C = (C_1, C_2)$ is a $p \times 2$ matrix. Look at $C^H L^2 C$.

$$C^H L^2 C = \begin{pmatrix} C_1^H \\ C_2^H \end{pmatrix} L^2 \begin{pmatrix} C_1 & C_2 \end{pmatrix} = \begin{pmatrix} C_1^H L^2 C_1 & C_1^H L^2 C_2 \\ C_2^H L^2 C_1 & C_2^H L^2 C_2 \end{pmatrix}$$

This is distributed as

$$C^H L^2 C \sim \mathbf{C}W_2(n, C^H U^H \Sigma U C, C^H U^H \delta U C)$$

This becomes particularly interesting when you choose $C_1^H C_2 = 0$. Then

$$C^H L^2 C = \begin{pmatrix} C_1^H L^2 C_1 & 0 \\ 0 & C_2^H L^2 C_2 \end{pmatrix}$$

However, notice that you do not get zeros in the population matrix.

$$C^H \Lambda^2 C = \begin{pmatrix} C_1^H \Lambda^2 C_1 & C_1^H \Lambda^2 C_2 \\ C_2^H \Lambda^2 C_1 & C_2^H \Lambda^2 C_2 \end{pmatrix}$$

Now, the matrix is an ordered set of 4 random variables. We have taken linear combinations to force two of the sample values to zero. Note that since $W = W^H$, we really have only 3 random variables.

We will pick up on this theme again when we discuss the density functions of sample eigenvalues in a later section. We next consider a projection theorem.

Theorem 58 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$, V *be a* k*-dimensional subspace of* $\mathbf{C}^n$, *and* P_V *a projection operator from* $\mathbf{C}^n$ *onto* V. *Then*

$$Z^H P_V Z \sim \mathbf{C}W_p(k, \Sigma, \mu^H P_V\ \mu)$$

This is a complexification of Arnold's theorem 17.7(a) [31].

Proof. This is a complexification of Arnold's proof. Let U be a unitary basis matrix for V. Then by theorem 41,

$$U^H Z \sim \mathbf{C}N_{k,p}(U^H \mu, I, \Sigma)$$

Then

$$(U^H Z)^H (U^H Z) = Z^H U U^H Z = Z^H P_V Z \sim \mathbf{C}W_p(k, \Sigma, \mu^H P_V\ \mu)$$

where

$$(U^H \mu)^H (U^H \mu) = \mu^H U U^H \mu = \mu^H P_V\ \mu$$

is the noncentrality parameter. Arnold comments that if A is an idempotent $n \times n$ matrix of rank k, then

$$Z^H A Z \sim \mathbf{C}W_p(k, \Sigma, \mu^H A \mu)$$

□

Lemma 17 *Let $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Then AZ and BZ are independent if $AB^H = 0$. This is a complexification of part 1 of Arnold's theorem 17.7(b) [31], which was stated without proof. This result differs from Arnold's in that independence does not imply $AB^H = 0$.*

Proof. Let $C = \begin{pmatrix} A \\ B \end{pmatrix}_{m \times n}$, where A is $q \times n$ and B is $(m-q) \times n$. Then by theorem 41,

$$Y = CZ \sim \mathbf{C}N_{m,p}(C\mu, CC^H, \Sigma)$$

We look at the characteristic function of Y.

$$\Phi_Y(T) = \Phi_{CZ}(T)$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(\begin{pmatrix} T_1^H & T_2^H \end{pmatrix} \begin{pmatrix} A\mu \\ B\mu \end{pmatrix} \right) \right] \right.$$

$$\left. -\frac{1}{4} \operatorname{tr}\left(\begin{pmatrix} T_1^H & T_2^H \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} I \begin{pmatrix} A^H & B^H \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \Sigma \right) \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(T_1^H A\mu + T_2^H B\mu \right) \right] - \frac{1}{4} \operatorname{tr}\left(\left(T_1^H A + T_2^H B \right) \left(A^H T_1 + B^H T_2 \right) \Sigma \right) \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(T_1^H A\mu + T_2^H B\mu \right) \right] \right.$$

$$\left. -\frac{1}{4} \operatorname{tr}\left(T_1^H AA^H T_1\Sigma + T_1^H AB^H T_2\Sigma + T_2^H BA^H T_1\Sigma + T_2^H BB^H T_2\Sigma \right) \right\}$$

$$= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(T_1^H A\mu \right) \right] - \frac{1}{4} \operatorname{tr}\left(T_1^H AA^H T_1\Sigma \right) \right\}$$

$$\times \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(T_2^H B\mu \right) \right] - \frac{1}{4} \operatorname{tr}\left(T_2^H BB^H T_2\Sigma \right) \right\}$$

$$\times \exp\left\{-\frac{1}{4}\operatorname{tr}\left(T_1^H A B^H T_2 \Sigma + T_2^H B A^H T_1 \Sigma\right)\right\}$$

If

$$\Phi_{CZ}(T) = \Phi_{AZ}(T_1)\Phi_{BZ}(T_2)$$

then

$$\operatorname{tr}\left(\left(T_1^H A B^H T_2 + T_2^H B A^H T_1\right)\Sigma\right) = 0$$

for all T_1, T_2. Consider $T_1 = \begin{pmatrix} T_{11} \\ T_{21} \end{pmatrix}$, $T_2 = \begin{pmatrix} T_{12} \\ T_{22} \end{pmatrix}$, $AB^H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then

$$\operatorname{tr}\left(T_1^H A B^H T_2 \Sigma\right) = \operatorname{tr}\left((T_{11}^* T_{12} + T_{21}^* T_{22})\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = 0$$

Therefore independence does not imply $AB^H = 0$.

If $AB^H = 0$, then

$$\Phi_{CZ}(T) = \Phi_{AZ}(T_1)\Phi_{BZ}(T_2)$$

□

Lemma 18 *Let $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Let B be nonnegative definite and $AB = 0$. Then AZ and $Z^H BZ$ are independent. This is a complexification of part 2 of Arnold's theorem 17.7(b) [31], which was stated without proof.*

Proof. Follow the proof of Arnold's theorem 1.13. Let $B = E^H E$ and let $\text{rank}(E_{r\times n}) = r$. Then $\text{rank}(B) = r$. Suppose $AB = 0$. Then $ABE^H = 0 = AE^H E E^H$. However, $\text{rank}(EE^H) = r$. Therefore $AE^H = 0$. By lemma 17, this implies AZ and EZ are independent. This, in turn, implies AZ and

$$(EZ)^H(EZ) = Z^H E^H EZ = Z^H BZ$$

are independent. □

Theorem 59 *Let $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Let A and B be nonnegative definite and $AB = 0$. Then $Z^H AZ$ and $Z^H BZ$ are independent. This is a complexification of part 3 of Arnold's theorem 17.7(b) [31], which was stated without proof.*

Proof. Let $A = D^H D$ where D is $s \times n$ and $\text{rank}(A) = s$. Let $B = E^H E$ where E is $r \times n$ and $\text{rank}(B) = r$. When $AB = 0$, then $DABE^H = 0$ which implies $DD^H DE^H EE^H = 0$. However, DD^H and EE^H are of full rank. Therefore, $DE^H = 0$ which implies DZ and EZ are independent by lemma 17. Then

$$(DZ)^H(DZ) = Z^H D^H DZ = Z^H AZ$$

is independent of

$$(EZ)^H(EZ) = Z^H E^H EZ = Z^H BZ$$

□

Theorem 60 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, \Xi, \Sigma)$ *where* Ξ *and* Σ *are positive definite and* $n \geq p$. *Then*

$$\Pr\{\text{rank}(Z) = p\} = 1$$

This is a complexification of Arnold's theorem 17.8(a) [31].

Proof. This is a complexification of Arnold's proof. Let $Z = (Z_1, \cdots, Z_p)$ and let $S_k(Z_1, \cdots, Z_k)$ be a subspace of $\mathbf{C}^n$ spanned by $(Z_1, \cdots, Z_k)$. Since the conditional distribution of $Z_{k+1} \mid (Z_1, \cdots, Z_k)$ is a nonsingular vector complex normal distribution by theorem 44 or theorem 47, then

$$\Pr\{X_{k+1} \in S_k(Z_1, \cdots, Z_k) \mid (Z_1, \cdots, Z_k)\} = 0$$

if $k < n$. Note that S_k is a subspace of dimension at most k. Therefore

$$\Pr\{Z_{k+1} \in S_k(Z_1, \cdots, Z_k)\}$$

$$= \mathcal{E}\{\Pr[Z_{k+1} \in S_k(Z_1, \cdots, Z_k) \mid (Z_1, \cdots, Z_k)]\} = 0$$

Finally,

$$\Pr(Z_1, \cdots, Z_p \text{ are linearly dependent}) \leq \sum_{k=1}^{p-1} \Pr[Z_{k+1} \in S_k(Z_1, \cdots, Z_k)] = 0$$

Therefore, $\Pr\{\text{rank}(Z) = p\} = 1$.□

Theorem 61 *Let* $Z \sim \mathbf{C}N_{n,p}(\mu, \Xi, \Sigma)$ *where* Ξ *and* Σ *are positive definite and* $n > p$. *Let* $a \in \mathbf{C}^n$ *such that* $a \neq 0$. *Let* (a, Z) *be the matrix* Z *augmented by the vector* a. *Then* $\Pr\{(a, Z) = p + 1\} = 1$. *This is a complexification of Arnold's theorem 17.8(b) [31], which was stated without proof.*

Proof. I have not proven this theorem. Since Arnold claims the theorem true in $\mathbf{R}^n$ and the proof depends on geometric concepts, then it is true in $\mathbf{C}^n$. This theorem is retained since it is possibly useful with updating algorithms in adaptive signal processing.

Corollary 18 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$. *If* $n \geq p$ *and* Σ *is positive definite, then* $\Pr\{W > 0\} = 1$. *This is a complexification of Arnold's corollary to his theorem 17.8 [31].*

Proof. This is a complexification of Arnold's proof. The rank of $W = Z^H Z$ is the same as the rank of Z. See Arnold's lemma A.9 and theorem A.3, with straight forward extensions to the complex case. By theorem 60, if Σ is positive definite and $n \geq p$, then $\Pr\{\text{rank}(W) = p\} = 1$. Since W is $p \times p$ of full rank p, it is nonsingular with probability 1. Hence $W > 0$ with probability 1.

Since W is invertible with probability 1 if $\Sigma > 0$ and $n \geq p$, then W has the *nonsingular complex Wishart* distribution. Otherwise, W has a *singular complex Wishart* distribution which is sometimes called a complex pseudo-Wishart distribution.

Lemma 19 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, 0) = \mathbf{C}W_p(n, \Sigma)$ *where* $\Sigma > 0$ *and* $n \geq p$. *Partition* W *and* Σ *such that* W_{11} *and* Σ_{11} *are both* $q \times q$ *matrices. Then*

$$W_{11.2} \sim \mathbf{C}W_q(n - p + q, \Sigma_{11.2})$$

This is a restatement and complexification of Arnold's lemma 17.9 [31] and it is also a complex version of Anderson theorem 7.3.6 [26].

Proof. The following is an expansion and complexification of Arnold's proof. Let $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}_{p\times p}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{p\times p}$. Let

$$T = W_{22}$$

$$U = W_{22}^{-1}W_{21}$$

$$V = W_{11} - W_{12}W_{22}^{-1}W_{21} = W_{11.2}$$

By theorem 55,

$$T = W_{22} \sim \mathbf{C}W_{p-q}(n, \Sigma_{22}, 0) = \mathbf{C}W_{p-q}(n, \Sigma_{22})$$

Let

$$(Y_{n\times q}, X_{n\times(p-q)}) \sim \mathbf{C}N_{n,p}(0, I, \Sigma)$$

and let $W = (Y, X)^H(Y, X)$. Then

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} = \begin{pmatrix} Y^HY & Y^HX \\ X^HY & X^HX \end{pmatrix}$$

By definition of the complex Wishart distribution, $W \sim \mathbf{C}W_p(n, \Sigma)$. Substituting into T, U, and V, we get

$$T = X^HX$$

$$U = (X^HX)^{-1}X^HY$$

$$V = Y^H(I - X(X^HX)^{-1}X^H)Y$$

Now, find the conditional distribution $(U, V) \mid X$. Since $(Y, X) \sim \mathbf{C}N_{n,p}(0, I, \Sigma)$ and Y is $n \times q$, then by theorem 44,

$$(Y \mid X) \sim \mathbf{C}N_{n,q}(X\Sigma_{22}^{-1}\Sigma_{21}, I, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \tag{D.54}$$

Now,

$$(U \mid X) = ((X^H X)^{-1} X^H Y \mid X)$$

which is distributed according to

$$\mathbf{C}N_{(p-q),q}\{(X^H X)^{-1} X^H X \Sigma_{22}^{-1}\Sigma_{21}, (X^H X)^{-1} X^H I[(X^H X)^{-1} X^H]^H,$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\}$$

or

$$(U \mid X) \sim \mathbf{C}N_{(p-q),q}(\Sigma_{22}^{-1}\Sigma_{21}, (X^H X)^{-1}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \tag{D.55}$$

by theorem 41 where $(X^H X)^{-H} = (X^H X)^{-1}$. Since the conditional distribution of $(U \mid X)$ depends on X only through $T = X^H X$, we can write

$$(U \mid T) \sim \mathbf{C}N_{(p-q),q}(\Sigma_{22}^{-1}\Sigma_{21}, T^{-1}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$= \mathbf{C}N_{(p-q),q}(\Sigma_{22}^{-1}\Sigma_{21}, T^{-1}, \Sigma_{11.2})$$

Consider

$$[I - X(X^H X)^{-1} X^H][I - X(X^H X)^{-1} X^H]$$

$$= I - X(X^H X)^{-1} X^H - X(X^H X)^{-1} X^H + X(X^H X)^{-1} X^H X(X^H X)^{-1} X^H$$

$$= I - 2X(X^H X)^{-1} X^H + X(X^H X)^{-1} X^H = I - X(X^H X)^{-1} X^H = P_V$$

and therefore this matrix is idempotent. X is of rank $p-q$, and so P_V is of rank $n-(p-q)$.

Consider

$$(X\Sigma_{22}^{-1}\Sigma_{21})^H[I - X(X^HX)^{-1}X^H](X\Sigma_{22}^{-1}\Sigma_{21})$$

$$= \Sigma_{21}^H\Sigma_{22}^{-H}X^H[I - X(X^HX)^{-1}X^H](X\Sigma_{22}^{-1}\Sigma_{21})$$

$$\Sigma_{21}^H\Sigma_{22}^{-H}X^HX\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{21}^H\Sigma_{22}^{-H}X^HX\Sigma_{22}^{-1}\Sigma_{21} = 0$$

where $\Sigma_{22}^{-H} = \Sigma_{22}^{-1}$. Since we know the distribution of $(Y \mid X)$ then by theorem 58

$$(V \mid X) = (Y^HP_VY \mid X) \sim \mathbf{C}W_q(n-p+q, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$= \mathbf{C}W_q(n-p+q, \Sigma_{11.2}) \tag{D.56}$$

Also note

$$[(X^HX)^{-1}X^H][I - X(X^HX)^{-1}X^H] = AP_V$$

$$= (X^HX)^{-1}X^H - (X^HX)^{-1}X^HX(X^HX)^{-1}X^H$$

$$= (X^HX)^{-1}X^H - (X^HX)^{-1}X^H = 0$$

where $A = (X^HX)^{-1}X^H$ and $P_V = I - X(X^HX)^{-1}X^H$. By lemma 18, $(AY \mid X) = (U \mid X)$ and $(Y^HP_VY \mid X) = (V \mid X)$ are independent. This implies $(U \mid T)$ is independent of $(V \mid T)$. In turn, this implies that

$$f(U, V \mid T) = f(V \mid T)f(U \mid X)$$

Thus

$$f(V \mid (U,T)) = \frac{f(V,U \mid T)}{f(U \mid T)} = f(V \mid T) = f(V \mid X)$$

Finally,

$$W_{11.2} = V \mid (U,T) \sim \mathbf{C}W_q(n-p+q, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$= \mathbf{C}W_q(n-p+q, \Sigma_{11.2})$$

This completes the proof. □

Corollary 19 *Let* $W \sim \mathbf{C}W_p(n,\Sigma,0) = \mathbf{C}W_p(n,\Sigma)$ *where* $\Sigma > 0$ *and* $n \geq p$. *Partition* W *and* Σ *such that* W_{11} *and* Σ_{11} *are both scalars. Let*

$$\sigma^2 = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11.2}$$

Then $\frac{2}{\sigma^2}W_{11.2} \sim \chi^2_{2(n-p+1)}(0)$. *This is a formalization and complexification of Arnold's corollary to his lemma 17.9 for the case of* $q = 1$ *[31]. Results of this special case are useful in test statistics of quadratic forms.*

Proof. This is an expansion and complexification of Arnold's proof. Let $q = 1$ in lemma 19 and let $\beta = \Sigma_{22}^{-1}\Sigma_{21}$. By equation D.54 and lemma 12 we have $(Y \mid X) \sim \mathbf{C}N_n(X\beta, \sigma^2 I)$. We started with $W \sim W_p(n,\Sigma)$ with $\Sigma > 0$ and $n \geq p$. Thus the matrix $(Y,X)_{n\times p}$ is of full row rank p, which implies $\text{rank}(X_{n\times(p-1)}) = p-1$.

Let $U = (X^H X)^{-1}X^H Y$ and

$$V = Y^H[I - X(X^H X)^{-1}X^H]Y$$

Then

$$(U \mid X) \sim \mathbf{C}N_{(p-1),1}(\beta, (X^H X)^{-1}, \sigma^2) = \mathbf{C}N_{p-1}(\beta, \sigma^2 (X^H X)^{-1})$$

by equation D.55 and lemma 12. Thus

$$(V \mid X) \sim CW_1(n-p+1, \sigma^2)$$

and

$$\frac{2}{\sigma^2}(V \mid X) \sim \chi^2_{2(n-p+1)}(0)$$

by equation D.56 and lemma 15.

Let $T = X^H X$. Then by lemma 19 and lemma 12,

$$V \mid (U,T) \sim \mathbf{C}W_1(n-p+1, \sigma^2)$$

and therefore

$$\frac{2}{\sigma^2} V \mid (U,T) \sim \chi^2_{2(n-p+1)}(0)$$

D.4 Distribution of Hotelling's T^2 for Complex Variables

Hotelling's T^2 is a classical statistic for testing means. It is a likelihood ratio test statistic. This result is provided because it is an easy result which naturally falls at this point in the general theory of complex multivariate analysis.

Definition 7 *Hotelling's T^2 for complex variables. Let Z and W be independent, $Z \sim \mathbf{C}N_p(\mu, \Sigma)$ and $W \sim \mathbf{C}W_p(n, \Sigma)$, where $n \geq p$ and $\Sigma > 0$. By corollary 18, $\Pr\{W > 0\} = 1$. Define*

$$F = \frac{n-p+1}{p} Z^H W^{-1} Z$$

Then

$$T^2 = \frac{np}{n-p+1} F$$

is called Hotelling's T^2. This is a complexification of Arnold's definition [31] in his equation 17.21 and accompanying discussion.

Unlike the case of real random variables, for complex variables the case of $p = 1$ does not yield the square of a random variable having a non-central t-distribution. The careful reader will realize that this is merely due to $\mathbf{C}^n \simeq \mathbf{R}^{2n}$. When $p = 1$, then

$$F = \frac{n|Z|^2}{W} = \left(\frac{|Z|}{\sqrt{W/n}} \right)^2 = T^2 \sim F_{2,2n}\left(\frac{2|\mu|^2}{\Sigma} \right)$$

In this section it is shown that F has an F-distribution even when $p > 1$. We prepare for our journey with the following theorem and corollary.

Theorem 62 *Let $A \in \mathbf{C}^{p \times k}$ and $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. Then*

$$A^H W A \sim \mathbf{C}W_k(n, A^H \Sigma A, A^H \delta A)$$

where $\delta = \mu^H \mu$.

Proof. Let $W = Z^H Z$ and $Y = ZA$. Then

$$Y^H Y = (ZA)^H (ZA) = A^H Z^H Z A = A^H W A$$

By theorem 41 we know $ZA \sim \mathbf{C}N_{n,k}(\mu A, I, A^H \Sigma A)$. Further,

$$(\mu A)^H (\mu A) = A^H \mu^H \mu A = A^H \delta A$$

and by lemma 14 we know $A^H W A \sim \mathbf{C}W_k(n, A^H \Sigma A, A^H \delta A)$. □

Theorem 63 *Let $Y \sim \mathbf{C}N_n(\mu, \Sigma)$ where Σ is positive definite. Then*

$$2Y^H \Sigma^{-1} Y \sim \chi^2_{2n}(2\mu^H \Sigma^{-1} \mu)$$

Proof. Let $Z = \Sigma^{-1/2} Y$. Then

$$Z^H Z = Y^H (\Sigma^{-1/2})^H \Sigma^{-1/2} Y = Y^H \Sigma^{-1} Y$$

Let $A \in \mathbf{C}^{m \times n}$ and $c \in \mathbf{C}^m$. By theorem 18,

$$\Phi_{AY+c}(t) = \exp\left[i \operatorname{Re}(t^H c)\right] \Phi_Y(A^H t)$$

$$= \exp\left[i \operatorname{Re}(t^H c)\right] \exp\left[i \operatorname{Re}([A^H t]^H \mu) - \frac{1}{4}(A^H t)^H \Sigma (A^H t)\right]$$

$$= \exp\left[i \operatorname{Re}(t^H c)\right] \exp\left[i \operatorname{Re}(t^H A \mu) - \frac{1}{4} t^H A \Sigma A^H t\right]$$

$$= \exp\left[i \operatorname{Re}\left(t^H [A\mu + c]\right) - \frac{1}{4} t^H (A \Sigma A^H) t\right]$$

Therefore

$$AY + c \sim \mathbf{C}N_m(A\mu + c, A\Sigma A^H)$$

which implies

$$Z \sim \mathbf{C}N_n\left(\Sigma^{-1/2}\mu, \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})^H\right) = \mathbf{C}N_n(\Sigma^{-1/2}\mu, I)$$

Then

$$Y^H\Sigma^{-1}Y = Z^HZ = W \sim \mathbf{C}W_1(n, I_1, \mu^H\Sigma^{-1}\mu)$$

by the definition of the noncentral complex Wishart distribution. Thus by lemma 15,

$$2Y^H\Sigma^{-1}Y \sim \chi^2_{2n}(2\mu^H\Sigma^{-1}\mu).$$

This completes the proof. □

With this preparation, we are ready to begin.

Theorem 64 *Let $W \sim \mathbf{C}W_p(n, \Sigma)$ where $n \geq p$ and $\Sigma > 0$. Let $a \in \mathbf{C}^n$ such that $a \neq 0$. Then*

$$\frac{2a^H\Sigma^{-1}a}{a^HW^{-1}a} \sim \chi^2_{2(n-p+1)}(0)$$

This is a complexification of Arnold's lemma 17.10. This is a good example that shows complexification involves more than merely changing transpose to Hermitian transpose.

Proof. This is an expansion and complexification of Arnold's proof. Let $a_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and partition W and Σ so that W_{11} and Σ_{11} are scalars. $W =$

$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Let $V = W^{-1}$. Then

$$V_{11} = a_0^H W^{-1} a_0 = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

By Arnold lemma A.2(b) [31], which applies to complex as well as to real matrices,

$$V_{11} = (W_{11} - W_{12} W_{22}^{-1} W_{21})^{-1}$$

Thus

$$a_0^H W^{-1} a_0 = (W_{11} - W_{12} W_{22}^{-1} W_{21})^{-1}$$

Similarly,

$$a_0^H \Sigma^{-1} a_0 = (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1}$$

For the special case of $k = 1$ and $\mu = 0$,

$$a_0^H W a_0 \sim \mathbf{C}W_1(n, a_0^H \Sigma a_0) = \mathbf{C}W_1(n, \Sigma_{11})$$

We will take advantage of the fact that $a_0^H W^{-1} a_0$ and $a_0^H \Sigma^{-1} a_0$ are scalars.

$$\frac{a_0^H \Sigma^{-1} a_0}{a_0^H W^{-1} a_0} = \frac{(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1}}{(W_{11} - W_{12} W_{22}^{-1} W_{21})^{-1}} = \frac{W_{11} - W_{12} W_{22}^{-1} W_{21}}{\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}$$

From corollary 19, let

$$V = W_{11} - W_{12} W_{22}^{-1} W_{21}$$

$$\sigma^2 = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$T = W_{22}$, and $U = W_{22}^{-1}W_{21}$. Lemma 19 established that $f(V) = f(V \mid (U,T))$ and $\frac{2V}{\sigma^2} \sim \chi^2_{2(n-p+1)}(0)$. Thus we get

$$2\frac{W_{11} - W_{12}W_{22}^{-1}W_{21}}{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}} = \frac{2V}{\sigma^2} \sim \frac{\sigma^2\chi^2_{2(n-p+1)}(0)}{\sigma^2} = \chi^2_{2(n-p+1)}(0)$$

or

$$\frac{2a_0^H\Sigma^{-1}a_0}{a_0^HW^{-1}a_0} \sim \chi^2_{2(n-p+1)}(0)$$

So, the lemma is true for $a = a_0$.

Now, let $a = Aa_0$ where A is invertible. The vector a is the first column of A. By theorem 54 for $B \in \mathbf{C}^{p\times p}$ and $W \sim \mathbf{C}W_p(n,\Sigma,\delta)$, then

$$BWB^H \sim \mathbf{C}W_p(n, B\Sigma B^H, B\delta B^H)$$

Let $B = A^{-1}$ and $\delta = 0$. Then

$$A^{-1}WA^{-H} \sim \mathbf{C}W_p(n, A^{-1}\Sigma A^{-H})$$

Thus

$$\frac{2a^H\Sigma^{-1}a}{a^HW^{-1}a} = \frac{2(Aa_0)^H\Sigma^{-1}(Aa_0)}{(Aa_0)^HW^{-1}(Aa_0)} = \frac{2a_0^H(A^H\Sigma^{-1}A)a_0}{a_0^H(A^HW^{-1}A)a_0}$$
$$= \frac{2a_0^H(A^{-1}\Sigma A^{-H})^{-1}a_0}{a_0^H(A^{-1}WA^{-H})^{-1}a_0} \sim \chi^2_{2(n-p+1)}(0)$$

by our proof with a_0. □

Theorem 65 *Let Z and W be independent where $Z \sim \mathbf{C}N_p(\mu,\Sigma)$, $W \sim \mathbf{C}W_p(n,\Sigma)$, $n \geq p$, and $\Sigma > 0$. Then*

$$F = \frac{n-p+1}{p}Z^HW^{-1}Z \sim F_{2p,2(n-p+1)}(2\mu^H\Sigma^{-1}\mu)$$

This is a complexification of Arnold's theorem 17.11 [31].

Proof. This is an expansion and complexification of Arnold's proof. Let $U = Z^H\Sigma^{-1}Z$ and $V = \frac{Z^H\Sigma^{-1}Z}{Z^HW^{-1}Z}$. If we fix the value of Z, then by theorem 64 we know

$$2(V \mid Z) \sim \chi^2_{2(n-p+1)}(0)$$

Thus V is independent of Z, and therefore also V is independent of U. Therefore $2V \sim \chi^2_{2(n-p+1)}(0)$. By theorem 63 we know $2U \sim \chi^2_{2p}(2\mu^H\Sigma^{-1}\mu)$. In the case of real variables, the result for V is similar to Muirhead's theorem 3.2.12. Continuing, we form the ratio for the F-statistic.

$$\frac{\frac{2}{2p}U}{\frac{2}{2(n-p+1)}V} = \frac{\frac{1}{p}Z^H\Sigma^{-1}Z}{\frac{1}{n-p+1}\frac{Z^H\Sigma^{-1}Z}{Z^HW^{-1}Z}} = \frac{n-p+1}{p}Z^HW^{-1}Z = \frac{n-p+1}{p}Z^HW^{-1}Z$$

$$\sim \frac{\frac{1}{2p}\chi^2_{2p}(2\mu^H\Sigma^{-1}\mu)}{\frac{1}{2(n-p+1)}\chi^2_{2(n-p+1)}(0)} = \frac{2(n-p+1)}{2p}\frac{\chi^2_{2p}(2\mu^H\Sigma^{-1}\mu)}{\chi^2_{2(n-p+1)}(0)} = F_{2p,2(n-p+1)}(2\mu^H\Sigma^{-1}\mu)$$

Therefore

$$F = \frac{n-p+1}{p}Z^HW^{-1}Z \sim F_{2p,2(n-p+1)}(2\mu^H\Sigma^{-1}\mu)$$

Maximum likelihood estimates of μ and Σ are statistically independent and can be the random variables used in this statistic. This distribution is useful in testing hypotheses about μ and for establishing confidence regions for μ. T^2 is the likelihood ratio for testing $H : \mu = \mu_0$. Discussion can be found beginning on p. 159 of Anderson [26].

Theorem 66 *Let $Z \sim \mathbf{C}N_p(\mu, \Sigma)$ such that $\Sigma > 0$, $W \sim \mathbf{C}W_p(n, \Xi)$ such that $\Xi > 0$ and $n \geq p$, and $Y \in \mathbf{C}^p$. Then*

$$\frac{n-p+1}{p} Z^H \Sigma^{-1} Z \frac{Y^H W^{-1} Y}{Y^H \Xi^{-1} Y} \sim F_{2p,2(n-p+1)}(2\mu^H \Sigma^{-1} \mu)$$

Proof. By theorem 62 we know

$$\frac{2Y^H \Xi^{-1} Y}{Y^H W^{-1} Y} \sim \chi^2_{2(n-p+1)}(0)$$

From theorem 65,

$$2Z^H \Sigma^{-1} Z \sim \chi^2_{2p}(2\mu^H \Sigma^{-1} \mu)$$

This implies

$$\frac{\frac{2}{2p} Z^H \Sigma^{-1} Z}{\frac{2}{2(n-p+1)} \frac{Y^H \Xi^{-1} Y}{Y^H W^{-1} Y}} \sim F_{2p,2(n-p+1)}(2\mu^H \Sigma^{-1} \mu)$$

which implies our result

$$\frac{n-p+1}{p} Z^H \Sigma^{-1} Z \frac{Y^H W^{-1} Y}{Y^H \Xi^{-1} Y} \sim F_{2p,2(n-p+1)}(2\mu^H \Sigma^{-1} \mu)$$

□

Corollary 20 *Let $Z \sim \mathbf{C}N_p(\mu, aR)$, $W \sim \mathbf{C}W_p(n, bR)$ such that $R > 0$ and $n \geq p$, and $a, b \in \mathbf{C}$. Then*

$$F = \frac{b(n-p+1)}{ap} Z^H W^{-1} Z \sim F_{2p,2(n-p+1)}(\frac{2}{a} \mu^H R^{-1} \mu)$$

Proof. Let $Y = Z$, $\Sigma = aR$, and $\Xi = bR$. Note that $\Sigma^{-1} = \frac{1}{a} R^{-1}$. Then

$$\frac{2(n-p+1)}{2p} Z^H (aR)^{-1} Z \frac{Z^H W^{-1} Z}{Z^H (bR)^{-1} Z} = \frac{2b(n-p+1)}{2ap} \left(\frac{Z^H R^{-1} Z}{Z^H R^{-1} Z} \right) Z^H W^{-1} Z$$

The final result follows immediately from this by applying theorem 66. □

Appendix E

DISTRIBUTIONS, PART II

This work was undertaken to determine the density function of the complex Wishart distribution because only a few reports of the density function exists in the literature, and these were not identical. I considered that use of the correct result is critical to the primary question of order determination. In preparing a background for this task, it was discovered that the pieces of needed knowledge were scattered throughout the literature using different notational conventions, did not form a complete theory, and occasionally contained minor errors which inevitably get through any editorial and publishing process. In writing the main part of this thesis, I intended to draw only on those portions of the general development of complex multivariate statistics as was absolutely needed. It became obvious that it was both needed and simpler to produce a well organized presentation of the material. Much of the material to follow has most likely resided in the minds of others, but I have not found it. Readers are encouraged to communicate their findings so that a history of this fascinating subject can be constructed.

What follows began as a complexification of Chapter 17 of Arnold's well organized text [31]. It has a very nice development of ideas, it uses matrix notation throughout, it uses norms and projections where it can do so profitably, and draws upon some group theoretic ideas. To develop the needed theory for

the change of complex variables, extensive adaptation was made to results of Chapter 2 and the Appendix of Muirhead's fine text [187]. Other references have been used where it was needed. An attempt has been made to accentuate the similarity of this work with the works of others. By reading the sources and examining the enclosed work, it is hoped that others may learn quickly how to make adaptations from real variables to complex variables.

E.1 Complex Wishart Density

The purpose of this appendix is to prove the form of the density function of the complex Wishart distribution. Several respectable references give conflicting expressions for this density function. It is shown in this appendix that Goodman [92] provided the correct form. Use of the correct form is critical to future work. Therefore, three different derivations are presented to gain confidence that the correct result is obtained. The first is a complexification of the derivation done by Arnold [31] which gives a proof by induction. This approach has not been previously applied to the complex case. The second derivation is the one by Goodman from his classic paper cited above. The third more general result is by Srivastava [256] which has the complex Wishart density as a special case. It is reassuring that we get the same answer in three different ways.

E.1.1 Arnold's Proof by Induction

This is a detailed complexification and extension of the derivation provided by Arnold in Section 17.6 of his text. The goal is to develop a complex version of Arnold's Theorem 17.12. We want to find the density function for the nonsingular central complex Wishart distribution. First, consider the simplified case of $\Sigma = I$.

Let $W \sim \mathrm{C}W_p(n, I),\ n \geq p$.

First, let $p = 1$. Then by Theorem 53, $2W \sim \chi^2_{2n}(0)$.

Anderson [26] gives the density of the noncentral χ^2 distribution with p degrees of freedom and noncentrality parameter τ^2 as

$$f(u; p, \tau^2) = \frac{1}{2^{p\backslash 2}} \exp\left[-\frac{1}{2}(\tau^2 + u)\right] u^{\frac{p}{2}-1} \left\{\sum_{\beta=0}^{\infty} \left(\frac{\tau^2}{4}\right)^{\beta} \left[\beta!\, \Gamma(\frac{1}{2}p + \beta)\right]^{-1} u^{\beta}\right\} du$$

Let $p = 2n,\ \tau^2 = 0,\ \beta = k$, and $u = x$. Then

$$x \sim \chi^2_{2n}(0) = 2^{-n} \exp[-\frac{x}{2}] x^{n-1} \left\{\sum_{k=0}^{\infty} 0^k [k!\, \Gamma(n+k)]^{-1} x^k\right\} dx$$

$$= 2^{-n} \frac{1}{\Gamma(n)} x^{n-1} \exp(-\frac{x}{2}) dx$$

Perform a change of variables $x = 2W$, which implies $dx = 2(dW)$. Then W has a density function given by

$$f(W) = 2^{-n} \left(\frac{1}{\Gamma(n)}\right) (2W)^{n-1} \exp(-W) 2(dW) = \left(\frac{1}{\Gamma(n)}\right) W^{n-1} \exp(-W)(dW)$$

For $p > 1$, we will prove by induction that

$$f(W) = \frac{|\det W|^{n-p} \exp[-\operatorname{tr} W]}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(n - i + 1)} (dW)$$

Then, the result will be extended to $\mathbf{C}W_p(n,\Sigma)$ by a change of variables.

Assume that

$$f(W_{n-1}) = \frac{|\det W|^{n-(p-1)} \exp[-\operatorname{tr} W_{n-1}]}{\pi^{(p-1)(p-2)/2} \prod_{i=1}^{p-1} \Gamma(n-i+1)}$$

Then, we can express W in terms of T, U, and V. We know conditional distributions involving T, U, and V from our proof of lemma 19. Recall that the assumptions were $W \sim \mathbf{C}W_p(n,\Sigma)$, $\Sigma > 0$, $n \geq p$ where W and Σ are partitioned such that W_{11} and Σ_{11} are $q \times q$. We defined $T = W_{22}$, $U = W_{22}^{-1}W_{21}$, and $V = W_{11} - W_{12}W_{22}^{-1}W_{21}$. Then

$$T \sim \mathbf{C}W_{p-q}(n, \Sigma_{22})$$

$$(U \mid T) \sim \mathbf{C}N_{p-q,q}(\Sigma_{22}^{-1}\Sigma_{21}, T^{-1}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$V \mid (U,T) \sim \mathbf{C}W_q(n-p+q, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

The joint density of T, U, V is then

$$f(T,U,V) = f(V \mid U,T)f(U,T) = f(V \mid U,T)f(U \mid T)f(T)$$

By a change of variables, we get

$$f(W) = f_{V|(U,T)}(W_{11} - W_{12}W_{22}^{-1}W_{21} \mid W_{22}^{-1}W_{21}, W_{22}) \times \qquad \text{(E.1)}$$

$$\times f_{U|T}(W_{22}^{-1}W_{21} \mid W_{22}) f_T(W_{22}) |J|$$

To evaluate the Jacobian, look closer at the change of variables. Suppose

$$y_1 = g_1(x_1,x_2,x_3) = W_{22} = T = g_1(T,U,V)$$

$$y_2 = g_2(x_1,x_2,x_3) = W_{21} = TU = g_2(T,U,V)$$

$$y_3 = g_3(x_1,x_2,x_3) = W_{11} = V + U^H T^H U = g_3(T,U,V)$$

The inverse transformations are

$$x_1 = f_1(y_1, y_2, y_3) = T = W_{22} = f_1(W_{22}, W_{21}, W_{11})$$

$$x_2 = f_2(y_1, y_2, y_3) = U = W_{22}^{-1}W_{21} = f_2(W_{22}, W_{21}, W_{11})$$

$$x_3 = f_3(y_1, y_2, y_3) = V = W_{11} - W_{12}W_{22}^{-1}W_{21} = f_3(W_{22}, W_{21}, W_{11})$$

The absolute value of the Jacobian is computed as

$$|\det J| = \left| \left\{ \det \left[\left(\frac{\partial f_j(y_1, y_2, y_3)}{\partial y_i} \right)_{ij} \right] \right\}^2 \right|$$

$$= \left| \left\{ \det \begin{bmatrix} \frac{\partial W_{22}}{\partial W_{22}} & \frac{\partial W_{22}^{-1}W_{21}}{\partial W_{22}} & \frac{\partial (W_{11} - W_{12}W_{22}^{-1}W_{21})}{\partial W_{22}} \\ \frac{\partial W_{22}}{\partial W_{21}} & \frac{\partial W_{22}^{-1}W_{21}}{\partial W_{21}} & \frac{\partial (W_{11} - W_{12}W_{22}^{-1}W_{21})}{\partial W_{21}} \\ \frac{\partial W_{22}}{\partial W_{11}} & \frac{\partial W_{22}^{-1}W_{21}}{\partial W_{11}} & \frac{\partial (W_{11} - W_{12}W_{22}^{-1}W_{21})}{\partial W_{11}} \end{bmatrix} \right\}^2 \right|$$

$$= \left| \left\{ \det \begin{bmatrix} I & * & * \\ 0 & W_{22}^{-1} & * \\ 0 & 0 & I \end{bmatrix} \right\}^2 \right| = \left| \left\{ \det W_{22}^{-1} \right\}^2 \right| = \left| \{ \det W_{22} \}^{-2} \right|$$

where the dots indicate terms not evaluated because they can be ignored when evaluating the determinant by expansion. Notice that we use $\left|\det W_{22}^{-1}\right|^2$ rather than $\left|\det W_{22}^{-1}\right|$ in evaluating the Jacobian because we are now doing a change of complex variables, in accordance with theorem 22. Thus, we can write

$$|\det J| = \left| \left\{ \det W_{22}^{-1} \right\}^2 \right| = \left| \det W_{22}^{-1} \right|^2$$

Restricting attention to the case where $q = 1$, for

$$\Sigma_n = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^H & \Sigma_{n-1} \end{pmatrix}$$

$$W_n = \begin{pmatrix} W_{11} & \Sigma_{12} \\ W_{12}^H & W_{n-1} \end{pmatrix}$$

we obtain

$$T \sim \mathbf{C}W_{p-1}(n, \Sigma_{n-1})$$

$$(U \mid T) \sim \mathbf{C}N_{p-1,1}(\Sigma_{n-1}^{-1}\Sigma_{12}^H, T^{-1}, \Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H)$$

$$V \mid (U,T) \sim \mathbf{C}W_1(n-p+1, \Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H)$$

The absolute value of the Jacobian is

$$|J(T,U,V \to W_{11}, W_{12}, W_{n-1})| = |\det W_{n-1}|^{-2}$$

Then equation E.1 becomes $f(W) =$

$$\frac{\left|W_{11} - W_{12}W_{n-1}^{-1}W_{12}^H\right|^{n-p+1-1} \operatorname{etr}\left[-(\Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H)^{-1}(W_{11} - W_{12}W_{n-1}^{-1}W_{12}^H)\right]}{\pi^{1(1-1)/2}\left|\Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H\right|^{n-p+1}\left[\prod_{i=1}^{1}\Gamma(n-p+1-1+i)\right]}$$

$$\times\frac{\operatorname{etr}\left[-W_{n-1}(W_{n-1}^{-1}W_{12}^H - \Sigma_{n-1}^{-1}\Sigma_{12}^H)(\Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H)^{-1}(W_{n-1}^{-1}W_{12}^H - \Sigma_{n-1}^{-1}\Sigma_{12}^H)^H\right]}{\pi^{1(p-1)}\left|W_{n-1}^{-1}\right|^1 \cdot \left|\Sigma_{11} - \Sigma_{12}\Sigma_{n-1}^{-1}\Sigma_{12}^H\right|^{p-1}}$$

$$\times\frac{|W_{n-1}|^{n-(p-1)} \operatorname{etr}\left[-\Sigma_{n-1}^{-1}W_{n-1}\right]}{\pi^{(p-1)(p-2)/2}|\Sigma_{n-1}|^n\left[\prod_{i=1}^{p-1}\Gamma(n-(p-1)+i)\right]} \times |W_{n-1}|^{-2}$$

To simplify the problem, let $\Sigma_n = I_n$. Then $f(W)$ simplifies to

$$f(W) = \frac{\left|W_{11} - W_{12}W_{n-1}^{-1}W_{12}^H\right|^{n-p} \cdot |W_{n-1}|^{n-p+1} \cdot |W_{n-1}|^{-2}}{\pi^{p-1+(p-1)(p-2)/2}|W_{n-1}|^{-1}}$$

$$\times\frac{\exp\left\{-\operatorname{tr}\left(W_{11} - W_{12}W_{n-1}^{-1}W_{12}^H\right) - \operatorname{tr}\left[W_{12}^H\left(W_{n-1}^{-1}W_{12}^H\right)^H\right] - \operatorname{tr}(W_{n-1})\right\}}{\Gamma(n-p+1)\Gamma(n-p+2)\Gamma(n-p+3)\cdots\Gamma(n)}$$

Observe that

$$p - 1 + \frac{(p-1)(p-2)}{2} = \frac{2p - 2 + p^2 - 3p + 2}{2} = \frac{p(p-1)}{2}$$

and

$$\left|W_{11} - W_{12}W_{n-1}^{-1}W_{12}^{H}\right| \cdot |W_{n-1}|^{n-p} = |W|^{n-p}$$

by the partitioned matrix determinant. Also,

$$\operatorname{tr}(W_{12}^{H}W_{12}W_{n-1}^{-1}) = \operatorname{tr}(W_{12}W_{n-1}^{-1}W_{12}^{H})$$

by property of the trace function. Thus

$$f(W) = \frac{|W|^{n-p}\exp\{-W_{11} - \operatorname{tr}(W_{n-1})\}}{\pi^{p(p-1)/2}\prod_{i=1}^{p}\Gamma(n-p+i)} = \frac{|W|^{n-p}\exp\{-W\}}{\pi^{p(p-1)/2}\prod_{i=1}^{p}\Gamma(n-p+i)}$$

We recognize this as the distribution function of $\mathbf{C}W_p(n, I)$. Thus by induction, we proved the result true for $p = 1$, assumed it was true for $p - 1$, and based on this assumption showed it true for p. Thus, it is true for all p.

Now, the result must be extended to general $\Sigma^H = \Sigma > 0$. To bridge the gap from $\mathbf{C}W_p(n, I)$ to $\mathbf{C}W_p(n, \Sigma)$, I must first make the transition from $\mathbf{C}N(0, I, I)$ to $\mathbf{C}N(0, I, \Sigma)$. Suppose that the $n \times p$ matrix random variable Y has the matrix complex normal distribution $\mathbf{C}N_{n,p}(0_{n\times p}, I_n, \Sigma_{p\times p})$. By theorem 51, the density of Y is given by

$$f(Y) = \pi^{-np}|\Sigma|^{-n}\exp\left[-\operatorname{tr}(Y\Sigma^{-1}Y^{H})\right]$$

Since Σ^{-1} is positive definite, it can be factored into $\Sigma^{-1} = TT^H$ where T is $p \times p$ lower triangular with positive real diagonal elements. Thus

$$f(Y) = \pi^{-np}|\Sigma|^{-n}\exp\left[-\operatorname{tr}(YTT^{H}Y^{H})\right]$$

Let $X = YT$, and thus $Y = XT^{-1}$. Then by lemma 5 the Jacobian is $J(Y \to X) = \prod_{i=1}^{p} t_{ii}^{-2n}$. Therefore

$$f(XT^{-1}) = \pi^{-np} |\Sigma|^{-n} \exp\left[-\operatorname{tr}(XX^H)\right]$$

Since I left this as $f(XT^{-1})$ rather than $f(X)$, no Jacobian was needed.

From this point, this is a complexification and expansion of Arnold's corollary to his theorem 17.12. By corollary 36 (Cholesky or Bartlett Decomposition) there exists a unique $p \times p$ lower triangular matrix L with positive diagonal elements such that $\Sigma = LL^H$. Let $B \sim \mathbf{C}W_p(n, I)$. Then

$$W = LBL^H \sim \mathbf{C}W_p(n, LIL^H) = \mathbf{C}W_p(n, \Sigma)$$

by theorem 54. The Jacobian is $J(B \to W) = (\det \Sigma)^{-p}$. Thus

$$f(W) = f_B(L^{-1}WL^{-H})J(B \to W)$$

and so

$$f(W) = \frac{\left|\det(L^{-1}WL^{-H})\right|^{n-p} \operatorname{etr}(-L^{-1}WL^{-H})}{\pi^{p(p-1)/2}(\det \Sigma)^p \prod_{i-1}^{p} \Gamma(m - i + 1)}(dW)$$

$$f(W) = \frac{(\det L)^{-(n-p)} |\det W|^{n-p} (\det L^H)^{-(n-p)} \operatorname{etr}(-L^{-H}L^{-1}W)}{(\det \Sigma)^p \mathbf{C}\Gamma_p(n)}(dW)$$

$$= \frac{(\det \Sigma)^{-(n-p)} |\det W|^{n-p} \operatorname{etr}(-\Sigma^{-1}W)}{(\det \Sigma)^p \mathbf{C}\Gamma_p(n)}(dW)$$

$$f(W) = \frac{|\det W|^{n-p} \operatorname{etr}(-\Sigma^{-1}W)}{(\det \Sigma)^n \mathbf{C}\Gamma_p(n)}(dW) \tag{E.2}$$

where $\mathbf{C}\Gamma_p(n)$ is the complex multivariate gamma function. This is the final form of the probability density function for the central complex Wishart distribution $\mathbf{C}W_p(n, \Sigma)$.

E.1.2 Goodman's Construction

The following derivation follows that by Goodman [92]. Goodman is the first to publish the density of the complex Wishart distribution.

Consider the matrix Laplace transform of $(\det W)^k$ where W is a random $p \times p$ Hermitian positive definite matrix variable $W^H = W > 0$. This is given by

$$\mathcal{I}(\Sigma) = \int_W |W|^k \exp[-\operatorname{tr}(\Sigma^{-1}W)](dW) \tag{E.3}$$

where the integral is taken over all $W = W^H > 0$.

Let $T^H T = W$ be a change of variables from W to T, where T is a complex upper triangular matrix with positive real elements on the diagonal. By theorem 27, the Jacobian of this transformation is

$$J(W \to T) = 2^p \prod_{i=1}^{p} t_{ii}^{2(p-i)+1} \tag{E.4}$$

Consider the special case where $\Sigma = \Lambda^2 = \operatorname{diag}(\lambda_1^2, \cdots, \lambda_p^2)$. Then $\Lambda^{-2}T^H T =$

$$\begin{pmatrix} \lambda_1^{-2} & & & \\ & \lambda_2^{-2} & & \\ & & \ddots & \\ & & & \lambda_p^{-2} \end{pmatrix} \times$$

$$\times \begin{pmatrix} t_{11}^2 & t_{11}t_{12} & \cdots & t_{11}t_{1p} \\ t_{12}^* t_{11} & |t_{12}|^2 + t_{22}^2 & \cdots & t_{12}^* t_{1p} + t_{22}t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{1p}^* t_{11} & t_{1p}^* t_{12} + t_{2p}^* t_{22} & \cdots & |t_{1p}|^2 + |t_{2p}|^2 + \cdots + |t_{pp}|^2 \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1^{-2} t_{11}^2 & \lambda_1^{-2} t_{11}t_{12} & \cdots & \lambda_1^{-2} t_{11}t_{1p} \\ \lambda_2^{-2} t_{12}^* t_{11} & \lambda_2^{-2}(|t_{12}|^2 + t_{22}^2) & \cdots & \lambda_2^{-2}(t_{12}^* t_{1p} + t_{22}t_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_p^{-2} t_{1p}^* t_{11} & \lambda_p^{-2}(t_{1p}^* t_{12} + t_{2p}^* t_{22}) & \cdots & \lambda_p^{-2}(|t_{1p}|^2 + |t_{2p}|^2 + \cdots + |t_{pp}|^2) \end{pmatrix}$$

Thus $\operatorname{tr}(\Lambda^{-2} T^H T) =$

$$\lambda_1^{-2} t_{11}^2 + \lambda_2^{-2}\left(|t_{12}|^2 + t_{22}^2\right) + \cdots + \lambda_p^{-2}\left(|t_{1p}|^2 + |t_{2p}|^2 + \cdots + t_{pp}^2\right)$$

In preparation for the next leap of faith, we note that

$$(\det W)^k = \left[\det(T^H T)\right]^k = \left|\det T^H\right|^k \cdot |\det T|^k = |\det T|^{2k}$$

since $\left|\det T^H\right| = |\det T|$. Substituting back into equation E.3 we get

$$\mathcal{I}(\Lambda^2) = \int |\det T|^{2k} \exp[-\operatorname{tr}(\Lambda^{-2} T^H T)] J(W \to T)(dT)$$

$$= \int \left(\prod_{i=1}^{p} t_{ii}\right)^{2k} \times$$

$$\times \exp\left[\lambda_1^{-2} t_{11}^2 - \lambda_2^{-2}(|t_{12}|^2 + t_{22}^2) - \cdots - \lambda_p^{-2}(|t_{1p}|^2 + |t_{2p}|^2 + \cdots + t_{pp}^2)\right] \times$$

$$\times 2^p \left(\prod_{i=1}^{p} t_{ii}^{2(p-i)+1}\right)(dT)$$

$$= 2^p \left[\prod_{i=1}^{p} \int t_{ii}^{2(k+p-i)+1} \exp\left(-\lambda_i^{-2} t_{ii}^2\right) dt_{ii}\right] \left[\prod_{i=2}^{p} \prod_{j=1}^{i-1} \int \exp\left(-\lambda_i^{-2} |t_{ji}|^2\right) dt_{ji}\right]$$

$$= 2^p \left[\prod_{i=1}^{p} \left(\frac{1}{2}\lambda_i^{2(k+p-i+1)}\Gamma(k+p-i+1)\right)\right] \left[\prod_{i=2}^{p}\prod_{j=1}^{i-1} \left(\lambda_i^2\pi\right)\right]$$

$$2^p 2^{-p} \left[\prod_{i=1}^{p} \left(\lambda_i^{2(k+p-i+1)}\Gamma(k+p-i+1)\right)\right] \pi^{p(p-1)/2} \left[\prod_{i=2}^{p} \lambda_i^{2(i-1)}\right]$$

where

$$\prod_{i=2}^{p}\prod_{j=1}^{i-1} \pi = \pi^{p(p-1)/2}$$

This is most easily seen by considering a triangular array of the constant π.

$$\begin{array}{ccccccc} & & & i \rightarrow & & & \\ \pi & \pi & \pi & \cdots & \pi & & \\ & \pi & \pi & \cdots & \pi & j & \\ & & \pi & \cdots & \pi & \downarrow & \\ & & & \ddots & \vdots & & \\ & & & & \pi & & \\ & & & & & p\times p & \end{array}$$

The columns are indexed by i and the rows are indexed by j. There are $\frac{p(p-1)}{2}$ elements in the array above the diagonal. Also note that $\lambda_i^{2(i-1)} = 1$ when $i = 1$. Thus

$$\prod_{i=2}^{p} \lambda_i^{2(i-1)} = \prod_{i=1}^{p} \lambda_i^{2(i-1)}$$

which implies

$$\mathcal{I}(\Lambda^2) = \pi^{p(p-1)/2} \prod_{i=1}^{p} \left(\lambda_i^{2(k+p-i+1+i-1)}\Gamma(k+p-i+1)\right)$$

$$= \pi^{p(p-1)/2} \prod_{i=1}^{p} \left(\lambda_i^{2(k+p)}\Gamma(k+p-i+1)\right)$$

$$= \pi^{p(p-1)/2} \left[\det \Lambda^2\right]^{k+p} \prod_{i=1}^{p} \Gamma(k+p-i+1)$$

since $\det(\Lambda^2) = \prod_{i=1}^{p} \lambda_i^2$.

$$\mathcal{I}(\Lambda^2) = \pi^{p(p-1)/2} \left[\det \Lambda^2\right]^{k+p} \prod_{i=1}^{p} \Gamma(k+i)$$

since

$$\prod_{i=1}^{p} \Gamma(k+p-i+1) = \Gamma(k+p)\Gamma(k+p-1)\cdots\Gamma(k+1) = \prod_{i=1}^{p} \Gamma(k+i)$$

Therefore

$$\mathcal{I}(\Lambda^2) = \int |W|^k \operatorname{etr}[-\Lambda^{-2}W](dW)$$

$$= \pi^{p(p-1)/2} \left[\det \Lambda^2\right]^{k+p} \Gamma(k+p)\Gamma(k+p-1)\cdots\Gamma(k+1)$$

Let $\Sigma = U^H \Lambda^2 U$ where $U^H U = I$. Note: $\Sigma = \Sigma^H$. Then

$$\int_{W>0} |\det W|^k \operatorname{etr}[-\Sigma^{-1}W](dW) = \int_{W>0} |\det W|^k \operatorname{etr}[-U^H \Lambda^{-2} U\ W](dW)$$

$$= \int_{W>0} |\det W|^k \operatorname{etr}[-\Lambda^{-2} U\ WU^H](dW)$$

Let $K = U\ WU^H$ which implies $W = U^H K U$. By corollary 7, $J(K \dashrightarrow W) = 1$ which implies $J(W \rightarrow K) = 1$. This gives us

$$\int_{W>0} |\det W|^k \operatorname{etr}[-\Lambda^{-2}W](dW) = \int_{K>0} \left|\det(U^H K U)\right|^k \operatorname{etr}[-\Lambda^{-2}K](dK)$$

$$= \int_{K>0} |K|^k \operatorname{etr}[-\Lambda^{-2}K](dK)$$

$$= \pi^{p(p-1)/2} \left[\det \Lambda^2\right]^{k+p} \Gamma(k+p)\Gamma(k+p-1)\cdots\Gamma(k+1)$$

$$= \pi^{p(p-1)/2} \left[\det \Sigma\right]^{k+p} \Gamma(k+p)\Gamma(k+p-1)\cdots\Gamma(k+1)$$

since

$$\det(\Sigma) = \det(U^H \Lambda^2 U) = \det(\Lambda^2)$$

Therefore

$$\frac{\int_{W>0} |\det W|^k \operatorname{etr}[-\Sigma^{-1}W](dW)}{\pi^{p(p-1)/2}[\det \Sigma]^{k+p} \prod_{i=1}^{p} \Gamma(k+i)} = 1$$

Thus

$$f(W;\Sigma) = \frac{|\det W|^k \operatorname{etr}[-\Sigma^{-1}W]}{\pi^{p(p-1)/2}[\det \Sigma]^{k+p} \prod_{i=1}^{p} \Gamma(k+i)}(dW)$$

is a density function on the space of Hermitian positive definite matrices. Let $k = n - p$. Then for $\mathbf{C}W_p(n,\Sigma)$ we have

$$f(W;n,p,\Sigma) = \frac{|\det W|^{n-p} \operatorname{etr}[-\Sigma^{-1}W]}{\pi^{p(p-1)/2}[\det \Sigma]^{n} \prod_{i=1}^{p} \Gamma(n-p+i)}(dW)$$

$$= \frac{|\det W|^{n-p} \operatorname{etr}[-\Sigma^{-1}W]}{\pi^{p(p-1)/2}[\det \Sigma]^{n} \prod_{i=1}^{p} \Gamma(n-i+1)}(dW) = f(W,\Sigma)$$

This agrees with Goodman equation (1.6) [92]. The function $f(W;\Sigma)$ is the probability density function for the central complex Wishart distribution. For the density function to exist, Σ must be nonsingular. Note that both W and Σ are Hermitian positive definite. W is a random variable. Σ, n, and p are fixed parameters. W is $p \times p$ of full rank. In equation E.3, k is an arbitrary complex constant with $\operatorname{Re}(k+i) > 0$ for $1 \leq i \leq p$. We finally let $k = n - p$ to obtain the density for the complex Wishart distribution where n is taken to be the number of samples of the multivariate complex normal distribution used in forming W. This relation becomes more apparent in other derivations of this density function. It can also be seen in the extension to the complex case of Arnold's discussion of the Wishart distribution. A more compact notation

for the density function is

$$f(W) = \frac{|\det W|^{n-p} \operatorname{etr}[-\Sigma^{-1}W]}{[\det \Sigma]^n \, \mathrm{C}\Gamma_p(n)}(dW)$$

where

$$\mathrm{C}\Gamma_p(n) = \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(n-i+1)$$

is the complex multivariate gamma function. A shorthand notation for this complex Wishart distribution is $W \sim \mathrm{C}W_p(n, \Sigma)$. This distribution has not been around and used long enough for a notation to become standard. Of course, we still do not have a universally accepted notation for the chi-square distribution yet, either.

E.1.3 Srivastava's Derivation

Srivastava [256] provided a derivation that obtained a more general result in the process. The following discussion expands Srivastava's work. He begins with a Lemma.

Lemma 20 *If Y is a matrix of complex elements of order $p \times m$ where $p \leq m$, and $rank(Y) = p$, then there exists a unique lower triangular matrix T with positive real diagonal elements and a matrix H such that $HH^H = I_p$ and $Y = TH$. The matrices live in the following spaces: $Y \in \mathbf{C}^{p \times m}$, $T \in \mathbf{C}^{p \times p}$,, and $H \in \mathbf{C}^{p \times m}$. Note that because H is not square and because $HH^H = I_p$, H is called* semi-unitary.

Proof. This is corollary 37, where $Y = A$, and $T = L$, with appropriate changes in dimensions.□

Density of a Quadratic

The following theorem is a generalization of Anderson's Lemma 13.3.1 (2nd. Edition, p. 533) [26] to a matrix Y with complex elements. The original citation is to Anderson (1st. Edition, 1958, p. 319).

Theorem 67 *(*Important*) If the density $f(YY^H)$ of $Y_{p\times m}$ is a function of YY^H, then the density of $B = YY^H$ is given by*

$$g(B) = \frac{|\det B|^{m-p} f(B)\pi^{p[m-\frac{1}{2}(p-1)]}}{\prod_{i=1}^{p}\Gamma(m-i+1)}(dB) = \frac{|\det B|^{m-p} f(B)}{\pi^{-pm}\mathbf{C}\Gamma_p(m)}(dB)$$

Note that when $Y \sim \mathbf{C}N_{m,p}(0_{m\times p}, I_m, I_p)$, then

$$p(Y) = \pi^{-mp}\operatorname{etr}(-YY^H) = f(YY^H)$$

Y can also be viewed as a random sample of size m from the complex multivariate standard normal distribution $\mathbf{C}N_p(0_p, I_p)$. Then $f(B)$ is the density of $B \sim \mathbf{C}W_p(n, I_p)$.

Proof. From lemma 20, we can write $Y = TH$. $T_{p\times p}$ is complex lower triangular with positive real diagonal elements. $HH^H = I_p$.

We shall now find the Jacobian of the transformation, $J(Y \to T, H)$. A basic property of Jacobians that is used is that if changing variables from X to Y, then the Jacobian $J(X \to Y)$ is the same as the Jacobian $J[(dX) \to (dY)]$

of the change of variables from (dX) to (dY). See Deemer and Olkin Property 5.B.3 [67]. We see that $Y = TH$ implies that $(dY) = (dT)H + T(dH)$. The object now is to define a series of transformations where the Jacobian of each individual transformation is more easily computed than the original transformation. We will take advantage of a property of Jacobians that says that

$$J(X \to Y) = J(X \to U)J(U \to V)J(V \to Y)$$

where U and V are any functions of X and Y such that none of the terms on the right vanish. See Deemer and Olkin Property 5.B.2 [67]. For completeness' sake, there are two more general properties of Jacobians that deserve to be mentioned. The first is

$$J(X \to Y) = \frac{1}{J(Y \to X)}$$

The second is a bit longer. If $X = F_1(U)$ and $Y = F_2(V)$ defines a transformation from variables (X, Y) to new variables (U, V), then

$$J[(X, Y) \to (U, V)] = J(X \to U)J(Y \to V)$$

The sequence of transformations that Srivastava chose are as follows. Given

$$(dY) = (dT)H + T(dH)$$

premultiply by T^{-1} to get

$$T^{-1}(dY) = T^{-1}(dT)H + (dH)$$

Let $U = T^{-1}(dY)$, resulting in

$$U = T^{-1}(dT)H + (dH)$$

Let $V = T^{-1}(dT)$ to produce $U = VH + (dH)$. The Jacobian is given by

$$J(Y \rightarrow T, H) = J(dY \rightarrow dT, dH)$$

$$= J(dY \rightarrow U)J(U \rightarrow V, dH)J[(V, dH) \rightarrow (dT, dH)]$$

To see where this comes from, notice that $U = T^{-1}(dY)$ implies $dY = TU$. Therefore we seek $J(dY \rightarrow U) = J_1$. We also see that $U = VH + (dH)$ implies that we seek $J[U \rightarrow (V, dH)] = J_2$. Treating V as a matrix of constants, this Jacobian is a function of only H. Let

$$J[U \rightarrow (V, dH)] = g(H) = J_2$$

The relation $V = T^{-1}(dT)$ implies we seek $J[(V, dH) \rightarrow (dT, dH)] = J_3$. Putting this all together, we see that

$Y \rightarrow dY = TU$	using $J_1(dY \rightarrow U)$
$TU = T[VH + (dH)]$	using $J_2(U \rightarrow H)$
$T[VH + (dH)] = TVH + T(dH) = TT^{-1}(dT)H + T(dH)$	using $J_3(V \rightarrow dT)$
$TT^{-1}(dT)H + T(dH) = (dT)H + T(dH) = d(TH) = (dY)$	by theorem 21

Consider $J_1 = J(dY \rightarrow U)$ where $dY = TU$. The matrix dY is a matrix whose elements are the differentials of $Y_{p \times m}$. Thus, dY is also of dimensions

$p \times m$. T is complex lower triangular of dimension $p \times p$ with positive real diagonal elements. U is $p \times m$. By lemma 3, we know that

$$J_1 = J(dY \rightarrow U) = \prod_{i=1}^{p} t_{ii}^{2m}$$

Consider

$$J_3 = J[(V, dH) \rightarrow (dT, dH)] = J(V \rightarrow dT)$$

where $V = T^{-1}(dT)$. $T_{p \times p}$ is lower triangular, so (dT) and T^{-1} are also lower triangular. Hence, V is also lower triangular. All have real diagonal elements. It is simpler to examine $(dT) = TV$ and $J(dT \rightarrow V)$, and then take the inverse of the Jacobian. By lemma 6,

$$J(dT \rightarrow V) = \prod_{i=1}^{p} t_{ii}^{2i-1}$$

Thus

$$J(V \rightarrow dT) = J_3 = \prod_{i=1}^{p} t_{ii}^{-2i+1}$$

As noted before, $J_2 = g(H)$ is a function of H. We avoid explicitly evaluating it by integrating it out in the next step to find the density of T.

The joint probability density of T and H is found by the change of variables

$$f(YY^H)dY = f[(TH)(TH)^H]J_1J_2J_3d(T, H) = f(TT^H)J_1J_2J_3d(T, H)$$

Thus

$$f(YY^H) = f(TT^H)\left(\prod_{i=1}^{p} t_{ii}^{2m}\right) g(h) \left(\prod_{i=1}^{p} t_{ii}^{-2i+1}\right)$$

$$= p(T,H) = f(TT^H)g(H)\prod_{i=1}^{p} t_{ii}^{2(m-i)+1}$$

Integrating out H to obtain the density of T alone, we get

$$p(T) = \int p(T,H)(dH) = \left(\prod_{i=1}^{p} t_{ii}^{2(m-i)+1}\right) f(TT^H) \int_H g(H)(dH) \qquad \text{(E.5)}$$

where the integral is over all H such that $HH^H = I_p$. Let $C_1 = \int_H g(H)(dH)$. This C_1 will be evaluated later, and will be shown to contain the information about the distribution of Y. At this point, it is worth pointing out that $f(YY^H)$ can by *any* function of YY^H, and we are merely doing changes of variables. This only takes on importance in probability when we later choose a function f of a quadratic which also turns out to be a probability density function. So, this derivation is really quite general. In the search for the complex Wishart density, we will choose $f(ZZ^H)$ such that $f(Z)$ is the complex matrix standard normal distribution.

Make the transformation

$$B = YY^H = (TH)(TH)^H = THH^HT^H = TT^H$$

By Khatri section (2.8) [137], which was proven as theorem 26,

$$J(B \to TT^H) = 2^p \prod_{i=1}^{p} t_{ii}^{2(p-i)+1}$$

T must be lower triangular for this to be true. Thus

$$J(TT^H \to B) = 2^{-p} \prod_{i=1}^{p} t_{ii}^{-2(p-i)-1}$$

Hence, the density of B is

$$p(B) = \left(\prod_{i=1}^{p} t_{ii}^{2(m-i)+1}\right) C_1 \left(2^{-p} \prod_{i=1}^{p} t_{ii}^{-2(p-i)-1}\right) f(B)$$

which simplifies to

$$p(B) = C_1 2^{-p} f(B) \prod_{i=1}^{p} t_{ii}^{2(m-p)} = C_1 2^{-p} f(B) \left|\det T\right|^{2(m-p)}$$

$$= C_1 2^{-p} f(B) \left|\det(TT^H)\right|^{m-p}$$

since T has real diagonal elements. Finally,

$$p(B) = C_1 2^{-p} \left|\det B\right|^{m-p} f(B) \tag{E.6}$$

This is Srivastava's main result.

Specialization to Complex Wishart

Srivastava's main result is now specialized to the case where $Y \sim \mathbf{C}N_{m,p}(0, I_m, I_p)$ by evaluating the constant C_1. The probability density for Y is given by theorem 51 as

$$p(Y) = \pi^{-mp} \operatorname{etr}[-YY^H] = \pi^{-mp} \operatorname{etr}[-(TH)(TH)^H]$$

$$= \pi^{-mp} \operatorname{etr}[-TT^H] = f(TT^H)$$

Thus, $f(B) = \pi^{-mp} \operatorname{etr}[-B]$.

Returning to equation E.5, we have

$$p(T) = C_1 f(TT^H) \prod_{i=1}^{p} t_{ii}^{2(m-i)+1} = C_1 \pi^{-mp} \operatorname{etr}(-TT^H) \prod_{i=1}^{p} t_{ii}^{2(m-i)+1}$$

Since $p(T)$ is a probability density, if we integrate over all T, we get unity.

$$1 = \int p(T)(dT) = C_1 \pi^{-mp} \int \text{etr}(-TT^H) \prod_{i=1}^{p} t_{ii}^{2(m-i)+1} (dT)$$

Concentrate on the integral. From the proof of theorem 26, note that

$$\text{tr}(TT^H) = t_{11}^2 + |t_{21}|^2 + t_{22}^2 + |t_{31}|^2 + |t_{32}|^2 + t_{33}^2 + \cdots + |t_{p-1,1}|^2 +$$

$$+ |t_{p-1,2}|^2 + \cdots + t_{p-1,p-1}^2 + |t_{p1}|^2 + |t_{p2}|^2 + \cdots + t_{pp}^2$$

The integral can thus be expanded as

$$\mathcal{I} = \int \text{etr}(-TT^H) \left(\prod_{i=1}^{p} t_{ii}^{2(m-i)+1} \right) (dT)$$

$$= \int e^{-t_{11}^2} e^{-|t_{21}|^2} e^{-t_{22}^2} e^{-|t_{31}|^2} \cdots e^{t_{pp}^2} t_{11}^{2(m-1)+1} t_{22}^{2(m-2)+1} \cdots t_{pp}^{2(m-p)+1} (dT)$$

$$= \left(\int t_{11}^{2(m-1)+1} e^{-t_{11}^2} dt_{11} \right) \left(\int t_{22}^{2(m-2)+1} e^{-t_{22}^2} dt_{22} \right) \cdots \left(\int t_{pp}^{2(m-p)+1} e^{-t_{pp}^2} dt_{pp} \right)$$

$$\times \left(\int e^{-|t_{21}|^2} dt_{21} \right) \left(\int e^{-|t_{31}|^2} dt_{31} \right) \cdots \left(\int e^{-|t_{p,p-1}|^2} dt_{p,p-1} \right)$$

$$= \left[\prod_{i=1}^{p} \left(\int t_{ii}^{2(m-i)+1} e^{-t_{ii}^2} dt_{ii} \right) \right] \left[\prod_{i=2}^{p} \prod_{j=1}^{i-1} \left(\int e^{-|t_{ij}|^2} dt_{ij} \right) \right]$$

By lemma 64, $\int e^{-|t_{ij}|^2} dt_{ij} = \pi$, and by lemma 65,

$$\int t_{ii}^{2(m-i)+1} e^{-t_{ii}^2} dt_{ii} = \frac{1}{2} \Gamma(m-i+1), \;\; m-i+1 > 0$$

Therefore,

$$\mathcal{I} = \left[\prod_{i=1}^{p} \left(\frac{1}{2} \Gamma(m-i+1) \right) \right] \left[\prod_{i=2}^{p} \prod_{j=1}^{i-1} \pi \right] = 2^{-p} \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)$$

where $m - p + 1 > 0$. Returning to the evaluation of C_1, we see

$$1 = C_1 \pi^{-mp} 2^{-p} \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)$$

$$= C_1 2^{-p} \pi^{p[\frac{1}{2}(p-1)-m]} \prod_{i=1}^{p} \Gamma(m-i+1)$$

Solving for C_1 yields

$$C_1 = \frac{1}{2^{-p}\pi^{p[\frac{1}{2}(p-1)-m]} \prod_{i=1}^{p} \Gamma(m-i+1)} = \frac{2^p \pi^{pm}}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)}$$

Substituting into equation E.6 yields theorem 67.

$$p(B) = \frac{2^p}{\pi^{-pm}\mathbf{C}\Gamma_p(m)} 2^{-p} |\det B|^{m-p} f(B) = \frac{|\det B|^{m-p} f(B)}{\pi^{-pm}\mathbf{C}\Gamma_p(m)}$$

Substituting $f(B) = \pi^{-mp} \operatorname{etr}(-B)$ gives us

$$p(B) = \frac{|\det B|^{m-p} \operatorname{etr}(-B)}{\mathbf{C}\Gamma_p(m)} = \frac{|\det B|^{m-p} \operatorname{etr}(-B)}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)}, \text{ for } m-p+1>0$$

This is Goodman's result when $\Sigma = I$.

Suppose that $Y \sim \mathbf{C}N_{m,p}(0_{m\times p}, I_m, \Sigma_{p\times p})$. This is the same as having a random sample of size m from the complex normal distribution $\mathbf{C}N_p(0_p, \Sigma_{p\times p})$. By theorem 51, the density of Y is

$$p(Y) = \pi^{-mp} |\Sigma|^{-m} \operatorname{etr}(-Y\Sigma^{-1}Y^H)$$

By corollary 36, since Σ^{-1} is positive definite, it can be factored into $\Sigma^{-1} = TT^H$ where T is $p \times p$ lower triangular with positive real diagonal elements. Thus

$$p(Y) = \pi^{-mp} |\Sigma|^{-m} \operatorname{etr}(-YTT^HY^H)$$

Let $X = YT$ which implies $Y = XT^{-1}$. By lemma 5, $J(Y \to X) = \prod_{i=1}^{p} t_{ii}^{-2m}$. Then $p(XT^{-1}) = \pi^{-mp} |\Sigma|^{-m} \operatorname{etr}(-XX^H)$.

We want to find the density function for $\mathbf{C}W_p(m, \Sigma)$. We get $\mathbf{C}W_p(m, \Sigma)$ variables by obtaining a random sample of size m from the complex multivariate normal distribution $\mathbf{C}N_p(0_p, \Sigma_{p\times p})$. This yields a complex matrix normal random variable $Y \sim \mathbf{C}N_{m,p}(0_{m\times p}, I_m, \Sigma_{p\times p})$.

By corollary 36 there exists a unique $p \times p$ lower triangular matrix L with positive diagonal elements such that $\Sigma = LL^H$. Let $B \sim \mathbf{C}W_p(m, I)$. Then

$$W = LBL^H \sim \mathbf{C}W_p(m, LIL^H) = \mathbf{C}W_p(m, \Sigma)$$

by theorem 54. By theorem 24, $J(B \to W) = (\det \Sigma)^{-p}$. Thus

$$f(W) = p(L^{-1}WL^{-H})J(B \to W)$$

$$= \frac{\left|\det(L^{-1}WL^{-H})\right|^{m-p} \operatorname{etr}(-L^{-1}WL^{-H})}{\pi^{p(p-1)/2}(\det \Sigma)^p \prod_{i=1}^{p} \Gamma(m-i+1)}$$

$$= \frac{(\det L)^{-(m-p)} \left|\det W\right|^{m-p} (\det L^H)^{-(m-p)} \operatorname{etr}(-L^{-H}L^{-1}W)}{(\det \Sigma)^p \mathbf{C}\Gamma_p(m)}$$

$$= \frac{(\det \Sigma)^{-(m-p)} \left|\det W\right|^{m-p} \operatorname{etr}(-\Sigma^{-1}W)}{(\det \Sigma)^p \mathbf{C}\Gamma_p(m)}$$

Therefore

$$f(W) = \frac{\left|\det W\right|^{m-p} \operatorname{etr}(-\Sigma^{-1}W)}{(\det \Sigma)^m \mathbf{C}\Gamma_p(m)}(dW)$$

This is the same answer obtained by Goodman [92]. The extension of Strivastava's result [256] to $\mathbf{C}W_p(m, \Sigma)$ was motivated by Arnold's introduction [31] to the proof of his Corollary to his Theorem 17.12, which is an extension of $W_p(m, I)$ to $W_p(m, \Sigma)$.

E.2 General Theorem on Density of Eigenvalues

Theorem 68 *(*Important*) If the Hermitian matrix X has a density of the form*

$$g(f_1, f_2, \cdots, f_p) \text{ where } f_1 > f_2 > \cdots > f_p$$

are the eigenvalues of X, then the joint density of the roots is given by

$$g(f_1, f_2, \cdots, f_p) \frac{\pi^{p(p-1)}}{\mathbf{C}\Gamma_p(p)} \prod_{i<j}^{p} (f_i - f_j)^2$$

This is a complexification of theorem 13.3.1 of Anderson (p. 532)[26], which is also similar to theorem 3.2.17 of Muirhead [187].

Proof. I have followed Anderson's general logic, substituting the Jacobians and other necessary changes to transform it to the complex case. Recall from theorem 7 that the joint density of the eigenvalues $\{f_i\}_1^p$ that satisfy $\det[A - f(A+B)] = 0$ where $A \sim \mathbf{C}W_p(m, I_p)$ and $B \sim \mathbf{C}W_p(n, I_p)$ is given by

$$g(F) = \left(\frac{\pi^{p(p-1)} \mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \right) \left[\prod_{i=1}^{p} f_i^{m-p} (1 - f_i)^{n-p} \right] \left[\prod_{i<j}^{p} (f_i - f_j)^2 \right]$$

Suppose we let $A = WW^H$, $G = CC^H = B + WW^H$, and $W = CU$. Then

$$0 = \det[A - f(A+B)] = \det[WW^H - fG]$$

$$= \det[CUU^H C^H - fCC^H] = \det(C) \det(UU^H - fI_p) \det(C^H)$$

Therefore, the roots of $\det[A - f(A+B)] = 0$ are also the roots of

$$\det(UU^H - fI_p) = 0$$

The general strategy is to first do a change of variables using the eigenvalue decomposition $X = CFC^H$ where $F = \text{diag}(f_1, \cdots, f_p)$ and $f_1 > \cdots > f_p$. We know from theorem 115 that we can do this. As we reasoned in theorem 7, we choose the phase θ_k of the scaling for each c_{k1} so that $e^{i\theta_k}c_{k1} \geq 0$ to force the transformation from X to (F, C) to be unique. Let the Jacobian of this transformation be $J[X \rightarrow (F, C)]$. Then the joint density of (F, C) is

$$g(f_1, \cdots, f_p)J[X \rightarrow F, C]$$

The marginal density of F is given by

$$g(f_1, \cdots, f_p)\int_C J[X \rightarrow F, C](dC) = g(X)\int_C J[X \rightarrow F, C](dC)$$

To evaluate $\int_C J[X \rightarrow F, C](dC)$, we pick a distribution for X for which we know the marginal density, $g(F)$. We then have

$$g(X)\int_C J[X \rightarrow F, C](dC) = g(F)$$

Thus

$$\int_C J[X \rightarrow F, C](dC) = \frac{g(F)}{g(X)}$$

We want to choose a distribution for X that will give us an answer easily. Let $X = UU^H$. In theorem 94 we constructed a random variable U with the

density function

$$g(U) = \frac{\mathbf{C}\Gamma_p(m+n)}{\pi^{mp}\mathbf{C}\Gamma_p(n)} \left|\det(I_p - UU^H)\right|^{n-p}$$

Using theorem 67, since $g(U)$ is a function of UU^H, then the density of $X = UU^H$ is given by

$$g(X) = \frac{|\det X|^{m-p}}{\pi^{-pm}\mathbf{C}\Gamma_p(m)} \frac{\mathbf{C}\Gamma_p(m+n)}{\pi^{mp}\mathbf{C}\Gamma_p(n)} |\det(I_p - X)|^{n-p}$$

$$= \frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} |\det X|^{m-p} |\det(I_p - X)|^{n-p}$$

Since the eigenvalues of X are $\{f_i\}_1^p$, the density of X is a function of its eigenvalues.

$$g(X) = \frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} \left|\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right|$$

where we know by lemma 54 that $I_p - X$ has eigenvalues $1 - f_i$. The joint density of (F, C) is then $g(f_1, \cdots, f_p)\, J(X \to F, C)$. The marginal density of F is given by

$$g(F) = g(f_1, \cdots, f_p) \int_C J(X \to F, C)(dC)$$

$$= \frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} \prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p} \int_C J(X \to F, C)(dC)$$

We know $g(F)$ from the beginning of this proof to be

$$g(F) = \frac{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right] \left[\prod_{i<j}^{p} (f_i - f_j)^2\right]$$

Solving for the integral, we find

$$\int_C J(X \to F, C)(dC) = \frac{\frac{\pi^{p(p-1)}\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right] \left[\prod_{i<j}^{p} (f_i - f_j)^2\right]}{\frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(m)\mathbf{C}\Gamma_p(n)} \left[\prod_{i=1}^{p} f_i^{m-p}(1-f_i)^{n-p}\right]}$$

$$= \frac{\pi^{p(p-1)}}{\mathbb{C}\Gamma_p(p)} \left[\prod_{i<j}^{p} (f_i - f_j)^2 \right]$$

Therefore, the density of the roots of X is

$$h(f_1, \cdots, f_p) = g(f_1, \cdots, f_p) \frac{\pi^{p(p-1)}}{\mathbb{C}\Gamma_p(p)} \left[\prod_{i<j}^{p} (f_i - f_j)^2 \right]$$

which finishes the proof. □

E.3 Joint Density of Eigenvalues of Complex Standard Wishart

Theorem 69 *Let $A \sim \mathbb{C}W_p(n, I_p)$. Then the joint density of the eigenvalues of A is given by*

$$h(l_1^2, \cdots, l_p^2) = \frac{\pi^{p(p-1)}}{\mathbb{C}\Gamma_p(n)\mathbb{C}\Gamma_p(p)} \exp\left[-\sum_{i-1}^{p} l_i^2 \right] \left[\prod_{i=1}^{p} l_i^{2(n-p)} \right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right]$$

This is a complexification of a theorem by Anderson (p. 534) [26]. It agrees with James [120] equation (95) for the case of $\Sigma = I_p$ and with Khatri [137] equation (7.1.7).

Proof. The density of A is given by

$$g(A) = \frac{|\det A|^{n-p} \operatorname{etr}(-A)}{\mathbb{C}\Gamma_p(n)}$$

By theorem 68, the joint density of the eigenvalues of A is given by

$$h(l_1^2, \cdots, l_p^2) = \frac{\left[\prod_{i=1}^{p} l_i^{2(n-p)} \right] \exp\left[-\sum_{i-1}^{p} l_i^2 \right]}{\mathbb{C}\Gamma_p(n)} \frac{\pi^{p(p-1)}}{\mathbb{C}\Gamma_p(p)} \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right]$$

$$= \frac{\pi^{p(p-1)}}{\mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)} \exp\left[-\sum_{i-1}^{p} l_i^2\right] \left[\prod_{i=1}^{p} l_i^{2(n-p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right]$$

which completes the proof.

E.4 Joint Density of Eigenvalues of Complex Wishart

Theorem 70 *(*Important*) Let* $A \sim \mathrm{C}W_p(n, \Sigma)$. *Then the joint density of the eigenvalues* $(l_1^2, \cdots, l_p^2)$ *of* A *is*

$$\frac{\pi^{p(p-1)}}{\mathrm{C}\Gamma_p(n)\mathrm{C}\Gamma_p(p)} \frac{{}_0\tilde{F}_0(-\Sigma^{-1}, A)}{|\det \Sigma|^n} \left[\prod_{i=1}^{p} l_i^{2(n-p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right]$$

This result was written down by inspection without derivation by James as his equation (95) [120] for the complex case citing the similarity of forms of the real case. My solution is written in terms of singular values rather than eigenvalues, and thus my l_i^2 corresponds to other authors' l_i. The proof that follows is done without making reference to the case for real variables.

Proof. We begin by following the theme of an earlier paper by James [117]. The primary concept is to recognize that the distribution being sought is invariant over some group. This leads to the approach of averaging over that group. James [118][120] gave an introduction to the group structure that justified his approach. A more complete construction of the group under study is given in section H.6 of this thesis. We begin by noting that the distribution

of the central complex Wishart variable A is

$$g(A) = \frac{|\det A|^{n-p} \operatorname{etr}(-\Sigma^{-1}A)}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)}(dA)$$

We want the distribution of the set D of eigenvalues of A. Let $A = U_1 D U_1^H$. Notice that $|\det A|^{n-p}$ is a function only of D.

The term we need to deal with is $\operatorname{etr}(-\Sigma^{-1}A)$. Recall that the similarity transformation $B = U^H A U$ leaves the eigenvalues unchanged where U is a unitary matrix, $U \in \mathbf{U}(p)$. In fact, a convex sum of distributions is again a distribution. Let $U_1, \cdots U_r$ be r fixed unitary matrices and let $\nu_1, \cdots, \nu_r$ be positive real numbers such that $\sum_{i=1}^{r} \nu_i = 1$. Then if A has the distribution

$$\frac{|\det A|^{n-p}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)} \sum_{i=1}^{r} \nu_i \operatorname{etr}(-\Sigma^{-1} U_i^H A U_i)(dA)$$

the distribution of D is unchanged. With a suitable choice of a sequence of sets of (U_i, ν_i), this function tends to

$$g(f_1, \cdots, f_p) = \frac{|\det A|^{n-p}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)} \int_{\mathbf{U}(p)} \operatorname{etr}(-\Sigma^{-1} U^H A U)(dU)(dA)$$

Notice that the function $\int_{\mathbf{U}(p)} \operatorname{etr}(-\Sigma^{-1} U^H A U)(dU)$, after it is evaluated, is only a function of Σ and the eigenvalues D of A. Only the elements of D are left as random variables. This ends the portion of the proof given in [117]. We now apply theorem 68. Thus, the density of D is given by

$$dF(D) = \frac{|\det A|^{n-p} \int_{\mathbf{U}(p)} \operatorname{etr}(-\Sigma^{-1} U^H A U)(dU)}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)} \frac{\pi^{p(p-1)}}{\mathbf{C}\Gamma_p(p)} \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2 \right] (dD)$$

Rearranging slightly, we get $dF(D) =$

$$\frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\int_{\mathbf{U}(p)} \operatorname{etr}(-\Sigma^{-1}U^H A U)(dU)\right] \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dD) \tag{E.7}$$

Recall that $|\det A|^{n-p} = \prod_{i=1}^{p} l_i^{2(n-p)}$. This result is the complexification of Muirhead theorem 3.2.18.

I supplied the next portion of this proof, up to the statement of the next corollary. Let us take a closer look at the integral. Let $A = U_1 D U_1^H$ and $\Sigma = P\Lambda^2 P^H$. Then

$$\begin{aligned}\int_{\mathbf{U}(p)} \operatorname{etr}(-\Sigma^{-1}U^H A U)(dU) &= \int_{\mathbf{U}(p)} \operatorname{etr}(-(P\Lambda^2 P^H)^{-1}U^H(U_1 D U_1^H)U)(dU)\\ &= \int_{\mathbf{U}(p)} \operatorname{etr}(-P\Lambda^{-2}P^H U^H U_1 D U_1^H U)(dU)\\ &= \int_{\mathbf{U}(p)} \operatorname{etr}(-\Lambda^{-2}P^H U^H U_1 D U_1^H U P)(dU)\end{aligned}$$

Suppose U and P are both members of the set $\mathbf{U}(p)$ of $p \times p$ unitary matrices. By definition, we know

$$U^H U = U U^H = P^H P = P P^H = I_p$$

This implies

$$(UP)^H(UP) = P^H U^H U P = I_p$$

Therefore the set $\mathbf{U}(p)$ is closed under matrix multiplication. Let $V = U_1^H U P \in \mathbf{U}(p)$. Then $(dU) = (dV)$ because U_1 and P are unitary. Therefore our integral is

$$\int_{\mathbf{U}(p)} \operatorname{etr}(-\Lambda^{-2}V^H D V)(dV) = \int_{\mathbf{U}(p)} \operatorname{etr}\left[-\sum_{j=1}^{p}\sum_{i=1}^{p} \frac{l_i^2}{\lambda_j^2}|v_{ij}|^2\right](dV)$$

where $\Lambda^{-2} = \text{diag}(\lambda_1^{-2}, \cdots, \lambda_p^{-2})$ and $D = \text{diag}(l_1^2, \cdots, l_p^2)$. To see this, observe

$$-\text{tr}(\Lambda^{-2}V^H DV) = -\text{tr}\left[\Lambda^{-2}\begin{pmatrix} v_1^H \\ \vdots \\ v_p^H \end{pmatrix} D \begin{pmatrix} v_1 & v_p \end{pmatrix}\right]$$

$$= -\text{tr}\left[\Lambda^{-2}\begin{pmatrix} v_1^H Dv_1 & \cdots & v_1^H Dv_p \\ \vdots & \ddots & \vdots \\ v_p^H Dv_1 & \cdots & v_p^H Dv_p \end{pmatrix}\right]$$

$$= -\text{tr}\begin{pmatrix} \lambda_1^{-2} v_1^H Dv_1 & \cdots & \lambda_1^{-2} v_1^H Dv_p \\ \vdots & \ddots & \vdots \\ \lambda_p^{-2} v_p^H Dv_1 & \cdots & \lambda_p^{-2} v_p^H Dv_p \end{pmatrix}$$

$$= -\text{tr}(\lambda_1^{-2} v_1^H Dv_1 + \cdots + \lambda_p^{-2} v_p^H Dv_p) = -[\lambda_1^{-2}\,\text{tr}(v_1^H Dv_1) + \cdots \lambda_p^{-2}\,\text{tr}(v_p^H Dv_p)]$$

Note that $v_j^H Dv_j$ is a scalar. Then

$$v_j^H Dv_j = \begin{pmatrix} v_{1j}^* & \cdots & v_{pj}^* \end{pmatrix}\begin{pmatrix} l_1^2 & & \\ & \ddots & \\ & & l_p^2 \end{pmatrix}\begin{pmatrix} v_{1j} \\ \vdots \\ v_{pj} \end{pmatrix}$$

$$= \sum_{i=1}^{p} v_{ij}^* l_i^2 v_{ij} = \sum_{i=1}^{p} l_i^2 \,|v_{ij}|^2$$

The trace becomes

$$-\sum_{j=1}^{p}\sum_{i=1}^{p} \lambda_j^{-2} l_i^2 \,|v_{ij}|^2$$

Here is an intermediate result.

Corollary 21 *Let $A \sim \mathbf{CW}_p(n, \Sigma = \sigma^2 I_p)$ have eigenvalue decomposition $A = U_1 D U_1^H$, and let $V \in \mathbf{U}(p)$. $D = \mathrm{diag}(l_1^2, \cdots, l_p^2)$. Then the joint density function of the sample eigenvalues is given by*

$$dF(D) = \frac{\pi^{p(p-1)}}{\sigma^{2pn}} \frac{\exp\left(-\frac{1}{\sigma^2}\sum_{i=1}^{p} l_i^2\right)}{\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\prod_{i=1}^{p} l_i^{2(n-p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD)$$

This result is the complexification of Muirhead [187] corollary 3.2.19. When $\sigma^2 = 1$, this result is Khatri's equation (7.1.7) [137]. When $\sigma^2 = 1$ and n is replaced by $\frac{n+p-1}{2}$, this is Krishnaiah and Schuurmann equation (3.1) [151].

Proof. This is a complexification of Muirhead's proof. $\Lambda^{-2} = \frac{1}{\sigma^2} I_p$ and

$$\int_{\mathbf{U}(p)} \mathrm{etr}(-\Lambda^{-2} V^H D V)(dV) = \int_{\mathbf{U}(p)} \mathrm{etr}(-\frac{1}{\sigma^2} V^H D V)(dV)$$

$$= \int_{\mathbf{U}(p)} \mathrm{etr}(-\frac{1}{\sigma^2} D V V^H)(dV) = \mathrm{etr}(-\frac{1}{\sigma^2} D) \int_{\mathbf{U}(p)} (dV) = \exp\left(-\frac{1}{\sigma^2}\sum_{i=1}^{p} l_i^2\right)$$

since $VV^H = I_p$. Substitute this into equation E.7 to obtain the result where $\det A = \prod_{i=1}^{p} l_i^2$. Note that the Haar measure has been normalized so that $\int_{\mathbf{U}(p)} (dV) = 1$. □

We build slightly on this intermediate result.

Corollary 22 *Let $\Sigma = \lambda^2 I_p$ and $A \sim \mathbf{CW}_p(n, \lambda^2 I_p)$. Let $S = \frac{1}{n} A$, and let $D_S = \mathrm{diag}(l_1^2, \cdots, l_p^2)$ be the eigenvalues of S. Then the joint density of the sample eigenvalues of S is $dF(D_S) =$*

$$\left(\frac{n}{\lambda^2}\right)^{pn} \frac{\pi^{p(p-1)}}{\mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \exp\left(-\frac{n}{\lambda^2}\sum_{i=1}^{p} l_i^2\right) \left[\prod_{i=1}^{p} l_i^{2(n-p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD_S)$$

This is the complexification of Muirhead corollary 9.4.2 [187], which was stated without proof.

Proof. $S \sim \mathbf{C}W_p(n, \frac{\lambda^2}{n} I_p)$ by lemma 16. Substitute $\sigma^2 = \frac{\lambda^2}{n}$ into the previous result.□

We know $E\{S\} = \lambda^2 I_p$ by theorem 52. If $A \sim \mathbf{C}W_p(n, \lambda^2 I_p, \delta)$, then $S \sim \mathbf{C}W_p(n, \frac{\lambda^2}{n} I_p, \frac{1}{n}\delta)$ and $E\{S\} = \lambda^2 I_p + \frac{1}{n}\delta$. The expression for $dF(D_S)$ would be considerably messier.

We return to developing the central density function of the sample eigenvalues of a complex Wishart matrix. The term we must evaluate is

$$\int_{\mathbf{U}(p)} \text{etr}(-\Sigma^{-1} U^H A U)(dU)$$

To do this, we draw from the work by Gross and Richards [96]. Observe that Σ^{-1} and A are both nonsingular Hermitian matrices, and that $\mathbf{U}(p)$ is a maximal compact subgroup of $GL(p, \mathbf{C})$.

We proceed:

$$\int_{\mathbf{U}(p)} \text{etr}(-\Sigma^{-1} U^H A U)(dU) = \int_{\mathbf{U}(p)} \exp[\text{tr}(-\Sigma^{-1} U^H A U)](dU)$$

$$= \int_{\mathbf{U}(p)} \sum_{d=0}^{\infty} \frac{1}{d!} [\text{tr}(-\Sigma^{-1} U^H A U)]^d (dU) = \sum_{d=0}^{\infty} \frac{1}{d!} \int_{\mathbf{U}(p)} \sum_{|m|=d} Z_m(-\Sigma^{-1} U^H A U)(dU)$$

where Z_m is the zonal polynomial of order m with matrix argument

$$(-\Sigma^{-1} U^H A U)$$

The summation over $|m| = d$ means that the sum extends over all

$$m_1 + m_2 + \cdots = d < \infty$$

where

$$m_1 \geq m_2 \geq \cdots \geq 0$$

are integers. By the splitting theorem (proposition 41) we decompose the integral.

$$\int_{\mathbf{U}(p)} \text{etr}(-\Sigma^{-1}U^H AU)(dU) = \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \int_{\mathbf{U}(p)} Z_m(-\Sigma^{-1}U^H AU)(dU)$$

$$= \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Sigma^{-1})Z_m(A)}{Z_m(I_p)}$$

Thus, in terms of zonal polynomials, $dF(D) =$

$$\frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Sigma^{-1})Z_m(A)}{Z_m(I_p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD)$$

By definition 90, this becomes

$$dF(D) = \frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[{}_0F_0(-\Sigma^{-1}, A)\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD)$$

This is the form as given in James [120] equation (95). In [119], James stated that the zonal polynomials for the case of real positive definite symmetric matrices are the same for the complex orthogonal group in the complex full linear group, and the real orthogonal group in the unitary group. A *contribution of this thesis* is applying Gross and Richards' work [96] which is valid for Hermitian positive definite matrices. In particular, we are working with the unitary group in the complex general linear group. The appearance of the expression is the same. Its meaning is now extended to include our signal processing problem. This is the complex version of Muirhead [187] theorem 9.4.1.

Proof of this result completely in the context of the complex Wishart distribution is one of the major contributions of this research. The key insights were provided by Gross and Richards [96]. Because of the great importance of that result and because the required working set of mathematics involved is uncommon to engineers, their paper is explained with commentary in appendix G.

From Gross and Richards equation (5.4.5) we know that for Hermitian matrix $X = X^H$, zonal polynomials have the property that the value of Z_m at X is uniquely determined by the eigenvalues of X. That is, $Z_m(X) = Z_m(\Lambda^2)$. We can equivalently say that $Z_m(U^H X U) = Z_m(X)$ for all $U \in \mathbf{U}(n)$. Thus, in our problem, we know $Z_m(A) = Z_m(D)$ by using U_1 and $Z_m(-\Sigma) = Z_m(\Lambda^{-2})$ by using U_2 where Λ^2 is the matrix of eigenvalues of Σ. From this, we get the form $dF(D) =$

$$\frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|m|=d} \frac{Z_m(-\Lambda^{-2}) Z_m(D)}{Z_m(I_p)}\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD) \tag{E.8}$$

By definition 90, this is

$$dF(D) = \frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[{}_0F_0(-\Lambda^{-2}, D)\right] \left[\prod_{i<j}^{p} (l_i^2 - l_j^2)^2\right] (dD) \tag{E.9}$$

where ${}_0F_0(-\Lambda^{-2}, D)$ is the hypergeometric function of two matrix arguments $(-\Lambda^{-2}, D)$ and of numerator and denominator orders equal to zero. It is the joint probability density function of the sample eigenvalues $D = \mathrm{diag}(l_1^2, \cdots, l_p^2)$ where the population eigenvalues are $\Lambda^2 = \mathrm{diag}(\lambda_1^2, \cdots, \lambda_p^2)$.

The next result is one that I think is wrong. In the unlikely possibility that it is right, it opens up new practical possibilities. Recall the definition that ${}_0F_0(-\Lambda^{-2}, D) = \text{etr}(-\Lambda^{-2}D)$. Then

$$dF(D) = \frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[\text{etr}(-\Lambda^{-2}D)\right] \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dD) \quad \text{(E.10)}$$

$$= \frac{|\det A|^{n-p}\,\pi^{p(p-1)}}{[\det \Sigma]^n \mathbf{C}\Gamma_p(n)\mathbf{C}\Gamma_p(p)} \left[-\sum_{k=1}^{p}\frac{l_k^2}{\lambda_k^2}\right] \left[\prod_{i<j}^{p}(l_i^2 - l_j^2)^2\right](dD) \quad \text{(E.11)}$$

What is wrong with this is, that in general, $\text{etr}(-\Sigma^{-1}A) \neq \text{etr}(-\Lambda^{-2}D)$. That gives a practical immediate answer. Stopping the questioning process here, though, eliminates the insights we need to find out why fundamentally the derivation fails. I think that some steps of the derivation perhaps should have been only one way implications ($\Rightarrow$) rather than equalities [$(\Leftarrow) \cup (\Rightarrow)$].

E.5 Distribution of $F(2n, 2n)$

I have supplied all of the work in this section.

One of the proposed tests to determine the number of significant eigenvalues of a complex Wishart matrix involves obtaining two independent sets of data from which two independent complex Wishart matrices are formed. When the number of data samples used in forming those matrices are identical and even, then a simple form of the cumulative distribution function for the relevant test statistic is derivable in closed form. The appropriate test statistic was shown in theorem 6 to be distributed according to the F distribution with

$2n_1$ and $2n_2$ degrees of freedom. Corollary 1 specialized the result to apply specifically to comparing linear combinations of sample eigenvalues.

Attention is drawn to the work of Lentner [164] who developed a set of expressions for arbitrary positive integer degrees of freedom. When his work is restricted to the special case considered here, the results can be shown to be identical. The contribution of the following work is that there is a simple form when the condition of even degrees of freedom can be reasonably met.

The problem was solved by direct integration of the probability density function. Reduction of the resulting expression to its final form was made possible through application of an identity from combinatorics.

Theorem 71 *Let n be an even positive integer. Then the cumulative distribution function for the $F(n,n)$ distribution is given by*

$$\Pr\{F \leq f\} = (f+1)^{1-n} \sum_{k=n/2}^{n-1} \binom{n-1}{k} f^k = \frac{f^{n/2}}{(f+1)^{n-1}} \bigwedge_{k=n/2}^{n-1} \left[\binom{n-1}{k} + f \right]$$

Discussion. The coefficients of f^k are the right half of the corresponding n-row of Pascal's Triangle. This provides an efficient means for testing the significance of principal components in the small sample case.

Proof. Hogg and Craig [109] (p. 146) define the cumulative F distribution in terms of the probability density function as follows.

$$\Pr\{F \leq f\} = \int_0^f f(\omega) d\omega, \quad 0 < f < \infty \tag{E.12}$$

where

$$g(f) = \underbrace{\frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}}_{\Phi(r_1,r_2)} \underbrace{\frac{f^{(r_1/2)-1}}{\left(1+\frac{r_1}{r_2}f\right)^{(r_1+r_2)/2}}}_{h(r_1,r_2,f)} \tag{E.13}$$

This is the probability that the random variable F is less than or equal to the value of f.

For the eigenvalue test, the form of the F statistic has the same number of degrees of freedom in the numerator and the denominator. This means that $r_1 = r_2 = n$. This simplifies equation E.12 to $g(f) = \Phi(n)f(n,f)$ where

$$\Phi(n) = \frac{(n-1)!}{[\Gamma(n/2)]^2}$$

and

$$h(n,f) = \frac{f^{(n-2)/2}}{(1+f)^n}$$

The resulting integral still must be altered to ease the integration. The strategy is to take advantage of the property of $g(f)$ being a probability density function. This allows the problem to be cast into an integral whose upper limit is infinite. An additional change of variables brings the lower limit to zero, which allows the application of complex integration. The first step, then, is

$$\Pr\{F \leq f\} = \int_0^f \Phi(n)h(n,\omega)d\omega = 1 - \Phi(n)\int_f^\infty h(n,\omega)d\omega$$

$$\Pr\{F \leq f\} = 1 - \Phi(n)K(n,f) \tag{E.14}$$

where $K(n,f)$ is the integral of $h(n,f)$. To make the lower limit of integration zero, change variables to let $x = \omega - f$. Then $\omega = x + f$ which implies

$d\omega = dx$. The limits change from $\omega \in [f, \infty)$ to $x \in [0, \infty)$. With this change, the integral becomes

$$K(n, f) = \int_0^\infty \frac{(x+f)^{(n-2)/2}}{(1+x+f)^n} dx$$

Let z be a complex variable and let

$$h(n, z) = \frac{(z+f)^{(n-2)/2}}{(1+z+f)^n}$$

Note that $h(n, z)$ is not an "even" function. *To make the integration more tractable, restrict n to the set of even positive integers.* That is, let $n = 2m$ where m is a positive integer. This gets rid of evaluating a square root in the numerator. The function becomes

$$h(2m, z) = (z+f+1)^{-2m}(z+f)^{m-1}$$

Note that $h(2m, z)$ has a pole of order $2m$ located at $z_0 = (f+1)e^{i\pi}$. Because $h(2m, z)$ is not an "even" function, it becomes advantageous to use the integration technique given by Hayek [103].

E.5.1 Integration

Consider the integral

$$\int_0^\infty h(2m, z) \log z \, dz$$

The path of complex integration is given in figure E.1.

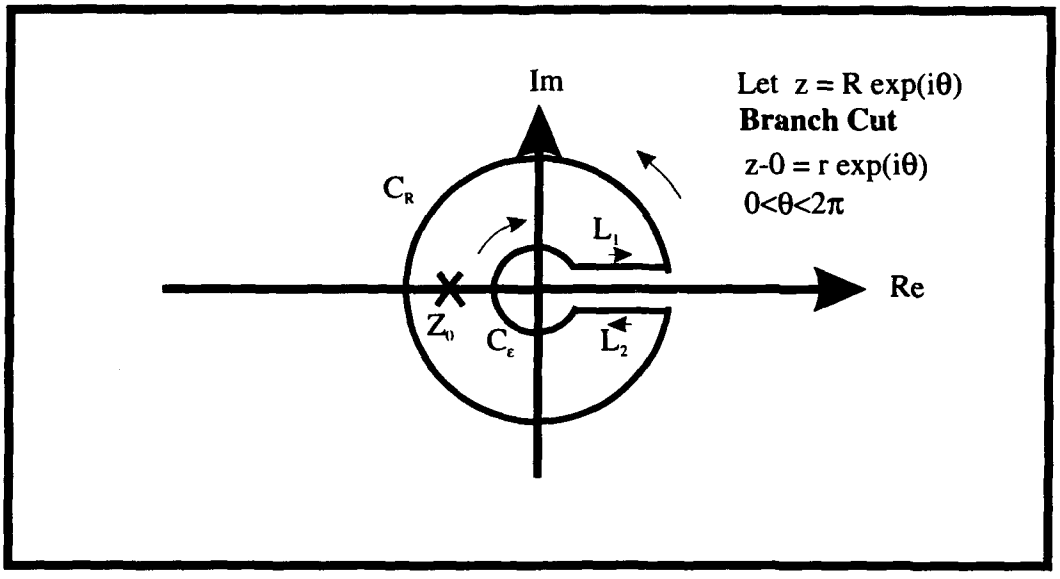


Figure E.1. Integration Contour to Get Cumulative Distribution Function

The integral is

$$\oint = \int_{C_R+L_2+C_\epsilon+L_1} = 2\pi i \text{ RES}\left[h(2m,z)\log z,\ z_0\right]_{2m}$$

$$= 2\pi i \text{ RES}\left[h(2m,z)\log z,\ (f+1)e^{i\pi}\right]_{2m}$$

Examine the integral of the function along the outer circle as the radius is allowed to become infinite.

$$\lim_{R\to\infty}\int_{C_R} = \lim_{R\to\infty}\left|\frac{\left(Re^{i\theta}+f\right)^{m-1}}{\left(Re^{i\theta}+f+1\right)^{2m}}\left(i\theta+\log R\right)iRe^{i\theta}\right| \tag{E.15}$$

$$= \lim_{R\to\infty}\left|\frac{R^{m-1}}{R^{2m}}R\log R\right| = \lim_{R\to\infty}\left|\frac{\log R}{R^m}\right| = 0, \quad m>0$$

Recall that we required m to be a positive integer in our hypothesis, so the condition on m is satisfied.

Table E.1. Change of Variables

Line	z	dz	log z
$L1$	R	dR	$\log R$
$L2$	$Re^{i2\pi}$	dR	$i2\pi + \log R$

Examine the integral of the function along the inner circle as its radius is allowed to vanish to zero.

$$\lim_{\epsilon\to 0}\int_{C_\epsilon} = \lim_{\epsilon\to 0}\left|\frac{\left(\epsilon e^{i\theta}+f\right)^{m-1}}{\left(\epsilon e^{i\theta}+f+1\right)^{2m}}\left(i\theta+\log\epsilon\right)i\epsilon e^{i\theta}\right| \tag{E.16}$$

$$= \lim_{\epsilon\to 0}\left|\frac{f^{m-1}}{(f+1)^{2m}}\left(i\theta+\log\epsilon\right)\epsilon\right| = \lim_{\epsilon\to 0}\left|\frac{f^{m-1}}{(f+1)^{2m}}\epsilon\log\epsilon\right|$$

$$= \lim_{\epsilon\to 0}\left|\frac{f^{m-1}}{(f+1)^{2m}}(-\epsilon)\right| = 0$$

The term $\log\epsilon$ dominates $i\theta$ as ϵ goes to zero. Invoking L'Hôpital's Rule on $\epsilon\log\epsilon$ demonstrates that this quantity goes to zero as the limit is applied. Note that there are no poles inside C_ϵ. The integral reduces to

$$\oint = \int_{L_1+L_2} = 2\pi i \text{ RES}\left[h(2m,z)\log z,\,(f+1)e^{i\pi}\right]_{2m} \tag{E.17}$$

The change of variables for evaluating the integrals are given in table E.1. The integrals which are evaluated are given in table E.2. Substituting these into equation E.17 yields

$$\oint = -i2\pi\int_0^\infty \frac{(R+f)^{m-1}}{(R+f+1)^{2m}}dR = 2\pi i \text{ RES}\left[h(2m,z)\log z,\,(f+1)e^{i\pi}\right]_{2m}$$

Rearranging to obtain the integral of original interest yields

$$\int_0^\infty \frac{(R+f)^{m-1}}{(R+f+1)^{2m}}dR = -\text{ RES}\left[h(2m,z)\log z,\,(f+1)e^{i\pi}\right]_{2m} \tag{E.18}$$

Table E.2. Line Integral Evaluation

Line	$\int_a^b \mathbf{h(2m,z)(\log z)\,dz}$
L1	$\int_0^\infty \frac{(R+f)^{m-1}}{(R+f+1)^{2m}}(\log R)\,dR$
L2	$\int_\infty^0 \frac{(R+f)^{m-1}}{(R+f+1)^{2m}}(i2\pi + \log R)\,dR$
	$= -\int_0^\infty \frac{(R+f)^{m-1}}{(R+f+1)^{2m}}(i2\pi + \log R)\,dR$

E.5.2 Evaluation of Residue

To complete the evaluation for the case of even values of n, the residue must be evaluated at the pole located at $z_0 = (f+1)e^{i\pi}$. Recall that this pole is of order $2m$. The residue is evaluated by

$$\text{RES}\left[h(2m,z)\log z,\,(f+1)e^{i\pi}\right]_{2m}$$

$$= \lim_{z\to(f+1)\exp(i\pi)} \frac{1}{(2m-1)!}\frac{d^{2m-1}}{dz^{2m-1}}\left[\left(z-(f+1)e^{i\pi}\right)^{2m} h(2m,z)\log z\right]$$

$$= \lim_{z\to(f+1)\exp(i\pi)} \frac{1}{(2m-1)!}\frac{d^{2m-1}}{dz^{2m-1}}\left[(z+f+1)^{2m}\,\frac{(z+f)^{m-1}}{(z+f+1)^{2m}}\log z\right]$$

$$= \lim_{z\to(f+1)\exp(i\pi)} \frac{1}{(2m-1)!}\frac{d^{2m-1}}{dz^{2m-1}}\left[(z+f)^{m-1}\log z\right] \qquad \text{(E.19)}$$

To evaluate this high order derivative, recall that

$$\frac{d^k}{dz^k} = \left(\frac{dR}{dz}\right)^k \frac{d^k}{dR^k}$$

Let $z = Re^{i\theta}$ be a change of variables. This implies $dz = e^{i\theta}dR$, $\left(\frac{dR}{dz}\right)^k = e^{-i\theta k}$, and $\frac{d^k}{dz^k} = e^{-i\theta k}\frac{d^k}{dR^k}$. Let $\theta = \pi$. Then

$$\frac{d^k}{dz^k} = (-1)^k \frac{d^k}{dR^k} \qquad \text{(E.20)}$$

Substituting equation E.20 into equation E.19 yields

$$\text{RES}\left[h(2m,z)\log z,\,(f+1)e^{i\pi}\right]_{2m}$$

$$=\lim_{R\to f+1}\frac{(-1)^{2m-1}}{(2m-1)!}\frac{d^{2m-1}}{dR^{2m-1}}\left[(f-R)^{m-1}\log\left(R\epsilon^{i\pi}\right)\right]$$

where the pole is being approached from the origin. This equals

$$\lim_{R\to f+1}\frac{-1}{(2m-1)!}\frac{d^{2m-1}}{dR^{2m-1}}\left[(f-R)^{m-1}\left(i\pi+\log R\right)\right]$$

$$=\lim_{R\to f+1}\frac{-1}{(2m-1)!}\left[i\pi\frac{d^{2m-1}}{dR^{2m-1}}(f-R)^{m-1}+\frac{d^{2m-1}}{dR^{2m-1}}(f-R)^{m-1}\log R\right] \tag{E.21}$$

Several other identities are provided below to aid the evaluation of the required derivatives.

$$\frac{d^k}{dx^k}(a+bx)^m\begin{cases}=b^k\frac{m!}{(m-k)!}(a+bx)^{m-k}, & k\le m\\ 0, & k>m\end{cases} \tag{E.22}$$

Equations E.23 and E.24 are taken from Tuma [268] (p. 86).

$$\frac{d^k}{dx^k}\ln x=x^{-k}(-1)^{k-1}(k-1)! \tag{E.23}$$

$$\frac{d^{2m-1}}{dR^{2m-1}}(uv)=\sum_{k=0}^{2m-1}\binom{2m-1}{k}u^{(k)}v^{(2m-1-k)} \tag{E.24}$$

Substituting these derivatives into equation E.21 yields

$$\lim_{R\to f+1}\frac{-1}{(2m-1)!}\frac{d^{2m-1}}{dR^{2m-1}}(f-R)^{m-1}\log R$$

Notice that the second condition of equation E.22 causes the first derivative term of equation E.21 to go to zero, and the second derivative summation limit

to be $m-1$. Therefore,

$$\mathrm{RES}[\cdot] = \lim_{R\to f+1} \frac{-1}{(2m-1)!} \left\{ \sum_{k=0}^{m-1} \binom{2m-1}{k} \underbrace{(-1)^k \frac{(m-1)!}{(m-1-k)!}(f-R)^{m-1-k}}_{u^{(k)}} \right.$$

$$\left. \times \underbrace{(-1)^{2m-2-k} \frac{(2m-2-k)!}{R^{2m-1-k}}}_{v^{(2m-1-k)}} \right\}$$

$$= \lim_{R\to f+1} \frac{-1}{(2m-1)!} \sum_{k=0}^{m-1} \frac{(2m-1)!}{k!(2m-1-k)!} \frac{(m-1)!}{(m-1-k)!}(f-R)^{m-1-k} \frac{(2m-2-k)!}{R^{2m-1-k}}$$

$$= \lim_{R\to f+1} -\sum_{k=0}^{m-1} \frac{(2m-2-k)!}{k!(2m-1-k)!} \frac{(m-1)!}{(m-1-k)!} \frac{(f-R)^{m-1-k}}{R^{2m-1-k}}$$

$$= \lim_{R\to f+1} -\sum_{k=0}^{m-1} \frac{1}{k!(2m-1-k)} \frac{(m-1)!}{(m-1-k)!} \frac{(f-R)^{m-1-k}}{R^{2m-1-k}} \qquad \text{(E.25)}$$

Applying the limit to equation E.25 on R yields

$$-\sum_{k=0}^{m-1} \frac{1}{k!(2m-1-k)} \frac{(m-1)!}{(m-1-k)!} \frac{(-1)^{m-1-k}}{(f+1)^{2m-1-k}}$$

$$= +\sum_{k=0}^{m-1} \frac{(m-1)!}{k!(m-1-k)!} \frac{(-1)^{m-k}}{(2m-1-k)(f+1)^{2m-1-k}}$$

$$= \sum_{k=0}^{m-1} \binom{m-1}{k} \frac{(-1)^{m-k}}{(2m-1-k)(f+1)^{2m-1-k}}, \quad m>0$$

Substituting this into equation E.18 gives us the integral required by equation E.14.

$$\Pr\{F \le f\}_{2m} = 1 - \Phi(2m)\, K(2m, f)$$

$$= 1 - \Phi(2m) \sum_{k=0}^{m-1} \binom{m-1}{k} \frac{(-1)^{m+1-k}}{(2m-1-k)(f+1)^{2m-1-k}}$$

$$\Pr\{F \le f\} = 1 - \frac{(2m-1)!}{[(m-1)!]^2} \sum_{k=0}^{m-1} \frac{(m-1)!}{k!(m-1-k)!} \frac{(-1)^{m+1-k}}{(2m-1-k)(f+1)^{2m-1-k}} \qquad \text{(E.26)}$$

E.5.3 Simplification

The reduction to a function of Pascal's Triangle is possible through cancelling factorials, performing a binomial expansion, and recognizing a combinatorial identity. Working on equation E.26, we get

$$\Pr\{F \leq f\} = 1 - \frac{(2m-1)!}{(m-1)!} \sum_{k=0}^{m-1} \frac{(-1)^{m+1-k}}{k!(m-1-k)!(2m-1-k)(f+1)^{2m-1-k}} \tag{E.27}$$

Pulling out the $(f+1)^{2m-1}$ term from the summation and reshuffling factorials gives us

$$1 - \frac{1}{(f+1)^{2m-1}} \sum_{k=0}^{m-1} \frac{(2m-1)!}{(m-1)!} \frac{1}{k!} \frac{1}{(m-1-k)!} \frac{(-1)^{m+1-k}}{(2m-1-k)} (f+1)^k \tag{E.28}$$

Applying the binomial expansion to $(f+1)^k$ yields

$$1 - \frac{1}{(f+1)^{2m-1}} \sum_{k=0}^{m-1} \frac{(2m-1)!}{(m-1)!} \frac{1}{k!(m-1-k)!} \frac{(-1)^{m+1-k}}{(2m-1-k)} \sum_{j=0}^{k} \binom{k}{j} f^j$$

$$= 1 - \frac{1}{(f+1)^{2m-1}} \sum_{k=0}^{m-1} \frac{(2m-1)!}{(m-1)!} \frac{(-1)^{m+1-k}}{k!(m-1-k)!(2m-1-k)} \sum_{j=0}^{k} \frac{k! f^j}{j!(k-j)!}$$

$$= 1 + \frac{1}{(f+1)^{2m-1}} \frac{(2m-1)!}{(m-1)!} \sum_{k=0}^{m-1} \frac{(-1)^{m-k}}{(m-1-k)!(2m-1-k)} \sum_{j=0}^{k} \frac{f^j}{j!(k-j)!} \tag{E.29}$$

The next stage of simplification requires a study of the expansion of equation E.29. The goal is to pull the term f out of the inner-most summation. Placing individual terms of equation E.29 summation into table E.3 allows recognition of a pattern that permits a change of indices.

Table E.3. Expansion of $\Pr[F \leq f]$ Summation

New	**Old**	**j = 0**	**j = 1**
$j = m-1$	$k = 0$	$\frac{(-1)^m f^0}{(m-1)!0!0!(2m-1)}$	
$j = m-2$	$k = 1$	$\frac{(-1)^{m-1} f^0}{(m-2)!0!1!(2m-2)}$	$\frac{(-1)^{m-1} f^1}{(m-2)!1!0!(2m-2)}$
$j = m-3$	$k = 2$	$\frac{(-1)^{m-2} f^0}{(m-3)!0!2!(2m-3)}$	$\frac{(-1)^{m-2} f^1}{(m-3)!1!1!(2m-3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$j = 1$	$k = m-2$	$\frac{(-1)^2 f^0}{1!0!(m-2)!(m+1)}$	$\frac{(-1)^2 f^1}{1!1!(m-3)!(m+1)}$
$j = 0$	$k = m-1$	$\frac{(-1)^1 f^0}{0!0!(m-1)!(m)}$	$\frac{(-1)^1 f^1}{0!1!(m-2)!(m)}$
	New	**k = 0**	**k = 1**

New	**Old**	**j = 2**	$\cdots$	**j = m − 1**
$j = m-3$	$k = 2$	$\frac{(-1)^{m-2} f^2}{(m-3)!2!0!(2m-3)}$		
$\vdots$	$\vdots$	$\vdots$		
$j = 0$	$k = m-1$	$\cdots$	$\cdots$	$\frac{(-1)^1 f^{m-1}}{0!(m-1)!0!(m)}$
	New	**k = 2**	$\cdots$	**k = m − 1**

Using the observation that the summation may be reordered allows equation E.29 to be reformed as

$$\Pr\{F \leq f\} = 1 + \frac{1}{(f+1)^{2m-1}} \frac{(2m-1)!}{(m-1)!} \sum_{k=0}^{m-1} \frac{f^k}{k!} \sum_{j=0}^{m-1-k} \frac{(-1)^{j+1}}{j!(m-1-k-j)!(m+j)} \tag{E.30}$$

Examine just the right most summation, with a " -1" factored out.

$$\begin{aligned}
&\sum_{j=0}^{m-1-k} \frac{(-1)^j}{j!(m-1-k-j)!(m+j)} \\
&= \frac{1}{m(m-1-k)!} \sum_{j=0}^{m-1-k} \frac{(m-1-k)!}{j!(m-1-k-j)!} \frac{(-1)^j m}{(m+j)} \\
&= \frac{1}{m(m-1-k)!} \sum_{j=0}^{m-1-k} \binom{m-1-k}{j} (-1)^j \frac{m}{(m+j)} \\
&= \frac{1}{m(m-1-k)!} \binom{2m-1-k}{m-1-k}^{-1}
\end{aligned} \tag{E.31}$$

The last step is made possible by an identity in Riordan [223] (p. 47). Substituting equation E.31 back into equation E.30 gives us

$$\begin{aligned}
\Pr\{F \leq f\} &= 1 - \frac{1}{(f+1)^{2m-1}} \frac{(2m-1)!}{(m-1)!} \sum_{k=0}^{m-1} \frac{f^k}{k!} \frac{1}{m(m-1-k)!} \frac{(m-1-k)!m!}{(2m-1-k)!} \\
&= 1 - \frac{1}{(f+1)^{2m-1}} \frac{(2m-1)!}{(m-1)!} \sum_{k=0}^{m-1} \frac{f^k}{k!} \frac{m(m-1)!}{m(2m-1-k)!} \\
&= 1 - \frac{1}{(f+1)^{2m-1}} \sum_{k=0}^{m-1} \frac{(2m-1)!\, f^k}{k!(2m-1-k)!} \\
&= 1 - \frac{1}{(f+1)^{2m-1}} \sum_{k=0}^{m-1} \binom{2m-1}{k} f^k \\
&= \frac{1}{(f+1)^{2m-1}} \left[(f+1)^{2m-1} - \sum_{k=0}^{m-1} \binom{2m-1}{k} f^k \right] \\
&= \frac{1}{(f+1)^{2m-1}} \left[\sum_{k=0}^{2m-1} \binom{2m-1}{k} f^k - \sum_{k=0}^{m-1} \binom{2m-1}{k} f^k \right]
\end{aligned}$$

which we obtained by the binomial expansion of $(f+1)^{2m-1}$. Thus

$$\Pr\{F \le f\} = \frac{1}{(f+1)^{2m-1}} \sum_{k=m}^{2m-1} \binom{2m-1}{k} f^k$$

Recall that $n = 2m$ for positive integer m. Then we obtain the final result that

$$\Pr\{F \le f\} = \frac{1}{(f+1)^{n-1}} \sum_{k=n/2}^{n-1} \binom{n-1}{k} f^k = \frac{f^{n\backslash 2}}{(f+1)^{n-1}} \bigwedge_{k=n/2}^{n-1} \left[\binom{n-1}{k} + f\right] \tag{E.32}$$

The coefficients of f^k come from the right half of Pascal's Triangle. Recall that Pascal's Triangle takes on the form given in table E.4. For example, when $n = 8$ we get

$$\Pr\{F \le f\}_8 = \left[35f^4 + 21f^5 + 7f^6 + f^7\right] \frac{1}{(f+1)^7}$$

$$= \frac{1}{(f+1)^7} \left\{\left(\left[\left(f+7\right) f + 21\right] f + 35\right) f^4\right\}$$

$$= \frac{f^4}{(f+1)^7} \bigwedge_{k=4}^{n-1} \left[\binom{7}{4} + f\right]$$

E.6 Ordered Versus Unordered Eigenvalues

I supplied this section.

When you study basic probability, an early exercise examines counting rules, permutations, and combinations. Olkin and Derman (p. 67, Counting Rule 5.3)[200] tells us that the number of distinguishable arrangements of n

Table E.4. Pascal's Triangle

n										
	1									
2	1									
	1	2	1							
4	1	3	3	1						
	1	4	6	4	1					
6	1	5	10	10	5	1				
	1	6	15	20	15	6	1			
8	1	7	21	35	35	21	7	1		
	1	8	28	56	70	56	28	8	1	
10	1	9	36	84	126	126	84	36	9	1
k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

items, n_1 of which are of type 1, n_2 of which are of type 2, ..., n_k of which are of type k, is given by

$$m = \frac{n!}{n_1! n_2! \cdots n_k!}$$

When you look at the density function of a vector random variable

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

you are including the specification of the order, an n-tuple $(z_1, z_2, \cdots, z_n)$. This is treated differently than $(z_2, z_1, \cdots . z_n)$. For our complex vector normal distribution, interchanging vector elements would yield a different covariance matrix.

We are interested in looking at an ordered set of eigenvalues. When we decompose our sample covariance matrix, we get something like

$$W = UL^2U^H$$

$$= \begin{pmatrix} U_1 & U_2 & \cdots & U_p \end{pmatrix} \begin{pmatrix} l_1^2 & & & \\ & l_2^2 & & \\ & & \ddots & \\ & & & l_p^2 \end{pmatrix} \begin{pmatrix} U_1^H \\ U_2^H \\ \vdots \\ U_p^H \end{pmatrix}$$

$$= \begin{pmatrix} l_1^2 U_1 & l_2^2 U_2 & \cdots & l_p^2 U_p \end{pmatrix} \begin{pmatrix} U_1^H \\ U_2^H \\ \vdots \\ U_p^H \end{pmatrix} = \sum_{k=1}^{p} l_k^2 U_k U_k^H = W$$

As you know, order of summation is unimportant (i.e., commutative). As long as you interchange positions of the eigenvectors to the same ordering as the eigenvalues, the sum of the products will remain the same. Let σ be our permutation of the index set $(1, 2, \cdots, k)$. Then

$$\sum_{j=1}^{p} l_{\sigma(j)}^2 U_{\sigma(j)} U_{\sigma(j)}^H = \sum_{k=1}^{p} l_k^2 U_k U_k^H$$

When $l_i^2 \neq l_j^2$ for all pairs of i and j, then there are $p!$ orderings of the paired eigenvalues and eigenvectors. Given that the density function for the ordered set of eigenvalues $(l_1^2, l_2^2, \cdots, l_p^2)$ is $g(L^2)$, then the density function for the unordered set of eigenvalues $(l_{\sigma(1)}^2, l_{\sigma(2)}^2, \cdots, l_{\sigma(p)}^2)$ is $p!\, g(L^2)$.

Appendix F

RESULTS FOR SIGNAL PROCESSING

This appendix provides results either directly related to known signal processing needs or are one step removed from those needs. Many of these results are a complexification of Arnold's section 17.7 [31]. It includes studies on forms involving the matrix complex normal distribution, the complex Wishart distribution, and functions on these forms. Most of the statistical work done in this thesis bears directly on making a path to the answer to the thesis question. With just a small amount of extra work, it was possible to produce results of value to other portions of the acoustic signal processing community. Many of those results are presented in this appendix. These forms include the trace, determinant, inverse, and some selected ratios. For completeness' sake, at the end of the chapter, some beamforming results by Tague will be presented to show the usefulness of these methods.

This appendix builds on itself as it goes along. Although you can get a highlight of results by looking at the theorem statements, the best way to find out how they live is to start from the beginning.

F.1 Trace Distributions

Theorem 72 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$ *and let* T *be an* $m \times r$ *complex matrix.*

Then

$$\mathrm{tr}(T^H Z) \sim \mathbf{C}N_1\left(\mathrm{tr}(T^H\mu), \mathrm{tr}(T^H \Xi T \Sigma)\right)$$

This is a complexification of Arnold's theorem 17.13(a) [31], which was stated without proof.

Proof. From equation D.9, recall the characteristic function for Z.

$$\Phi_Z(T) = \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[T^H\mu]\right) - \frac{1}{4}\,\mathrm{tr}[T^H \Xi T \Sigma]\right]$$

Let $u = \mathrm{tr}(T^H Z)$. Then

$$\Phi_u(t) = \mathcal{E}\{\exp[i\,\mathrm{Re}(t^H u)]\}$$

where t is a scalar and therefore commutes. Then

$$\Phi_u(t) = \mathcal{E}\{\exp[i\,\mathrm{Re}(t^H\,\mathrm{tr}[T^H Z])]\}$$

$$= \mathcal{E}\{\exp[i\,\mathrm{Re}(\mathrm{tr}[(Tt)^H Z])]\} = \Phi_Z(tT) = \Phi_{\mathrm{tr}(T^H Z)}(t)$$

since $u = \mathrm{tr}(T^H Z)$.

Alternately, by theorem 19

$$\Phi_{tr(T^H Z)} = \Phi_{T^H Z}(tI_r) = \Phi_Z(TtI_r) = \Phi_Z(tT)$$

with the last step justified by theorem 18.

Let $\tau = tT$. Then $\Phi_Z(tT) = \Phi_Z(\tau)$. Note that τ is a matrix of the same dimensions as T. The transform variable t is a scalar.

$$\Phi_Z(\tau) = \exp\left[i\,\mathrm{Re}\left(\mathrm{tr}[\tau^H\mu]\right) - \frac{1}{4}\,\mathrm{tr}[\tau^H \Xi \tau \Sigma]\right]$$

$$= \exp\left[i \operatorname{Re}\left(\operatorname{tr}[(tT)^H \mu]\right) - \frac{1}{4} \operatorname{tr}[(tT)^H \Xi (tT) \Sigma]\right]$$

$$= \exp\left[i \operatorname{Re}\left(t^H \operatorname{tr}[T^H \mu]\right) - \frac{1}{4} |t|^2 \operatorname{tr}[T^H \Xi T \Sigma]\right]$$

This is the characteristic function of a scalar complex normal random variable with mean $\operatorname{tr}(T^H \mu)$ and variance $\operatorname{tr}(T^H \Xi T \Sigma)$. Therefore,

$$\operatorname{tr}(T^H Z) \sim \mathbf{C}N_1\left(\operatorname{tr}[T^H \mu], \operatorname{tr}[T^H \Xi T \Sigma]\right)$$

□

Lemma 21 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, I, I)$. *Then*

$$\frac{2}{\sigma^2} \operatorname{tr}\left[Z^H Z\right] \sim \chi^2_{2mr}\left[\frac{2}{\sigma^2} \operatorname{tr}(\mu^H \mu)\right]$$

and when $\sigma^2 = 1$ *this is*

$$2 \operatorname{tr}\left[Z^H Z\right] \sim \chi^2_{2mr}\left[2 \operatorname{tr}(\mu^H \mu)\right]$$

This is a special case of the complexification of Arnold's theorem 17.13(b) [31] (which was stated without proof), and will be used in the proof of the more general case.

Proof. Let $Z = (Z_{ij})$, $\mu = (\mu_{ij})$ where the Z_{ij} are independent and $Z_{ij} \sim \mathbf{C}N_1(\mu_{ij}, 1)$. Then $Z \sim \mathbf{C}N_{m,r}(\mu, I, I)$ and $Z^H Z \sim \mathbf{C}W_r(m, I, \delta)$ where $\delta = \mu^H \mu$. Consider $Z^H Z$ directly.

$$\operatorname{tr}(Z^H Z) = \operatorname{tr}\left[\begin{pmatrix} Z_1^H \\ \vdots \\ Z_r^H \end{pmatrix} \begin{pmatrix} Z_1 & \cdots & Z_r \end{pmatrix}\right]$$

where $Z_{m\times r} = (Z_1, \cdots, Z_r)$ and $Z_i = \begin{pmatrix} Z_{1i} \\ \vdots \\ Z_{mi} \end{pmatrix}$.

$$\mathrm{tr}(Z^H Z) = \sum_{i=1}^{r} Z_i^H Z_i = \sum_{i=1}^{r} (Z_{1i}^* \cdots Z_{mi}^*) \begin{pmatrix} Z_{1i} \\ \vdots \\ Z_{mi} \end{pmatrix} = \sum_{i=1}^{r} \sum_{j=1}^{m} |Z_{ji}|^2$$

Note that this is the same answer as you get when you vectorize Z. Let

$$\tilde{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_r \end{pmatrix} = \begin{pmatrix} Z_{11} \\ Z_{21} \\ \vdots \\ Z_{m1} \\ \vdots \\ Z_{1r} \\ \vdots \\ Z_{mr} \end{pmatrix}$$

Then $\mathrm{tr}(Z^H Z) = \tilde{Z}^H \tilde{Z}$. Similarly, $\mathrm{tr}(\mu^H \mu) = \tilde{\mu}^H \tilde{\mu}$ when μ is vectorized in the same way. Then $\tilde{Z}_{mr\times 1} \sim \mathbf{C}N_{mr}(\tilde{\mu}, I)$, where $\sigma^2 = 1$. By theorem 53 and lemma 15 we know

$$\frac{2}{\sigma^2} \tilde{Z}^H \tilde{Z} \sim \chi^2_{2mr} \left(\frac{2}{\sigma^2} \tilde{\mu}^H \tilde{\mu} \right)$$

or

$$\tilde{Z}^H \tilde{Z} \sim \mathbf{C}W_1(mr, \sigma^2, \delta)$$

where

$$\tilde{\delta} = \tilde{\mu}^H \tilde{\mu} = \mathrm{tr}(\mu^H \mu)$$

Therefore,

$$\frac{2}{\sigma^2}\,\mathrm{tr}(Z^H Z) \sim \chi^2_{2mr}\left(\frac{2}{\sigma^2}\,\mathrm{tr}(\mu^H \mu)\right) = \chi^2_{2mr}\left(2\,\mathrm{tr}(\mu^H \mu)\right)$$

where $\sigma^2 = 1$. □

Theorem 73 *Let* $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$, *with Hermitian positive definite* Ξ *and* Σ. *Then*

$$2\,\mathrm{tr}\left[(Z-\mu)^H \Xi^{-1}(Z-\mu)\Sigma^{-1}\right] \sim \chi^2_{2mr}[2\,\mathrm{tr}(\mu^H \Xi^{-1}\mu\Sigma^{-1})]$$

This is a complexification of the first part of Arnold's theorem 17.13(b) [31], which was stated without proof. The real case for my result differs from Arnold's result.

Proof. Let $\Xi = AA^H$ and $\Sigma = B^H B$, which we know by theorem 119. Then

$$\mathrm{tr}\left[(Z-\mu)^H \Xi^{-1}(Z-\mu)\Sigma^{-1}\right] = \mathrm{tr}\left[(Z-\mu)^H (AA^H)^{-1}(Z-\mu)(B^H B)^{-1}\right]$$

$$= \mathrm{tr}\left[(Z-\mu)^H A^{-H} A^{-1}(Z-\mu)B^{-1}B^{-H}\right] = \mathrm{tr}\left[B^{-H}(Z-\mu)^H A^{-H} A^{-1}(Z-\mu)B^{-1}\right]$$

$$= \mathrm{tr}\left(\left[A^{-1}(Z-\mu)B^{-1}\right]^H \left[A^{-1}(Z-\mu)B^{-1}\right]\right) = \mathrm{tr}(Z^H Z)$$

where $Z = A^{-1}(Z-\mu)B^{-1}$. By theorem 41, $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$, which implies

$$Z = A^{-1}_{m\times m}(Z-\mu)_{m\times r} B^{-1}_{r\times r}$$

$$\sim \mathbf{C}N_{m,r}(A^{-1}\mu B^{-1}, A^{-1}\Xi A^{-H}, B^{-H}\Sigma B^{-1}) = \mathbf{C}N_{m,r}(A^{-1}\mu B^{-1}, I, I)$$

Then by lemma 21 we know

$$2\,\mathrm{tr}(Z^H Z) \sim \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\left(A^{-1}\mu B^{-1}\right)^H\left(A^{-1}\mu B^{-1}\right)\right\}\right]$$

$$= \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\left(A^{-1}\mu B^{-1}\right)\left(A^{-1}\mu B^{-1}\right)^H\right\}\right]$$

$$= \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{A^{-1}\mu B^{-1}B^{-H}\mu^H A^{-H}\right\}\right] = \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\mu(B^H B)^{-1}\mu^H (AA^H)^{-1}\right\}\right]$$

$$= \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\mu\Sigma^{-1}\mu^H\Xi^{-1}\right\}\right] = \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\mu^H\Xi^{-1}\mu\Sigma^{-1}\right\}\right]$$

Therefore

$$2\,\mathrm{tr}\left[(Z-\mu)^H\Xi^{-1}(Z-\mu)\Sigma^{-1}\right] \sim \chi^2_{2mr}\left[2\,\mathrm{tr}\left\{\mu^H\Xi^{-1}\mu\Sigma^{-1}\right\}\right]$$

which concludes the proof. □

Proposition 40 *Let* $Z_{m\times r} \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$, *with Hermitian positive definite* $\Xi_{m\times m}$ *and* $\Sigma_{r\times r}$. *Let*

$$\Xi^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Define the trace of a rectangular matrix to be the trace of that matrix when made square by augmentation with an appropriately chosen zero matrix. Then

$$\mathrm{tr}(\underbrace{Z^H\Xi^{-1}}_{r\times m}) = \mathrm{tr}(\underbrace{\Xi^{-H}Z}_{m\times r}) = \mathrm{tr}(\Xi^{-1}Z)$$

$$= \begin{cases} \mathbf{C}N_1\left(\mathrm{tr}\left(T^H\Xi^{-1}\mu\right), \mathrm{tr}\left(C_{11}\Sigma\right)\right) & \doteq \mathbf{C}N_1[\mathrm{tr}(\Xi^{-1}\mu), \mathrm{tr}(C_{11}\Sigma)] \quad \textit{for } m \geq r \\ \mathbf{C}N_1\left(\mathrm{tr}\left(T^H\Xi^{-1}\mu\right), \mathrm{tr}\left(\Xi^{-1}\Sigma_{11}\right)\right) & \doteq \mathbf{C}N_1[\mathrm{tr}(\Xi^{-1}\mu), \mathrm{tr}(\Xi^{-1}\Sigma_{11})] \quad \textit{for } m \leq r \end{cases}$$

Note that the argument of the trace function here is not a square matrix. The trace function is usually defined only for square matrices. Finally,

$$2\,\mathrm{tr}(Z^H\Xi^{-1}Z\Sigma^{-1}) \sim \chi^2_{2mr}\left[2\,\mathrm{tr}(\mu^H\Xi^{-1}\mu\Sigma^{-1})\right]$$

This is a complexification and extension of errata to Arnold's theorem 17.13(b) [31], which was stated without proof.

Proof. First note that $\Xi = \Xi^H$. Since $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$, then by theorem 41 we know $\Xi^{-1}Z \sim \mathbf{C}N_{m,r}(\Xi^{-1}\mu, \Xi^{-1}, \Sigma)$. Consider $\mathrm{tr}(\underbrace{T^H\Xi^{-1}Z}_{r\times r})$ where T is $m \times r$. We consider two cases, based on the comparison of m and r.

First, let $m \geq r$. Define $T = \begin{bmatrix} I_r \\ 0_{(m-r)\times r} \end{bmatrix}$. Then

$$T^H\Xi^{-1}Z = \begin{bmatrix} I_r & 0_{r\times(m-r)} \end{bmatrix}\Xi^{-1}Z = \begin{bmatrix} \Xi^{-1}Z & 0 \end{bmatrix}_{r\times r}$$

Also,

$$T^H\Xi^{-1}T = \begin{bmatrix} I_r & 0_{r\times(m-r)} \end{bmatrix}\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}\begin{bmatrix} I_r \\ 0_{(m-r)\times r} \end{bmatrix} = (I_r, 0)\begin{pmatrix} C_{11} \\ C_{12}^H \end{pmatrix} = C_{11}$$

Therefore by theorem 41

$$T^H\Xi^{-1}Z \sim \mathbf{C}N_{r,r}(T^H\Xi^{-1}\mu, C_{11}, \Sigma)$$

Notice that tr $\left[T^H \Xi^{-1} Z\right]$ is the sum of the elements down the main diagonal (identified by x in the matrix below) of the $m \times r$ rectangular array $\Xi^{-1} Z$.

$$\begin{bmatrix} x & y & y & y & 0 & 0 \\ y & x & y & y & 0 & 0 \\ y & y & x & y & 0 & 0 \\ y & y & y & x & 0 & 0 \\ y & y & y & y & 0 & 0 \\ y & y & y & y & 0 & 0 \end{bmatrix}$$

Let us define this as tr$[\Xi^{-1} Z]$. Now, consider tr$(T^H \Xi^{-1} T \Sigma) =$ tr$(C_{11} \Sigma)$. From our study on characteristic functions, recall from theorem 19 that $\Phi_{trX}(t) = \Phi_X(tI)$. Now, by equation D.9,

$$\Phi_{T^H \Xi^{-1} Z}(\tau) = \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(\tau^H T^H \Xi^{-1} \mu\right)\right] - \frac{1}{4} \operatorname{tr}\left(\tau^H C_{11} \tau \Sigma\right)\right\}$$

Now, let $\tau = tI$ to obtain the characteristic function of the trace of $T^H \Xi^{-1} Z$. We get

$$\begin{aligned} \Phi_{\operatorname{tr}(T^H \Xi^{-1} Z)}(t) &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(t^* I^H T^H \Xi^{-1} \mu\right)\right] - \frac{1}{4} \operatorname{tr}\left(I^H t^* C_{11} t I \Sigma\right)\right\} \\ &= \exp\left\{ i \operatorname{Re}\left[\operatorname{tr}\left(t^* T^H \Xi^{-1} \mu\right)\right] - \frac{1}{4} \operatorname{tr}\left(t^* t C_{11} \Sigma\right)\right\} \\ &= \exp\left\{ i \operatorname{Re}\left[t^* \operatorname{tr}\left(T^H \Xi^{-1} \mu\right)\right] - \frac{1}{4} |t|^2 \operatorname{tr}\left(C_{11} \Sigma\right)\right\} \end{aligned}$$

This is the characteristic function of the distribution

$$\mathbf{C}N_1\left(\operatorname{tr}\left(T^H \Xi^{-1} \mu\right), \operatorname{tr}\left(C_{11} \Sigma\right)\right)$$

By our definition for the trace of a rectangular matrix, we can call this

$$\mathbf{C}N_1\left(\mathrm{tr}\left(\Xi^{-1}\mu\right), \mathrm{tr}\left(C_{11}\Sigma\right)\right)$$

Now, let $r \geq m$. Define $T = \begin{bmatrix} I_m & 0_{m\times(r-m)} \end{bmatrix}$. Then

$$T^H\Xi^{-1}Z = \begin{bmatrix} I_m \\ 0_{(r-m)\times m} \end{bmatrix}\Xi^{-1}Z = \begin{bmatrix} \Xi^{-1}Z \\ 0 \end{bmatrix}$$

We further note that

$$T^H\Xi^{-1}T = \begin{bmatrix} I_m \\ 0_{(r-m)\times m} \end{bmatrix}\Xi^{-1}\begin{bmatrix} I_m & 0_{m\times(r-m)} \end{bmatrix}$$

$$= \begin{bmatrix} I_m \\ 0_{(r-m)\times m} \end{bmatrix}\begin{bmatrix} \Xi^{-1} & 0_{m\times(r-m)} \end{bmatrix} = \begin{pmatrix} \Xi^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

By theorem 41 we know

$$T^H\Xi^{-1}Z \sim \mathbf{C}N_{r,r}\left(T^H\Xi^{-1}\mu, \begin{pmatrix} \Xi^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \Sigma\right)$$

We also observe that

$$T^H\Xi^{-1}T\Sigma = \begin{pmatrix} \Xi^{-1} & 0 \\ 0 & 0 \end{pmatrix}\Sigma = \begin{pmatrix} \Xi^{-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \Xi^{-1}\Sigma_{11} & \Xi^{-1}\Sigma_{12} \\ 0 & 0 \end{pmatrix}$$

From this we see that $\mathrm{tr}(T^H\Xi^{-1}T\Sigma) = \mathrm{tr}(\Xi^{-1}\Sigma_{11})$. Once again examining the characteristic function, we now change only the definition of T and we observe

$$\Phi_{T^H\Xi^{-1}Z}(\tau) = \exp\left\{i\,\mathrm{Re}\left[\mathrm{tr}\left(\tau^H T^H \Xi^{-1}\mu\right)\right] - \frac{1}{4}\,\mathrm{tr}\left(\tau^H T^H \Xi^{-1} T \tau \Sigma\right)\right\}$$

$$= \exp\left\{i\,\mathrm{Re}\left[\mathrm{tr}\left(\tau^H \begin{bmatrix} \Xi^{-1}\mu \\ 0 \end{bmatrix}\right)\right] - \frac{1}{4}\,\mathrm{tr}\left[\tau^H \begin{pmatrix} \Xi^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tau\Sigma\right]\right\}$$

by equation D.9. To obtain the characteristic function of $\mathrm{tr}(T^H\Xi^{-1}\Sigma)$ we once again let $\tau = tI$. This gives us

$$\Phi_{T^H\Xi^{-1}Z}(tI) = \exp\left\{i\,\mathrm{Re}\left[\mathrm{tr}\left(t^* \begin{bmatrix} \Xi^{-1}\mu \\ 0 \end{bmatrix}\right)\right] - \frac{1}{4}\,|t|^2\,\mathrm{tr}\begin{bmatrix} \Xi^{-1}\Sigma_{11} & \Xi^{-1}\Sigma_{12} \\ 0 & 0 \end{bmatrix}\right\}$$

$$= \exp\left\{i\,\mathrm{Re}\left[\mathrm{tr}\left(t^* \begin{bmatrix} \Xi^{-1}\mu \\ 0 \end{bmatrix}\right)\right] - \frac{1}{4}\,|t|^2\,\mathrm{tr}\left[\Xi^{-1}\Sigma_{11}\right]\right\}$$

This is the characteristic function of $\mathbf{C}N_1\left(\mathrm{tr}\begin{pmatrix} \Xi^{-1}\mu \\ 0 \end{pmatrix}, \mathrm{tr}\left(\Xi^{-1}\Sigma_{11}\right)\right)$.

You can observe that this theorem can be generalized easily by application of the Singular Value Decomposition.

Now we want the distribution of $\mathrm{tr}\left(Z^H\Xi^{-1}Z\Sigma^{-1}\right)$. Recall that $Z \sim \mathbf{C}N_{m,r}(\mu, \Xi, \Sigma)$. Let $\Xi = AA^H$ and $\Sigma = B^HB$. We do this by theorem 119. Then

$$\mathrm{tr}\left(Z^H\Xi^{-1}Z\Sigma^{-1}\right) = \mathrm{tr}\left(Z^H(AA^H)^{-1}Z(B^HB)^{-1}\right) = \mathrm{tr}\left(Z^HA^{-H}A^{-1}ZB^{-1}B^{-H}\right)$$

$$= \mathrm{tr}\left(B^{-H}Z^HA^{-H}A^{-1}ZB^{-1}\right) = \mathrm{tr}\left((A^{-1}ZB^{-1})^H(A^{-1}ZB^{-1})\right) = \mathrm{tr}(Y^HY)$$

where $Y = A^{-1}ZB^{-1}$. Note that

$$Y \sim \mathbf{C}N_{m,r}(A^{-1}\mu B^{-1}, A^{-1}\Xi A^{-H}, B^{-H}\Sigma B^{-1})$$

$$= \mathbf{C}N_{m,r}(A^{-1}\mu B^{-1}, I, I)$$

by theorem 41. By lemma 21 we know

$$2\,\mathrm{tr}(Y^H Y) = 2\,\mathrm{tr}\left(Z^H \Xi^{-1} Z \Sigma^{-1}\right) \sim \chi^2_{2mr}\left[2\,\mathrm{tr}\left(\mu^H \Xi^{-1}\mu\Sigma^{-1}\right)\right]$$

Theorem 74 *Let $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ with Hermitian nonnegative definite Σ. Then $2\,\mathrm{tr}(\Sigma^{-1}W) \sim \chi^2_{2np}[2\,\mathrm{tr}(\Sigma^{-1}\delta)]$. This is a complexification of Arnold's theorem 17.14 [31], which was stated without proof, and it also generalizes the complex version of the distributional result of Muirhead's theorem 3.2.20 [187].*

Proof. Because Σ is nonnegative definite, there exists a decomposition $\Sigma = B^H B$ by theorem 119. From the definition of W, let $W = Z^H Z$ where $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. B is $p \times p$, Z is $n \times p$. Then

$$\mathrm{tr}(\Sigma^{-1}W) = \mathrm{tr}[(B^H B)^{-1}(Z^H Z)] = \mathrm{tr}(B^{-1}B^{-H}Z^H Z)$$

$$= \mathrm{tr}(B^{-H}Z^H Z B^{-1}) = \mathrm{tr}((ZB^{-1})^H (ZB^{-1}))$$

Let $Y = ZB^{-1}$. The new variable Y is $n \times p$. Then

$$Y \sim CN_{n,p}(\mu B^{-1}, I, B^{-H}\Sigma B^{-1}) = CN_{n,p}(\mu B^{-1}, I, I)$$

by theorem 41. We note that

$$\mathrm{tr}\left[(\mu B^{-1})^H (\mu B^{-1})\right] = \mathrm{tr}(B^{-H}\mu^H \mu B^{-1}) = \mathrm{tr}(B^{-1}B^{-H}\mu^H \mu) = \mathrm{tr}(\Sigma^{-1}\delta)$$

where $\delta = \mu^H \mu$ and $\Sigma = (B^H B)^{-1}$. Then

$$2\,\mathrm{tr}(Y^H Y) = 2\,\mathrm{tr}(\Sigma^{-1} W) \sim \chi^2_{2np}[2\,\mathrm{tr}(\Sigma^{-1}\delta)]$$

by theorem 21.

Lemma 22 *Let* $W \sim \mathbf{C}W_p(n, \Sigma, \delta)$ *where* Σ *has the eigenvalue decomposition* $\Sigma = U\Lambda^2 U^H$. *Then*

$$2\,\mathrm{tr}(\Lambda^{-2} U^H W U) \sim \chi^2_{2np}[2\,\mathrm{tr}(\Lambda^{-2} U^H \delta U)]$$

Proof. Note that $\Lambda^2 = U^H \Sigma U$. By the definition of the complex Wishart distribution, let $W = Z^H Z$ where $Z \sim \mathbf{C}N_{n,p}(\mu, I, \Sigma)$. By theorem 54 we know

$$U^H W U \sim \mathbf{C}W_p(n, \Lambda^2, U^H \delta U)$$

By theorem 74, we get the result

$$2\,\mathrm{tr}(\Lambda^{-2} U^H W U) \sim \chi^2_{2np}[2\,\mathrm{tr}(\Lambda^{-2} U^H \delta U)]$$

□

F.2 Characteristic Function of the Complex Wishart Distribution

Theorem 75 *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$, $\Sigma > 0$. *Let* $\breve{W} = 2W - \Delta(W)$ *where* $\Delta(W)$ *is a diagonal matrix consisting of the elements on the main diagonal of*

W. Then $\check{W}$ has the characteristic function

$$\Phi_{\check{W}}(T) = [\det(I_p - i\Sigma T)]^{-n}$$

where $T^H = T \in \mathbf{C}^{p\times p}$ and $\check{W}$ has the joint distribution of the random variables

$$(W_{11}, W_{22}, \cdots, W_{pp}, 2W_{R12}, 2W_{I12}, \cdots, 2W_{R(p-1),p}, 2W_{I(p-1),p})$$

This is Goodman equation 1.7 [92], Eaton proposition 8.3(iii), (p. 305) [74], and the complexification of Arnold's theorem 17.15 [31].

Proof. As a preamble, note explicitly that we are *not* obtaining the characteristic function of W. This is not the characteristic function of the joint distribution of the random variables

$$(W_{11}, \cdots, W_{pp}, W_{R12}, W_{I12}, \cdots, W_{R(p-1),p}, W_{I(p-1),p})$$

This tradition was also honored by other authors in deriving transforms for the real Wishart distribution. The characteristic function for $\check{W}$ is useful for studying some transformation of variables, but great attention to detail is necessary if it is to be useful for computing expected values of moments. This is because $\frac{\partial}{\partial t_{ij}} t_{ij}^*$ does not exist when $i \neq j$. This is a result of imposing the condition $T = T^H$ which is used to justify the existence of an eigenvalue decomposition in equation F.1.

This is a complexification and expansion of Eaton's proof. Let $Z \sim \mathbf{C}N_p(0, I)$ and $C \in \mathbf{C}^{p\times p}$. Then

$$X = CZ \sim \mathbf{C}N_p(0, CC^H) = \mathbf{C}N_p(0, \Sigma)$$

where $\Sigma = CC^H$ by theorem 119. Let $\{X_j\}_1^n$ be a random sample of size $n \geq p$. Then

$$W = \sum_{j=1}^{n} X_j X_j^H \sim \mathbf{C}W_p(n, \Sigma)$$

where X_j is a column vector. We want to find the characteristic function for W. Note that $W = W^H > 0$. Thus, let the argument of the characteristic function have the property $T = T^H > 0$. This will make the answer come out in a nice form.

It turns out that we are *not* deriving the characteristic function of W, but rather we are deriving the characteristic function of a related matrix variable which I will call $\check{W}$. The transformation matrix T in my proof is called A in Eaton's proof.

$$\Phi_W(T) = \mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(T^H \check{W}))]\} = \mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(T\check{W}))]\}$$

since $T^H = T$.

$$= \mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(T \sum_{j=1}^{n} X_j X_j^H))]\} = \mathcal{E}\{\prod_{j=1}^{n} \exp[i \operatorname{Re}(\operatorname{tr}(T X_j X_j^H))]\}$$

$$= \prod_{j=1}^{n} \mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(T X_j X_j^H))]\}$$

since the $\{X_j\}$ are independent. Further, since the $\{X_j\}$ are identically distributed, for your favorite j we can say this equals

$$\left[\mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(T X_j X_j^H))]\}\right]^n = \left[\mathcal{E}\{\exp[i \operatorname{Re}(\operatorname{tr}(X_j^H T X_j))]\}\right]^n$$

$$= \left[\mathcal{E}\{\exp[i \operatorname{Re}(X_j^H T X_j)]\}\right]^n$$

because we recognize that $X_j^H T X_j$ is a scalar and thus equal to its trace. We drop the subscript j and use the lower case x to signify our generic independent identically distributed vector. Let $x = Cz$ and $B = C^H T C$. Then

$$\Phi_W(T) = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(z^H C^H T C z)]\}\right]^n = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(z^H B z)]\}\right]^n$$

Note that $B^H = B > 0$, so by theorem 115 it has an eigenvalue decomposition

$$B = \Gamma \Lambda^2 \Gamma^H \tag{F.1}$$

Thus

$$\Phi_W(T) = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(z^H \Gamma \Lambda^2 \Gamma^H z)]\}\right]^n = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(y^H \Lambda^2 y)]\}\right]^n$$

where $y = \Gamma^H z$. Note that

$$y = \Gamma^H z \sim \mathbf{C}N_p(0, \Gamma^H \Gamma) = \mathbf{C}N_p(0, I)$$

This is the same distribution that z has. Thus, we can write

$$\Phi_W(T) = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(z^H \Lambda^2 z)]\}\right]^n = \left[\mathcal{E}\{\exp[i\,\mathrm{Re}(\sum_{k=1}^{p} \lambda_k^2 |z_k|^2)]\}\right]^n$$

where $\{z_k\}_1^p$ are the elements of complex vector z, and are independently distributed as $\mathbf{C}N_1(0,1)$. Therefore,

$$\Phi_W(T) = \left[\prod_{k=1}^{p} \mathcal{E}\{\exp[i\lambda_k^2 |z_k|^2]\}\right]^n$$

since the λ_k^2 and $|z_k|^2$ are real-valued. We continue by expressing the expected value in its integral form.

$$\Phi_W(T) = \{\prod_{k=1}^{p} \int_C \exp[i\lambda_k^2 |z_k|^2] \frac{1}{\pi} \exp[-|z_k|^2] dz_k\}^n$$

$$= \pi^{-np} \prod_{k=1}^{p} \{ \int_C \exp[-(1 - i\lambda_k^2)|z_k|^2] dz_k \}^n = \pi^{-np} \prod_{k=1}^{p} \left(\frac{\pi}{1 - i\lambda_k^2} \right)^n$$

$$= \pi^{-np} \pi^{np} \prod_{k=1}^{p} \left(1 - i\lambda_k^2\right)^{-n} = \left[det \left(I - i\Lambda^2\right) \right]^{-n}$$

where $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_p^2)$ and $\det(I - i\Lambda^2) \neq 0$. Since $B^H = B > 0$, all its eigenvalues are real by corollary 34. Therefore $\lambda_k^2 \neq -i$ for any value of k, so the determinant always exists.

I have lost the pedigree of the proof that λ_j^2 cannot be pure imaginary. However, it is important, so it is presented. Suppose there exist some $\lambda_j^2, \lambda_k^2 \in \mathbf{R}$ such that $(1 - i\lambda_j^2)(1 - i\lambda_k^2) = 0$. Then

$$1 - i(\lambda_j^2 + \lambda_k^2) - \lambda_j^2 \lambda_k^2 = 0$$

which implies

$$i = \frac{1 - \lambda_j^2 \lambda_k^2}{\lambda_j^2 + \lambda_k^2}$$

This is impossible, so there can never be such λ_j^2, λ_k^2.

Continuing,

$$\Phi_W(T) = [det(I_p)]^{-n} \left[det \left(I_p - i\Lambda^2\right) \right]^{-n} = \left[det(\Gamma\Gamma^H) \right]^{-n} \left[det \left(I_p - i\Lambda^2\right) \right]^{-n}$$

$$= [det(\Gamma)]^{-n} \left[det \left(I_p - i\Lambda^2\right) \right]^{-n} \left[det(\Gamma^H) \right]^{-n}$$

$$= \left[det \left(\Gamma\Gamma^H - i\Gamma\Lambda^2\Gamma^H\right) \right]^{-n} = [det(I_p - iB)]^{-n}$$

since $B = \Gamma\Lambda^2\Gamma^H$. We also recall that $B = C^H T C$ which gives us

$$\Phi_W(T) = \left[det \left(I_p - iC^H T C\right) \right]^{-n} = \left[det \left(I_p - iCC^H T\right) \right]^{-n}$$

by Lemma 47. Then

$$\Phi_{\check{W}}(T)= [det\,(I_p - i\Sigma T)]^{-n}$$

where $\Sigma = CC^H$. □

F.3 Functions of a Wishart Matrix

Example 6 $\mathcal{E}\{W\} = n\Sigma$. *This result is known by many people for the case of the real Wishart matrix. The point of this example is the use of the characteristic function of* $\check{W}$ *to compute the expected value of* W. *It is not quite the trivial exercise one might expect from experience with univariate statistics. Blame this example on me.*

Proof. Recall that the characteristic function corresponding to the joint density of

$$(W_{11}, W_{22}, \cdots, W_{pp}, 2W_{R12}, 2W_{I12}, \cdots, 2W_{R(p-1),p}, 2W_{I(p-1),p})$$

is given by

$$\Phi_{\check{W}}(T) = [\det\,(I_p - i\Sigma T)]^{-n}$$

where $T = T^H$. From the properties of characteristic functions, we recall for the differential operator

$$D(T_{jk}) = \left(\frac{\partial}{\partial T_{Rjk}} + i\frac{\partial}{\partial_{Ijk}}\right)$$

that

$$\mathcal{E}\{W\} = -iD(T)\Phi_X(T)\Big|_{T=0}$$

when $D(T)\Phi_X(T)$ exists. So,

$$D(T)[det\,(I_p - i\Sigma T)]^{-n} = -n\,[det\,(I_p - i\Sigma T)]^{-(n+1)}D(T)[det\,(I_p - i\Sigma T)]$$

From theorem 17, we recall

$$D(T)[det\,(I_p - i\Sigma T)]\Big|_{T=0} = i\,[\Delta(\Sigma) - 2\Sigma]$$

where $\Delta(\Sigma)$ is a diagonal matrix of the elements on the main diagonal of Σ.

Therefore

$$-iD(T)[det\,(I_p - i\Sigma T)]^{-n}\Big|_{T=0} = ini\,[\Delta(\Sigma) - 2\Sigma]$$

$$= n[2\Sigma - \Delta(\Sigma)] = \mathcal{E}\{\check{W}\}$$

What we really want is $\mathcal{E}\{W\}$ obtained from the joint density of

$$(W_{11}, \cdots, W_{pp}, W_{R12}, W_{I12}, \cdots, W_{R(p-1),p}, W_{I(p-1),p})$$

$$\mathcal{E}\{\check{W}\} = \mathcal{E}\{2W - \Delta(W)\} = n[2\Sigma - \Delta(\Sigma)]$$

Note that $\Delta(\check{W}) = \Delta(W)$, thus $\mathcal{E}\{\Delta(\check{W})\} = \mathcal{E}\{\Delta(W)\}$. Then

$$\mathcal{E}\{\Delta(\check{W})\} = n[2\Delta(\Sigma) - \Delta(\Sigma)] = n\Delta(\Sigma) = \mathcal{E}\{\Delta(W)\}$$

Therefore $\mathcal{E}\{W\} = n\Sigma$.

Theorem 76 *Let $W \sim \mathbf{C}W_p(n, \Sigma)$ and let $a \in \mathbf{C}^p$ be a fixed vector of complex numbers. Then the characteristic function of the quadratic $a^H\check{W}a$ is given by*

$$\Phi_{a^H\check{W}a}(t) = \left[\det(1 - ia^H\Sigma at)\right]^{-n}$$

where $\check{W} = 2W - \Delta(W)$ and $\Delta(W)$ is a diagonal matrix whose diagonal entries are the elements on the main diagonal of W. I supplied this result.

Proof. Let $a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}$ and

$$W = \begin{pmatrix} W_{11} & \cdots & W_{1p} \\ \vdots & \ddots & \vdots \\ W_{p1} & \cdots & W_{pp} \end{pmatrix} = (W_1, \cdots, W_p)$$

Then

$$a^H\check{W}a = \sum_{k=1}^{p} a^H\check{W}_k a_k = \sum_{j=1}^{p}\sum_{k=1}^{p} a_j^*\check{W}_{jk}a_k$$

Let the $(j,k)^{th}$ element of A^* be $a_j^*a_k = (A^*)_{jk}$. Then $A = aa^H$ and by theorem 20

$$\Phi_{a^H\check{W}a}(t) = \Phi_{\check{W}}(At) = \Phi_{\check{W}}(aa^Ht)$$

Applying Goodman equation 1.7 [92] for $\Phi_W(T) = [\det(I_p - i\Sigma T)]^{-n}$ we get

$$\Phi_{a^H W a}(t) = \left[\det(I_p - i\Sigma a a^H t)\right]^{-n}$$

where $t \in \mathbf{C}$ is a scalar. Treating Σa as a $p \times 1$ matrix and a^H as a $1 \times p$ matrix in Eaton lemma 1.35, we get

$$\Phi_{a^H W a}(t) = \left[\det(1 - ia^H \Sigma a t)\right]^{-n}$$

□

Theorem 77 *Let $W \sim \mathbf{C}W_p(n, \Sigma)$, $\Sigma > 0$. If $n \geq p$, then*

$$\frac{\det(W)}{\det(\Sigma)} = \det(\Sigma^{-1} W)$$

has the same distribution as $\prod_{i=1}^{p} U_i$ where the U_i are independent and $2U_i \sim \chi^2_{2(n-i+1)}$. This is a complexification of Arnold's theorem 17.15(b) [31], which was stated without proof. Goodman [93] gives an alternative proof. It is also a complexification of theorem 7.5.3 of Anderson [26].

Proof. The proof presented here follows the hints given in problem 17.13(a) of Arnold [31] applied as to the complex Wishart case. Let $\Sigma^{-1} = CC^H$. This exists by theorem 121. Then

$$\det(\Sigma^{-1} W) = \det(CC^H W) = \det(C^H W C)$$

Since $W \sim \mathbf{C}W_p(n, \Sigma)$, then by theorem 54 we know

$$C^H W C \sim \mathbf{C}W_p(n, C^H \Sigma C) = \mathbf{C}W_p(n, C^H C^{-H} C^{-1} C)$$

Thus, $C^H W C \sim \mathbf{C}W_p(n, I_p)$. Let $V = C^H W C$. Note that the partitioned form of V is $V = \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} \end{pmatrix}$. By lemma 46,

$$\det V = (\det V_{22}) \left(\det(V_{11} - V_{12} V_{22}^{-1} V_{21})\right) = \det(V_{22}) \det(V_{11})$$

By lemma 19, with V_{11} being $\det(z)$, we know that with $\sigma^2 = 1$,

$$2U_1 = 2(V_{11} \mid V_{22}) \sim \chi^2_{2(n-p+1)}(0)$$

By theorem 55, $V_{22} \sim \mathbf{C}W_{p-1}(n, I_{p-1})$.

Now, partition V_{22} in the same manner that V was partitioned. Then

$$2U_2 = 2(V_{22} \mid V_{33}) \sim \chi^2_{2(n-p+2)}(0)$$

and $V_{33} \sim \mathbf{C}W_{p-2}(n, I_{p-2})$. repeating this process through the p^{th} entry, we observe that $2^p \det(V) = \prod_{i=1}^{p} (2U_i)$ where $2U_i \sim \chi^2_{2(n-p+i)}(0)$. By reversing the index, we get $2U_i \sim \chi^2_{2(n-i+1)}(0)$, since

$$\{2U_1, \cdots, 2U_p\} \sim \{\chi^2_{2(n-p+1)}, \chi^2_{2(n-p+2)}, \cdots, \chi^2_{2n}\}$$

Note that this theorem says that the distribution of $\det(W)$ can be considered as

$$\det(W) \sim 2^{-p} [\det \Sigma] \prod_{i=1}^{p} \chi^2_{2(n-i+1)}$$

and

$$\frac{2^p \det(W)}{\det(\Sigma)} = 2^p \det(\Sigma^{-1} W) \sim \prod_{i=1}^{p} \chi^2_{2(n-i+1)}$$

□

Theorem 78 . *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$, $\Sigma > 0$. *If* $n \geq p$, *then*

$$\mathcal{E}\{\det(W)\} = p!\binom{n}{p}\det(\Sigma)$$

This is a complexification of Arnold's theorem 17.15(c)(i) [31], which was stated without proof.

Proof.

$$\mathcal{E}\{\det(W)\} = \mathcal{E}\left\{\frac{\det(W)}{\det(\Sigma)}\det(\Sigma)\right\} = [\det\Sigma]\,\mathcal{E}\left\{\frac{\det(W)}{\det(\Sigma)}\right\}$$

By theorem 77, $\det(\Sigma^{-1}W)$ has the same distribution as $\prod_{i=1}^{p} U_i$ where the U_i are independent and $2U_i \sim \chi^2_{2(n-i+1)}$. By property of the χ^2 distribution, $\mathcal{E}\{2U_i\} = 2(n-i+1)$. Thus, $\mathcal{E}\{U_i\} = n-i+1$. Continuing,

$$\mathcal{E}\{\prod_{i=1}^{p} U_i\} = \prod_{i=1}^{p}\mathcal{E}\{U_i\}$$

since the U_i are independent. So, we have

$$\mathcal{E}\{\det(\Sigma^{-1}W)\} = \mathcal{E}\{\prod_{i=1}^{p} U_i\} = n(n-1)\cdots(n-p+1)$$

which implies

$$\mathcal{E}\{\det(W)\} = \det(\Sigma)\mathcal{E}\{\det(\Sigma^{-1}W)\} = n(n-1)\cdots(n-p+1)\det(\Sigma)$$

$$= \frac{n!}{(n-p)!}\det(\Sigma) = p!\binom{n}{p}\det(\Sigma)$$

□

Theorem 79 *If* $W \sim \mathbf{C}W_p(n, \Sigma), \Sigma > 0$ *and* $n \geq p$, *then*

$$\mathcal{E}\{[\det W]^k\} = [\det \Sigma]^k \, \mathbf{C}\Gamma_p(n+k)/\mathbf{C}\Gamma_p(n)$$

$$\mathcal{E}\{[\det W]^2\} = [\det \Sigma]^2 \, [p!]^4 \binom{n+1}{p}\binom{n}{p}$$

$$\text{var}\{\det W\} = [\det \Sigma]^2 \, [p!]^2 \binom{n}{p}\binom{p}{n+1-p}$$

This is a complexification and generalization of Anderson's lemma (p. 264) [26].

Proof. This is a complexification and generalization of Anderson's proof. By theorem 77, $\det(\Sigma^{-1}W)$ has the same distribution as $\prod_{i=1}^{p} U_i$ where the U_i are independent and

$$2U_i \sim \chi^2_{2(n-i+1)}$$

From Patil et al. (p. 35) [204], we know that if $x \sim \chi^2_m$, then

$$\mathcal{E}\{x^k\} = \frac{2^k \Gamma(\frac{1}{2}m + k)}{\Gamma(\frac{1}{2}m)}$$

Thus

$$\mathcal{E}\{(2U_i)^k\} = \frac{2^k \Gamma(n+k-i+1)}{\Gamma(n-i+1)}$$

which implies

$$\mathcal{E}\{U_i{}^k\} = \frac{\Gamma(n+k-i+1)}{\Gamma(n-i+1)}$$

Since the U_i are independent, then

$$\mathcal{E}\left\{\left(\prod_{i=1}^{p} U_i\right)^k\right\} = \prod_{i=1}^{p} \frac{\Gamma(n+k-i+1)}{\Gamma(n-i+1)}$$

which implies

$$\mathcal{E}\{[\det W]^k\} = [\det \Sigma]^k \prod_{i=1}^{p} \frac{\Gamma(n+k-i+1)}{\Gamma(n-i+1)}$$

$$= [\det \Sigma]^k \frac{\mathbf{C}\Gamma_p(n+k)}{\pi^{p(p-1)/2}} \frac{\pi^{p(p-1)/2}}{\mathbf{C}\Gamma_p(n)} = [\det \Sigma]^k \frac{\mathbf{C}\Gamma_p(n+k)}{\mathbf{C}\Gamma_p(n)}$$

In the special case of $k = 1$, then

$$\mathcal{E}\{[\det W]\} = [\det \Sigma] \frac{\Gamma(n+1)\Gamma(n)\cdots\Gamma(n+2-p)}{\Gamma(n)\Gamma(n-1)\cdots\Gamma(n+1-p)} = [\det \Sigma] \frac{\Gamma(n+1)}{\Gamma(n+1-p)}$$

$$= [\det \Sigma] \frac{n!}{(n-p)!} = [\det \Sigma]\, p! \binom{n}{p}$$

This is the same answer we got in theorem 78.

When $k = 2$,

$$\mathcal{E}\{[\det W]^2\} = [\det \Sigma]^2 \frac{\Gamma(n+2)\Gamma(n+1)\cdots\Gamma(n+3-p)}{\Gamma(n)\Gamma(n-1)\cdots\Gamma(n+1-p)}$$

$$= [\det \Sigma]^2 \frac{\Gamma(n+2)\Gamma(n+1)}{\Gamma(n+2-p)\Gamma(n+1-p)} = [\det \Sigma]^2 \frac{(n+1)!n!}{(n+1-p)!(n-p)!}$$

Therefore,

$$\mathcal{E}\{[\det W]^2\} = [\det \Sigma]^2 [p!]^2 \binom{n+1}{p} \binom{n}{p}$$

The variance of $\det W$ is

$$\operatorname{var}\{\det W\} = \mathcal{E}\{[\det W]^2\} - [\mathcal{E}\{[\det W]\}]^2$$

$$= [\det \Sigma]^2 [p!]^2 \left[\binom{n+1}{p} \binom{n}{p} - \binom{n}{p}^2 \right] = [\det \Sigma]^2 [p!]^2 \binom{n}{p} \left(\frac{p}{n+1-p} \right)$$

□

Theorem 80 *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$ *where* Σ *has eigenvalue decomposition* $\Sigma = \Gamma\Lambda^2\Gamma^H$. *Let* $\check{W} = 2W - \Delta(W)$ *and* $\Delta(W)$ *is a diagonal matrix whose diagonal entries are the elements on the main diagonal of* W. *Then the characteristic function of* $\operatorname{tr}\check{W}$ *is*

$$\Phi_{\operatorname{tr}\check{W}}(t) = \prod_{k=1}^{p} \lambda_k^{-2n}(\lambda_k^{-2} - it)^{-n} = \left[\det(I_p - i\Lambda^2 t)\right]^{-n}$$

This is similar to equation (5.58) of Goodman [92]

Proof. This proof is essentially due to Goodman (p. 169)[92],

$$\Phi_{\operatorname{tr}\check{W}}(t) = \Phi_{\check{W}}(I_p t) = \Phi_{\check{W}}(T)$$

where t is a scalar. By Goodman equation 1.7, $\Phi_{\check{W}}(T) = [\det(I_p - i\Sigma T)]^{-n}$ where $T \in \mathbf{C}^{p\times p}$. Then

$$\Phi_{\operatorname{tr}\check{W}}(t) = [\det(I_p - i\Sigma I_p t)]^{-n} = [\det(I_p - i\Sigma t)]^{-n}$$

Using the eigenvalue decomposition, we get

$$\Phi_{\operatorname{tr}\check{W}}(t) = \left[\det(I_p - i\Gamma\Lambda^2\Gamma^H t)\right]^{-n} = \left[\det(\Gamma\Gamma^H - i\Gamma\Lambda^2\Gamma^H t)\right]^{-n}$$

$$= [\det\Gamma]^{-n}\left[\det(I_p - i\Lambda^2 t)\right]^{-n}\left[\det\Gamma^H\right]^{-n}$$

$$= \left[\det\Gamma\Gamma^H\right]^{-n}\left[\det(I_p - i\Lambda^2 t)\right]^{-n} = \left[\det(I_p - i\Lambda^2 t)\right]^{-n}$$

Since a common use of a characteristic function involves setting $t = 0$, we with to preserve evidence of dependence on Λ^2. So,

$$\Phi_{\operatorname{tr}\check{W}}(t) = \left[\det\Lambda^2\right]^{-n}\left[\det(\Lambda^{-2} - iI_p t)\right]^{-n} = \prod_{k=1}^{p} \lambda_k^{-2n}(\lambda_k^{-2} - it)^{-n}$$

□

Note that since $T^H = T = I_p t$ where $t \in \mathbf{C}$, we know $t \in \mathbf{R}$. Therefore, $\frac{\partial}{\partial t}\Phi_{\mathrm{tr}\,\check{W}}(t)$ exists.

Note that $\mathrm{tr}\,\check{W}$ is a function of only those elements on the diagonal:

$$W_{11}, W_{22}, \cdots, W_{pp}$$

Consequently, we can work with the characteristic function of the joint distribution of

$$(W_{11}, \cdots, W_{pp}, 2W_{R12}, \cdots, 2W_{I(p-1),p})$$

and still get an answer to the question being asked about $\mathrm{tr}\,W$.

Theorem 81 *Let $W \sim \mathbf{C}W_p(n, \Sigma)$. Then $\mathcal{E}\{\mathrm{tr}\,W\} = n\,\mathrm{tr}\,\Sigma$. This is the complexification of theorem 17.15(e) of Arnold [31].*

Proof. This proof is an application of the concepts developed in section B.4. By theorem 80, the characteristic function of $\mathrm{tr}\,\check{W}$ is given by

$$\Phi_{\mathrm{tr}\,\check{W}}(t) = \left[\det(I_p - i\Lambda^2 t)\right]^{-n}$$

where Λ^2 is the diagonal matrix of eigenvalues of Σ and $t \in \mathbf{R}$. Taking the derivative, we obtain the following.

$$\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-n} = -n\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)}\frac{\partial}{\partial t}\det(I_p - i\Lambda^2 t)$$

$$\frac{\partial}{\partial t}\det(I_p - i\Lambda^2 t) = \frac{\partial}{\partial t}\prod_{k=1}^{p}(1 - i\lambda_k^2 t)$$

We apply the chain rule.

$$\frac{\partial}{\partial t}\det(I_p - i\Lambda^2 t) = \sum_{l=1}^{p}(-i\lambda_l^2)\prod_{\substack{k=1\\k\neq l}}^{p}(1 - i\lambda_k^2 t)$$

Putting the problem all together,

$$\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-n} = in\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)}\sum_{l=1}^{p}\lambda_l^2\prod_{\substack{k=1\\k\neq l}}^{p}(1 - i\lambda_k^2 t)$$

We evaluate at $t = 0$, and we obtain

$$\left.\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-n}\right|_{t=0} = in\sum_{l=1}^{p}\lambda_l^2 = in\,\mathrm{tr}\,\Lambda^2 = in\,\mathrm{tr}\,\Sigma$$

Recall that

$$\mathcal{E}\{\mathrm{tr}\,W\} = -i\left.\frac{\partial}{\partial t}\Phi_{\mathrm{tr}\,W}(t)\right|_{t=0} = (-i)\,i\,n\,\mathrm{tr}\,\Sigma = n\,\mathrm{tr}\,\Sigma$$

Theorem 82 *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$. *Then*

$$\mathcal{E}\{(\mathrm{tr}\,W)^2\} = n^2(\mathrm{tr}\,\Sigma)^2 + n(\mathrm{tr}\,\Sigma^2)$$

and $\mathrm{var}(\mathrm{tr}\,W) = n(\mathrm{tr}\,\Sigma^2)$.

Proof. This proof is an application of concepts developed in section B.4. By theorem 80, the characteristic function of $\mathrm{tr}\,\dot{W}$ is

$$\Phi_{\mathrm{tr}\,W}(t) = \left[\det(I_p - i\Lambda^2 t)\right]^{-n}$$

where Λ^2 is the diagonal matrix of eigenvalues of Σ. The first derivative with respect to t is

$$\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-n} = in\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)} \sum_{l=1}^{p} \lambda_l^2 \prod_{\substack{k=1 \\ k\neq l}}^{p} (1 - i\lambda_k^2 t)$$

Apply the chain rule for the second derivative.

$$\frac{\partial^2}{\partial t^2}\left[\det(I_p - i\Lambda^2 t)\right]^{-n}$$

$$= in\left[\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)}\right] \sum_{l=1}^{p} \lambda_l^2 \prod_{\substack{k=1 \\ k\neq l}}^{p} (1 - i\lambda_k^2 t)$$

$$+in\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)} \left[\frac{\partial}{\partial t} \sum_{l=1}^{p} \lambda_l^2 \prod_{\substack{k=1 \\ k\neq l}}^{p} (1 - i\lambda_k^2 t)\right]$$

$$\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)} = -(n+1)\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+2)} \sum_{l=1}^{p} (-i)\lambda_l^2 \prod_{\substack{k=1 \\ k\neq l}}^{p} (1 - i\lambda_k^2 t)$$

$$\frac{\partial}{\partial t} \sum_{l=1}^{p} \lambda_l^2 \prod_{\substack{k=1 \\ k\neq l}}^{p} (1 - i\lambda_k^2 t) = \sum_{l=1}^{p} \lambda_l^2 \sum_{\substack{m=1 \\ m\neq l}}^{p} (-i\lambda_m^2) \prod_{\substack{k=1 \\ k\neq l \\ k\neq m}}^{p} (1 - i\lambda_k^2 t)$$

We evaluate at $t = 0$ since we want

$$\left.\frac{\partial^2}{\partial t^2}\Phi_{\mathrm{tr}\,W}(t)\right|_{t=0}$$

$$\left.\frac{\partial}{\partial t}\left[\det(I_p - i\Lambda^2 t)\right]^{-(n+1)}\right|_{t=0} = i(n+1)(\mathrm{tr}\,\Sigma)$$

$$\frac{\partial}{\partial t}\sum_{l=1}^{p}\lambda_l^2\prod_{\substack{k=1\\k\neq l}}^{p}(1-i\lambda_k^2 t)\Bigg|_{t=0} = -i\sum_{l=1}^{p}\lambda_l^2\sum_{\substack{m=1\\m\neq l}}^{p}\lambda_m^2 = -i\sum_{l=1}^{p}\lambda_l^2\left(-\lambda_l^2+\operatorname{tr}\Sigma\right)$$

$$= -i\sum_{l=1}^{p}\left(-\lambda_l^4+\lambda_l^2\operatorname{tr}\Sigma\right) = i\operatorname{tr}(\Sigma^2) - i(\operatorname{tr}\Sigma)^2$$

Assembling all the parts, we get

$$\frac{\partial^2}{\partial t^2}\left[\det(I_p - i\Lambda^2 t)\right]^{-n} = \left(in\left[i(n+1)(\operatorname{tr}\Sigma)\right]\sum_{l=1}^{p}\lambda_l^2\right)+\left(in[i]\left[\operatorname{tr}(\Sigma^2)-(\operatorname{tr}\Sigma)^2\right]\right)$$

$$= -n(n+1)(\operatorname{tr}\Sigma)^2 - n\operatorname{tr}(\Sigma^2) + n(\operatorname{tr}\Sigma)^2 = -n^2(\operatorname{tr}\Sigma)^2 - n\operatorname{tr}(\Sigma^2)$$

Then

$$\mathcal{E}\{(\operatorname{tr}W)^2\} = (-i)^2\frac{\partial^2}{\partial t^2}\Phi_{\operatorname{tr}W}(t)\Bigg|_{t=0} = n^2(\operatorname{tr}\Sigma)^2 + n\operatorname{tr}(\Sigma^2)$$

The variance of $\operatorname{tr}\breve{W}$ is obtained from

$$\operatorname{var}(\operatorname{tr}W) = \mathcal{E}\{(\operatorname{tr}W)^2\} - [\mathcal{E}\{\operatorname{tr}W\}]^2$$

From theorem 81 we have $[\mathcal{E}\{\operatorname{tr}W\}]^2 = n^2(\operatorname{tr}\Sigma)^2$. Thus

$$\operatorname{var}(\operatorname{tr}W) = n^2(\operatorname{tr}\Sigma)^2 + n\operatorname{tr}(\Sigma^2) - n^2(\operatorname{tr}\Sigma)^2 = n\operatorname{tr}(\Sigma^2)$$

□

Theorem 83 *Let $A \sim \mathbf{C}W_p(n,\Sigma)$. Let $A = T^HT$ where T is upper triangular with positive real values on the diagonal. Then the probability density of T is*

$$f(T) = \frac{2^p\operatorname{etr}(-\Sigma^{-1}T^HT)}{[\det\Sigma]^n\,\mathbf{C}\Gamma_p(n)}\prod_{k=1}^{p}t_{kk}^{2(n-k)+1}$$

This is Goodman equation 5.51 [92].

Proof. The density for A is given by

$$g(A) = \frac{[\det A]^{n-p} \operatorname{etr}(-\Sigma^{-1}A)}{[\det \Sigma]^n \, \mathbf{C}\Gamma_p(n)}(dA)$$

The Jacobian for the change of variables from A to T is given by Goodman equation 5.25 [92] and by theorem 27 as

$$J(A \rightarrow T) = 2^p \prod_{k=1}^{p} t_{kk}^{2(p-k)+1}$$

Performing the change of variables gives us

$$f(T) = \left[\frac{\left[\det(T^H T)\right]^{n-p} \operatorname{etr}(-\Sigma^{-1}T^H T)}{[\det \Sigma]^n \, \mathbf{C}\Gamma_p(n)} \right] 2^p \prod_{k=1}^{p} t_{kk}^{2(p-k)+1}(dT)$$

Note that

$$\left[\det(T^H T)\right]^{n-p} = [\det(T)]^{2(n-p)} = \prod_{k=1}^{p} t_{kk}^{2(n-p)}$$

The final result is by observing that

$$2(n-p) + 2(p-k) + 1 = 2(n-k) + 1$$

□

Theorem 84 *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$ *where* $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ *and* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. *Let* $V = W_{11} - W_{12}W_{22}^{-1}W_{21}$ *and* $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$,

where W_{11} and Σ_{11} are $q \times q$. Then $V \sim \mathbf{C}W_q(n-p+q, \Sigma_{11.2})$ and V is independent of W_{12} and W_{22}. Also, $W_{22} \sim \mathbf{C}W_{p-q}(n, \Sigma_{22})$, and $(W_{12} \mid W_{22}) \sim \mathbf{C}N_{q,(p-q)}(\Sigma_{12}\Sigma_{22}^{-1}W_{22}, \Sigma_{11.2}, W_{22})$. This is a complexification of Muirhead's theorem 3.2.10 [187].

Proof. This is a complexification and expansion of Muirhead's proof. Let $B = \Sigma^{-1/2}W(\Sigma^{-1/2})^H$. Note that $W = W^H$, $\Sigma = \Sigma^H$, where $\Sigma^{1/2}$ is the positive definite square root of Σ. $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^H$ by theorem 119. Perform the following change of variables. Let $V = W_{11} - W_{12}W_{22}^{-1}W_{21}$, $B_{12} = W_{12}$, $B_{22} = W_{22}$. Recall that $B_{21} = W_{21} = W_{12}^H$. Thus,

$$(dW) = (dW_{11}) \bigwedge (dW_{12}) \bigwedge (dW_{22}) = (dV) \bigwedge (dB_{12}) \bigwedge (dB_{22})$$

Recall that

$$\det W = (\det W_{22})\,[\det(W_{11} - W_{12}W_{22}^{-1}W_{21})] = (\det B_{22}) \det V$$

and

$$\det \Sigma = (\det \Sigma_{22}) \det \Sigma_{11.2}$$

Let

$$C = \Sigma^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{11} is $q \times q$. Then

$$\operatorname{tr}(\Sigma^{-1}W) = \operatorname{tr}\left(\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} V + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \right)$$

$$= \mathrm{tr}(C_{11}V + C_{11}B_{12}B_{22}^{-1}B_{21} + C_{12}B_{21}) + \mathrm{tr}(C_{21}B_{12} + C_{22}B_{22})$$

$$= \mathrm{tr}(C_{11}V) + \mathrm{tr}(C_{11}B_{12}B_{22}^{-1}B_{21}) + \mathrm{tr}(C_{12}B_{21}) + \mathrm{tr}(C_{21}B_{12}) + \mathrm{tr}(C_{22}B_{22})$$

Observe that

$$\mathrm{tr}[C_{11}(B_{12} + C_{11}^{-1}C_{12}B_{22})B_{22}^{-1}(B_{12} + C_{11}^{-1}C_{12}B_{22})^H]$$

$$+ \mathrm{tr}[B_{22}(C_{22} - C_{21}C_{11}^{-1}C_{12})] + \mathrm{tr}[C_{11}V]$$

$$= \mathrm{tr}[C_{11}B_{12}B_{22}^{-1}(B_{12} + C_{11}^{-1}C_{12}B_{22})^H] + \mathrm{tr}[C_{12}(B_{12} + C_{11}^{-1}C_{12}B_{22})^H]$$

$$+ \mathrm{tr}[B_{22}(C_{22} - C_{21}C_{11}^{-1}C_{12})] + \mathrm{tr}[C_{11}V]$$

$$= \mathrm{tr}[C_{11}B_{12}B_{22}^{-1}B_{12}^H] + \mathrm{tr}[C_{11}B_{12}B_{22}^{-1}B_{22}^HC_{12}^HC_{11}^{-H}] + \mathrm{tr}[C_{12}B_{12}^H]$$

$$+ \mathrm{tr}[C_{12}B_{22}{}^HC_{12}^HC_{11}^{-H}] + \mathrm{tr}[B_{22}C_{22}] - \mathrm{tr}[B_{22}C_{21}C_{11}^{-1}C_{12}] + \mathrm{tr}[C_{11}V]$$

Recall that $B_{22} = B_{22}^H$, $B_{21} = B_{12}^H$, $C_{21} = C_{12}^H$, $C_{11}^{-1} = C_{11}^{-H}$. The expansion continues as

$$\mathrm{tr}[C_{11}B_{12}B_{22}^{-1}B_{21}] + \mathrm{tr}[C_{11}B_{12}C_{21}C_{11}^{-1}] + \mathrm{tr}[C_{12}B_{21}]$$

$$+ \mathrm{tr}[C_{12}B_{22}C_{21}C_{11}^{-1}] + \mathrm{tr}[B_{22}C_{22}] - \mathrm{tr}[B_{22}C_{21}C_{11}^{-1}C_{12}] + \mathrm{tr}[C_{11}V]$$

Recall that $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$. This allows us to produce the expansion

$$\mathrm{tr}[C_{11}B_{12}B_{22}^{-1}B_{21}] + \mathrm{tr}[B_{12}C_{21}] + \mathrm{tr}[C_{12}B_{21}] + \mathrm{tr}[C_{12}B_{22}C_{21}C_{11}^{-1}]$$

$$+ \mathrm{tr}[B_{22}C_{22}] - \mathrm{tr}[C_{12}B_{22}C_{21}C_{11}^{-1}] + \mathrm{tr}[C_{11}V]$$

$$= \mathrm{tr}[C_{11}B_{12}B_{22}^{-1}B_{21}] + \mathrm{tr}[B_{12}C_{21}] + \mathrm{tr}[C_{12}B_{21}] + \mathrm{tr}[B_{22}C_{22}] + \mathrm{tr}[C_{11}V]$$

$$= \text{tr}[C_{11}V] + \text{tr}[C_{11}B_{12}B_{22}^{-1}B_{21}] + \text{tr}[C_{12}B_{21}] + \text{tr}[C_{21}B_{12}] + \text{tr}[C_{22}B_{22}]$$

This is the same expression we had for $\text{tr}[\Sigma^{-1}W]$. From the partitioned matrix inverse, we know that

$$C_{11} = \Sigma_{11.2}^{-1} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$$

$$C_{22} - C_{21}C_{11}^{-1}C_{12} = \Sigma_{22}^{-1}$$

$$C_{11}^{-1}C_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$$

To see this, look at the inverse from both directions. From

$$C = \Sigma^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{12}\Sigma_{11.2}^{-1} & \Sigma_{22.1}^{-1} \end{pmatrix}$$

We observe $C_{11} = \Sigma_{11.2}^{-1}$ and $C_{22} - C_{21}C_{11}^{-1}C_{12}$

$$= \Sigma_{22.1}^{-1} - \Sigma_{22}^{-1}\Sigma_{12}\Sigma_{11.2}^{-1}\Sigma_{11.2}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}$$

$$= \Sigma_{22}^{-1}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\Sigma_{22.1}^{-1} = \Sigma_{22}^{-1}$$

From

$$\Sigma = C^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} C_{11.2}^{-1} & -C_{11}^{-1}C_{12}C_{22.1}^{-1} \\ -C_{22}^{-1}C_{21}C_{11.2}^{-1} & C_{22.1}^{-1} \end{pmatrix}$$

we observe

$$-\Sigma_{12}\Sigma_{22}^{-1} = C_{11}^{-1}C_{12}C_{22.1}^{-1}C_{22.1} = C_{11}^{-1}C_{12}$$

Recall that the complex Wishart density function is given by

$$f(W) = \frac{|\det W|^{n-p}\,\text{etr}[-\Sigma^{-1}W]}{[\det \Sigma]^{n}\mathbf{C}\Gamma_p(n)}$$

Note that $|\det W| = \det W$ since W *is positive definite.* Substituting in our results, we get

$$f(W) = \frac{|(\det B_{22})(\det V)|^{n-p}}{[(\det \Sigma_{22})(\det \Sigma_{11.2})]^n \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(n-i+1)}$$

$$\times \operatorname{etr}[-C_{11}V - B_{22}(C_{22} - C_{21}C_{11}^{-1}C_{12})$$

$$-C_{11}(B_{12} + C_{11}^{-1}C_{12}B_{22})B_{22}^{-1}(B_{12} + C_{11}^{-1}C_{12}B_{22})^H]$$

$$= \frac{(\det V)^{n-p+q-q} \operatorname{etr}\left(-\Sigma_{11.2}^{-1}V\right)}{(\det \Sigma_{11.2})^{n-p+q} \pi^{q(q-1)/2} \prod_{i=1}^{q} \Gamma(n-p+q-i+1)}$$

$$\times \frac{(\det B_{22})^{n-p+q} \operatorname{etr}\left(-\Sigma_{22}^{-1}B_{22}\right)}{(\det \Sigma_{22})^n \pi^{(p-q)(p-q-1)/2} \prod_{i=1}^{p-q} \Gamma(n-i+1)}$$

$$\times \frac{\operatorname{etr}\left[-\Sigma_{11.2}^{-1}(B_{12} - \Sigma_{12}\Sigma_{22}^{-1}B_{22})B_{22}^{-1}(B_{12} - \Sigma_{12}\Sigma_{22}^{-1}B_{22})^H\right]}{\pi^{(p-q)q}(\det \Sigma_{11.2})^{p-q}(\det B_{22})^q}$$

Note that the exponents of π obey

$$\frac{1}{2}q(q-1) + \frac{1}{2}(p-q)(p-q-1) + (p-q)q = \frac{1}{2}q(q-1) + \frac{1}{2}(p-q)(p+q-1)$$

$$= \frac{1}{2}q(q-1) + \frac{1}{2}(p-q)(q-1) + \frac{1}{2}(p-q)p$$

$$= \frac{1}{2}p(q-1) + \frac{1}{2}p(p-q) = \frac{1}{2}p(p-1)$$

Also note that

$$\left[\prod_{i=1}^{q} \Gamma(n-p+q-i+1)\right]\left[\prod_{i=1}^{p-q} \Gamma(n-i+1)\right]$$

$$= \Gamma(n-p+q)\Gamma(n-p+q-1)\cdots\Gamma(n-p+1)\Gamma(n)\Gamma(n-1)\cdots\Gamma(n-p+q+1)$$

$$= \Gamma(n)\Gamma(n-1)\cdots\Gamma(n-p+1) = \prod_{i=1}^{p} \Gamma(n-i+1)$$

Thus, $f(W)$ is the product of three density functions. The first one is the density function for V. It is distributed $\mathbf{C}W_q(n-p+q, \Sigma_{11.2})$. The second term is the density function for $B_{22} = W_{22}$. It is distributed $\mathbf{C}W_{p-q}(n, \Sigma_{22})$. The last term is the conditional density of $B_{12} = W_{12}$, given that $B_{22} = W_{22}$ is fixed. It is distributed

$$\mathbf{C}N_{q,(p-q)}(\Sigma_{12}\Sigma_{22}^{-1}W_{22}, \Sigma_{11.2}, W_{22})$$

In conclusion, V is independent of W_{22} and $(W_{12} \mid W_{22})$, and therefore is independent of W_{12}. □

Corollary 23 *Let* $W \sim \mathbf{C}W_p(n, \Sigma)$ *and let* $X = W_{22} - W_{21}W_{11}^{-1}W_{12}$ *and* $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ *where* W_{22} *and* Σ_{22} *are* $(p-q) \times (p-q)$. *Then* $X \sim \mathbf{C}W_{p-q}(n-q, \Sigma_{22.1})$ *and* X *is independent of* W_{21} *and* W_{11}. *Also,* $W_{11} \sim \mathbf{C}W_q(n, \Sigma_{11})$ *and* $(W_{21} \mid W_{11}) \sim \mathbf{C}N_{(p-q),q}(\Sigma_{21}\Sigma_{11}^{-1}W_{11}, \Sigma_{22.1}, W_{11})$. *This is a corollary to a complexification of Muirhead's theorem 3.2.10 [187].*

Proof. This follows the general logic of Muirhead's proof of theorem 3.2.10, modified by the different partition of interest. Let $B = \Sigma^{-1/2}W(\Sigma^{-1/2})^H$. We perform the change of variables $X = W_{22} - W_{21}W_{11}^{-1}W_{12}$, $B_{12} = W_{12}$, $B_{11} = W_{11}$. Recall $B_{21} = W_{21} = W_{12}^H$. Then

$$(dW) = (dW_{11}) \bigwedge (dW_{12}) \bigwedge (dW_{22}) = (dB_{11}) \bigwedge (dB_{12}) \bigwedge (dX)$$

Note that

$$\det W = (\det W_{11}) \det(W_{22} - W_{21}W_{11}^{-1}W_{12})$$

by lemma 45. thus

$$\det W = (\det W_{11}) \det X$$

and

$$\det \Sigma = (\det \Sigma_{22}) \det \Sigma_{22.1}$$

Let

$$C = \Sigma^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{11} is $q \times q$. Then

$$\operatorname{tr}(\Sigma^{-1}W) = \operatorname{tr}\left(\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & X + B_{21}B_{11}^{-1}B_{12} \end{bmatrix} \right)$$

$$= \operatorname{tr}(C_{11}B_{11} + C_{12}B_{21}) + \operatorname{tr}(C_{21}B_{12} + C_{22}X + C_{22}B_{21}B_{11}^{-1}B_{12})$$

$$C = \Sigma^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1} & \Sigma_{22.1}^{-1} \end{pmatrix}$$

$$C_{22} = \Sigma_{22.1}^{-1}$$

$$C_{11} - C_{12}C_{22}^{-1}C_{21} = \Sigma_{11.2}^{-1} - (-\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1})\Sigma_{22.1}(-\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1})$$

$$= \Sigma_{11.2}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}$$

$$= (I - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\Sigma_{11.2}^{-1} = \Sigma_{11}^{-1}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\Sigma_{11.2}^{-1} = \Sigma_{11}^{-1}$$

$$\Sigma = C^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} C_{11.2}^{-1} & -C_{11}^{-1}C_{12}C_{22.1}^{-1} \\ -C_{22}^{-1}C_{21}C_{11.2}^{-1} & C_{22.1}^{-1} \end{pmatrix}$$

Note that

$$-\Sigma_{21}\Sigma_{11}^{-1} = C_{22}^{-1}C_{21}C_{11.2}^{-1} = C_{22}^{-1}C_{21}$$

Then

$$\operatorname{tr}\left[C_{22}\left(B_{21}+C_{22}^{-1}C_{21}B_{11}\right)B_{11}^{-1}\left(B_{21}+C_{22}^{-1}C_{21}B_{11}\right)^{H}\right]$$

$$+\operatorname{tr}\left[B_{11}\left(C_{11}-C_{12}C_{22}^{-1}C_{21}\right)\right]+\operatorname{tr}\left(C_{22}X\right)$$

$$=\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}\left(B_{21}+C_{22}^{-1}C_{21}B_{11}\right)^{H}\right]+\operatorname{tr}\left[C_{21}\left(B_{21}+C_{22}^{-1}C_{21}B_{11}\right)^{H}\right]$$

$$+\operatorname{tr}\left[B_{11}\left(C_{11}-C_{12}C_{22}^{-1}C_{21}\right)\right]+\operatorname{tr}\left(C_{22}X\right)$$

$$=\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]+\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{11}^{H}C_{21}^{H}C_{22}^{-H}\right]$$

$$+\operatorname{tr}\left[C_{21}B_{21}^{H}\right]+\operatorname{tr}\left[C_{21}B_{11}^{H}C_{21}^{H}C_{22}^{-H}\right]$$

$$+\operatorname{tr}\left[B_{11}C_{11}\right]-\operatorname{tr}\left[B_{11}C_{12}C_{22}^{-1}C_{21}\right]+\operatorname{tr}\left(C_{22}X\right)$$

Recall that $B_{11}=B_{11}^{H}$, $B_{12}=B_{21}^{H}$, $C_{12}=C_{21}^{H}$, $C_{22}^{-1}=C_{22}^{-H}$. We use this to simplify the notion to

$$\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]+\operatorname{tr}\left[C_{22}B_{21}C_{12}C_{22}^{-1}\right]+\operatorname{tr}\left[C_{21}B_{12}\right]+\operatorname{tr}\left[C_{21}B_{11}C_{12}C_{22}^{-1}\right]$$

$$+\operatorname{tr}\left[B_{11}C_{11}\right]-\operatorname{tr}\left[B_{11}C_{12}C_{22}^{-1}C_{21}\right]+\operatorname{tr}\left[C_{22}X\right]$$

$$=\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]+\operatorname{tr}\left[B_{21}C_{12}\right]+\operatorname{tr}\left[C_{21}B_{12}\right]+\operatorname{tr}\left[C_{21}B_{11}C_{12}C_{22}^{-1}\right]$$

$$+\operatorname{tr}\left[B_{11}C_{11}\right]-\operatorname{tr}\left[B_{11}C_{12}C_{22}^{-1}C_{21}\right]+\operatorname{tr}\left[C_{22}X\right]$$

$$=\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]+\operatorname{tr}\left[B_{21}C_{12}\right]+\operatorname{tr}\left[C_{21}B_{12}\right]+\operatorname{tr}\left[B_{11}C_{11}\right]+\operatorname{tr}\left[C_{22}X\right]$$

$$=\operatorname{tr}\left[C_{22}X\right]+\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]+\operatorname{tr}\left[C_{21}B_{12}\right]+\operatorname{tr}\left[C_{12}B_{21}\right]+\operatorname{tr}\left[C_{11}B_{11}\right]$$

$$\operatorname{tr}\left(\Sigma^{-1}W\right)=\operatorname{tr}\left[C_{11}B_{11}\right]+\operatorname{tr}\left[C_{12}B_{21}\right]+\operatorname{tr}\left[C_{21}B_{12}\right]$$

$$+\operatorname{tr}\left[C_{22}X\right]+\operatorname{tr}\left[C_{22}B_{21}B_{11}^{-1}B_{21}^{H}\right]$$

$$C_{22} = \Sigma_{22.1}^{-1} = \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}$$

$$C_{11} - C_{12}C_{22}^{-1}C_{21} = \Sigma_{11}^{-1}$$

$$C_{22}^{-1}C_{21} = -\Sigma_{21}\Sigma_{11}^{-1}$$

We recognize some of the pieces as follows.

$$|\det W|^{n-p} = |(\det B_{11})(\det X)|^{n-p}$$

and

$$\exp\left[-\operatorname{tr}\left(\Sigma^{-1}W\right)\right]$$

$$= \exp\left\{-\operatorname{tr}\left[C_{22}X\right] - \operatorname{tr}\left[B_{11}\left(C_{11} - C_{12}C_{22}^{-1}C_{21}\right)\right]\right.$$

$$\left. - \operatorname{tr}\left[C_{22}\left(B_{21} + C_{22}^{-1}C_{21}B_{11}\right)B_{11}^{-1}\left(B_{21} + C_{22}^{-1}C_{21}B_{11}\right)^{H}\right]\right\}$$

$$[\det \Sigma]^n = [(\det \Sigma_{11})(\det \Sigma_{22.1})]^n$$

$$\mathbf{C}\Gamma_p(n) = \pi^{p(p-1)/2}\prod_{i=1}^{p}\Gamma(n-i+1)$$

We expect

$$\mathbf{C}W_p(n,\Sigma) = \mathbf{C}W_{p-q}(n-q,\Sigma_{22.1})\cdot\mathbf{C}W_q(n,\Sigma_{11})\cdot\mathbf{C}N_{(p-q),q}(\Sigma_{21}\Sigma_{11}^{-1}W_{11},\Sigma_{22.1},W_{11})$$

We look at the density functions.

$$f\left[\mathbf{C}W_p(n,\Sigma)\right] = \frac{|\det W|^{n-p}\operatorname{etr}\left[-\Sigma^{-1}W\right]}{(\det\Sigma)^n\,\mathbf{C}\Gamma_p(n)}$$

$$f\left[\mathbf{C}W_{p-q}(n-q,\Sigma_{22.1})\right] = \frac{|\det X|^{n-q-p+q}\operatorname{etr}\left[-\Sigma_{22.1}^{-1}X\right]}{(\det\Sigma_{22.1})^{n-q}\,\mathbf{C}\Gamma_{p-q}(n-q)}$$

$$f\left[\mathbf{C}W_q(n,\Sigma_{11})\right] = \frac{|\det B_{11}|^{n-q}\operatorname{etr}\left[-\Sigma_{11}^{-1}B_{11}\right]}{(\det\Sigma_{11})^n\,\mathbf{C}\Gamma_q(n)}$$

$$f\left[\mathbf{C}N_{(p-q),q}(\Sigma_{21}\Sigma_{11}^{-1}W_{11}, \Sigma_{22.1}, W_{11})\right]$$

$$= \frac{\text{etr}\left[-(\Sigma_{22.1})^{-1}\left(B_{21} - \Sigma_{21}\Sigma_{11}^{-1}W_{11}\right)W_{11}^{-1}\left(B_{21} - \Sigma_{21}\Sigma_{11}^{-1}W_{11}\right)^{H}\right]}{\pi^{q(p-q)}\,|\Sigma_{22.1}|^{q}\,|W_{11}|^{p-q}}$$

To show the claimed product is true, we observe

$$[\det \Sigma_{22.1}]^{n-q}\,[\det \Sigma_{11}]^{n}\,[\Sigma_{22.1}]^{q} = [\det \Sigma]^{n}$$

$$\frac{|\det X|^{n-p}\,|\det B_{11}|^{n-q}}{|\det B_{11}|^{p-q}} = |\det X|^{n-p}\,|\det B_{11}|^{n-p} = |\det W|^{n-p}$$

$$\mathbf{C}\Gamma_{p-q}(n-q)C\Gamma_{q}(n)\pi^{q(p-q)}$$

$$= \left[\pi^{(p-q)(p-q-1)/2}\right]\left[\prod_{i=1}^{p-q}\Gamma(n-q-i+1)\right]\left[\pi^{q(q-1)/2}\right]\left[\prod_{i=1}^{q}\Gamma(n-i+1)\right]\pi^{q(p-q)}$$

Exponents of π are

$$\frac{1}{2}(p-q)(p-q-1) + \frac{1}{2}q(q-1) + \frac{2}{2}q(p-q)$$

$$= \frac{1}{2}(p-q)(p-q-1+2q) + \frac{1}{2}q(q-1) = \frac{1}{2}[(p-q)(p+q-1) + q(q-1)]$$

$$= \frac{1}{2}[p(p-q) + (p-q)(q-1) + q(q-1)] = \frac{1}{2}[p(p-q) + p(q-1)] = \frac{1}{2}p(p-1)$$

The product of the Gamma functions is

$$\left[\prod_{i=1}^{p-q}\Gamma(n-q-i+1)\right]\left[\prod_{i=1}^{q}\Gamma(n-i+1)\right]$$

$$= \left[\prod_{i=q+1}^{p}\Gamma(n-q-i+q+1)\right]\left[\prod_{i=1}^{q}\Gamma(n-i+1)\right]$$

$$= \prod_{i=1}^{p}\Gamma(n-i+1)$$

□

Theorem 85 *Let* $W \sim \mathbf{C}W_p(n, \Sigma), \Sigma > 0$, *and let* $A = W^{-1}$ *and* $G = \Sigma^{-1}$. *Then the density function of* A *is*

$$A \sim \mathbf{C}IW_p(n, G) = \frac{|\det A|^{-(n+p)} \operatorname{etr}(-GA^{-1}) [\det G]^n}{\mathbf{C}\Gamma_p(n)}(dA)$$

This is the Complex Inverted Wishart Distribution. This is a complex generalization of Mardia et al. equation 3.8.2 [171], and also a complexification of theorem 7.7.1 of Anderson [26]. The real variables case is also reported in Siskind [247].

Proof. This is a complexification of Anderson's proof. Recall that the density for the complex Wishart distribution is given by

$$f_W(W) = \frac{|\det W|^{n-p} \operatorname{etr}(-\Sigma^{-1}W)}{\pi^{p(p-1)/2} [\det \Sigma]^n \prod_{i=1}^{p} \Gamma(n-i+1)}(dW)$$

The Jacobian for the complex change of variables $W = A^{-1}$ where $W^H = W > 0$ is given in theorem 40 to be $J(W \to A) = |\det A|^{-2p}$. Thus $f_A(A) = f_W(A^{-1})J(W \to A)$.

$$f_A(A) = \frac{|\det A^{-1}|^{n-p} \operatorname{etr}(-GA^{-1}) |\det A|^{-2p}}{[\det G^{-1}]^n \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(n-i+1)}$$

Note that $-n + p - 2p = -n - p = -(n+p)$.

$$f_A(A) = \frac{[\det G]^n |\det A|^{-(n+p)} \operatorname{etr}(-GA^{-1})}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(n-i+1)} = \frac{|A|^{-(n+p)} \operatorname{etr}(-GA^{-1}) [\det G]^n}{\mathbf{C}\Gamma_p(n)}$$

□

Theorem 86 *Let $W \sim CW_p(n, \Sigma), \Sigma > 0$, $n > p$. Then*

$$\mathcal{E}\{\left|\det W^{-1}\right|\} = \frac{1}{p!\binom{n-1}{p}\left[\det \Sigma\right]}$$

This is a complexification of Arnold's theorem 17.15(c)(ii) [31], which was stated without proof.

Proof. This proof was motivated by Mardia et al. [171] (p. 487) equation B.3.6. First, recall for the χ^2_m distribution that if $x \sim \chi^2_m$ then

$$\mathcal{E}\{x^k\} = 2^k \frac{\Gamma\left(\frac{m}{2} + k\right)}{\Gamma\left(\frac{m}{2}\right)}$$

When $k = -1$ then

$$\mathcal{E}\{\frac{1}{x}\} = \frac{\Gamma\left(\frac{m}{2} - 1\right)}{2\Gamma\left(\frac{m}{2}\right)} = \frac{\Gamma\left(\frac{m}{2} - 1\right)}{2\Gamma\left(\frac{m}{2} - 1\right)\Gamma\left(\frac{m}{2} - 1\right)} = \frac{1}{2\left(\frac{m}{2} - 1\right)}$$

This implies $\mathcal{E}\{\frac{1}{x}\} = \frac{1}{m-2}$.

By theorem 78, we know $\mathcal{E}\left\{\frac{|\det W|}{[\det \Sigma]}\right\}$ has the same distribution as $\prod_{i=1}^{p} U_i$ where U_i are independent and $2U_i \sim \chi^2_{2(n-i+1)}$. Thus, $\mathcal{E}\left\{\frac{|\det W|}{[\det \Sigma]}\right\}$ has the same distribution as $\prod_{i=1}^{p} \frac{1}{U_i}$ where $2U_i \sim \chi^2_{2(n-i+1)}$.

$$\mathcal{E}\left\{\frac{1}{2U_i}\right\} = \frac{1}{2(n-i+1)-2} = \frac{1}{2(n-i)}$$

implies $\mathcal{E}\left\{\frac{1}{U_i}\right\} = \frac{1}{n-i}$. Therefore,

$$\mathcal{E}\left\{\left|det(\Sigma W^{-1})\right|\right\} = \prod_{i=1}^{p} \frac{1}{(n-i)}$$

$$\mathcal{E}\left\{\left|det(W^{-1})\right|\right\} = \mathcal{E}\left\{\left|\frac{[det\,\Sigma]}{|det\,W|}\frac{1}{[det\,\Sigma]}\right|\right\} = \frac{1}{[\det \Sigma]}\mathcal{E}\left\{\left|det(\Sigma W^{-1})\right|\right\}$$

$$= \frac{1}{(n-1)(n-2)\cdots(n-p)\,[\det \Sigma]} = \frac{1}{p!\binom{n-1}{p}\,[\det \Sigma]}$$

□

Theorem 87 *Let* $W \sim \mathbf{C}W_p(n,\Sigma), \Sigma > 0, n > p$. *Then* $\mathcal{E}\{W^{-1}\} = \frac{1}{n-p}\Sigma^{-1}$. *This is a complexification of Arnold's theorem 17.15(d) [31], which was stated without proof.*

Proof. Let $V \sim \mathbf{C}W_p(n,\Lambda^2)$ where $\Lambda^2 = \mathrm{diag}(\lambda_1^2,\cdots,\lambda_p^2)$. Let a be a column vector with zeros in every position except for a 1 in position i. Let V^{ii} be the element in position (i,i) of V^{-1}. Then by theorem 64,

$$2\frac{a^H\Lambda^{-2}a}{a^HV^{-1}a} = 2\frac{\left(\frac{1}{\lambda_i^2}\right)}{V^{ii}} = \frac{2}{\lambda_i^2V^{ii}} \sim \chi^2_{2(n-p+1)}(0)$$

Then

$$\mathcal{E}\left\{\frac{1}{2}\lambda_i^2V^{ii}\right\} = \frac{1}{2(n-p+1)-2} = \frac{1}{2(n-p)}$$

which implies $\mathcal{E}\{V^{ii}\} = \frac{1}{n-p}\frac{1}{\lambda_i^2}$. Thus $\mathcal{E}\{V^{-1}\} = \frac{1}{n-p}\Lambda^{-2}$ where $n > p$.

Let $\Sigma = \Gamma\Lambda^2\Gamma^H$, $W = \Gamma V\Gamma^H$. By theorem 54,

$$W \sim \mathbf{C}W_p(n,\Gamma\Lambda^2\Gamma^H) = \mathbf{C}W_p(n,\Sigma)$$

Then

$$\Gamma\mathcal{E}\left\{V^{-1}\right\}\Gamma^H = \mathcal{E}\left\{\Gamma V^{-1}\Gamma^H\right\} = \mathcal{E}\left\{(\Gamma V\Gamma^H)^{-1}\right\} = \mathcal{E}\left\{W^{-1}\right\}$$

$$= \frac{1}{n-p}\Gamma V^{-1}\Gamma^H = \frac{1}{n-p}\Sigma^{-1}$$

where $n > p$.

Note: $\mathcal{E}\{|W^{-1}|\} \neq |\mathcal{E}\{W^{-1}\}|$. In fact,

$$\left|\mathcal{E}\left\{W^{-1}\right\}\right| = \frac{(n-1)!}{(n-p)!}\mathcal{E}\{\left|W^{-1}\right|\} = (p-1)!\binom{n-1}{p-1}\mathcal{E}\{\left|W^{-1}\right|\}$$

□

Theorem 88 *Let* $W \sim \mathbf{C}W_p(n,\Sigma), \Sigma > 0$. *Then* $\mathcal{E}\{\operatorname{tr} W\} = n \operatorname{tr}\Sigma$. *If* $n > p$ *then* $\mathcal{E}\{\operatorname{tr} W^{-1}\} = \frac{1}{n-p}\operatorname{tr}\Sigma^{-1}$. *This is a complexification of Arnold's theorem 17.15(e) [31], which was stated without proof.*

Proof. By theorem 52, $\mathcal{E}\{W\} = n\Sigma$. The trace function is merely a linear combination of elements on the diagonal of a matrix. Expectation is a linear operator. Therefore

$$\mathcal{E}\{\operatorname{tr} W\} = \operatorname{tr}\mathcal{E}\{W\} = \operatorname{tr}[n\Sigma] = n \operatorname{tr}\Sigma$$

By theorem 87, if $n > p$ then $\mathcal{E}\{W^{-1}\} = \frac{1}{n-p}\Sigma^{-1}$. Therefore,

$$\mathcal{E}\{\operatorname{tr} W^{-1}\} = \frac{1}{n-p}\operatorname{tr}\Sigma^{-1}$$

□

F.4 Tague and Styan Properties

This section is included to *demonstrate the usefulness* of the statistical theory developed during this thesis research. This is all work by Tague [264], slightly reordered in places and with the derivation of some constants expanded. It is also included to collect work in the literature into a unified presentation.

Theorem 89 *Let* $W \sim \mathbf{C}W_p(k, I)$, $U \in \mathbf{U}(p)$, *and* $A \in \mathbf{C}^{p\times p}$. *Then*

$$g(A) = \mathcal{E}\{\text{etr}(AW)\} = \mathcal{E}\{\text{etr}(AU^H WU\} = g(UAU^H)$$

This is from Tague [264].

Proof. By definition of an expected value, we define $g(A)$ as follows.

$$g(A) = \int_{W>0} \text{etr}(AW) f_W(W)(dW) = \int_{W>0} \text{etr}(AW) \frac{|\det W|^{k-p} \text{etr}(-W)}{\mathbf{C}\Gamma_p(k)} (dW)$$

Now consider what happens under unitary similarity transformation.

$$g(UAU^H) = \int_{W>0} \text{etr}(UAU^H W) f_W(W)(dW)$$

$$= \int_{W>0} \text{etr}(AU^H WU) f_W(W)(dW)$$

by property of the trace function. Now, perform a change of variables $Y = U^H WU$,which has the inverse relation $W = UYU^H$. By corollary 7, the Jacobian of this transformation is 1. Thus

$$g(UAU^H) = \int_{Y>0} \text{etr}(AY) \frac{\left|\det(UYU^H)\right|^{k-p} \text{etr}(-UYU^H)}{\mathbf{C}\Gamma_p(k)} (dY)$$

$$= \int_{Y>0} \text{etr}(AY) \frac{\left|\det(U)\det(Y)\det(U^H)\right|^{k-p} \text{etr}(-U^H UY)}{\mathbf{C}\Gamma_p(k)} (dY)$$

We note that $U^H U = I$ because $U \in \mathbf{U}(p)$ and also $\left|\det(U)\det(U^H)\right| = 1$. Thus we have

$$g(UAU^H) = \int_{Y>0} \text{etr}(AY) \frac{|\det(Y)|^{k-p} \text{etr}(-Y)}{\mathbf{C}\Gamma_p(k)} (dY) = g(A)$$

□

Theorem 90 *Let* $W \sim \mathbf{C}W_p(k, \Sigma), \Sigma > 0$, *with deterministic matrix* $A \in \mathbf{C}^{p\times p}$. *Then* $\mathcal{E}\{WAW\} = k^2\Sigma A\Sigma + k\,\mathrm{tr}(A\Sigma)\Sigma$. *This result was proven for the complex case by Tague [264], motivated by Styan's treatment [262] of the problem for the real case. The result is not a simple extension of the real case.*

Proof. Let $W \sim \mathbf{C}W_p(k, I)$. Recall from lemma 58 that for random $W \in \mathbf{C}^{p\times p}$ and fixed $T \in \mathbf{C}^{p\times p}$ that, using a moment generating function argument,

$$\mathcal{E}\{W_{ij}W_{lm}\} = b_1\delta_{ij}\delta_{lm} + b_2\delta_{im}\delta_{jl}$$

where δ_{jk} is the delta function $\delta_{jk} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$ and b_1, b_2 are constants. By lemma 25

$$(WAW)_{im} = \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}W_{ij}W_{lm}$$

and thus

$$\mathcal{E}\{(WAW)_{im}\} = \mathcal{E}\left\{\sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}W_{ij}W_{lm}\right\} = \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}\mathcal{E}\left\{W_{ij}W_{lm}\right\}$$

$$= \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}\left(b_1\delta_{ij}\delta_{lm} + b_2\delta_{im}\delta_{jl}\right) = A_{im}b_1 + b_2\delta_{im}\sum_{j=1}^{p} A_{jj}$$

$$= b_1A_{im} + \delta_{im}b_2\,\mathrm{tr}(A)$$

Then for the whole matrix,

$$\mathcal{E}\{WAW\} = b_1A + b_2I\,\mathrm{tr}(A)$$

where $W \sim \mathbf{C}W_p(k, I)$.

We need to evaluate b_1 and b_2. Suppose we let $A = e_i e_i^H$ where e_i is the i^{th} standard basis vector, which is a zero vector with a 1 added to the i^{th} element. Then

$$\mathcal{E}\{WAW\} = b_1 e_i e_i^H + b_2 I$$

Further,

$$e_i^H \mathcal{E}\{WAW\} e_i = \mathcal{E}\{e_i^H W e_i e_i^H W e_i\} = b_1 + b_2 = \mathcal{E}\{W_{ii}^2\}$$

By corollary 17, $W_{ii} \sim \mathbf{C}W_1(k, I_1)$. By theorem 53, then $2W_{ii} \sim \chi_{2k}^2(0)$. From properties of the Chi-square distribution, theorem 142, we know $\mathcal{E}\{W_{ii}^2\} = k(k+1)$. Recall that it is $(2W_{ii})$ that has the $\chi_{2k}^2(0)$ distribution, not W_{ii}. Substituting into our earlier result,

$$\mathcal{E}\{W_{ii}^2\} = b_1 + b_2 = k^2 + k$$

Now consider off-diagonal elements of WAW. Let $A = e_j e_j^H$ where $i \neq j$. Since $\Sigma = I$, then theorem 56 tells us that the set of $\{W_{ii}\}$ are independent random variables. Thus

$$e_i^H \mathcal{E}\{WAW\} e_j = \mathcal{E}\{W_{ii} W_{jj}\} = \mathcal{E}\{W_{ii}\}\mathcal{E}\{W_{jj}\} = k \cdot k = k^2$$

and also

$$e_i^H \mathcal{E}\{WAW\} e_j = e_i^H \{b_1 e_i e_j^H + b_2 I \operatorname{tr}(e_i e_j^H)\} e_j = b_1 + b_2 \operatorname{tr}(e_i e_j^H) = b_1$$

Therefore $b_1 = k^2$ which implies $k^2 + b_2 = k^2 + k$ which gives us $b_2 = k$. Then

$$\mathcal{E}\{WAW\} = k^2 A + k I \operatorname{tr}(A) \qquad \text{(F.2)}$$

where $W \sim \mathbf{C}W_p(k, I)$.

We now want to consider the case when $W \sim \mathbf{C}W_p(k, \Sigma)$ for $\Sigma^H = \Sigma > 0$. By theorem 119, the decomposition $\Sigma = GG^H$ exists. G^{-1} and $(G^{-1})^H$ exist since $\Sigma > 0$. By theorem 54,

$$G^{-1}WG^{-H} \sim \mathbf{C}W_p(k, G^{-1}\Sigma G^{-H}) = \mathbf{C}W_p(k, I)$$

Applying equation F.2,

$$\mathcal{E}\{G^{-1}WG^{-H}BG^{-1}WG^{-H}\} = k^2B + kI\,\mathrm{tr}(B)$$

for $B \in \mathbf{C}^{p\times p}$. Then

$$G\mathcal{E}\{G^{-1}WG^{-H}BG^{-1}WG^{-H}\}G^H = \mathcal{E}\{WG^{-H}BG^{-1}W\}$$

$$= k^2GBG^H + kGIG^H\,\mathrm{tr}(B) = k^2GBG^H + kGG^H\,\mathrm{tr}(B)$$

Let $A = G^{-H}BG^{-1}$. Then $B = G^HAG$, and

$$\mathcal{E}\{WAW\} = k^2GG^HAGG^H + kGG^H\,\mathrm{tr}(G^HAG) = k^2\Sigma A\Sigma + k\Sigma\,\mathrm{tr}(A\Sigma)$$

which is the main result. □

Corollary 24 *Let $W \sim \mathbf{C}W_p(k, \Sigma)$. Then $\mathcal{E}\{W^2\} = k^2\Sigma^2 + k\Sigma\,\mathrm{tr}(\Sigma)$. This simple special case was first produced for the real variables case by Styan, and then rederived for the complex case by Tague.*

Proof. Let $A = I$ in theorem 90. □

Corollary 25 *Let* $W \sim \mathbf{C}W_p(k, \Sigma)$ *and* $a \in \mathbf{C}^p$. *Then* $\mathcal{E}\{Waa^HW\} = k^2\Sigma aa^H\Sigma + ka^H\Sigma a\Sigma$.

Proof. Let $A = aa^H$ in theorem 90. Note that $\mathrm{tr}(aa^H) = \mathrm{tr}\, a^H\Sigma a) = a^H\Sigma a$ is a scalar. □

Corollary 26 *Let* $W \sim \mathbf{C}W_p(k, \Sigma)$ *and* $a \in \mathbf{C}^p$. *then* $\mathrm{var}(Wa) = ka^H\Sigma a\Sigma$.

Proof. $\mathcal{E}\{Wa\} = \mathcal{E}\{W\}a = k\Sigma a$ by theorem 52. By definition,

$$\mathrm{var}(Wa) = \mathcal{E}\{(Wa)(Wa)^H\} - \mathcal{E}\{Wa\}\mathcal{E}\{(Wa)^H\}$$

$$= \mathcal{E}\{Waa^HW\} - k^2\Sigma aa^H\Sigma = k^2\Sigma aa^H\Sigma + ka^H\Sigma a\Sigma - k^2\Sigma aa^H\Sigma$$

from corollary 25. The final result is

$$\mathrm{var}(Wa) = ka^H\Sigma a\Sigma$$

□

Theorem 91 *Let* $W^{-1} = V \sim \mathbf{C}IW_p(k, I)$, $U \in \mathbf{U}(n)$, *and* $A \in \mathbf{C}^{p\times p}$. *Then* $g(A) = \mathcal{E}\{\mathrm{etr}(AV)\} = \mathcal{E}\{\mathrm{etr}(AU^HVU)\} = g(UAU^H)$. *This property was taken from Tague [264] with permission.*

Proof. This is analogous to Tague's proof for the case of $Z \sim \mathbf{C}W_p(k, I)$. We define $g(A)$ as follows.

$$g(A) = \int_{V>0} \mathrm{etr}(AV)\frac{|\det V|^{-(k+p)}\,\mathrm{etr}(-V^{-1})}{\mathbf{C}\Gamma_p(k)}(dV)$$

by the definition of an expected value, where the density function comes from theorem 85 for the complex inverted Wishart distribution. Now we consider what happens to g as we use UAU^H as the argument. Following the definition, then

$$g(UAU^H) = \int_{V>0} \text{etr}(UAU^H V) f_V(V)(dV) = \int_{V>0} \text{etr}(AU^H VU) f_V(V)(dV)$$

since $\text{tr}(XY) = \text{tr}(YX)$ as a general property of the trace function. Now perform a change of variables $Y = U^H VU$. The inverse transform is $V = UYU^H$. The Jacobian of the transformation is 1, by corollary 7. Thus

$$g(UAU^H) = \int_{Y>0} \text{etr}(AY) \frac{\left|\det\left(UYU^H\right)\right|^{-(k+p)} \text{etr}\left[-(UYU^H)^{-1}\right]}{\mathbf{C}\Gamma_p(k)}(dY)$$

$$= \int_{Y>0} \text{etr}(AY) \frac{\left|\det(Y)\right|^{-(k+p)} \text{etr}\left[-UY^{-1}U^H\right]}{\mathbf{C}\Gamma_p(k)}(dY)$$

$$= \int_{Y>0} \text{etr}(AY) \frac{\left|\det(Y)\right|^{-(k+p)} \text{etr}\left[-Y^{-1}\right]}{\mathbf{C}\Gamma_p(k)}(dY) = g(A)$$

□

Theorem 92 *Let $W_1 \sim \mathbf{C}W_p(k_1, \Sigma)$ and $W_2 \sim \mathbf{C}W_p(k_2, \Sigma)$. If $k_1 > p$, then $\mathcal{E}\left\{W_2 W_1^{-1} W_2\right\} = \frac{k_2(k_2+p)}{k_1-p}\Sigma$. This is a complexification by Tague [264] of a corollary Styan [262] provided for the case of real variables.*

Proof. First note that

$$\mathcal{E}\left\{W_2 W_1^{-1} W_2\right\} = \mathcal{E}_{W_1}\left\{\mathcal{E}\left\{W_2 W_1^{-1} W_2 \mid W_1\right\}\right\}$$

When W_1 is fixed, then theorem 90 tells us

$$\mathcal{E}\left\{W_2 W_1^{-1} W_2 \mid W_1\right\} = k_2^2 \Sigma W_1^{-1} \Sigma + k_2 \operatorname{tr}(W_1^{-1}\Sigma)\Sigma$$

from theorem 87, $\mathcal{E}\left\{W_1^{-1}\right\} = \frac{1}{k_1 - p}\Sigma^{-1}$. We require $k_1 > p$ to ensure the denominator is not zero. Also, $k_1 > p$ ensures W_1^{-1} exists. Continuing, taking the expectation with respect to W_1, we obtain

$$\begin{aligned}\mathcal{E}\left\{W_2 W_1^{-1} W_2\right\} &= k_2^2 \Sigma \left(\frac{1}{k_1 - p}\right)\Sigma^{-1}\Sigma + k_2 \operatorname{tr}\left[\left(\frac{1}{k_1 - p}\right)\Sigma^{-1}\Sigma\right]\Sigma \\ &= \left(\frac{k_2^2}{k_1 - p}\right)\Sigma + \left(\frac{k_2}{k_1 - p}\right)\operatorname{tr}(I_p)\Sigma = \frac{k_2(k_2 + p)}{k_1 - p}\Sigma\end{aligned}$$

Theorem 93 *If W_1 and W_2 are independent, $W_i \sim \mathbf{C}W_p(n_i, \Sigma)$, then*

$$W_1 + W_2 \sim \mathbf{C}W_p(n_1 + n_2, \Sigma)$$

This is a complexification of Arnold's theorem 17.15(f), which was stated without proof.

Proof. From theorem 75, the characteristic function of the associated random variable $\check{W}_i$ is

$$\Phi_{W_i}(T) = \left[\det\left(I_p - i\Sigma T\right)\right]^{-n_i}$$

Since W_1 and W_2 are independent, the characteristic function of the distribution of the sum is the product of the individual characteristic functions. Thus

$$\Phi_{W_1 + W_2}(T) = \Phi_{W_1}(T)\Phi_{W_2}(T) = \left[\det\left(I_p - i\Sigma T\right)\right]^{-n_1}\left[\det\left(I_p - i\Sigma T\right)\right]^{-n_2}$$

$$= [\det(I_p - i\Sigma T)]^{-(n_1+n_2)}$$

This is the characteristic function of the associated random variable corresponding to the complex Wishart distribution $\mathbf{C}W_p(n_1 + n_2, \Sigma)$. □

Theorem 94 *Let $A \sim \mathbf{C}W_p(n, I_p)$ and $B \sim \mathbf{C}N_{p,m}(0, I_p, I_m)$ be independent complex random variables. Let $Z = A + BB^H = CC^H$ and $B = CU$. Then $Z \sim \mathbf{C}W_p(m + n, I_p)$, and the density of U is given by*

$$g(U) = \frac{\mathbf{C}\Gamma_p(m+n)}{\pi^{mp}\mathbf{C}\Gamma_p(n)} \left|\det\left(I - UU^H\right)\right|^{n-p}$$

This is a complex version of the derivation given by Anderson (p. 302) [26] for the real variables case.

Proof. The concept of solving for the joint distribution of Z and U and the recognizing their independence is due to Anderson. From theorem 93, since $BB^H \sim \mathbf{C}W_p(m, I_p)$, we know

$$Z \sim \mathbf{C}W_p(n + m, I_p)$$

The joint distribution of A and B is

$$f(A, B) = \mathbf{C}W_p(n, I) \cdot \mathbf{C}N_{m,p}(0, I_m, I_p) = \frac{|\det A|^{n-p} \operatorname{etr}(-A)}{\mathbf{C}\Gamma_p(n)} \frac{\operatorname{etr}(-BB^H)}{\pi^{mp}}$$

Note that the density of Z is

$$g(Z) = \frac{|\det Z|^{m+n-p} \operatorname{etr}(-Z)}{\mathbf{C}\Gamma_p(m+n)}$$

We want to find the joint density of Z and U. Begin with $f(A, B)$ and manipulate it until we have a function that includes $g(A + BB^H)$ as a factor.

$$f(A,B) = \frac{\left|\det\left(A+BB^H\right)\right|^{m+n-p} \operatorname{etr}\left[-\left(A+BB^H\right)\right]}{\mathbf{C}\Gamma_p(m+n)} \frac{\mathbf{C}\Gamma_p(m+n)}{\mathbf{C}\Gamma_p(n)} \times$$

$$\times \frac{|\det A|^{n-p}}{|\det\left(A+BB^H\right)|^{n-p}} \frac{1}{|\det\left(A+BB^H\right)|^m \pi^{mp}}$$

Note that the term

$$\frac{\left|\det\left(A+BB^H\right)\right|^{m+n-p} \operatorname{etr}\left[-\left(A+BB^H\right)\right]}{\mathbf{C}\Gamma_p(m+n)} = \frac{|\det Z|^{m+n-p} \operatorname{etr}(-Z)}{\mathbf{C}\Gamma_p(m+n)}$$

$$= \mathbf{C}W_p(m+n, I)$$

already accounts for the change of variables from $(A + BB^H)$ to Z. Thus we only need to include the Jacobian for the change of variables from B to U. $J(B \to U) = |\det C|^{2m}$. Thus $g(Z, U) =$

$$\frac{|\det Z|^{m+n-p} \operatorname{etr}(-Z)}{\mathbf{C}\Gamma_p(m+n)} \frac{\mathbf{C}\Gamma_p(m+n)}{\pi^{mp}\mathbf{C}\Gamma_p(n)} \frac{|\det A|^{n-p}}{|\det(A+BB^H)|^{n-p}} \frac{|\det C|^{2m}}{|\det(A+BB^H)|^m}$$

where we still have the substitutions to complete. For this, we still have an identity to compute. From

$$Z = A + BB^H = CC^H$$

we have

$$A = Z - BB^H = CC^H - (CU)(CU)^H = C(I - UU^H)C^H$$

Then

$$\frac{|\det A|}{|\det(A+BB^H)|} = \frac{\left|\det\left\{C(I-UU^H)C^H\right\}\right|}{|\det(CC^H)|} = \left|\det(I-UU^H)\right|$$

and

$$\frac{|\det C|^2}{|\det(A + BB^H)|} = 1$$

We substitute our identities to obtain the joint density of Z and U.

$$g(Z,U) = \frac{|\det Z|^{m+n-p} \operatorname{etr}(-Z)}{\mathbf{C}\Gamma_p(m+n)} \frac{\mathbf{C}\Gamma_p(m+n)\left|\det(I - UU^H)\right|^{n-p}}{\mathbf{C}\Gamma_p(n)\pi^{mp}}$$

We note that Z and U are independent by the Neyman-Fisher factorization theorem. We know the density of Z is $\mathbf{C}W_p(m+n, I)$, and thus the density of U is

$$g(U) = \frac{\mathbf{C}\Gamma_p(m+n)}{\pi^{mp}\mathbf{C}\Gamma_p(n)} \left|\det(I - UU^H)\right|^{n-p}$$

□

Theorem 95 *Let* $W \sim \mathbf{C}W_p(n, \Sigma), \Sigma > 0, n \geq p$, *and* $\operatorname{rank}(A_{q\times q}) = q$. *Then* $(AW^{-1}A^H)^{-1} \sim \mathbf{C}W_q(n-p+q, (A\Sigma^{-1}A^H)^{-1})$. *This is a complexification of Arnold's theorem 17.15(g), which was stated without proof. This is also similar to Muirhead's theorem 3.2.11 for the real variables case.*

Proof. This follows Muirhead's proof, except that this version accounts for the structure of complex variables.

By theorem 119, there exists a positive definite complex matrix $\Sigma^{1/2}$ such that $\Sigma = (\Sigma^{1/2})(\Sigma^{1/2})^H$. Thus, $\Sigma^{-1} = \Sigma^{-H/2}\Sigma^{-1/2}$. Let $B = \Sigma^{-1/2}W\Sigma^{-H/2}$, which implies that $W = \Sigma^{1/2}B\Sigma^{H/2}$. By theorem 54,

$$B \sim \mathbf{C}W_q(n, \Sigma^{-1/2}\Sigma\Sigma^{-H/2}) = \mathbf{C}W_q(n, \Sigma^{-1/2}\Sigma^{1/2}\Sigma^{H/2}\Sigma^{-H/2}) = \mathbf{C}W_q(n, I_q)$$

Let $R = A\Sigma^{-H/2}$, which implies $A = R\Sigma^{H/2}$. Then

$$(AW^{-1}A^H)^{-1} = \left[\left(R\Sigma^{H/2}\right)\left(\Sigma^{1/2}B\Sigma^{H/2}\right)^{-1}\left(R\Sigma^{H/2}\right)^H\right]^{-1}$$

$$\left[R\Sigma^{H/2}\Sigma^{-H/2}B^{-1}\Sigma^{-1/2}\Sigma^{1/2}R^H\right]^{-1} = (RB^{-1}R^H)^{-1}$$

By theorem 125, R can be written as $R = L(I_q, 0)H$ where H is a $p \times p$ unitary matrix, and $L_{q\times q}$ is a positive definite matrix. Then

$$(AW^{-1}A^H)^{-1} = (RB^{-1}R^H)^{-1} = \left[\left(L\left\{I_q, 0\right\}H\right)B^{-1}\left(L\left\{I_q, 0\right\}H\right)^H\right]^{-1}$$

$$= \left[L(I_q, 0)HB^{-1}H^H\begin{pmatrix} I_q \\ 0 \end{pmatrix}L^H\right]^{-1}$$

$$L^{-H}\left[(I_q, 0)HB^{-1}H^H\begin{pmatrix} I_q \\ 0 \end{pmatrix}\right]^{-1}L^{-1} = L^{-H}\left[\begin{pmatrix} I_q & 0 \end{pmatrix}C^{-1}\begin{pmatrix} I_q \\ 0 \end{pmatrix}\right]^{-1}L^{-1}$$

where

$$C = HB^{-1}H^H \sim \mathbf{C}W_p(n, HI_pH^H) = \mathbf{C}W_p(n, I_p)$$

by theorem 54 since H is unitary.

Let $D = C^{-1} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$ and $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ where C_{11} and D_{11} are $q \times q$. Then

$$(AW^{-1}A^H)^{-1} = (RB^{-1}R^H)^{-1} = L^{-H}(C_{11}^{-1})^{-1}L^{-1} = L^{-H}D_{11}^{-1}L^{-1}$$

Recall from lemma 34 (the partitioned matrix right inverse) that $D_{11}^{-1} = C_{11} - C_{12}C_{22}^{-1}C_{21}$. D_{11}^{-1} corresponds to V of theorem 84. By that theorem, $D_{11}^{-1} \sim$

$\mathbf{C}W_q(n-p+q, I_p)$. Then by theorem 54,

$$(AW^{-1}A^H)^{-1} = L^{-H}D_{11}^{-1}L^{-1} \sim \mathbf{C}W_q(n-p+q, L^{-H}I_qL^{-1})$$

$$= \mathbf{C}W_q(n-p+q, (LL^H)^{-1})$$

Consider $(A\Sigma^{-1}A^H)^{-1}$. Expanding,

$$(A\Sigma^{-1}A^H)^{-1} = (A\Sigma^{-H/2}\Sigma^{-1/2}A^H)^{-1} = (R\Sigma^{H/2}\Sigma^{-H/2}\Sigma^{-1/2}(R\Sigma^{H/2})^H)^{-1}$$

$$= (R\Sigma^{H/2}\Sigma^{-H/2}\Sigma^{-1/2}\Sigma^{1/2}R^H)^{-1} = (RR^H)^{-1}$$

$$= \left[L\begin{pmatrix} I_q & 0 \end{pmatrix} H \left(L\begin{pmatrix} I_q & 0 \end{pmatrix} H\right)^H\right]^{-1}$$

$$= \left[L\begin{pmatrix} I_q & 0 \end{pmatrix} HH^H \begin{pmatrix} I_q \\ 0 \end{pmatrix} L^H\right]^{-1} = \left[L\begin{pmatrix} I_q & 0 \end{pmatrix} I_p \begin{pmatrix} I_q \\ 0 \end{pmatrix} L^H\right]^{-1}$$

$$= (LI_qL^H)^{-1} = (LL^H)^{-1}$$

Therefore,

$$(AW^{-1}A^H)^{-1} \sim \mathbf{C}W_q(n-p+q, (A\Sigma^{-1}A^H)^{-1})$$

□

Lemma 23 . *Let* $W^{-1} \sim \mathbf{C}IW_p(k, I_p)$, $k > p+1$ *and* $A \in \mathbf{C}^{p\times p}$. *Then*

$$\mathcal{E}\{W^{-1}AW^{-1}\} = d_1A + d_2I\,\mathrm{tr}(A)$$

where

$$d_1 = \frac{1}{(k-p+1)(k-p-1)}$$

and

$$d_2 = \frac{1}{(k-p+1)(k-p)(k-p-1)}$$

This is Styan's corollary 4 (part iii) [262] and theorem 3 which has been complexified by Tague [264].

Proof. By theorem 91, W^{-1} satisfies the property $g(A) = g(UAU^H)$ for unitary $U \in U(p)$. By lemma 58,

$$\mathcal{E}\{W^{ij}W^{lm}\} = d_1\delta_{ij}\delta_{lm} + d_2\delta_{im}\delta_{jl}$$

where W^{ii} is the element (i,j) of W^{-1}. Recall that

$$(W^{-1}AW^{-1})_{im} = \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}W^{ij}W^{lm}$$

and thus

$$\mathcal{E}\{(W^{-1}AW^{-1})_{im}\} = \mathcal{E}\left\{\sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}W^{ij}W^{lm}\right\} = \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}\mathcal{E}\left\{W^{ij}W^{lm}\right\}$$

$$= \sum_{j=1}^{p}\sum_{l=1}^{p} A_{jl}\left(d_1\delta_{ij}\delta_{lm} + d_2\delta_{im}\delta_{jl}\right) = A_{im}d_1 + d_2\delta_{im}\sum_{j=1}^{p} A_{jj} = A_{im}d_1 + \delta_{im}d_2\,\mathrm{tr}(A)$$

Then

$$\mathcal{E}\{W^{-1}AW^{-1}\} = d_1A + d_2I\,\mathrm{tr}(A)$$

where $W^{-1} \sim \mathbf{C}IW_p(k, I)$.

By lemma 40,

$$W^{ii} = e_i^H(Y^HY)^{-1}e_i + \frac{\left|e_i^H(Y^HY)^{-1}Y^HZ\right|^2}{Z^H\left\{I - Y(Y^HY)^{-1}Y^H\right\}Z}$$

where W is partitioned as

$$W = \begin{pmatrix} Y^H Y & Y^H Z \\ Z^H Y & Z^H Z \end{pmatrix}$$

with Z being a column vector, and e_I is the standard basis vector consisting of all zeros, except a 1 in position i. Also,

$$W^{pp} = \frac{1}{Z^H \{I - Y(Y^H Y)^{-1} Y^H\} Z}$$

This product is

$$W^{ii} W^{pp} = e_i^H (Y^H Y)^{-1} e_i W^{pp} + \left| e_i^H (Y^H Y)^{-1} Y^H Z \right|^2 (W^{pp})^2$$

We want to find $\mathcal{E}\{W^{ii} W^{pp}\}$.

Since $W^{-1} \sim \mathbf{C}IW_p(k, I)$, then $W \sim \mathbf{C}W_p(k, I)$. By theorem 64, elements on the diagonal of W^{-1} have the distributional property

$$\frac{2}{W^{jj}} \sim \chi^2_{2(k-p+1)}(0)$$

In corollary 23, if we let

$$X = W_{22} - W_{21} W_{11}^{-1} W_{12} = Z^H Z - Z^H Y (Y^H Y)^{-1} Y^H Z$$

we then see that X is independent of $W_{11} = Y^H Y$, and hence X is independent of Y. X is $(W^{pp})^{-1}$. Therefore, W^{pp} is independent of Y. Then

$$\mathcal{E}\{W^{ii} W^{pp}\} = \mathcal{E}\left\{ e_i^H (Y^H Y)^{-1} e_i W^{pp} + \left| e_i^H (Y^H Y)^{-1} Y^H Z \right|^2 (W^{pp})^2 \right\}$$

$$= \mathcal{E}\left\{ e_i^H (Y^H Y)^{-1} e_i \right\} \mathcal{E}\{W^{pp}\} + \mathcal{E}\left\{ \left| e_i^H (Y^H Y)^{-1} Y^H Z \right|^2 (W^{II})^2 \right\}$$

Now look at the first term. From theorem 87 we know $\mathcal{E}\{W^{pp}\} = \frac{1}{k-p}$. By theorem 55,

$$Y^H Y = W_{11} \sim \mathbf{C}W_{p-1}(k, I_{p-1})$$

Again, applying theorem 87, we know

$$\mathcal{E}\left\{(Y^H Y)^{-1}\right\} = \frac{1}{k-p+1} I_{p-1}$$

Substituting these results we find

$$e_i^H \mathcal{E}\left\{(Y^H Y)^{-1}\right\} e_i \mathcal{E}\{W^{pp}\} = \left(\frac{1}{k-p+1}\right)\left(\frac{1}{k-p}\right)$$

for $k > p$.

Now, consider the computation

$$\mathcal{E}\left\{(W^{pp})^2 \left|e_i^H (Y^H Y)^{-1} Y^H Z\right|^2\right\} = \mathcal{E}_Y\left\{\mathcal{E}\left\{(W^{pp})^2 \left|e_i^H (Y^H Y)^{-1} Y^H Z\right|^2 \,\middle|\, Y\right\}\right\}$$

Since

$$\frac{2}{W^{pp}} \sim \chi^2_{2(k-p+1)}(0)$$

from the χ^2 distribution we know

$$\mathcal{E}\left\{\left(\frac{W^{pp}}{2}\right)^2\right\} = \frac{1}{[2(k-p+1)-2]\,[2(k-p+1)-4]}$$

$$= \frac{1}{[2(k-p)]\,[2(k-p)-2]} = \frac{1}{[4(k-p)]\,[k-p-1]}$$

Therefore

$$\mathcal{E}\left\{(W^{pp})^2\right\} = \frac{1}{[k-p]\,[k-p-1]} \qquad \text{(F.3)}$$

for $k > p+1$. Continuing,

$$\mathcal{E}\left\{\left|e_i^H(Y^HY)^{-1}Y^HZ\right|^2 \;\middle|Y\;\right\} = e_i^H(Y^HY)^{-1}Y^H\mathcal{E}\left\{ZZ^H\right\}Y(Y^HY)^{-1}e_i$$

Since $W \sim \mathbf{C}W_p(k, I)$, then $Z \sim \mathbf{C}N_k(0, I)$ by theorem 55 and theorem 53.

By theorem 52, $\mathcal{E}\left\{ZZ^H\right\} = I$. Thus

$$e_i^H(Y^HY)^{-1}Y^HY(Y^HY)^{-1}e_i = e_i^H(Y^HY)^{-1}e_i$$

We substitute this back in to obtain

$$\mathcal{E}_Y\left\{\mathcal{E}\left\{(W^{pp})^2\left|e_i^H(Y^HY)^{-1}Y^HZ\right|^2 \;\middle|Y\;\right\}\right\}$$

$$= \frac{1}{[k-p]\,[k-p-1]}e_i^H\mathcal{E}\left\{(Y^HY)^{-1}\right\}e_i = \frac{1}{[k-p+1]\,[k-p]\,[k-p-1]}e_i^HIe_i$$

Adding our results, we get

$$\mathcal{E}\{W^{ii}W^{pp}\} = \left(\frac{1}{k-p+1}\right)\left(\frac{1}{k-p}\right) + \frac{1}{[k-p+1]\,[k-p]\,[k-p-1]}$$

$$= \frac{1}{[k-p+1]\,[k-p]}\left[1 + \frac{1}{k-p-1}\right] = \frac{1}{[k-p+1]\,[k-p-1]} \qquad \text{(F.4)}$$

Now, calculate $\mathcal{E}\{W^{-1}AW^{-1}\}$. Consider

$$\mathcal{E}\left\{\left(W^{ii}\right)^2\right\} = \mathcal{E}\left\{e_i^HW^{-1}e_ie_i^HW^{-1}e_i\right\} = \frac{1}{(k-p)\,(k-p-1)}$$

when $k > p+1$. If $A = e_ie_i^H$, then

$$\frac{1}{(k-p)\,(k-p-1)} = e_i^H\left[d_1e_ie_i^H + d_2\,\mathrm{tr}(e_ie_i^H)I\right]e_i = d_1 + d_2 \qquad \text{(F.5)}$$

Also,

$$\mathcal{E}\left\{W^{ii}W^{jj}\right\} = \mathcal{E}\left\{e_i^HW^{-1}e_ie_j^HW^{-1}e_j\right\}$$

Now let $A = e_i e_j^H$. Then we find

$$\mathcal{E}\left\{W^{ii}W^{jj}\right\} = e_i^H \left[d_1 e_i e_j^H + d_2 \operatorname{tr}(e_i e_j^H) I\right] e_j = d_1$$

From equation F.4, $d_1 = \frac{1}{(k-p+1)(k-p-1)}$. Substitute into equation F.5 to get

$$d_2 = \frac{1}{[k-p][k-p-1]} - \frac{1}{[k-p+1][k-p-1]}$$

$$= \left[\frac{1}{k-p} - \frac{1}{k-p+1}\right]\left(\frac{1}{k-p-1}\right) = \frac{1}{[k-p+1][k-p][k-p-1]}$$

Therefore, $W^{-1} \sim \mathbf{C}IW_p(k, I_p)$ and $k > p+1$,

$$\mathcal{E}\{W^{-1}AW^{-1}\} = \frac{1}{[k-p+1][k-p-1]}A + \frac{1}{[k-p+1][k-p][k-p-1]}[\operatorname{tr} A]\, I$$

□

Theorem 96 *Let $W^{-1} \sim \mathbf{C}IW_p(k, \Sigma)$ and $A \in \mathbf{C}^{p\times p}$. Then for $k > p+1$ we have*

$$\mathcal{E}\{W^{-1}AW^{-1}\} = \frac{\Sigma^{-1}A\Sigma^{-1}}{(k-p+1)(k-p-1)} + \frac{[\operatorname{tr}(A\Sigma^{-1})]\Sigma^{-1}}{(k-p+1)(k-p)(k-p-1)}$$

This is Tague's complexification [264] of Styan's corollary 14 [262].

Proof. We start with the result for $V^{-1} \sim \mathbf{C}IW_p(k, I_p)$ where

$$\mathcal{E}\left\{V^{-1}BV^{-1}\right\} = d_1 B + d_2 I \operatorname{tr}(B)$$

Let $\Sigma = GG^H$. Then $W \sim \mathbf{C}W_p(k, \Sigma)$ implies $V = G^{-1}WG^{-H} \sim \mathbf{C}W_p(k, I)$ by theorem 54. Thus

$$\mathcal{E}\left\{G^H W^{-1} G B G^H W^{-1} G\right\} = d_1 B + d_2 I \operatorname{tr}(B)$$

Multiply this by G^{-H} and postmultiply by G^{-1}. We get

$$G^{-H}G^{H}\mathcal{E}\left\{W^{-1}GBG^{H}W^{-1}\right\}GG^{-1} = d_1G^{-H}BG^{-1} + d_2G^{-H}IG^{-1}\,\mathrm{tr}(B)$$

which we rewrite as

$$\mathcal{E}\left\{W^{-1}GBG^{H}W^{-1}\right\} = d_1G^{-H}BG^{-1} + d_2\Sigma^{-1}\,\mathrm{tr}(B)$$

Let $A = GBG^{H}$, which implies $B = G^{-1}AG^{-H}$. Then

$$\mathcal{E}\{W^{-1}AW^{-1}\} = d_1\Sigma^{-1}A\Sigma^{-1} + d_2\Sigma^{-1}\,\mathrm{tr}(G^{-1}AG^{-H})$$

$$= d_1\Sigma^{-1}A\Sigma^{-1} + d_2\Sigma^{-1}\,\mathrm{tr}(\Sigma^{-1}A)$$

since $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$. Recall that

$$d_1 = \frac{1}{[k-p+1]\,[k-p-1]}$$

and

$$d_2 = \frac{1}{[k-p+1]\,[k-p]\,[k-p-1]}$$

□

Corollary 27 *Let $W \sim \mathbf{C}W_p(k,\Sigma)$ and $k > p+1$. then*

$$\mathcal{E}\left\{\left(W^{-1}\right)^2\right\} = \mathcal{E}\left\{W^{-1}W^{-1}\right\}$$

$$= \frac{(\Sigma^{-1})^2}{[k-p+1]\,[k-p-1]} + \frac{\Sigma^{-1}\,\mathrm{tr}\,(\Sigma^{-1})}{[k-p+1]\,[k-p]\,[k-p-1]}$$

This is Styan's corollary 16 [262] which was complexified by Tague [264].

Proof. Let $A = I$ in theorem 96.

Corollary 28 *Let* $W \sim \mathbf{C}W_p(k, \Sigma)$, $a \in \mathbf{C}^p$ *and* $k > p + 1$. *Then*

$$\text{var}(W^{-1}a) = \frac{\Sigma^{-1}aa^H\Sigma^{-1}}{[k-p+1][k-p]^2[k-p-1]} + \frac{a^H\Sigma^{-1}a\Sigma^{-1}}{[k-p+1][k-p][k-p-1]}$$

This was done originally by Tague [264], motivated by the work of Styan [262]. Tague also produced results for the real variables case.

Proof. Define

$$\text{var}(W^{-1}) = \mathcal{E}\left\{W^{-1}aa^HW^{-1}\right\} - \mathcal{E}\left\{W^{-1}a\right\}\mathcal{E}\left\{a^HW^{-1}\right\}$$

In theorem 96, let $A = aa^H$. Then

$$\mathcal{E}\left\{W^{-1}aa^HW^{-1}\right\} = \frac{\Sigma^{-1}aa^H\Sigma^{-1}}{[k-p+1][k-p-1]} + \frac{a^H\Sigma^{-1}a\Sigma^{-1}}{[k-p+1][k-p][k-p-1]}$$

where

$$\text{tr}(aa^H\Sigma^{-1}) = \text{tr}(a^H\Sigma^{-1}a) = a^H\Sigma^{-1}a$$

which is a scalar. The numerator of the last term could also be $\Sigma^{-1}a\Sigma^{-1}a^H$.

$$\left|\mathcal{E}\left\{W^{-1}a\right\}\right|^2 = \mathcal{E}\left\{W^{-1}a\right\}\mathcal{E}\left\{a^HW^{-1}\right\} = \left[\mathcal{E}\left\{W^{-1}\right\}a\right]\left[a^H\mathcal{E}\left\{W^{-1}\right\}\right]$$

By theorem 87 then

$$\left[\mathcal{E}\left\{W^{-1}\right\}a\right]\left[a^H\mathcal{E}\left\{W^{-1}\right\}\right] = \frac{1}{[k-p]^2}\Sigma^{-1}aa^H\Sigma^{-1}$$

Note that

$$\frac{1}{[k-p+1][k-p-1]} - \frac{1}{[k-p]^2}$$

$$= \frac{k^2 - 2kp + p^2 - [k^2 - kp - k - kp + p^2 + p + k - p - 1]}{(k-p+1)(k-p)^2(k-p-1)}$$

$$= \frac{k^2 - 2kp + p^2 - k^2 + 2kp - p^2 + 1}{(k-p+1)(k-p)^2(k-p-1)} = \frac{1}{(k-p+1)(k-p)^2(k-p-1)}$$

The result follows from these pieces substituted back into $\text{var}(W^{-1}a)$. □

Corollary 29 *Let* $W \sim \mathbf{C}W_p(k, \Sigma)$ *and* $k > p+1$. *Then*

$$\mathcal{E}\left\{\text{tr}\left[\left(W^{-1}\right)^2\right]\right\} = \frac{\text{tr}\left[\left(\Sigma^{-1}\right)^2\right]}{[k-p+1][k-p-1]} + \frac{[\text{tr}\,(\Sigma^{-1})]^2}{[k-p+1][k-p][k-p-1]}$$

This is a variation by Tague [264] on Styan's corollary 16 [262].

Proof. $\mathcal{E}\left\{\text{tr}\left[(W^{-1})^2\right]\right\} = \text{tr}\left[\mathcal{E}\left\{W^{-1}W^{-1}\right\}\right]$. The result follows immediately from corollary 27. □

Corollary 30 *Let* $W \sim \mathbf{C}W_p(k, I)$ *and* $k > p+1$. *Then*

$$\mathcal{E}\left\{\left[\text{tr}\left(W^{-1}\right)\right]^2\right\} = \frac{p[p(k-p)+1]}{[k-p+1][k-p][k-p-1]}$$

This is a variation by Tague [264] on Styan's corollary 16 [262].

Proof.

$$\mathcal{E}\left\{\left[\text{tr}\left(W^{-1}\right)\right]^2\right\} = \mathcal{E}\left\{\left[\text{tr}\left(W^{-1}\right)\right]\left[\text{tr}\left(W^{-1}\right)\right]\right\}$$

$$= \mathcal{E}\left\{\left[\sum_{i=1}^{p} e_i^H W^{-1} e_i\right]\left[\sum_{j=1}^{p} e_j^H W^{-1} e_j\right]\right\} = \sum_{i=1}^{p}\sum_{j=1}^{p} e_i^H \mathcal{E}\left\{W^{-1} e_i e_j^H W^{-1}\right\} e_j$$

Note that

$$e_i^H \mathcal{E}\left\{W^{-1} e_i e_j^H W^{-1}\right\} e_j = \begin{cases} \mathcal{E}\left\{(W^{ii})^2\right\} & i = j \\ \mathcal{E}\left\{W^{ii} W^{jj}\right\} & i \neq j \end{cases}$$

$\mathcal{E}\left\{(W^{ii})^2\right\}$ occurs p times, and $\mathcal{E}\{W^{ii}W^{jj}\}$ occurs $p(p-1)$ times. Using equations F.3 and F.4, we get

$$\mathcal{E}\left\{\left[\mathrm{tr}\left(W^{-1}\right)\right]^2\right\} = \frac{p}{[k-p]\,[k-p-1]} + \frac{p(p-1)}{[k-p+1]\,[k-p-1]}$$
$$= \frac{p(k-p+1)+p(p-1)(k-p)}{(k-p+1)(k-p)(k-p-1)}$$

The numerator is simplified as follows.

$$pk - p^2 + p + p(pk - p^2 - k + p) = pk - p^2 + p + p^2k - p^3 - pk + p^2$$

$$= p + p^2k - p^3 = p + p^2(k-p) = p[1 + p(k-p)]$$

The result follows from this. □

Corollary 31 *Let $W \sim \mathbf{CW}_p(k, I)$ and $k > p+1$. Then*

$$\mathrm{var}\left[\mathrm{tr}\left(W^{-1}\right)\right] = \frac{kp}{[k-p+1]\,[k-p]^2\,[k-p-1]}$$

This is Styan's corollary 17 [262] which was complexified by Tague [264].

Proof.

$$\mathrm{var}\left[\mathrm{tr}\left(W^{-1}\right)\right] = \mathcal{E}\left\{\left[\mathrm{tr}\left(W^{-1}\right)\right]^2\right\} - \left[\mathcal{E}\left\{\mathrm{tr}\left(W^{-1}\right)\right\}\right]^2$$

By theorem 88, $\mathcal{E}\{\mathrm{tr}\,(W^{-1})\} = \frac{p}{k-p}$. Using corollary 30, we get

$$\mathrm{var}\left[\mathrm{tr}\left(W^{-1}\right)\right] = \frac{p[p(k-p)+1]}{(k-p+1)(k-p)(k-p-1)} - \frac{p^2}{(k-p)^2}$$

Looking at the numerator of the difference, we get

$$p[1 + p(k-p)](k-p) - p^2(k-p+1)(k-p-1)$$

$$= p(1 + pk - p^2)(k - p) - p^2(k - p + 1)(k - p - 1)$$

$$= p(k + pk^2 - p^2k - p - p^2k + p^3) - p^2(k^2 - pk - k - pk + p^2 + p + k - p - 1)$$

$$= pk + p^2k^2 - p^3k - p^2 - p^3k + p^4 - p^2k^2 + 2p^3k - p^4 + p^2 = pk$$

Placing this over the common denominator $(k-p+1)(k-p)^2(k-p-1)$ yields the result. □

F.5 Tague Example: Signal-to-Noise Ratio

Let $x(t) \in \mathbf{C}^p$ be a random output of a sensor array at time t, which is the sum of signal $s(t)$ passed through a narrowband beamformer and random noise $n(t)$. Explicitly, $x(t) = ds(t) + n(t)$. The complex vector $d \in \mathbf{C}^p$ of unit length is the narrowband steering vector. The random noise $n(t)$ is assumed to have distribution $\mathbf{C}N_p(0, R_N)$. The random signal $s(t)$ is assumed to have distribution $\mathbf{C}N_1(0, \sigma_s^2)$.

Consider a beamformer whose output $y(t)$ is given by $y(t) = w^H x(t)$ where $w = \hat{R}_N^{-1} d$, and

$$\hat{R}_N = \frac{1}{k} \sum_{m=1}^{k} n_m n_m^H$$

We assume that each noise measurement is mutually independent of the signal and other noise measurements. The solution for w is the optimum Wiener solution and $w = \hat{R}_N^{-1} d$ is the Wiener-Hopf equation. This is a common example, and it is discussed by Monzingo and Miller, Chapter 3 [185].

The problem we want to solve is to find the signal-to-noise ratio of the beamformer output. We begin by noting

$$y(t) = w^H x(t) = \left(\hat{R}_N^{-1} d\right)^H x(t) = \left(\hat{R}_N^{-1} d\right)^H [ds(t) + n(t)] \tag{F.6}$$

The expected value of the power at the beamformer output is

$$\mathcal{E}\left\{\|y(t)\|_2^2\right\} = \mathcal{E}\left\{y^H(t)y(t)\right\} \tag{F.7}$$

$$= \mathcal{E}\left\{\left[\left(\hat{R}_N^{-1} d\right)^H [ds(t) + n(t)]\right]^H \left[\left(\hat{R}_N^{-1} d\right)^H [ds(t) + n(t)]\right]\right\}$$

$$= \mathcal{E}\left\{\left[\left[d^H s^*(t) + n^H(t)\right]\left(\hat{R}_N^{-1} d\right)\right]\left[\left(\hat{R}_N^{-1} d\right)^H [ds(t) + n(t)]\right]\right\}$$

$$= \mathcal{E}\left\{d^H s^*(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} ds(t) + n^H(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} ds(t)\right. \tag{F.8}$$

$$\left. + d^H s^*(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} n(t) + n^H(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} n(t)\right\}$$

Now we invoke the assumption that $s(t)$ and $n(t)$ are statistically independent. Then $\mathcal{E}\left\{\|y(t)\|_2^2\right\}$

$$= \mathcal{E}\left\{s^*(t)s(t)\right\} d^H \mathcal{E}\left\{\hat{R}_N^{-1} dd^H \hat{R}_N^{-H}\right\} d + \mathcal{E}\left\{n^H(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} d\right\} \mathcal{E}\left\{s(t)\right\} \tag{F.9}$$

$$+ \mathcal{E}\left\{s^*(t)\right\} d^H \mathcal{E}\left\{\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} n(t)\right\} + \mathcal{E}\left\{n^H(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} n(t)\right\}$$

Observing that $\mathcal{E}\left\{s(t)\right\} = 0$ and likewise $\mathcal{E}\left\{s^*(t)\right\} = 0$, we simplify this to

$$\mathcal{E}\left\{\|y(t)\|_2^2\right\} = \mathcal{E}\left\{|s(t)|^2\right\} d^H \mathcal{E}\left\{\hat{R}_N^{-1} dd^H \hat{R}_N^{-H}\right\} d + \mathcal{E}\left\{n^H(t)\hat{R}_N^{-1} dd^H \hat{R}_N^{-H} n(t)\right\} \tag{F.10}$$

Recall that $s(t)$ has zero mean, and thus $\mathcal{E}\left\{s^*(t)s(t)\right\} = \sigma_s^2$. Also note that $n^H(t)\hat{R}_N^{-1} d$ and $d^H \hat{R}_N^{-H} n(t)$ are scalars and therefore commute. The quantity

$\hat{R}_N^{-1}d$ is a column vector, so $\hat{R}_N^{-1}dd^H\hat{R}_N^{-H} = \left\|\hat{R}_N^{-1}d\right\|_2^2$ is a matrix norm. Using the observations, we get

$$\mathcal{E}\left\{\|y(t)\|_2^2\right\} = \sigma_s^2 d^H \mathcal{E}\left\{\left\|\hat{R}_N^{-1}d\right\|_2^2\right\} d + \mathcal{E}\left\{d^H\hat{R}_N^{-H}n(t)n^H(t)\hat{R}_N^{-1}d\right\} \tag{F.11}$$

$$= \sigma_s^2 d^H \mathcal{E}\left\{\left\|\hat{R}_N^{-1}d\right\|_2^2\right\} d + d^H \mathcal{E}\left\{\left\|\hat{R}_N^{-H}n(t)\right\|_2^2\right\} d$$

We note that $k\hat{R}_N \sim \mathbf{C}W_p(k, R_N)$. We apply theorem 96 using $A = dd^H$. Thus

$$\mathcal{E}\left\{\frac{1}{k^2}\hat{R}_N^{-1}dd^H\hat{R}_N^{-H}\right\} = \mathcal{E}\left\{\frac{1}{k^2}\hat{R}_N^{-1}dd^H\hat{R}_N^{-1}\right\}$$

since $\hat{R}_N = \hat{R}_N^H$, and so we get

$$\mathcal{E}\left\{\frac{1}{k^2}\hat{R}_N^{-1}dd^H\hat{R}_N^{-1}\right\} = \frac{R_N^{-1}dd^HR_N^{-1}}{[k-p+1][k-p-1]} + \frac{[\mathrm{tr}(dd^HR_N^{-1})]R_N^{-1}}{[k-p+1][k-p][k-p-1]} \tag{F.12}$$

This implies

$$\mathcal{E}\left\{\hat{R}_N^{-1}dd^H\hat{R}_N^{-1}\right\} = \frac{k^2}{[k-p+1][k-p-1]}R_N^{-1}dd^HR_N^{-1} \tag{F.13}$$

$$+\frac{k^2}{[k-p+1][k-p][k-p-1]}[\mathrm{tr}(dd^HR_N^{-1})]R_N^{-1}$$

To perform the next step, note that $\mathrm{tr}(dd^HR_N^{-1}) = d^HR_N^{-1}d$ is a scalar. This allows us to say

$$d^H\,\mathrm{tr}(dd^HR_N^{-1})R_N^{-1}d = \mathrm{tr}(dd^HR_N^{-1})d^HR_N^{-1}d = (d^HR_N^{-1}d)^2 \tag{F.14}$$

Then

$$d^H\mathcal{E}\left\{\hat{R}_N^{-1}dd^H\hat{R}_N^{-1}\right\}d \tag{F.15}$$

$$= \frac{k^2}{[k-p+1][k-p-1]}(d^H R_N^{-1} d)^2 + \frac{k^2}{[k-p+1][k-p][k-p-1]}(d^H R_N^{-1} d)^2$$

Since R_N^{-1} is Hermitian we know $d^H R_N^{-1} d$ is real which implies $(d^H R_N^{-1} d)^2 = \left| d^H R_N^{-1} d \right|^2$. Note that

$$\frac{1}{[k-p+1][k-p-1]} + \frac{1}{[k-p+1][k-p][k-p-1]}$$

$$= \frac{k-p+1}{[k-p+1][k-p][k-p-1]} = \frac{1}{[k-p][k-p-1]}$$

Then

$$\sigma_s^2 d^H \mathcal{E}\left\{ \left\| \hat{R}_N^{-1} d \right\|_2^2 \right\} d = \sigma_s^2 (d^H R_N^{-1} d)^2 \frac{k^2}{[k-p][k-p-1]} \tag{F.16}$$

We evaluate the second remaining term of $\mathcal{E}\left\{ \|y(t)\|_2^2 \right\}$ in stages. This is the noise component.

$$d^H \mathcal{E}\left\{ \hat{R}_N^{-H} n(t) n^H(t) \hat{R}_N^{-1} \right\} d = d^H \mathcal{E}\left\{ \hat{R}_N^{-H} \left[\mathcal{E}\left\{ n(t) n^H(t) \right\} \right] \hat{R}_N^{-1} \right\} d$$

where we note $\hat{R}_N^{-1}$ is Hermitian and $n(t)$ is independent of the noise samples used to construct $\hat{R}_N$. Thus we get

$$d^H \mathcal{E}\left\{ \hat{R}_N^{-H} n(t) n^H(t) \hat{R}_N^{-1} \right\} d = d^H \mathcal{E}\left\{ \hat{R}_N^{-1} R_N \hat{R}_N^{-1} \right\} d$$

With $k > p+1$, we now apply theorem 96.

$$\mathcal{E}\left\{ \hat{R}_N^{-1} R_N \hat{R}_N^{-1} \right\} = \frac{k^2}{[k-p+1][k-p-1]} R_N^{-1} R_N R_N^{-1} \tag{F.17}$$

$$+ \frac{k^2}{[k-p+1][k-p][k-p-1]} \left[\mathrm{tr}\left(R_N R_N^{-1} \right) \right] R_N^{-1}$$

$$= \frac{k^2}{[k-p+1][k-p-1]} R_N^{-1} + \frac{k^2 p}{[k-p+1][k-p][k-p-1]} R_N^{-1}$$

where $\mathrm{tr}\left(R_N R_N^{-1}\right) = \mathrm{tr}\,(I) = p$

$$= \frac{k^2(k-p+p)}{[k-p+1][k-p][k-p-1]} R_N^{-1} = \frac{k^3}{[k-p+1][k-p][k-p-1]} R_N^{-1}$$

Therefore

$$d^H \mathcal{E}\left\{\hat{R}_N^{-1} R_N \hat{R}_N^{-1}\right\} d = \frac{k^3}{[k-p+1][k-p][k-p-1]} d^H R_N^{-1} d \qquad \text{(F.18)}$$

We now compute the signal-to-noise ratio.

$$\frac{\mathcal{E}\left\{\left|w^H ds(t)\right|^2\right\}}{\mathcal{E}\left\{\left|w^H n(t)\right|^2\right\}} = SNR = \frac{\sigma_s^2 d^H \mathcal{E}\left\{\left\|\hat{R}_N^{-1} d\right\|_2^2\right\} d}{d^H \mathcal{E}\left\{\hat{R}_N^{-1} R_N \hat{R}_N^{-1}\right\} d} \qquad \text{(F.19)}$$

$$= \frac{\frac{k^2}{[k-p][k-p-1]} \sigma_s^2 (d^H R_N^{-1} d)^2}{\frac{k^3}{[k-p+1][k-p][k-p-1]} (d^H R_N^{-1} d)}$$

$$= \frac{k-p+1}{k} \sigma_s^2 (d^H R_N^{-1} d)$$

Appendix G

ZONAL POLYNOMIAL COMMENTARY

G.1 History of Development

Zonal polynomials are special functions, just as Bessel, Legendre, Tschebyshev, Hankel, and trigonometric functions are special functions. Zonal polynomials are important to this work because they are used to evaluate a factor term in the probability density function of the sample eigenvalues of a Wishart matrix.

At first blush, zonal polynomials appear to the casual reader of this thesis to have *very little to do* with the content of this thesis. However, the fundamental contribution to advancing the order determination problem hinges on the existence and properties of zonal polynomials. The properties of these functions are still objects of current research. Muirhead [188] reports that zonal polynomials have (as of the time articles were written for the Encyclopedia of Statistical Sciences published in 1988) been defined only for symmetric matrices. He gives a suggestion of how to extend the definition to Hermitian matrices. Future progress in the small sample order determination problem must build upon these concepts.

The application of zonal polynomials to the related problem of finding the probability density function of sample eigenvalues was first made by A. T. James equation (94) [120] in 1964. James derived his result for the case of a

real Wishart matrix, and stated the result for the symmetric complex Wishart matrix by observing similarity of forms of other results. In 1987, Gross and Richards [96] derived zonal polynomials for the real, complex Hermitian, and quaternion cases simultaneously by studying invariants in a group representation setting.

Classical work in acoustic signal processing has been done using forms resulting in Hermitian Wishart matrices, which I have merely called Complex Wishart matrices. Indeed, much of my development could be recast in terms of Complex Symmetric Wishart matrices with accompanying background theory, but at the expense of losing use of properties of an inner product space and thus also access to the use of the concept of an adjoint. The very important contribution by Gross and Richards [96] justified the application of the *form* of the results for the joint density of sample eigenvalues previously written down by inspection by A. T. James [120]. The meaning of the *detail* of James' results is different.

Gross and Richards [96] point out that zonal polynomials are spherical functions for the Gelfand pair (G, K). In a general setting, spherical functions are studied by Helgason, Chapter IV [105]. Readers of Helgason or Gross and Richards will profit by first preparing a background in Lie theory.

Muirhead [187] develops zonal polynomials for the real variables case in a manner easily followed by engineers. Because the Laplacian operator is un-

conditionally applied in his development, the approach does not work for the complex variable case without further restrictions. Once the general development by Gross and Richards [96] is accepted which bypasses the problem of unconditionally applying the Laplacian, then Muirhead's results apply either directly or with consideration for differences in the real dimension and structure of the real and complex variables cases. Takemura [265] provided a 7 page development of complex zonal polynomials that relied heavily on an earlier development of real zonal polynomials in that monograph.

G.2 Gross and Richards' Development

This section is a review of the development of zonal polynomials done by Gross and Richards [96]. Although much of this discussion is directly from their paper, I have generally omitted proofs and ventured comments that would allow an engineer to more easily follow their paper, provided they have read the various surveys of algebra and analysis contained in this thesis. Part of the contribution here is in helping identify which spaces are objects of study at any particular point. I also attempt to highlight what is important, and to provide concrete examples at various points.

Gross and Richards' work is very important. In one treatment, they develop zonal polynomials of matrix argument for real, complex Hermitian, and quaternion fields. Terminology used by them comes out of the study of Lie

groups and group representation theory. The mathematical dictionary of most frequent value in reading this paper is the one edited by Itô [115], published by MIT Press in four volumes, beginning in 1987. A new mathematical dictionary that has volumes through "Sp" published (1992) that is excellent is a translation of Vinogradov's *Soviet Mathematical Encyclopaedia*, edited by Hazewinkel [104]. Smaller dictionaries have proven to not be useful for reading Gross and Richards' work. We are still in need of a dictionary to translate the technical language of algebraists into the language that engineers understand, and vice versa.

We are interested in group representation theory because it allows us to connect a practical result we need with an intuitive abstraction about the nature of the problem we are dealing with. This approach allows us to understand properties of our problem that otherwise may escape notice.

We need to evaluate $\text{etr}(-\Sigma^{-1}A)$ in the development of the joint density of the eigenvalues of A. We have gotten to another form whose evaluation will get us closer to the answer we need. It is

$$\int_{\mathbf{U}(p)} \text{etr}(-\Sigma^{-1}U^H AU)(dU) \tag{G.1}$$

Our journey will lead us to express the exponential in terms of zonal polynomials. When this is done, we can take advantage of a splitting, or decomposition, property that separates into the product of a function of $(-\Sigma^{-1})$ times the same function of A.

In examining our trace function, we observe that its value depends only on the sum of the eigenvalues of the argument of the trace function. If we looked at the set of all polynomial functions of the argument and attempt to select an expression for the trace function, we find that we are dealing with concepts of invariant spaces. This is our clue to consider the wonderful world of group representation theory. This leads to the development of zonal polynomials which form the basis for the space of polynomials we are interested in.

Gross and Richards [96] begin their development of zonal polynomials by considering the structure of the algebra of polynomials defined on the General Linear Group $G = GL(n, F)$ consisting of all nonsingular $n \times n$ matrices whose elements are taken from the field F. This field F may be real ($\mathbf{R}$), Complex ($\mathbf{C}$), or Quaternion ($\mathbf{H}$). This set of polynomials is identified by the symbol $\mathbf{P}(G)$. Pay close attention to the various modifications to this notation to indicate different sets. Group representation theory first comes into play by the definition of the function R, that takes its argument from group G, and acts as a transformation on the linear space $\mathbf{P}(G)$. R is defined by the action

$$R(a)\varphi(x) = \varphi(xa) \tag{G.2}$$

where $\varphi \in \mathbf{P}(G)$ and $a, x \in G$. R is called the *right regular representation* of group G on the linear space $\mathbf{P}(G)$. Note that

$$R(b)[R(a)\varphi(x)] = R(b)\varphi(xa) = \varphi(xab) = R(ab)\varphi(x) \tag{G.3}$$

Thus $R(ab) = R(b)R(a)$. Since a function obeying $f(xy) = f(x)f(y)$ is

called a "homomorphism", we will call a function with the opposite order $f(xy) = f(y)f(x)$ a "heteromorphism" (a term first used by Tom Concannon, 1989). We note that Vilenkin [271] defines a group representation as being a homomorphism, but careful following of his nonabelian examples show them to be heteromorphisms. The structure of group representation theory can be recast as heteromorphisms without damage to its beauty.

Gross and Richards partition $\mathbf{P}(G)$ into sets of polynomials $\mathbf{P}_d(G)$ defined on G that are homogeneous of degree d. Thus

$$\mathbf{P}(G) = \bigoplus_{d=0}^{\infty} \mathbf{P}_d(G) \tag{G.4}$$

They define a scalar product on $\mathbf{P}(G)$ defined in terms of a differential operator. Using this, they show that the subspaces $\{\mathbf{P}_d(G)\}$ are mutually orthogonal and together span $\mathbf{P}(G)$. They avoid nagging issues of differentiability in the field of definition of the argument by instead considering the fields in their isomorphic real fields. They do not therefore require their polynomial functions to be holomorphic. They selected the inner product

$$< \psi, \varphi >= D(\bar{\psi})\varphi(x)\,|_{x=0} \tag{G.5}$$

because it has the property of forming a weighted sum of products of coefficients having the same product of indeterminates, raised to identical powers.

Zonal polynomials are defined as homogeneous harmonic polynomials on the surface of a sphere. Assumption of differentiability is routine and is not

an issue of central concern for real variables. It is a main concern for us. We note that polynomials in z are differentiable with respect to z. Problems arise with polynomials that include terms such as z^*.

Suppose, instead, that we treat polynomials as N-tuples in the way defined by Broida and Williamson (p. 253) [47]. Then a polynomial is represented as a vector of infinite length with N entries being non-zero. One may then define an inner product on the vector of the coefficients of a polynomial.

Define a compound index α the same way Gross and Richards did.

$$\alpha \underset{=}{\text{def}} (\alpha_1, \alpha_2, \cdots \alpha_N) \tag{G.6}$$

We let a term of a polynomial of N variables $a_\alpha X^\alpha$ be given by

$$a_\alpha X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_N^{\alpha_N} \tag{G.7}$$

We can establish a collating sequence for $(\alpha_1, \alpha_2, \cdots \alpha_N)$ to linearize our multi-dimensional array of coefficients $\{a_\alpha\}$. For some fixed d, we can use a counting sequence to establish an ordering for all α such that

$$|\alpha| \underset{=}{\text{def}} \alpha_1 + \alpha_2 + \cdots + \alpha_N = d \tag{G.8}$$

where $\alpha_k \geq 0$ is an integer. For example, let $b = d+1$ be the base of a number system. Then an ordering of $|\alpha| = d$ can be given by the number $\sum_{k=1}^{N} \alpha_k b^{k-1}$. For fixed N and fixed d, there are $\binom{N-1+d}{N-1}$ elements.

Define

$$\alpha! \underset{=}{\text{def}} (\alpha_1!)(\alpha_2!) \cdots (\alpha_N!) \tag{G.9}$$

We may define the scalar product of two polynomials $\psi = \sum_{\alpha} b_\alpha X^\alpha$ and $\varphi = \sum_{\alpha} a_\alpha X^\alpha$ by

$$< \psi, \varphi >= \sum_{\alpha} \alpha!\, b_\alpha^* a_\alpha \tag{G.10}$$

I have switched the standard mathematician's order of arguments in the inner product to make use of the notation common to engineers with vectors where $< x, y >= x^H y$. Note that as long as $N < \infty$, I could just as easily define the inner product by

$$< \psi, \varphi >= \sum_{\alpha} b_\alpha^* a_\alpha \tag{G.11}$$

since the finiteness guarantees convergence. I chose to retain the $\alpha!$ to maintain the same notation used by Gross and Richards. Note that I explicitly have *not* used an operator that is necessarily a differential operator. This scalar product obeys the properties of an inner product. Let

$$\|\varphi\| =< \varphi, \varphi >^{\frac{1}{2}} \tag{G.12}$$

be the inner product space norm of φ. Define the distance between ψ and φ by

$$d(\psi, \varphi) = \|\psi - \varphi\| \tag{G.13}$$

We now have a metric space.

The purpose of the inner product defined by Gross and Richards was to define orthogonality. The inner product defined above provides the same orthogonality results. They use the inner product to demonstrate the property

that $\{\mathbf{P}_d(G)\}$ is a set of mutually orthogonal spaces of homogeneous polynomials that together span $\mathbf{P}(G)$. The Broida and Williamson construct permits the desired observation without grappling with differentiability. Numerically, there is no difference between the inner product definitions. Weighting schemes not using the $\alpha!$ term in the sum can also produce the required properties. The requirement is to obtain a convergent series, particularly when N becomes unbounded, and to maintain satisfaction of the properties of an inner product. This means that differentiability, and thus harmonicity, are not required properties of the polynomials developed by Gross and Richards. Recall that zonal polynomials are characterized by the statement that they are homogeneous harmonic polynomials defined on the surface of an n-dimensional sphere. Homogeneous polynomials defined on the surface of an n-dimensional sphere that are harmonic are special cases of the set of polynomials developed by Gross and Richards. The property of harmonicity is an additional benefit when you select $\alpha!$ as the weighting term because this permits interpretation of the inner product as a differential operator as done by Gross and Richards. Their properties for representations R and τ continue to hold when weights are selected to produce a finite-valued inner product.

Let K be a maximal compact subgroup of G. When $F = \mathbf{C}$, then $K = \mathbf{U}(n)$ is the set of unitary $n \times n$ matrices

$$K = \{k : kk^H = I_n,\ k \in G) \tag{G.14}$$

Gross and Richards use the algebraist's convention of using 1 rather than I to denote the identity element under multiplication. They show that the right regular representation R of group G acting on the linear space of polynomials $\mathbf{P}(G)$ is unitary when the argument of R is restricted to elements of $k \in K \subset G$. Thus

$$< R(k)\psi, R(k)\varphi >=< \psi, \varphi > \tag{G.15}$$

Gross and Richards construct their practical result, which is a set of polynomials which they know how to compute. Let

$$vcu = x \tag{G.16}$$

be the LDU decomposition of x when x is represented by a square $n \times n$ matrix. See Stewart (p.132) [259] for a discussion of the LDU decomposition. The set of lower triangular matrices with ones on the main diagonal is V. We note that $v \in V$. The set of upper triangular matrices with ones on the main diagonal is U. We note that $u \in U$. Let $(\Omega, \oplus, \odot)$ be a ring with identity elements $e_\oplus$ and $e_\odot$. Let $\omega \in \Omega$. Then element ω of the ring is called *nilpotent* of order k if k is the smallest positive integer such that

$$\omega^k = \underbrace{\omega \odot \omega \odot \ldots \odot \omega}_{k \text{ times}} = e_\oplus$$

is the additive identity element of the ring. If you removed the ones from the main diagonals of v and u, then the new elements constructed from v and u would be nilpotent. The product of any n such nilpotent elements constructed

from elements of V, which we call $(V - I)$, or the product of any n such nilpotent elements constructed from elements of U, which we call $(U - I)$, is the zero matrix. A linear transformation is called *unipotent* if it has the form $I + A$ where A is nilpotent [116]. Thus, V and U are unipotent. The set of $n \times n$ nonsingular diagonal matrices is C. Let $c \in C$. The triple (U, C, V) is called the *standard bitriangular structure* for G.

Let

$$m = (m_1, m_2, \cdots m_n) \tag{G.17}$$

such that

$$m_1 \geq m_2 \geq \cdots \geq m_n \geq 0 \tag{G.18}$$

The set of polynomials $\mathbf{P}^{2m}(G)$ is defined as a set of all polynomials having the property

$$\varphi(vcx) = \mu_{2m}(c)\varphi(x) \tag{G.19}$$

where

$$\mu_{2m}(c) = |c_1|^{2m_1} |c_2|^{2m_2} \cdots |c_n|^{2m_n} \tag{G.20}$$

This function, $\mu_{2m}(c)$, is called the *character* of C. The set $\mathbf{P}^{2m}(G)$ is invariant under right translation by G. Let π_{2m} be R when R is restricted to being applied only to $\varphi \in \mathbf{P}^{2m}$. Then

$$\pi_{2m}(a)\varphi(x) = \varphi(xa) \tag{G.21}$$

describes the action of the right regular representation of G on the linear space $\mathbf{P}^{2m}(G)$.

Consider the structural relationship between the leading principal submatrices of an LDU decomposition and the corresponding leading principal submatrix of the original matrix x. Looking at an example will provide a foundation for understanding some definitions and properties to follow.

$$\left(\begin{array}{ccccccc} 1 & \vdots & 0 & \vdots & 0 & \vdots & 0 \\ \cdots & & & \vdots & & \vdots & \\ v_{21} & & 1 & \vdots & 0 & \vdots & 0 \\ \cdots & \cdots & \cdots & & & \vdots & \\ v_{31} & & v_{32} & & 1 & \vdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & & \\ v_{41} & & v_{42} & & v_{43} & & 1 \end{array}\right)\left(\begin{array}{ccccccc} c_1 & \vdots & & \vdots & & \vdots & \\ \cdots & & & \vdots & & \vdots & \\ & & c_2 & \vdots & & \vdots & \\ \cdots & \cdots & \cdots & & & \vdots & \\ & & & & c_3 & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & & \\ & & & & & & c_4 \end{array}\right) \times$$

$$\times \left(\begin{array}{ccccccc} 1 & \vdots & u_{12} & \vdots & u_{13} & \vdots & u_{14} \\ \cdots & & & \vdots & & \vdots & \\ 0 & & 1 & \vdots & u_{23} & \vdots & u_{24} \\ \cdots & \cdots & \cdots & & & \vdots & \\ 0 & & 0 & & 1 & \vdots & u_{34} \\ \cdots & \cdots & \cdots & \cdots & \cdots & & \\ 0 & & 0 & & 0 & & 1 \end{array}\right) \tag{G.22}$$

$$= \begin{pmatrix} c_1 & \vdots & c_1u_{12} & \vdots & c_1u_{13} & \vdots \\ \cdots & & & \vdots & & \vdots \\ v_{21}c_1 & & v_{21}c_1u_{12} + c_2 & \vdots & v_{21}c_1u_{13} + c_2u_{23} & \vdots \\ \cdots & \cdots & \cdots & & & \vdots \\ v_{31}c_1 & & v_{31}c_1u_{12} + v_{32}c_2 & & v_{31}c_1u_{13} + v_{32}c_2u_{23} + c_3 & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ v_{41}c_1 & & v_{41}c_1u_{12} + v_{42}c_2 & & v_{41}c_1u_{13} + v_{42}c_2u_{23} + v_{43}c_3 & \end{pmatrix} \tag{G.23}$$

$$\begin{pmatrix} c_1u_{14} \\ \\ v_{21}c_1u_{14} + c_2u_{24} \\ \\ v_{31}c_1u_{14} + v_{32}c_2u_{24} + c_3u_{34} \\ \\ v_{41}c_1u_{14} + v_{42}c_2u_{24} + v_{43}c_3u_{34} + c_4 \end{pmatrix}$$

What you need to notice here is the partitioning of the matrix into succeedingly smaller matrices anchored in the upper left corner. These matrices are called *leading principal submatrices.* Let A_k denote the k^{th} leading principal submatrix of square matrix A. Then for the LDU decomposition of x given by $vcu = x$, we observe that it is also true that $v_kc_ku_k = x_k$. Define

$$\Delta_k(x) = (\det x_k)^\eta \tag{G.24}$$

where $\eta = 1$ if $F = \mathbf{R}$ or $F = \mathbf{C}$, and $\eta = \frac{1}{2}$ for $F = \mathbf{H}$. With this notation, let

$$\varphi_{2m}(x) = |\Delta(x)|^{2m_n} \prod_{j=1}^{n-1} |\Delta_j(x)|^{2(m_j - m_{j+1})} \geq 0 \qquad \text{(G.25)}$$

Note that $\varphi_{2m} \in \mathbf{P}^{2m}(G)$.

φ_{2m} and π_{2m} turn out to have special properties. π_{2m} is an *irreducible representation* of G on $\mathbf{P}^{2m}(G)$. π_{2m} is the subrepresentation of R on the subspace $\mathbf{P}^{2m}(G)$. The function φ_{2m} has a group theoretic definition as *the highest weight vector of* π_{2m}. φ_{2m} is the element of $\mathbf{P}^{2m}(G)$, unique up to scalar multiples, for which

$$\pi_{2m}(cu)\,\varphi_{2m}(x) = \mu_{2m}(c)\,\varphi_{2m}(x) \qquad \text{(G.26)}$$

for all $(c, u) \in C \times U$. From this we know that $\mathbf{P}^{2m}(G)$ is the span of right translates of φ_{2m} under group G. This form looks like the familiar equation $Ax = \lambda x$ that defines the eigenvalues and eigenvectors of A.

We observe that $\mathbf{P}^{2m}(G) \subset \mathbf{P}_d(G)$ where $d = 2\,|m|$.

Gross and Richards also engage in intuitive abstraction. Recall that in our motivating problem we are dealing with the trace function. The trace function has the property that it yields the sum of eigenvalues of the matrix argument. Further, we are concerned with the trace of the product of Hermitian matrices. Let A and B be Hermitian positive definite. Matrix A may then be factored as $A = CC^H$. By a property of the trace function,

$$\mathrm{tr}(BA) = \mathrm{tr}(BCC^H) = \mathrm{tr}(C^H BC)$$

Matrix $D = C^H BC$ is also Hermitian. Thus, we are interested in properties of $\mathrm{tr}(s)$ where $s \in S$ and S is the set of Hermitian matrices. We know that somehow we need to work with invariant subspaces.

Let

$$I(G) = \{p(kx) = p(x)\} = \mathbf{P}(G)^K \tag{G.27}$$

be the set of all polynomials on G that are left-invariant under translation by an element of K. Let $I(G)_d \subset I(G)$ be the set of nonzero left K-invariant polynomials that are homogeneous of degree d. Note that $-I \in K$ since $(-I)(-I)^H = I$. Thus $p(-x) = p(x)$, which implies that only polynomials homogeneous of *even* degree are nonzero in $I(G)$. Therefore

$$I(G) = \bigoplus_d I(G)_{2d} \tag{G.28}$$

Let $\varphi \in \mathbf{P}(G)$. Define the spherical transformation $\Xi : \varphi \to \varphi^\#$ by

$$\varphi^\#(x) = \int_K \varphi(kx)dk \tag{G.29}$$

for all $x \in G$. This is the orthogonal projection of $\mathbf{P}(G)$ onto $I(G)$. The key observation by Gross and Richards that links the practical with the abstract is their Theorem 3.4 (presented next), which relies on Schur's lemma. It is the ordering of the $\{m_i\}$ in this theorem that establishes the ordering of the eigenvalues in the density function used for theorem 70.

Theorem 97 *Let $d \geq 0$ and $m = (m_1, \cdots, m_n)$ with $m_1 \geq \cdots \geq m_n \geq 0$ and $|m| = d$. Then*

1. The restriction Ξ_{2m} of Ξ to $\mathbf{P}^{2m}(G)$ is an isomorphism of $\mathbf{P}^{2m}(G)$ onto a subspace $I^{2m}(G)$ of $I(G)$, and

2.

$$I(G)_{2d} = \bigoplus_{|m|=d} I^{2m}(G) \tag{G.30}$$

is the decomposition of $I(G)_{2d}$ into irreducible subspaces.

Each space

$$I^{2m}(G) = \Xi(\mathbf{P}^{2m}(G)) \tag{G.31}$$

is an irreducible right invariant subspace of $I(G)$. The dimension of $I(G)_{2d}$ is

$$\binom{N-1+d}{N-1} = \sum_{|m|=d} \deg \pi_{2m} \tag{G.32}$$

Let ρ be the subrepresentation of R on the subspace $I(G)$ of $\mathbf{P}(G)$. Then

$$\rho_{2m}(a)\,\varphi^{\#}(x) = \Xi_{2m}\,\pi_{2m}(a)\,\Xi_{2m}^{-1}\,\varphi^{\#}(x) \tag{G.33}$$

ρ_{2m} is the irreducible representation of signature $2m$ that acts by right translation on the space $I^{2m}(G)$. Note that this looks like a basis change.

The most important fact Gross and Richards highlight is the relationship between $I^{2m}(G)$ and $\mathbf{P}^{2m}(G)$. Because we know how to compute $\varphi \in \mathbf{P}^{2m}(G)$, we can find the corresponding $\varphi^{\#} \in I^{2m}(G)$.

Let $\mathbf{P}(S)$ be the algebra of all polynomials on the set of Hermitian matrices $S = \{x = x^H\}$. Let $\mathbf{P}_d(S)$ be the subspace of $\mathbf{P}(S)$ of polynomials homogeneous of degree d. Let $\mathbf{P}(S)$ use the same inner product used on $\mathbf{P}(G)$.

Gross and Richards define a different representation on S, selected to preserve the structure of the argument of $q \in \mathbf{P}(S)$. Let

$$\tau(a)\, q(s) = q(a^H s a) \tag{G.34}$$

for $a \in G,\ s \in S$. Note that equation G.34 is slightly different in form than the one presented by Gross and Richards in their equation 4.2(2). When the argument of τ is restricted to K then τ is unitary. This means

$$< \tau(k) q_1(s), \tau(k) q_2(s) > = < q_1(s), q_2(s) > \tag{G.35}$$

Define a mapping $\Omega : I(G) \rightarrow \mathbf{P}(S)$ by

$$p(x) = q(x^H x) \tag{G.36}$$

for $p \in I(G)$ and $q \in \mathbf{P}(S)$. Note that

$$\Omega\, \rho(a)\, p(x) = \tau(a)\, \Omega\, p(x) = q(a^H x^H x a) \tag{G.37}$$

for all $a \in G$. We also can see $\Omega : I(G)_{2d} \rightarrow \mathbf{P}_d(S)$. Ω is an isomorphism. If we call the restriction of Ω to $I^{2m}(G)$ by the notation Ω_{2m}, then

$$\mathbf{P}^m(S) = \Omega_{2m}(I^{2m}(G)) \tag{G.38}$$

and

$$\tau_{2m}(a)\, q(s) = \Omega_{2m}\, \rho_{2m}(a)\, \Omega_{2m}^{-1}\, q(s) \tag{G.39}$$

τ_{2m} is the irreducible representation of G with signature $2m$ acting in subspace $\mathbf{P}^m(S)$ of $\mathbf{P}_d(S)$, and we get

$$\mathbf{P}_d(S) = \bigoplus_{|m|=d} \mathbf{P}^m(S) \tag{G.40}$$

and

$$\tau = \bigoplus_{|m|=d} \tau_{2m} \tag{G.41}$$

If we let

$$q_m(s) = \Delta(s)^{m_n} \prod_{j=1}^{n-1} \Delta_j(s)^{m_j - m_{j+1}} = \varphi_{2m}(x) \tag{G.42}$$

where $s = x^H x$, $|m| = d$, then q_m is homogeneous of degree d and

$$\tau(cu)\, q_m = \mu_{2m}(c)\, q_m$$

Note that $q_m \in \mathbf{P}^m(S)$. It is the *highest weight vector* of τ_{2m}.

Since we have an isomorphism between $\mathbf{P}^m(S)$ and $I^{2m}(G)$, we know there exists some K-invariant polynomial in $\mathbf{P}^m(S)$. Let

$$f_m(s) = \int_K q_m(k^H s k)\, dk \tag{G.43}$$

This is similar to the spherical transformation we did earlier. This f_m is unique up to constant multiples in $\mathbf{P}^m(S)$. Note that

$$f_m(k^H s k) = f_m(s) \tag{G.44}$$

for all $s \in S$ and $k \in K$. Gross and Richards point out that since any Hermitian matrix s can be diagonalized by some element $k \in K$, then f_m is uniquely determined by its restriction to diagonal matrices in S, which all have real entries. As a polynomial in the n diagonal entries which is homogeneous of degree d, f_m is invariant under the action of the symmetric group on n letters. The dimension of the space of such polynomials is the number of partitions

m of the positive integer d. There can be only one linearly independent K-invariant member of $\mathbf{P}^m(S)$.

Definition 8 *A non-zero K-invariant polynomial f_m in $\mathbf{P}^m(S)$ is called a* zonal polynomial *of S of weight m.*

These zonal polynomials of different signatures m form a basis $\{f_m\}$ for the set of K-invariant polynomials on S, $\mathbf{P}(S)^K$. Thus

$$\mathbf{P}(S)^K = \bigoplus_m c_m f_m \tag{G.45}$$

Gross and Richards use the convention by workers in analysis that constants are subsumed into a general constant at each step of a derivation without indexing or other distinguishment. Gross and Richards point out that zonal polynomials corresponding to different signatures are orthogonal. They are in different subspaces.

Recall that I used a definition for inner product different than Gross and Richards. They used the properties of inner products to demonstrate some important properties of zonal polynomials in their Lemma 5.2. For this reason, a proof of Lemma 5.2 is given with the new inner product. Their Lemma 5.2 remains valid.

Theorem 98 *For any $d \geq 0$,*

$$(\operatorname{tr} s)^d = \sum_{|m|=d} \alpha_m f_m(s)$$

where

$$\alpha_m = (d!)\,\|f_m\|^{-2} > 0$$

for all m. This is a modified Gross and Richards Lemma 5.2.

Proof. $(\mathrm{tr}\, s)^d$ is homogeneous of degree d, so it is a linear combination of all $f_m(s)$ for $|m| = d$, and thus

$$(\mathrm{tr}\, s)^d = \sum_{|m|=d} \alpha_m f_m(s) \tag{G.46}$$

for some suitable choices of α_m.

By definition of the exponential function, we know

$$\exp[\mathrm{tr}\, s] = \sum_{d=0}^{\infty} \frac{1}{d!}(\mathrm{tr}\, s)^d = \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|n|=d} \alpha_n f_n(s) \tag{G.47}$$

Consider the inner product $< f_m(s), e^{\mathrm{tr}\, s} >$.

$$< f_m(s), \exp[\mathrm{tr}\, s] >= \left\langle f_m(s),\ \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|n|=d} \alpha_n\, f_n(s) \right\rangle \tag{G.48}$$

$$= \sum_{d=0}^{\infty} \frac{1}{d!} \sum_{|n|=d} \alpha_n < f_m, f_n >= \frac{1}{d!} \alpha_m \|f_m\|^2$$

where $d = |m|$.

Suppose we normalize the coefficients of polynomial f_m so that

$$\frac{1}{d!} \alpha_m \|f_m\|^2 = 1 \tag{G.49}$$

If we do this, then

$$\alpha_m = d!\,\|f_m\|^{-2} > 0 \tag{G.50}$$

We know $\alpha_m > 0$ because $\|f_m\| > 0$ unless $f_m = 0$ for all s, and $d! \neq 0$.□

Definition 9 *We define Z_m to be a zonal polynomial with a different normalization, so that $Z_m = \alpha_m f_m$ where $\alpha_m = d! \, \|f_m\|^{-2}$.*

This means

$$(\operatorname{tr} s)^d = \sum_{|m|=d} Z_m(s) \tag{G.51}$$

and also

$$Z_m(k^H s k) = Z_m(s) \tag{G.52}$$

which means that Z_m is a function only of the eigenvalues of s.

The last result needed from Gross and Richards' paper is their Proposition 5.5.

Proposition 41 *For any $s, t \in S$,*

$$\int_K Z_m(sk^{-1}tk)dk = \frac{Z_m(s)Z_m(t)}{Z_m(I_n)}$$

where dk is the normalized Haar measure on K.

The integral is known as the "splitting property" for zonal polynomials. In K we know since $k^H k = I$ that $k^{-1} = k^H$. The integral can be written as

$$\int_K Z_m(sk^{-1}tk)\, dk \tag{G.53}$$

The proof is done by showing

$$\int_K p(ykx^H)\, dk = \frac{p(x)\, p(y)}{p(I)} \tag{G.54}$$

where $p = Z(x^H x)$. Note in following Gross and Richards' proof that all functions in $I(G)^{2m}$ are also *left* K-invariant as well as *right* K-invariant. Thus, as a function of x, then $\int_K p(ykx^H)\, dk$ is a left K-invariant element of $I(G)^{2m}$.

This splitting property, for the case of complex variables, fills in the steps that justify James equation (92) [120], and is the complex analog of his equation (23). The Z_m of Gross and Richards is the C_m of James.

Appendix H

SOME GROUP THEORY

The purpose of this appendix is to provide a minimal background for concepts from group theory, group representation theory, and topological group theory necessary to follow the material used in development of zonal polynomials. This body of theory is not in the usual preparation of acousticians, engineers, or statisticians, yet the future of acoustic signal processing is grounded in these concepts. The material presented here is barely enough to provide some basic definitions. Fluent use of these concepts requires a 2-3 course graduate sequence. With judicious topic selection, an applied graduate course could be constructed for engineers that could be learned in one semester. We conclude with an example which establishes some group invariance properties of the vector complex normal distribution which justifies our use of the zonal polynomial approach.

H.1 Basic Group Theory

Definition 10 *A* group G is a set G, *together* with an operator $\Box$, *that obeys the rules below.*

1. $\Box$ is a *binary* operator such that if $a \in G$ and $b \in G$, then $a \Box b \in G$.

2. $\square$ is an *associative* operator. If $a,\ b,\ c \in G$, then

$$(a \square b) \square c = a \square (b \square c)$$

Thus, it makes sense to write $a \square b \square c$.

3. There is an element $e_\square \in G$ such that for all $a \in G$ we have

$$a \square e_\square = e_\square \square a = a$$

It is a theorem of group theory that there is only one such element in G. We call this element the *identity element* of $(G, \square)$.

4. For each element $a \in G$, there is an element $b \in G$ such that

$$a \square b = b \square a = e_\square$$

It is a theorem of group theory that for any element $a \in G$, there is only one element $b \in G$ for which this is true. We call b the *inverse* of a, and we write $b = a^{-1}$.$\diamondsuit$

To remind us of this association, we can denote it by $(G, \square)_g$. The subscript g identifies $(G, \square)$ as a group. When the context is unambiguously referring to the group, the notation may be simplified to $(G, \square)$, or simply G. The symbol $\square$ was chosen to decouple our normal concepts of operators so that we can more easily think of general operators. The concepts of a group extend well beyond our usual addition of real numbers, or multiplication on the set

of real numbers when the zero element is removed. You may choose any other symbol, and the rules still apply. Another definition is important to this thesis is that of a subgroup.

Definition 11 *Let* $(G, \square)$ *be a group. Then* $(H, \square)$ *is a* subgroup *of* $(G, \square)$ *if* $H \subset G$ *and if* $(H, \square)$ *is a group.*

H.2 Group Representation Theory

H.2.1 Group Representation Definition

Definition 12 *A* representation *of group* $(G, \circ)$ *is some other group* $(B, \square)$ *related to* $(G, \circ)$ *by some homomorphism* φ. *Thus,* $\varphi(g)$ *is a representation of* $g \in G$ *if*

$$\varphi(g \circ h) = \varphi(g) \square \varphi(h)$$

for all $g, h \in G$, *where* $\varphi : G \rightarrow B$, $\varphi(g) \in B$, *and* $\varphi(h) \in B$.

Let $I \in B$ be the identity element in $(B, \square)$, and let $e \in G$ be the identity element in $(G, \circ)$. Then

$$\varphi(g) = \varphi(e \circ g) = \varphi(e)\varphi(g)$$

$$\varphi(g)\varphi^{-1}(g) = I = \varphi(e) \square \varphi(g) \square \varphi^{-1}(g) = \varphi(e)$$

Therefore $I = \varphi(e)$ means that the identity in $(G, \circ)$ maps to the identity in $(B, \square)$. A related result is

$$I = \varphi(e) = \varphi(g \circ g^{-1}) = \varphi(g) \square \varphi(g^{-1})$$

$$\varphi^{-1}(g) = \varphi^{-1}(g) \square I = \varphi^{-1}(g) \square \varphi(g) \square \varphi(g^{-1}) = \varphi(g^{-1})$$

Therefore $\varphi^{-1}(g) \in B$ is mapped from $g^{-1} \in G$.

H.2.2 Homomorphism Familiar Examples

A common example of a group homomorphism is

$$\exp(x + y) = \exp(x) \cdot \exp(y) \tag{H.1}$$

In this example, x and y belong to the set of complex numbers which forms a group structure using ordinary addition as the group operator. The set of numbers $\{e^x\}$ are complex numbers without zero, which forms a group under ordinary multiplication. In this example, $\varphi(x) = e^x$. Saracino (pp.106-108) [231] gives other examples.

Let $(A, \circ)$ be the group of invertible square matrices of complex numbers, $GL(n, \mathbf{C})$. Then $\det(a)$ defines a group homomorphism from $(A, \circ)$ to the set of complex numbers (except zero) under multiplication. Note that

$$\det(a_1 \cdot a_2) = \det(a_1) \times \det(a_2) \tag{H.2}$$

Let $(B, \oplus)$ be the group of square matrices of complex numbers with matrix addition as the group operator. Then $\mathrm{tr}(b)$ defines a group homomorphism

from $(B, \oplus)$ to the set of complex numbers (including zero) under addition. Note that

$$\mathrm{tr}(b_1 \oplus b_2) = \mathrm{tr}(b_1) + \mathrm{tr}(b_2) \tag{H.3}$$

H.2.3 Homomorphism Theorems

The structure imposed by group homomorphisms permit many powerful and useful insights. A listing of theorems (without proof) from Saracino [231] is presented below.

Theorem 99 *Let $\varphi : G \to H$ and $\psi : H \to K$ be homomorphisms. Then $\psi \circ \varphi : G \to K$ is a homomorphism, where $\circ$ is composition of functions. This is Saracino theorem 12.1(i).*

Theorem 100 *Let $\varphi : G \to H$ be a homomorphism. Then*

1. $\varphi(e_G) = e_H$ where e_G and e_H are the identity elements in groups G and H.

2. For any $x \in G$ and any integer n, then $\varphi(x^n) = [\varphi(x)]^n$.

3. The notation $o(x) = n$ means the order of element x is n. This means $x^n = e_G$. If $o(x) = n$, then $o[\varphi(x)]$ divides n.

This is Saracino theorem 12.4.

Theorem 101 *Let $\varphi : G \to K$ be a homomorphism. If H is a subgroup of G, then $\varphi(H)$ is a subgroup of K.*

$$\varphi(H) = \{k \in K \mid k \text{ is } \varphi(h) \text{ for some } h \in H\}$$

This is Saracino theorem 12.6(i).

Definition 13 *Let H be a subgroup of G. Then H is called a* normal *subgroup if $ghg^{-1} \in H$ for every $g \in G$ and for every $h \in H$. We denote this by $H \lhd G$.*

Definition 14 *An* automorphism *is a homomorphism $\varphi : G \to G$ that maps a group back onto itself.*

Theorem 102 *A subgroup H of a group G is* characteristic *if $\varphi(H) \subseteq H$ for every automorphism φ of G. Every characteristic subgroup is normal. (The converse is false.) This is Saracino problem 12.23.*

Definition 15 *If $H \lhd G$, then G/H denotes the set of right (=left) cosets of H in G.*

$$G/H = \{Ha \mid a \in G\} \tag{H.4}$$

where

$$Ha = \{ha \mid h \in H\} \tag{H.5}$$

Definition 16 *If $\varphi : G \to K$ is a homomorphism, then the* kernel *of φ is*

$$\ker(\varphi) = \varphi^{-1}(\{e_K\}) = \{g \in G \mid \varphi(g) = e_K\}$$

Theorem 103 *For any homomorphism $\varphi : G \to K$, then*

$$\ker(\varphi) \lhd G$$

This is Saracino theorem 13.1.

Theorem 104 *(Fundamental Theorem on Group Homomorphisms). Let $\varphi : G \to K$ be a homomorphism from G* onto *K. Then K is isomorphic to $G/\ker(\varphi)$. We use the notation*

$$K \cong G/\ker(\varphi) = \{\ker(\varphi)a \mid a \in G\}$$

This is Saracino theorem 13.2.

Theorem 105 *(Second Isomorphism Theorem). Let H and K be subgroups of G, and let $K \lhd G$. Then*

$$H/(H \cap K) \cong HK/K$$

This is Saracino theorem 13.4.

Theorem 106 *(Third Isomorphism Theorem). Let $H \lhd K \lhd G$ and $H \lhd G$. Then*

$$(K/H) \lhd (G/H)$$

and

$$(G/H)/(K/H) \cong (G/K)$$

This is Saracino theorem 13.5.

H.2.4 Transformation Groups

We are now ready to discuss groups of functions that transform members of some space. We call such groups by the name *transformation groups.* This material is taken from Wijsman's monograph [288].

Wijsman (p.15) assigns a technical meaning to the word "action".

Definition 17 *An* action *of group G on arbitrary space A to the* left *is any function*

$$\psi : G \times A \mapsto A$$

with the following properties:

1. For every $g \in G$, $\psi(g, \cdot) : A \to A$ is bijective.

2. $\psi(e, a) = a$ for every $a \in A$.

3. $$\psi(g_2, \psi(g_1, a)) = \psi(g_2 g_1, a)$$

   for every $g_1, g_2 \in G$ and $a \in A$.

Definition 18 *If $ga = a$ for every $g \in G$ and $a \in A$, then the action of G is said to be* trivial.

With the previous discussion in mind, let us examine the properties of the mapping $\psi(g, a) = g[a] = ga$.

1. The range of ψ is

$$R(\psi) = \{\psi(g,a), a \in A\} = \{ga, a \in A\} = \{a, a \in A\} = A$$

Therefore, ψ is an *onto* (surjective) function. Suppose $\psi(g,a_1) = \psi(g,a_2)$. Since $\psi(g,a) = ga = a$ for all $a \in A$, we know that

$$\psi(g,a_1) = a_1 = a_2 = \psi(g,a_2)$$

for any $a_1, a_2 \in A$. Therefore, ψ is *one-to-one* (injective). Since ψ is one-to-one and onto, we call it bijective.

2. $\psi(e,a) = ea = a$ for all $a \in A$ where $e = (I_m, I_n)$ is the group identity element. We do not have to recompute for $e \in G$ since we already established $\psi(g,a) = a$ for all $g \in G$, which includes $g = e$.

3.

$$\psi(g_2, \psi(g_1,a)) = \psi(g_2,a) = a \in A$$

Then

$$\psi(g_1g_2,a) = \psi(g_3,a) = a$$

since $g_1g_2 = g_3 \in G$. Therefore

$$\psi(g_2, \psi(g_1,a)) = \psi(g_1g_2,a)$$

We thus declare that $\psi(g,a) = ga$ is an *action* of G on A to the left, and this action is trivial.

Definition 19 *For a given action of G on A, the* orbit *of $a \in A$ is defined as*

$$Ga \underset{=}{\text{def}} \{ga : g \in G\}$$

This defines a partitioning of A. For our situation, $Ga = \{a\}$. Each partition contains only one point, a.

Definition 20 *The abstract space whose points are the G-orbits is called the* orbit space *under G, and denoted A/G. In our case,*

$$A/G = \{Ga, a \in A\} = \{a, a \in A\} = A$$

Definition 21 *The* orbit projection, *$\pi : A \mapsto A/G$, assigns each $a \in A$ its orbit. $\pi(a) = Ga = a$ where $a \in A$.*

Let $B \subset A$ and $g \in G$. Then

$$gB \underset{=}{\text{def}} \{ga, a \in B\}$$

defines the *g-translate* of B. For our case, $gB = B$.

Definition 22 *The* saturation *of B is*

$$GB = \{gB : g \in G\}$$

For our case, $GB = B$.

Definition 23 *A set $B \subset A$ such that $gB = B$ for all $g \in G$ is called* invariant. *It coincides with its saturation.*

Definition 24 *Define the* isotropy subgroup *(or* stability *subgroup) of G at a for arbitrary $a \in A$ to be*

$$G_a \stackrel{\text{def}}{=} \{g \in G : ga = a\}$$

In our case, $G_a = G$.

Definition 25 *In addition to G and A, suppose we have space C and a function $\varphi : A \to C$. For $g \in G$, the $g-$* translate of φ, *written $g\varphi$, is defined by*

$$(g\varphi)(a) \stackrel{\text{def}}{=} \varphi(g^{-1}a)$$

for all $a \in A$.

Since G is a group and $g^{-1} \in G$, then $g^{-1}a = a$. Thus

$$\varphi(g^{-1}a) = \varphi(a) \tag{H.6}$$

for all $a \in A$. Thus $g\varphi = \varphi$ for all $g \in G$. Therefore, φ is *invariant* under left action of G on A, since each $a \in A$ is a different orbit.

Definition 26 *If an invariant function assumes different values on distinct orbits, it is called* maximal invariant. *This says for our case that if $\varphi(a_1) \neq \varphi(a_2)$ for $a_1 \neq a_2$, then φ is a maximal invariant.*

Definition 27 *A function $\nu : A \to C$ is called* equivariant *if*

$$g(\nu(a)) = \nu(ga)$$

for all $g \in G$ and $a \in A$. In our case, $\nu(ga) = \nu(a)$ for all $g \in G$. Recall

$$(g\nu)(a) = \nu(g^{-1}a) = \nu(a) \tag{H.7}$$

Let

$$g(\nu(a)) \underline{\underline{\text{def}}} (g\nu)(a)$$

Therefore

$$g(\nu(a)) = \nu(ga)$$

for all $g \in G$ and $a \in A$. Thus ν is equivariant.

H.3 Topology and Basic Measure Theory

This section contains a brief highlighting of the main points of measure theory and topology as preparation for the introduction to topological group theory. The source for this material is Rudin [230]. Mastery of these concepts is necessary for any new serious work in signal processing.

H.3.1 Topology

Definition 28 *A* topological space *(G, τ) is a set G upon which a collection of subsets τ of G is defined with the following properties.*

1. τ contains the empty set Φ and also G.

2. τ is closed under finite intersections. For $\{\tau_k\}_{k=1}^n \in \tau$, then $\bigcap_{k=1}^{n} \tau_k \in \tau$.

3. τ is closed under arbitrary unions. For $\{\tau_\alpha\} \in \tau$ where the range of index α is possibly uncountable, then $\bigcup_\alpha \tau_\alpha \in \tau$.

Definition 29 *The subsets of G that are members of τ are called the* open *sets with respect to τ. The set complement*

$$G \backslash X = X^c = Y$$

of any set $X \in \tau$ is called a closed *set.*

Definition 30 *Let φ map topological space (G, τ) into topological space (B, σ). Then φ is called* continuous *if $\varphi^{-1}(X) \in \tau$ for any $X \in \sigma$, where $\varphi^{-1}(X)$ is the preimage of X under the mapping φ.*

Definition 31 *If $g \in G$ and $X \subset G$, where X is not necessarily in τ, then X is a* neighborhood *of g if there is a $Y \in \tau$ such that $g \in Y \subset X$.*

If (G, τ) and (B, σ) are topologies, then $(G \times B, \tau \times \sigma)$ is a topology. However, note that it is possible to define a topology ν on $G \times B$ where $\nu \not\subset \tau \times \sigma$. This means that it is possible to define a mapping that is continuous with respect to ν, but which is not continuous with respect to $\tau \times \sigma$, even though the domain and range of φ are the same in both cases.

Definition 32 *Let (G, τ) be a topological space. Let $K \subset G$, where K is not necessarily in τ. Let $\{B_\alpha\}$ be an arbitrary collection of subsets of G such that $K \subset \bigcup_\alpha B_\alpha$. Then:*

1. The collection $\{B_\alpha\}$ is called a *cover* of K.

2. If all of the B_α are in τ, then the collection $\{B_\alpha\}$ is called an *open cover* of K.

3. Let the index set β be a subset of the index set α, such that $K \subset \bigcup_\beta B_\beta$. Then for the general case, $\{B_\beta\}$ is called a *subcover* of K with respect to $\{B_\alpha\}$.

4. When each $B_\beta \in \tau$, then $\{B_\beta\}$ is called an *open subcover.*

5. If the number of subsets in the collection $\{B_\beta\}$ is finite, then that collection is a *finite subcover*, and is a *finite open subcover* when each $B_\beta \in \tau$.

Definition 33 *Set K is called* compact *if every open cover of K contains a finite open subcover of K.*

Definition 34 *If $B \subset G$, then the* closure $\bar{B}$ *of B is the smallest closed set with respect to τ that contains B. Thus $B \subset \bar{B} \subset G$.*

Definition 35 *G is called* locally compact *if every $g \in G$ has a neighborhood B whose closure $\bar{B}$ is compact.*

Definition 36 *G is called* Hausdorff *if for all $g, h \in G$, $g \neq h$, that g and h have disjoint neighborhoods. The elements $g \in G$, when G is Hausdorff, are called* separable.

H.3.2 Measure Theory

Definition 37 *A measurable space* $(G, \mathcal{M})$ *is a set* G *and a collection of subsets* $\mathcal{M}$ *of* G *where* $\mathcal{M}$ *is called a* σ-algebra, *and* $\mathcal{M}$ *has the following properties.*

1. $\mathcal{M}$contains the empty set Φ.

2. If $B \in \mathcal{M}$, then the set complement $G \backslash B = B^c$ is also in $\mathcal{M}$.

3. $\mathcal{M}$ is closed under countable intersections. For $\{B_k\}_{k \in \mathbf{N}} \in \mathcal{M}$, then

$$\bigcap_{k \in \mathbf{N}} B_k \in \mathcal{M}$$

Definition 38 *Any set* $B \in \mathcal{M}$ *is called a* measurable set.

Definition 39 *Let* φ *map measurable space* $(G, \mathcal{M})$ *into topological space* (B, σ). *Then* φ *is called a* measurable function *if* $\varphi^{-1}(x) \in \mathcal{M}$ *for every* $x \in \sigma$.

Theorem 107 *Let* η *be an arbitrary family of subsets of* G. *With the empty set* Φ *and the subset collection* η, *construct a* σ*-algebra* $\mathcal{M}$. $\mathcal{M}$ *is called the* σ*-algebra generated by* η. *Then there exists a* σ*-algebra* $\mathcal{M}_0$ *containing* η *that has fewer subsets of* G *than any other constructed* $\mathcal{M}$. *If* $\mathcal{F}$ *is the family of all* σ*-algebras* $\mathcal{M}$ *which contain* η, *then* $\mathcal{M}_0 = \bigcap_{\mathcal{M} \in \mathcal{F}} \mathcal{M}$. *(Ocneanu [196]).*

Definition 40 *Let* (G, τ) *be a topological space. Using* τ, *which can be considered an arbitrary collection of subsets of* G, *generate a* σ*-algebra* $\mathcal{M}$. *This* $\mathcal{M}$

is called the Borel σ-algebra, $\mathcal{B}$, *with respect to* τ. *Subsets that are contained in* $\mathcal{B}$ *are called* Borel sets. *If* φ *is a measurable function where the* σ*-algebra is a Borel* σ*-algebra, then* φ *is called a* Borel measurable function.

Proposition 42 *The composition of continuous functions yields a continuous function.*

Proposition 43 *A continuous function of a measurable function produces a measurable function.*

Definition 41 *Let* $(G, \mathcal{M})$ *be a measurable space with* σ*-algebra* $\mathcal{M}$. *Let* $(\mathcal{F}, \oplus, \square)$ *be a field. A function* $\mu : \mathcal{M} \rightarrow \mathcal{F}$ *is called a* measure *if:*

1. Let 0 be the identity element of $\oplus$ in $\mathcal{F}$. Then $\mu(\Phi) = 0$ where Φ is the empty set.

2. Let $\{B_k\}_{k \in \mathbf{N}}$ be a countable collection of disjoint measurable sets, where the index k takes on values in the set of natural numbers, $\mathbf{N} = \{0, 1, 2, \cdots\}$. Then

$$\mu\left(\bigcup_k B_k\right) = \bigoplus_k \mu(B_k)$$

This property is called "countable additivity".

Definition 42 *Let* c *be a constant in* $\mathcal{F}$, *and let* $\mathbf{1}$ *be the identity element of* $\square$ *in* $\mathcal{F}$. *Let* $\oplus$ *and* $\square$ *be arithmetic addition and multiplication on the sets to be discussed. When* $\mathcal{F} = [0, \infty] \subset \mathbf{R}$, *then* μ *is called a* positive measure *or*

(more usually) just a measure. *When* $\mathcal{F} = [0,1] = \mathbf{R}$ *and* $\mu(G) = \mathbf{1}$, *then* μ *is called a* probability measure. *When* G *is a locally compact Hausdorff space and* $\mathcal{M}$ *is a Borel* σ*-algebra, then* μ *is called a* Borel measure.

Definition 43 *Let* G *be a locally compact Hausdorff space. Let* σ*-algebra* $\mathcal{M}$ *contain all the Borel sets in* G. *(Thus* $\mathcal{B} \subset \mathcal{M}$.*) Let* μ *be a positive measure. Then*

1. For $A \subset G$, the measure defined by

$$\mu(A) = \inf\{\mu(U) : A \subset U, U \text{ open}\}$$

   is called an *outer measure* of A. A measure μ with this property for all $A \in \mathcal{M}$ is called *outer regular.*

2. For $A \subset G$, the measure defined by

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ compact}\}$$

   is called an *inner measure* of A. A measure μ with this property for every open set A and for every $A \in \mathcal{M}$ with $\mu(A) < \infty$ is called *inner regular.*

3. If μ is both inner regular and outer regular, then μ is called *regular.*

Definition 44 *A* Radon measure *on* G *is a Borel measure which is finite on compact sets, outer regular on all Borel sets, and inner regular on all open sets.*

Definition 45 *Let G be a locally compact Hausdorff topological group. A* left Haar measure *on G is a nonzero Radon measure μ which satisfies $\mu(gA) = \mu(A)$ for all $g \in G$ and for all open A. Similarly, a* right Haar measure *satisfies $\mu(Ag) = \mu(A)$.*

A few other often encountered terms in measure theory are defined below.

Definition 46 *If $(G, \mathcal{M}, \mu)$ is a measure space, then a set $B \in \mathcal{M}$ is called a* null set *if $\mu(B) = 0$.*

Definition 47 *A measure μ whose domain $\mathcal{M}$ contains all subsets of null sets is called* complete.

Theorem 108 *Let $G = \mathbf{R}$ and let $\mathcal{M}$ be the Borel σ-algebra defined on G. Let $\varphi : G \to \mathbf{R}$ be any increasing, right continuous function. Then there is a unique measure*

$$\mu_\varphi((a, b]) = \varphi(b) - \varphi(a)$$

for all $a, b \in G$. If ν is another such function, then $\mu_\varphi = \mu_\nu$ if and only if $\varphi - \nu$ is a constant. This is Folland Theorem 1.16 [85].

Definition 48 *The completion of this measure, when $\varphi(g) = g$ for all $g \in G$ is called the* Lebesgue measure, *and this measure is usually denoted by m. Lebesgue measure on $\mathbf{R}^n$ is the completion of the n-fold product of the Lebesgue measure on $\mathbf{R}$.*

Definition 49 *Now we return to measure space* $(G, \mathcal{M}, \mu)$ *and group* $(\mathcal{F}, \oplus)$ *with*

$$\mu : \mathcal{M} \rightarrow \mathcal{F}$$

1. When

$$\mathcal{F} = \{-\infty\} \cup \mathbf{R} \cup \{+\infty\}$$

and $\oplus$ is addition, then μ is called a *signed measure.*

2. When $\mathcal{F} = \mathbf{C}$, the set of complex numbers, with $\oplus$ being addition, then μ is called a *complex measure.*

3. When $|\mu(G)| < \infty$, then μ is called a *finite* or *bounded* measure.

Definition 50 *Let* $\mathcal{M}$ *be the power set* $\mathbf{P}(G)$ *of* G. *The power set* $\mathbf{P}(G)$ *is the collection of all possible subsets of* G.

$$\mathbf{P}(G) = \{x \mid x \in G\}$$

In particular, each element $g \in G$ *is a member of* $\mathbf{P}(G)$. *Two special cases are important. Let* $\mu(G) = f$ *where* $g \in G$ *and* $f \in \mathcal{F}$.

1. When $\mathcal{F} = \mathbf{N}\{0, 1, 2, \cdots\}$, and $f = 1$ where $\oplus$ is addition, then μ is called a *counting measure.*

2. Let $\mu(g_0) = f$ for one specific $g_0 \in G$, and $\mu(G) = E$ for $g \in G$, $g \neq g_0$, where $f \in \mathcal{F}$, $f \neq E$. When $\mathcal{F} = \mathbf{N}$, $f = 1$, and $E = 0$, then μ is called

the Dirac measure, atomic measure, or the point mass at g_0. The group operator $\oplus$ is addition.

H.4 Topological Group Theory

This discussion is taken from Folland (p. 312 ff) [85] or it is motivated by Folland. We begin with a definition.

Definition 51 *A topological group $(G, \circ, \tau)$ is a group $(G, \circ)$ with a topology τ defined on G such that the group operator $\circ$ and the group inverse mapping $g \mapsto g^{-1}$ are continuous with respect to τ.*

Let (G, τ) be a topology. Then $(G \times G, \tau \times \tau)$ is also a topology. Let $\circ$ be a mapping $\circ : G \times G \mapsto G$ where $\circ$ is the group operator of group $(G, \circ)$. Since $\circ$ is a continuous operator, then for any $D \in \tau$ we know that its preimage under $\circ$ is some $A \times B \in \tau \times \tau$. The awkward formal notation is $\circ^{-1}(D) \in \tau \times \tau$ for any $D \in \tau$. Because we carefully constructed $\tau \times \tau$, we know that $A \in \tau$ and $B \in \tau$. In nicer notation, we say $A \circ B \mapsto D$ is continuous with respect to τ.

Definition 52 *Let $A, B \subset G$ and $g \in G$. Then*

1. $gA = g \circ A = \{g \circ a \mid a \in A\}$.

2. $Ag = A \circ g = \{a \circ g \mid a \in A\}$.

3. $A^{-1} = \{a^{-1} \mid a \in A\}$ where a^{-1} is the element in G that is the group inverse of a.

4. $AB = A \circ B = \{a \circ b \mid a \in A, b \in B\}$.

Definition 53 *If $A \subset G$ and $A = A^{-1}$, then A is called* symmetric.

Note that A^{-1} is the set of *group element inverses*

$$\{g^{-1} \mid g \in A\} \underline{\text{def}} \; A^{-1}$$

Note that the conditions $A \subset G$ and $A = A^{-1}$ are not sufficient to define A to be a subgroup of G. For example, let A be all elements of G except the group identity element e. Then $A \subset G$ and $A = A^{-1}$ and yet A is not a group. Note that AA^{-1} is not a set consisting of only the group identity element.

$$AA^{-1} = \{a \circ b^{-1} \mid a, b \in A \subset G\}$$

In the general case of A being a subset of G (but not a subgroup of G), we cannot guarantee that $a^{-1}, b^{-1} \in A$. Being a subset is different than being a subgroup. We cannot even claim that

$$\{a \circ b \mid a, b \in A\} \subset A$$

When $A = A^{-1}$, then

$$A^{-1} = \{a^{-1} \mid a \in A\} = A$$

implies $a^{-1} \in A$. We cannot claim

$$\{a \circ b^{-1} \mid a, b \in A\} \subset A$$

even though $\{b^{-1} \mid b \in A\} = A$. In our example of $A = G\backslash\{e\}$, when $a = b$, then $a \circ b^{-1} = e \notin A$.

Some basic properties of topological groups are given by Folland, as follows.

Proposition 44 *This is due to Folland [85]. Let G be a topological group. Then for $(G, \circ, \tau)$:*

1. The topology of G is translation invariant. Thus, if $U \in \tau$ and $g \in G$, then $g \circ U \in \tau$ and $U \circ g \in \tau$.

2. For every neighborhood A of e, there is a symmetric neighborhood B of e with $B \subset A$.

3. For every neighborhood A of e, there is a neighborhood B of e with $BB \subset A$.

4. If $(H, \circ)$ is a subgroup of $(G, \circ)$, then so is its closure $(\bar{H}, \circ)$.

5. Every open subgroup $(A, \circ)$, $A \in \tau$, of $(G, \circ)$ is also closed.

6. If A and B are compact subsets of G, then AB is also a compact subset of G.

I have lost record of the pedigree of the following definitions.

Definition 54 *Let $(G, \circ, \tau)$ and $(H, \circ, \sigma)$ be topological groups having the same group operator $\circ$. H is a* topological subgroup *of G if $\sigma \subset \tau$. Usually, $\sigma = \tau \cap H$.*

Definition 55 *Let H be a topological subgroup of G. If H is compact, then H is a* compact subgroup *of G.*

Definition 56 *Let H be a compact subgroup of G. Then H is a* maximal compact subgroup *of G if there is no other compact subgroup A of G that contains H. Note that G can possibly have more than one maximal compact subgroup.*

Definition 57 *Let $a, b \in G$ be fixed elements. Let φ be a continuous function on the topological group G. Let $\|\cdot\|$ be a norm on this space of functions. Let $g \in G$ be an arbitrary element. Then*

1.

$$(L_a\varphi)(g) \underset{=}{\text{def}} \varphi(a^{-1} \circ g)$$

is called the *left translate* of φ through a.

$$L_{a \circ b} = L_a L_b$$

φ is called *left uniformly continuous* if for every $\epsilon > 0$ there is a neighborhood V of e such that $\|L_a\varphi - \varphi\| < \epsilon$ for $a \in V \subset G$, where e is the group identity in G.

2.

$$(R_a\varphi)(g) \underset{=}{\text{def}} \varphi(g \circ a)$$

is called the *right translate* of φ through a. $R_{aob} = R_a R_b$. φ is called *right uniformly continuous* if for every $\epsilon > 0$ there is a neighborhood V of e such that $\|R_a\varphi - \varphi\| < \epsilon$ for $a \in V \subset G$.

Proposition 45 *If $\varphi \in C_c(G)$, the space of continuous complex-values functions $\varphi(g)$ on G with compact support, then φ is left and right uniformly continuous.*

We continue with Folland's discussion of the Haar measure.

Proposition 46 *Let G be a locally compact Hausdorff topological group. Let*

$$C_c^+ \underset{=}{\text{def}} \{\varphi \in C_c(G) : \varphi \geq 0 \text{ and } \|\varphi\| > 0\}$$

Then

1. A Radon measure μ on G is a left Haar measure if and only if

$$\tilde{\mu}(A) \underset{=}{\text{def}} \mu(A^{-1}),\ A \subset G \text{ open}$$

is a right Haar measure.

2. A nonzero Radon measure μ on G is a left Haar measure if and only if

$$\int \varphi d\mu = \int L_a \varphi d\mu$$

for all $\varphi \in C_c^+$ and $a \in G$.

3. If μ is a left Haar measure on G, then $\mu(U) > 0$ for all nonempty open $U \subset G$, and $\int \varphi d\mu > 0$ for all $f \in C_c^+$.

4. If μ is a left Haar measure on G, then $\mu(G) < \infty$ if and only if G is compact.

Theorem 109 *Every locally compact Hausdorff topological group possesses a left Haar measure. This is Folland [85] theorem 10.5.*

Theorem 110 *If μ and ν are left Haar measures on G, then there exists $c > 0$ such that $\mu = c\nu$. This is Folland [85] theorem 10.14.*

Proposition 47 *Left and right Haar measures are mutually absolutely continuous. This is Folland [85] theorem 10.18.*

Theorem 111 *Let G be a compact Hausdorff topological group. There there is a unique real-valued function I, called the Haar integral, defined for continuous real-valued functions φ on G, such that:*

1. $I(\varphi_1 + \varphi_2) = I(\varphi_1) + I(\varphi_2)$.
2. $I(c\varphi) = cI(\varphi)$, where $c \in \mathbf{R}$.
3. If $\varphi(g) \geq 0$ for all $g \in G$, then $I(\varphi) \geq 0$.
4. $I(e) = 1$.
5. $I(R_a\varphi) = I(\varphi) = I(L_a\varphi)$ for all $a \in G$. Often the notation $\int \varphi(g)dg$ is used for $I(\varphi)$. For example, this property can be written as

$$\int \varphi(g \circ a)dg = \int \varphi(g)dg = \int (a^{-1} \circ g)dg$$

This is Bredon [43] theorem 3.1.

H.5 Matrix Groups

There are some standard groups in the theory of matrix groups. The following were taken from Curtis [64]. You are already familiar with the notation $\mathbf{R}$ for the set of all real numbers and $\mathbf{C}$ for the set of all complex numbers. The next level of generalization is the set of quaternions, denoted $\mathbf{H}$. For the sake of generality, we let $K \in \{\mathbf{R}, \mathbf{C}, \mathbf{H}\}$.

Definition 58 *$M_n(K)$ is the vector space of set K^n over field $(K, +, \cdot)$. It distributes the field over the set as*

$$b(x + y) = bx + by$$

where $b \in K$ and $x, y \in K^n$. This is called an algebra.

Definition 59 *If $\mathcal{A}$ is an algebra, then $x \in \mathcal{A}$ is a* unit *if there is some $y \in \mathcal{A}$ such that*

$$xy = yx = 1$$

I.e., x is a unit if it has a multiplicative inverse. If $\mathcal{A}$ is an algebra with associative multiplication and $U \subset A$ is the set of units in $\mathcal{A}$, then U is a group under multiplication.

The group of units in the algebra consisting of $n \times n$ matrices in field K is denoted by $GL(n, K)$, and is called a General Linear group. Another notation

Table H.1. Notation for Types of Orthogonal Groups

Notation	**Group Name**
$\mathcal{O}(n, \mathbf{R}) = \mathcal{O}(n)$	Orthogonal group
$\mathcal{O}(n, \mathbf{C}) = \mathbf{U}(n)$	Unitary group
$\mathcal{O}(n, \mathbf{H}) = Sp(n)$	Symplectic group

used is $M_n(K) = GL(n, K)$. Often, when the notation $M_n(\cdot)$ is used, it refers to the set of *nonsingular* $n \times n$ matrices. So, when $K = \mathbf{C}$, we can use $M_n(\mathbf{C})$.

Some special groups are the class of orthogonal groups $\mathcal{O}(n, K)$, defined by

$$\mathcal{O}(n, K) = \{A \in M_n(K) \mid < Ax, Ay > = < x, y > \text{ for all } x, y \in K^n\}$$

When $A \in \mathcal{O}(n, K)$ then $AA^H = I_n$ and $A^H A = I_n$. Here, the notation specializes. The notation for orthogonal matrix groups is given in table H.1.

The determinant of $A \in \mathcal{O}(n)$ is $\det A \in [-1, 1]$, and of $B \in \mathbf{U}(n)$ is

$$\det B \in \{e^{i\theta}, \theta \in \mathbf{R}\}$$

When attention is restricted to those cases where $\det A = 1$ and $\det B = 1$, special groups are formed. Notation for these special groups is given in table H.2.

Let $A = A^H \in M_n(\mathbf{C})$. Can we make a group from the set of Hermitian matrices? Let

$$H = \{A \in M_n(\mathbf{C}) \mid A = A^H\}$$

Table H.2. Special Orthogonal Groups

Notation	**Group Name**
$\{A \in \mathcal{O}(n) \mid \det A = 1\} = SO(n)$	Special Orthogonal Group
$\{B \in \mathbf{U}(n) \mid \det B = 1\} = SU(n)$	Special Unitary Group

Then H is the set of all $n \times n$ Hermitian matrices. (Recall, $\mathbf{H}$ is the set of quaternions. The font style is significant in the notation.) Let us see if $(H\backslash 0, \cdot)$ is a group where the operation is ordinary matrix multiplication. The set $H\backslash 0$ is the set H with the zero matrix removed. Suppose

$$A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 2 & 1-i \\ 1+i & 1 \end{pmatrix}$$

Both A and B are in $H\backslash 0$. However,

$$C = A \cdot B = \begin{pmatrix} 1+i & 1 \\ 1-i & -i \end{pmatrix}$$

is not in $H\backslash 0$. Thus, $(H\backslash 0, \cdot)$ cannot be a group because it is not closed under matrix multiplication.

Lemma 24 *If A and B are Hermitian, then AB is Hermitian if and only if $AB = BA$. This is Nomizu exercise 8.4.1 (p. 237) [193].*

Proof. Let $A = A^H$, $B = B^H$. Then

$$(AB)^H = B^H A^H = BA$$

The statement that AB is Hermitian means that $(AB)^H = AB$. Thus AB is Hermitian if and only if $AB = BA$, or equivalently if $AB - BA = 0$, which is a form familiar to those who have studied Lie Theory.

To further explore, suppose we consider the set of Hermitian nonnegative definite matrices. From Johnson [124] we have the example

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & -3 \\ -3 & 10 \end{pmatrix}$$

and

$$AB = \begin{pmatrix} -8 & 27 \\ -27 & 91 \end{pmatrix}$$

Both A and B are Hermitian (symmetric) positive definite. However, AB is not definite, positive or negative. In looking at $x^H ABx$, if $x_1^2 > \frac{91}{8} x_2^2$ then $x^H ABx < 0$. If $x_1^2 < \frac{91}{8} x_2^2$, then $x^H ABx > 0$.

Here are two more examples of structures that fail to form a group. Let $X = X^H$ be the set of all Hermitian matrices that are nonsingular. Let $A, B \in X$. Now, define $B \circ A = B^H AB$. The operator $\circ$ is a binary operator,

and $B^H AB \in X$.

$$(B^H AB)^H = B^H A^H B = B^H AB$$

Let us see if it is associative.

$$C \circ (B \circ A) = C \circ (B^H AB) = C^H B^H ABC = CBABC \text{ if } B = B^H,\ C = C^H$$

$$(C \circ B) \circ A = (C^H BC) \circ A = C^H BCAC^H BC$$

$$= CBCACBC \text{ if } B = B^H,\ C = C^H$$

Thus

$$C \circ (B \circ A) \neq (C \circ B) \circ A$$

and therefore $(X, \circ)$ is not a group because $\circ$ is not an associative operator.

Similarly, $A \square B = B^H AB$ is not a group because

$$C \square (B \square A) = C \square (A^H BA)$$

$$= ABACABA, \text{ for } A = A^H,\ B = B^H,\ C = C^H$$

$$(C \square B) \square A = A^H (C \square B) A = A^H (B^H CB) A$$

$$= ABCBA, \text{ for } A = A^H,\ B = B^H,\ C = C^H$$

We will want to take advantage of properties of topological groups. In preparation, the following remarks are taken from Bredon (p.5 ff) [43]. The following groups are topological groups with the relative topology from $M_n(\mathcal{F})$: $GL(n, \mathcal{F})$, $SL(n, \mathcal{F})$, $\mathcal{O}(n, \mathcal{F})$, and

$$SO(n, \mathcal{F}) = \mathcal{O}(n, \mathcal{F}) \cap SL(n, \mathcal{F})$$

where the field $\mathcal{F}$ may be taken over $\mathbf{C}$ or $\mathbf{R}$. $\mathbf{U}(n) = \mathcal{O}(n, \mathbf{C})$ is closed in $M_n(\mathbf{C}) \approx \mathbf{C}^{n^2}$. For matrix $U \in \mathbf{U}(n)$, U^H is a continuous function of U and $UU^H = I_n$. Thus $\mathbf{U}(n)$ is bounded in $M_n(\mathbf{C})$. Thus $\mathbf{U}(n)$ is compact. Because $SU(n)$ is also a closed subgroup of $\mathbf{U}(n)$, then $SU(n)$ is also compact. In fact, Gross and Richards (Section 1.3)[96] remark that $\mathcal{O}(n, \mathcal{F})$ is a maximal compact subgroup of $GL(n, \mathcal{F})$.

Let $A, B \in \mathcal{O}(n, K)$. Then

$$(AB)^H(AB) = (AB)(AB)^H = I_n$$

Thus $\mathcal{O}(n, K)$ is closed under matrix multiplication.

H.6 Group Invariance Property of the Vector Complex Normal Distribution

In this section we establish a group invariance property of the vector complex normal distribution. The work done here is a slight generalization of that done by James [120], and I think a necessary background for understanding his paper which revolutionized thinking about the statistical distribution of sample eigenvalues. What is special about the approach given now is the application of the invariance argument to the distribution rather than just some factor terms of the density function. This is a step towards incorporating a measure-theoretic approach with the group invariance ideas which will lead

to an ability to deal with a larger class of distributions and perhaps to an ability to incorporate these ideas into sensitivity analyses.

We begin with some abstractions. We are going to define a set G whose elements are *pairs of matrices.* A special operation will be defined which will provide a rule for combining one set element with another set element. We will then see that this set, with the operator, forms a group. The next step will be to define a set A upon which elements of G will act. We will establish that we have defined a *transformation group* which justifies our use of the machinery of some topological group theory. We then apply our findings to the vector complex normal density function to study invariance properties. In the process, we also see the distinction between a mapping and a change of variables crystalized. This section is really the link between the application and the abstract mathematical work required.

H.6.1 Construction of a Group

This work is supplied by me.

Define a set G by

$$G \underset{=}{\text{def}} \{g = (L, U) \mid L \in M_m(\mathbf{C}), UU^H = I_n\} \tag{H.8}$$

and a binary operator $\circ$ by

$$g_2 \circ g_1 \underset{=}{\text{def}} (L_2, U_2) \circ (L_1, U_1) \underset{=}{\text{def}} (L_2 L_1, U_1 U_2) \tag{H.9}$$

The identity element is $e = (I_m, I_n)$. The inverse element g^{-1} of g is given by $g^{-1} = (L^{-1}, U^H)$. Then the claim to be tested is that $(G, \circ)$ is a group. To show this, we must show that $\circ$ is a binary associative operator, that with this operator there is an associated element e in G that is an identity element, and that each element in G has an inverse element in G associated with this operator.

The operator $\circ$ was defined to be a binary operator between group elements g_1 and g_2. We now examine if $\circ$ is associative, as implied by the claim that $(G, \circ)$ is a group. To do this, we have to show

$$(g_3 \circ g_2) \circ g_1 = g_3 \circ (g_2 \circ g_1)$$

for arbitrary elements g_1, g_2, g_3 in set G. We apply the definitions and observe the results.

$$(g_3 \circ g_2) \circ g_1 = ((L_3, U_3) \circ (L_2, U_2)) \circ (L_1, U_1)$$

$$= (L_3 L_2, U_2 U_3) \circ (L_1, U_1) = (L_3 L_2 L_1, U_1 U_2 U_3)$$

$$g_3 \circ (g_2 \circ g_1) = (L_3, U_3) \circ ((L_2, U_2) \circ (L_1, U_1))$$

$$= (L_3, U_3) \circ (L_2 L_1, U_1 U_2) = (L_3 L_2 L_1, U_1 U_2 U_3)$$

From the equality of the two approaches, we can conclude that $\circ$ is an associative operator.

We now examine the element e to see if it really is an identity element. To

do this, we must show that $g \circ e = e \circ g = g$.

$$g \circ e = (L, U) \circ (I_m, I_n) = (LI_m, I_n U) = (L, U) = g$$

$$e \circ g = (I_m, I_n) \circ (L, U) = (I_m L, U I_n) = (L, U) = g$$

From this we can conclude that $e = (I_m, I_n)$ is the group identity element.

Next, we verify that each element of G has an inverse which is also a member of G. To do this, we must show that for an arbitrary $g \in G$ that there is an element $g^{-1} \in G$ such that $g \circ g^{-1} = e$.

$$g \circ g^{-1} = (L, U) \circ (L^{-1}, U^H) = (LL^{-1}, U^H U) = (I_m, I_n) = e$$

where $UU^H = I_n$ implies $U^{-1} = U^H$, which in turn implies $U^H U = I$.

$$g^{-1} \circ g = (L^{-1}, U^H) \circ (L, U) = (L^{-1} L, U U^H) = (I_m, I_n) = e$$

Thus each element g has an inverse. We conclude that we have a group.

By the theorems from group theory, we know that e is unique, and that for each element $g \in G$ that the associated $g^{-1} \in G$ is unique.

H.6.2 Action Set Definition

This work is supplied by me.

We next define a set A and a rule for operating on A by elements of group $(G, \circ)$. Let A be the set of all multivariate complex normal probability distribution (or measures), $P_z(\mu, \Sigma)$ where $\Sigma^H = \Sigma > 0$. Denote the multivariate

random variable by Z, the mean by μ, and the covariance matrix by Σ. You can consider (μ, Σ) as an index that selects a particular distribution from set A. Define the *action* on elements of A by elements of group G by the set of simultaneous mappings $g[a]$, where $a \in A$ is one specific $P_Z(\mu, \Sigma)$, by

$$(L, U)[P_Z(\mu, \Sigma)] = P_{LZU}(L\mu U, L\Sigma L^H) \tag{H.10}$$

The traditional expression of this action is the set of simultaneous mappings

$$\left.\begin{aligned} Z &\mapsto LZU = Y \\ \mu &\mapsto L\mu U = M \\ \Sigma &\mapsto L\Sigma L^H = S \end{aligned}\right\} \tag{H.11}$$

The density function exists when Σ is nonsingular. For $Z \sim \mathbf{C}N_{p,n}(\mu, \Sigma, I)$ we have by theorem 51

$$dF_Z(Z) = \frac{1}{\pi^{np}\,|\Sigma|^n}\,\mathrm{etr}[-(Z-\mu)^H\Sigma^{-1}(Z-\mu)](dZ) = a \tag{H.12}$$

This represents just one element $a \in A$.

H.6.3 Invariance Demonstration

This work is supplied by me.

Now, pick an arbitrary element $g = (L, U) \in G$. When we apply g to a we are picking a new element $b \in A$. Note that this is a mapping, not a change of variables.

$$b = g[a] = (L, U)[dF_Z(Z)] = \frac{1}{\pi^{np}\,|S|^n}\,\mathrm{etr}[-(Y-M)^H S^{-1}(Y-M)](dY) \tag{H.13}$$

$$= \frac{1}{\pi^{np}\,|L\Sigma L^H|^n}\,\mathrm{etr}[-(Y-L\mu U)^H(L\Sigma L^H)^{-1}(Y-L\mu U)](dY)$$

$$= \frac{1}{\pi^{np}\,|\Sigma|^n\,|L|^{2n}} \times$$

$$\times\,\mathrm{etr}[-U^H(L^{-1}YU^{-1}-\mu)^H L^H L^{-H}\Sigma^{-1}L^{-1}L(L^{-1}YU^{-1}-\mu)U](dY)$$

$$= \frac{1}{\pi^{np}\,|\Sigma|^n\,|L|^{2n}}\,\mathrm{etr}[-(L^{-1}YU^{-1}-\mu)^H\Sigma^{-1}(L^{-1}YU^{-1}-\mu)](dY) \qquad \text{(H.14)}$$

This is the new point $b \in A$. We wish to investigate the invariance properties associated with this mapping. To be able to compare, we now change variables from Y to Z, while remaining at the location b. Let

$$Y = LZU \qquad \text{(H.15)}$$

By theorem 34 we know the Jacobian of this transformation is

$$J(Y \to Z) = |\det L|^{2n}\,|\det U|^{2p} = |\det L|^{2n}$$

since $UU^H = I$ implies $|\det U|^{2p} = 1$. So,

$$b = \frac{1}{\pi^{np}\,|\Sigma|^n\,|L|^{2n}} \times$$

$$\times\,\mathrm{etr}[-(L^{-1}LZUU^{-1}-\mu)^H\Sigma^{-1}(L^{-1}LZUU^{-1}-\mu)]\,|J(Y \to Z)|\,(dZ)$$

$$= \frac{1}{\pi^{np}\,|\Sigma|^n\,|L|^{2n}}\,\mathrm{etr}[-(Z-\mu)^H\Sigma^{-1}(Z-\mu)]\,|L|^{2n}\,(dZ) \qquad \text{(H.16)}$$

$$= \frac{1}{\pi^{np}\,|\Sigma|^n}\,\mathrm{etr}[-(Z-\mu)^H\Sigma^{-1}(Z-\mu)](dZ) = a \qquad \text{(H.17)}$$

Thus, A is invariant under action by group $(G, \circ)$. We have shown that for any $a \in A$ that $ga = a$ for all $g \in G$.

Appendix I

ELEMENTARY HILBERT SPACE THEORY

This is a very terse review of the fundamental definitions of Hilbert space theory. The following definitions are taken from Rudin [230], slightly modified to fit our applications.

Definition 60 *A complex vector space H is called an* inner product space *(or* unitary space*) if to each ordered pair of vectors x and y in H there is associated a complex number $< x, y >$, called the* inner product *or* scalar product *of x and y such that the following rules hold.*

(a) $< y, x > = < x, y >^*$. The asterisk denotes complex conjugate.

(b) $< z, x + y > = < z, x > + < z, y >$ if x, y, and z are in H.

(c) $< x, \alpha y > = \alpha < x, y >$ if x and y are in H and α is a scalar.

(d) $< x, x > \geq 0$ for all x in H.

(e) $< x, x >= 0$ only if $x = 0$.

Note that (b) and (c) say that $< x, y >$ is linear in the second argument. This is the change I have made from the usual mathematician's definition. The reason for doing this is to be able to use the natural notation of the Hermitian transpose in the definition of an inner product. For example, $< x, y >= x^H y$ is a valid inner product.

Definition 61 *Define the inner product space* norm *of x to be*

$$\|x\|_H =< x, x >^{1/2}$$

using the non-negative square root.

Definition 62 *Define the* distance *(or* metric*) between x and y to be*

$$d_H(x, y) = \|x - y\|_H$$

Then all the axioms of a metric space are satisfied. Thus, our inner product space H is now also a metric space.

Definition 63 *If every Cauchy sequence, using $d_H(x, y)$, converges in H, then this metric space is* complete. *When this is true, our complete inner product space H is called a* Hilbert space.

Example 7 *Vectors in C^n with inner product $< x, y >= x^H y$ form a Hilbert space.*

Example 8 *If μ is any positive measure, $L^2(\mu)$ is a Hilbert space with inner product $< f, g >= \int_x f^* g \, d\mu$. Note that*

$$\|f\|_H =< f, f >^{1/2} = \{\int_x |f|^2 \, d\mu\}^{1/2} = \|f\|_2$$

Appendix J

COMPLEX VECTORS

In this section we will define a vector space and by way of an example show some of the problems associated with treating a vector space in n-dimensional complex numbers $\mathbf{C}^n$ as a vector space in $2n$-dimensional real numbers $\mathbf{R}^{2n\times 2}$. We first define a field and then a vector space.

J.1 Abstract Field

A *vector space* requires a special kind of group called an abelian group. An abelian group is a group that also obeys the following rule (Group Rule number 5).

5. For all $a \in G$ and $b \in G$, then $a\Box b = b\Box a$. The order of the elements is not important. When this is true, we call the operator $\Box$ a *commutative* operator. We say that a and b *commute* under $\Box$.

Let $G/\{e_\Box\}$ denote the set G with the element $e_\Box$ removed. Now, equip $G/\{e_\Box\}$ with an operator $\circ$ such that $(G/\{e_\Box\}, \circ)_g$ is an abelian group. We now have defined two groups, $(G, \Box)$ and $(G/\{e_\Box\}, \circ)$. When we glue these two groups together using rules (6) and (7) below, we get a *field.* We denote this by $\mathcal{F} = (G, \Box, \circ)$ when the context is clear. Group Rules (6) and (7) are called the *distributive laws.*

6. For all $a, b, c \in G$, then $a \circ (b \square c) = (a \circ b) \square (a \circ c)$.

7. For all $a, b, c \in G$, then $(b \square c) \circ a = (b \circ a) \square (c \circ a)$.

We are familiar with two common fields: (i) the field over the real numbers $(\mathbf{R}, +, \cdot)$ where $+$ is ordinary addition and $\cdot$ is ordinary multiplication, and $(\mathbf{C}, +_c, \cdot_c)$ where $+_c$ is complex addition and $\cdot_c$ is complex multiplication.

J.2 Abstract Vector Space

A nonempty set V, together with operators Δ and $\star$ is said to be a vector space over a field $\mathcal{F} = (G, \square, \circ)$ if the following rules (a) through (j) hold.

(a) For each $x \in V$ and $y \in V$, then $x \Delta y \in V$.

(b) For each $x \in V$ and $a \in \mathcal{F}$, then $a \star x \in V$.

(c) $x \Delta y = y \Delta x$ for all $x, y \in V$.

(d) $(x \Delta y) \Delta z = x \Delta (y \Delta z)$ for all $x, y, z \in V$.

(e) There is an identity element $e_\Delta \in V$ such that $e_\Delta \, \Delta \, x = x$ for all $x \in V$.

(f) There is an inverse element $y \in V$ for each $x \in V$ such that $x \Delta y = e_\Delta$.

(g) For all $x, y \in V$ and $a \in \mathcal{F}$, we have the distributive law that

$$a \star (x \Delta y) = (a \star x) \Delta (a \star y) \in V$$

(h) For all $x, y \in V$ and $a \in \mathcal{F}$, we have the distributive law that

$$(a \square b) \star x = (a \star x) \Delta (b \star x) \in V$$

(i) For all $x \in V$ and $a, b \in \mathcal{F}$, we have the associative law that

$$a \star (b \star x) = (a \circ b) \star x$$

Note the operators! There is a change.

(j) There is an identity element $e_\circ$ where $e_\circ \star x = x$ for all $x \in V$ and $e_\circ$ is the identity in $G/\{e_\square\}$ under $\circ$.

We denote the vector space by

$$\mathcal{V} = (V, \Delta, \star, \mathcal{F})$$

or

$$\mathcal{V} = (V, \Delta, \star, G, \square, \circ)$$

J.3 Complex Vector Space

It is occasionally claimed that a vector space in $\mathbf{C}^n$ is merely a vector space over $\mathbf{R}^{2n}$. It is not quite that simple. For example, consider Broida and Williamson's problem 2.2.4 [47]. The vectors $u = (1 + i, 2i)$ and $v = (1, 1 + i)$ in $\mathbf{C}^2$ are linearly dependent over $\mathbf{C}$, but linearly independent over $\mathbf{R}$. In $\mathbf{C}$,

$$iu + (1 - i)v = 0$$

Consider the elements $u, v \in V$ as in a vector space over $\mathbf{R}$. This means that the scalars come from $\mathbf{R}$. In this case, there are no scalars $a, b \in \mathbf{R} = \mathcal{F}$

such that $au + bv = 0$. Therefore $(V, \Delta, \star) = (\mathbf{C}^2, +, \cdot)$ is not a vector space over the field of real numbers $\mathcal{F} = (\mathbf{R}, +, \cdot)$.

Consider representing an element in V given by $(a + ib, c + id)$ in $\mathbf{C}^2$ as (a, b, c, d) in $\mathbf{R}^4$. Then u and v are expressed as $(1, 0, 1, 2) = u$ and $(1, 1, 0, 1) = v$. These are linearly independent whether the field $\mathcal{F}$ is taken to be $\mathbf{R}$ or $\mathbf{C}$. What is missing is the accounting for the structure of $\mathbf{C}$ under complex multiplication. There is a way of representing $\mathcal{V}$ where $V = \mathbf{C}^n$ and $\mathcal{F} = \mathbf{C}$, with $\hat{V} \in \mathbf{R}^{2n\times 2}$ and $\hat{\mathcal{F}} \in R^{2\times 2}$. The complex number $x + iy \in \mathbf{C}$ is isomorphic to the matrix

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \in \mathbf{R}^{2\times 2}$$

Using this form, matrix addition and matrix multiplication yield the same answers as complex addition and complex multiplication in the scalar case. In our vector space, care must be taken in defining $a \star x$ where $a \in \hat{\mathcal{F}} \subset \mathbf{R}^{2\times 2}$ and $x \in \hat{V} \subset \mathbf{R}^{2n\times 2}$. For

$$Z_i = \begin{pmatrix} X_i & -Y_i \\ Y_i & X_i \end{pmatrix}$$

$$X = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

and

$$a = \begin{pmatrix} c & -b \\ b & c \end{pmatrix}$$

we define

$$a \star x = \begin{pmatrix} a \odot Z_1 \\ \vdots \\ a \odot Z_n \end{pmatrix}$$

where $\odot$ is matrix multiplication of matrices in $\mathbf{R}^{2\times 2}$. Then from our example,

$$u = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & -2 \\ 2 & 0 \end{pmatrix}$$

$$v = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

$$u - v = w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & -1 \\ 1 & -1 \end{pmatrix}$$

and

$$-iw = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \star \begin{pmatrix} 0 & -1 \\ 1 & 0 \\ -1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} = v$$

We are now back to a structure where u and v are linearly dependent.

J.3.1 Verifying if a Field is Defined

It might be worthwhile just to verify that we really do have a field and a vector space under this structure. Let

$$a = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}$$

$$b = \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix}$$

and

$$c = \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix}$$

Define $\square$ to be matrix addition and $\circ$ to be matrix multiplication. Let G be the set of all 2×2 matrices of the form

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

where $x, y \in \mathbf{R}$. Then we examine each of the properties.

1a.

$$a\square b = \begin{pmatrix} a_1 + b_1 & -(a_2 + b_2) \\ a_2 + b_2 & a_1 + b_1 \end{pmatrix} \in G$$

Thus $\square$ is well defined on G.

2a.

$$(a\square b)\square c = \begin{pmatrix} (a_1 + b_1) + c_1 & -[(a_2 + b_2) + c_2] \\ (a_2 + b_2) + c_2 & (a_1 + b_1) + c_1 \end{pmatrix}$$

$$= \begin{pmatrix} a_1 + (b_1 + c_1) & -[a_2 + (b_2 + c_2)] \\ a_2 + (b_2 + c_2) & a_1 + (b_1 + c_1) \end{pmatrix} = a\square(b\square c)$$

Therefore $\square$ is associative.

3a. The identity element of G under $\square$ is

$$e_\square = \begin{pmatrix} 0 & -0 \\ 0 & 0 \end{pmatrix}$$

which is the zero matrix.

4a. The inverse element of

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

is

$$\begin{pmatrix} -x & y \\ -y & -x \end{pmatrix}$$

5a. G is abelian under $\square$, which is inherited from addition of real numbers in $(\mathbf{R}, +)$.

We also need to verify that $(G/\{e_\square\}, \circ)$ is an abelian group. Let $\hat{G} = G/\{e_\square\}$.

1b.

$$a \circ b = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 - a_2b_2 & -[a_2b_1 + a_1b_2] \\ a_2b_1 + a_1b_2 & a_1b_1 - a_2b_2 \end{pmatrix} \in \hat{G}$$

Thus $a \circ b$ is well defined in $\hat{G}$.

2b.

$$(a \circ b) \circ c = \begin{pmatrix} a_1b_1 - a_2b_2 & -[a_2b_1 + a_1b_2] \\ a_2b_1 + a_1b_2 & a_1b_1 - a_2b_2 \end{pmatrix} \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix}$$

$$= \begin{pmatrix} (a_1b_1 - a_2b_2)c_1 - (a_2b_1 + a_1b_2)c_2 & -[(a_2b_1 + a_1b_2)c_1 + (a_1b_1 - a_2b_2)c_2] \\ (a_2b_1 + a_1b_2)c_1 + (a_1b_1 - a_2b_2)c_2 & (a_1b_1 - a_2b_2)c_1 - (a_2b_1 + a_1b_2)c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1(b_1c_1 - b_2c_2) - a_2(b_2c_1 + b_1c_2) & -[a_1(b_2c_1 + b_1c_2) + a_2(b_1c_1 - b_2c_2)] \\ a_1(b_2c_1 + b_1c_2) + a_2(b_1c_1 - b_2c_2) & a_1(b_1c_1 - b_2c_2) - a_2(b_2c_1 + b_1c_2) \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} b_1c_1 - b_2c_2 & -[b_2c_1 + b_1c_2] \\ b_2c_1 + b_1c_2 & b_1c_1 - b_2c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \left[\begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix} \right] = a \circ (b \circ c)$$

Therefore $\hat{G}$ is associative under $\circ$.

3b. The identity element is

$$e_o = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

4b. The inverse of

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

is

$$\frac{1}{x^2+y^2}\begin{pmatrix} x & y \\ -y & x \end{pmatrix}$$

5b.

$$b \circ a = \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix}\begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} = \begin{pmatrix} b_1a_1 - b_2a_2 & -[b_2a_1 + b_1a_2] \\ b_2a_1 + b_1a_2 & b_1a_1 - b_2a_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 - a_2b_2 & -[a_1b_2 + a_2b_1] \\ a_1b_2 + a_2b_1 & a_1b_1 - a_2b_2 \end{pmatrix} = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}\begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} = a \circ b$$

Thus $\hat{G}$ is abelian under $\circ$.

6.

$$a \circ (b \square c) = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}\left[\begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} + \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix}\right]$$

$$= \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}\begin{pmatrix} b_1 + c_1 & -[b_2 + c_2] \\ b_2 + c_2 & b_1 + c_1 \end{pmatrix}$$

$$= \begin{pmatrix} a_1(b_1 + c_1) - a_2(b_2 + c_2) & -[a_1(b_2 + c_2) + a_2(b_1 + c_1)] \\ a_1(b_2 + c_2) + a_2(b_1 + c_1) & a_1(b_1 + c_1) - a_2(b_2 + c_2) \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 - a_2b_2 + a_1c_1 - a_2c_2 & -[a_1b_2 + a_2b_1 + a_1c_2 + a_2c_1] \\ a_1b_2 + a_2b_1 + a_1c_2 + a_2c_1 & a_1b_1 - a_2b_2 + a_1c_1 - a_2c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 - a_2b_2 & -[a_1b_2 + a_2b_1] \\ a_1b_2 + a_2b_1 & a_1b_1 - a_2b_2 \end{pmatrix} + \begin{pmatrix} a_1c_1 - a_2c_2 & -[a_1c_2 + a_2c_1] \\ a_1c_2 + a_2c_1 & a_1c_1 - a_2c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} + \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix}$$

$$= (a \circ b)\square(a \circ c)$$

Thus the first distributive law in $\mathcal{F}$ is satisfied.

7.

$$(b\square c) \circ a = \begin{pmatrix} b_1 + c_1 & -[b_2 + c_2] \\ b_2 + c_2 & b_1 + c_1 \end{pmatrix} \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}$$

$$= \begin{pmatrix} (b_1 + c_1)a_1 - (b_2 + c_2)a_2 & -[(b_1 + c_1)a_2 + (b_2 + c_2)a_1] \\ (b_1 + c_1)a_2 + (b_2 + c_2)a_1 & (b_1 + c_1)a_1 - (b_2 + c_2)a_2 \end{pmatrix}$$

$$= \begin{pmatrix} b_1a_1 - b_2a_2 + c_1a_1 - c_2a_2 & -[b_1a_2 + b_2a_1 + c_1a_2 - c_2a_1] \\ b_1a_2 + b_2a_1 + c_1a_2 - c_2a_1 & b_1a_1 - b_2a_2 + c_1a_1 - c_2a_2 \end{pmatrix}$$

$$= \begin{pmatrix} b_1a_1 - b_2a_2 & -[b_1a_2 + b_2a_1] \\ b_1a_2 + b_2a_1 & b_1a_1 - b_2a_2 \end{pmatrix} + \begin{pmatrix} c_1a_1 - c_2a_2 & -[c_1a_2 + c_2a_1] \\ c_1a_2 + c_2a_1 & c_1a_1 - c_2a_2 \end{pmatrix}$$

$$= \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} + \begin{pmatrix} c_1 & -c_2 \\ c_2 & c_1 \end{pmatrix} \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}$$

$$= (b \circ a)\square(c \circ a)$$

Thus the second distributive law in $\mathcal{F}$ is satisfied.

Therefore, $(G,\square,\circ)$ is a field. If we let an element of V be

$$x = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

where

$$Z_i = \begin{pmatrix} x_i & -y_i \\ y_i & x_i \end{pmatrix}$$

define the operator Δ to be matrix addition.

(a) Then for every $x, y \in V$ we know $x \Delta y \in V$.

(b) From examining $a \circ b$, we know that the product of two matrices of the form of Z_i will again have that form. Thus

$$a \star x = \begin{pmatrix} a \circ Z_1 \\ \vdots \\ a \circ Z_n \end{pmatrix}$$

is in the same form as

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

Therefore $a \star x \in V$.

(c) Because matrix addition is abelian, we know $x \Delta y = y \Delta x$ for all $x, y \in V$.

(d) Because matrix addition is associative, we know

$$(x \Delta y) \Delta z = x \Delta (y \Delta z)$$

for all $x, y, z \in V$.

(e) The identity element $e_\Delta \in V$ is the $2n \times 2$ zero matrix 0 for all $x \in V$.

(f) The inverse element of $x \in V$ is $-x$ where

$$-x = \begin{pmatrix} -Z_1 \\ \vdots \\ -Z_n \end{pmatrix}$$

and

$$-Z_i = \begin{pmatrix} -x_i & y_i \\ -y_i & -x_i \end{pmatrix}$$

(g)

$$a \star (x \Delta y) = a \star \begin{pmatrix} Z_{1x} + Z_{1y} \\ \vdots \\ Z_{nx} + Z_{ny} \end{pmatrix}$$

$$= \begin{pmatrix} a \circ (Z_{1x} + Z_{1y}) \\ \vdots \\ a \circ (Z_{nx} + Z_{ny}) \end{pmatrix} = \begin{pmatrix} (a \circ Z_{1x}) + (a \circ Z_{1y}) \\ \vdots \\ (a \circ Z_{nx}) + (a \circ Z_{ny}) \end{pmatrix}$$

by the first distributive law in $\mathcal{F}$. This, in turn, equals

$$\begin{pmatrix} a \circ Z_{1x} \\ \vdots \\ a \circ Z_{nx} \end{pmatrix} + \begin{pmatrix} a \circ Z_{1y} \\ \vdots \\ a \circ Z_{ny} \end{pmatrix} = (a \star x) \Delta (a \star y)$$

Thus the first distributive law in V is satisfied.

(h)

$$(a\Box b)\star x = (a\Box b)\star \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

$$= \begin{pmatrix} (a\Box b)\circ Z_1 \\ \vdots \\ (a\Box b)\circ Z_n \end{pmatrix} = \begin{pmatrix} (a\circ Z_1)\Box(b\circ Z_1) \\ \vdots \\ (a\circ Z_n)\Box(b\circ Z_n) \end{pmatrix}$$

by the second distributive law in $\mathcal{F}$. This, in turn, equals

$$\begin{pmatrix} a\circ Z_1 \\ \vdots \\ a\circ Z_n \end{pmatrix} \Delta \begin{pmatrix} b\circ Z_1 \\ \vdots \\ b\circ Z_n \end{pmatrix} = (a\star x)\Delta(b\star x)$$

Thus the second distributive law in V is satisfied.

(i)

$$a\star(b\star x) = a\star \begin{pmatrix} b\circ Z_1 \\ \vdots \\ b\circ Z_n \end{pmatrix}$$

$$= \begin{pmatrix} a\circ(b\circ Z_1) \\ \vdots \\ a\circ(b\circ Z_n) \end{pmatrix} = \begin{pmatrix} (a\circ b)\circ Z_1 \\ \vdots \\ (a\circ b)\circ Z_n \end{pmatrix}$$

by associativity of $\circ$ in $\mathcal{F}$, which gives us $(a\circ b)\star x$. Thus the associative law in V is satisfied.

(j) The identity element

$$e_o = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

in $\hat{G}$ under o satisfies the scalar identity in V :

$$e_o \star x = \begin{pmatrix} e_o \circ Z_1 \\ \vdots \\ e_o \circ Z_n \end{pmatrix} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = x$$

Conclude that we successfully defined a linear vector space $(\hat{V}, \Delta, \star, \mathcal{F}) = V$.

We have shown that the vector space

$$(\mathbf{C}^n, +, \cdot, \mathbf{C}, +, \cdot)$$

is isomorphic to

$$(\mathbf{R}^{2n\times 2}, +, \cdot, \mathbf{R}^{2\times 2}, +, \cdot)$$

where the operators $+$ and $\cdot$ are context sensitive. One valid observation is that developing complex vector theory using only real numbers is really more complex than sticking with complex numbers. We have shown that $\mathbf{C}^n$, *as a vector space*, is *not* isomorphic to $\mathbf{R}^{2n}$. This omits a lot of structure present in $\mathbf{C}^n$. As a curiosity, we also showed that under proper conditions, special kinds of matrices can be vectors and *scalars*.

J.4 Construction of Vector Space in $\mathbf{R}^{2n}$ Isomorphic to $\mathbf{C}^n$

I want to construct a vector space in $\mathbf{R}^{2n}$ that is isomorphic to a vector space in $\mathbf{C}^n$ where the vector space $\mathbf{C}^n$ over the field $\mathbf{C}$ has the usual complex addition and multiplication operators.

Let elements of G_R be in $\mathbf{R}^2$, such as

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

which correspond to elements of G_C in $\mathbf{C}$ such as $a + ib$. Then the addition operator $\square_R$ defined by

$$\begin{pmatrix} a \\ b \end{pmatrix} \square \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d \end{pmatrix}$$

where $+$ is ordinary addition corresponds to $\square_C$ defined by

$$(a + ib)\square_C(c + id) = (a + c) + i(b + d)$$

The multiplication operator $\circ_R$ defined by

$$\begin{pmatrix} a \\ b \end{pmatrix} \circ_R \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a \cdot c - b \cdot d \\ b \cdot c + a \cdot d \end{pmatrix}$$

where $+$ and $-$ are ordinary addition and subtraction, and $\cdot$ is ordinary multiplication, which corresponds to $\circ_C$ defined by

$$(a + ib) \circ_c (c + id) = (ac - bd) + i(bc + ad)$$

Let $V_R \subset \mathbf{R}^{2n}$ have typical elements like

$$\begin{pmatrix} a_1 \\ b_1 \\ \vdots \\ a_n \\ b_n \end{pmatrix}$$

to correspond to $V_C \subset \mathbf{C}^n$ with elements that look like

$$\begin{pmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{pmatrix}$$

Define operator Δ_R to be ordinary real element-wise addition between elements of V_R, and Δ_C to be ordinary complex element-wise addition between elements of V_C. Finally, define the operator $\star_R$ between elements of V_R and G_R to be

$$\begin{pmatrix} x \\ y \end{pmatrix} \star_R \begin{pmatrix} a_1 \\ b_1 \\ \vdots \\ a_n \\ b_n \end{pmatrix} = \begin{pmatrix} x \cdot a_1 - y \cdot b_1 \\ x \cdot b_1 + y \cdot a_1 \\ \vdots \\ x \cdot a_n - y \cdot b_n \\ x \cdot b_n + y \cdot a_n \end{pmatrix}$$

where $+, -, \cdot$ are the usual real arithmetic operators, which corresponds to the

operator $\star_C$ defined by

$$(x+iy)\star_C \begin{pmatrix} a_1+ib_1 \\ \vdots \\ a_n+ib_n \end{pmatrix} = \begin{pmatrix} (x\cdot a_1 - y\cdot b_1)+i(x\cdot b_1+y\cdot a_1) \\ \vdots \\ (x\cdot a_n - y\cdot b_n)+i(x\cdot b_n+y\cdot a_n) \end{pmatrix}$$

With these definitions, then vector space V_R is isomorphic to V_C. In this sense, $\mathbf{R}^{2n} \simeq \mathbf{C}^n$.

J.5 A More General Vector Space

In this part, we define a vector space in a way that buys us more freedom that any of the previous definitions. Vector spaces usually require that the vector be constructed from elements of the field of scalars, such that $\mathcal{F} \subset S$ and $V \subset S^n$. For example, our experience is most frequently with the structure $\mathcal{F} \times \mathcal{F}^n$. However, this is not a necessary restriction.

Let $(\mathcal{F},+,\cdot)$ be a field with additive identity 0 and multiplicative identity 1. Let $(V,\bowtie)$ be an abelian group with identity $\hat{0}$. A vector space is the Cartesian product $\mathcal{F} \times V$ with elements $x = (a,v)$ and operators $\oplus$ and $\odot$, denoted by

$$\mathcal{V} = (\mathcal{F} \times V, +, \cdot, \bowtie, \oplus, \odot)$$

or by

$$\mathcal{V} = (\mathcal{F} \times V, \oplus, \odot)$$

with the following properties.

(1) $\oplus : (\mathcal{F} \times V\ , \mathcal{F} \times V\ ) \to \mathcal{F} \times V.$

(2) $\odot : (\mathcal{F}, \mathcal{F} \times V) \to \mathcal{F} \times V.$

(3) $x \oplus y = y \oplus x$ for all $x, y \in \mathcal{F} \times V.$

(4) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ for all $x, y, z \in \mathcal{F} \times V.$

(5) [Key point] Let $0 = \{(0, v), (a, 0)\}$ for all $v \in V$ and $a \in \mathcal{F}$. Then

(a) $0 \odot x = a \odot e$ for all $x \in \mathcal{F} \times V$, $a \in \mathcal{F}$, $e \in 0$.

(b) $e \oplus x = x$ for all $x \in \mathcal{F} \times V$, $e \in 0$.

(6) [Key point] Let $-X = \{(-a, v), (a, \ominus v)\}$ where $-a$ is the inverse of a in $(\mathcal{F}, +, \cdot)$ under $+$, and $\ominus v$ is the inverse of v in $(V, \bowtie)$. Then for each $x = (a, v) \in \mathcal{F} \times V$ there exist $-X$ such that $x \oplus (-x) \in 0$ for all $-x \in -X$ associated with x.

(7) $a \odot (x \oplus y) = (a \odot x) \oplus (a \odot y)$ for all $a \in \mathcal{F}$ and for all $x, y \in \mathcal{F} \times V.$

(8) $(a + b) \odot x = (a \odot x) \oplus (b \odot x)$ for all $a, b \in \mathcal{F}$ and for all $x \in \mathcal{F} \times V.$

(9) $1 \odot x = x$ for all $x \in \mathcal{F} \times V.$

(10) $(a \cdot b) \odot x = a \odot (b \odot x)$ for all $a, b \in \mathcal{F}$ and for all $x \in \mathcal{F} \times V.$

Appendix K

COMPLEX MATRIX ALGEBRA

The results contained here are mecl anical in nature. Proofs of the partitioned matrix determinant are interesting enough to go though once, but not to commit to memory. Other theorem statements should be read, but the proofs are mundane and time is better spent on other material. These proofs are included because I had to do them to extend results from the real case to the complex case.

K.1 Basic Definitions

Definition 64 *Matrix A is* Hermitian *if it equals the transpose of its complex conjugate,* A^{H}, *which is called the* Hermitian transpose.

Definition 65 *Matrix A is called* positive semidefinite *or* nonnegative definite *if* $x^H Ax \geq 0$ *for all* x.

Definition 66 *Matrix A is called* positive definite *if* $x^H Ax > 0$ *for all nonzero* x.

Definition 67 *Matrix A is called* negative semidefinite *or* nonpositive definite *if* $x^H Ax \leq 0$ *for all* x.

Definition 68 *Matrix A is called* negative definite *if* $x^H Ax < 0$ *for all nonzero* x.

Definition 69 *Matrix A is called* definite *if it is either positive definite or negative definite.*

Definition 70 *Matrix A is called* indefinite *if* $x^H Ax > 0$ *for some* x, *and* $x^H Ax < 0$ *for some other* x. *It is also possible for* $x^H Ax = 0$ *for some* x.

Definition 71 *Let* $A \in \mathbf{C}^{n \times n}$. *Then A is called* orthogonal *if* $AA^H = A^H A = D$ *is a diagonal matrix, not necessarily the identity matrix.*

Definition 72 *Let* $A \in \mathbf{C}^{n \times n}$. *Then A is called* complex orthogonal *if* $AA^T = A^T A = D$ *is a diagonal matrix, not necessarily the identity matrix.*

Definition 73 *Let* $A \in \mathbf{C}^{n \times n}$. *Then A is called* orthonormal *if* $AA^T = A^T A = I_n$.

Definition 74 *Let* $A \in \mathbf{C}^{n \times n}$. *Then A is* unitary *if* $AA^H = A^H A = I_n$, *and thus* $A^{-1} = A^H$.

Most authors call such a matrix orthonormal, using the same terminology they use for real matrices. Many authors refer to this as orthogonal. Thus a complex square matrix with orthonormal columns is called a unitary matrix. (Stewart, p. 259.) [259].

Example 9 *From Nomizu (p. 250) [193], the following example illustrates that orthogonal and unitary are not the same concepts. Let*

$$Y = \begin{bmatrix} i & \sqrt{2} \\ \sqrt{2} & -i \end{bmatrix}$$

Then $Y^T Y = I$, *but*

$$Y^H Y = \begin{bmatrix} 3 & -i2\sqrt{2} \\ i2\sqrt{2} & 3 \end{bmatrix}$$

Thus Y *is complex orthogonal, but* Y *is not unitary.*

$$X = \begin{bmatrix} e^{i\omega} & 0 \\ 0 & e^{i\omega} \end{bmatrix}$$

for $\omega \neq m\pi$ *is unitary but not complex orthogonal. However, an orthogonal real matrix is unitary.*

Definition 75 *Let* $A \in \mathbf{C}^{m\times n}$. *Then* A *is* subunitary *if* $AA^H = I_m$ *and* $m < n$, *or if* $A^H A = I_n$ *and* $m > n$. *Here* A^{-1} *is undefined.*

Definition 76 *Matrix* A *is called* skew-symmetric *if* $A = -A^T$.

The covariance matrix of the complex normal distribution, when expressed in $\mathbf{R}^{2n\times 2n}$ is skew-Symmetric. When expressed in $\mathbf{C}^{n\times n}$, the covariance matrix of the complex normal distribution is Hermitian.

Definition 77 *Matrix* A *is called* skew-Hermitian *if* $A = -A^H$.

Properties. The elements on the main diagonal are either pure imaginary or zero. The imaginary part of A is symmetric. $\text{Im}(A) = [\text{Im}(A)]^T$. As for the real part, $\text{Re}(A) = -[\text{Re}(A)]^T$. The diagonal of $\text{Re}(A)$ is zero. A is skew-Hermitian if $-iA$ is Hermitian.

Suppose $A = -A^H$. Consider BAB^H. Take its Hermitian transpose.

$$(BAB^H)^H = BA^HB^H = -BAB^H$$

Thus, BAB^H is also skew-Hermitian.

Example 10 *Let*

$$\Phi = \begin{pmatrix} a+ib & e+if \\ g+ih & c+id \end{pmatrix}$$

Then

$$-\Phi^H = \begin{pmatrix} -a+ib & -g+ih \\ -e+if & -c+id \end{pmatrix}$$

For $\Phi = -\Phi^H$, *then we must have* $a = -a = 0$, $c = -c = 0$, $g = -e$, *and* $f = h$. *Thus we have*

$$\Phi = \begin{pmatrix} ib & e+if \\ -e+if & id \end{pmatrix}$$

Proposition 48 *If A is positive semidefinite, then A is Hermitian.*

Proof. A is positive semidefinite (or, A is nonnegative definite) means that $x^HAx \geq 0$ for all x. Since $x^HAx \geq 0$, it must be real. So $(x^HAx)^H = x^HAx$. This implies

$$x^HA^Hx - x^HAx = x^H(A^H - A)x = 0$$

for *all* x. Thus $A = A^H$. □

K.2 Trace Identities

The various trace identities that have been worked out are ones that have been required one or more times in subsequent work. This is particularly true for those used in the evaluation of moments from characteristic functions.

Lemma 25 *Let* $W, A \in \mathbf{C}^{n\times n}$, *where* W *and* A *have no required structure. Then*

$$WAW = \left(\left(\sum_{l=1}^{n} \sum_{j=1}^{n} A_{lj} W_{il} W_{jm} \right)_{im} \right)$$

Proof. Let $B = CAT$ where $A, B, C, T \in \mathbf{C}^{n\times n}$. Partition C and T as $T = (T_1, T_2, \cdots, T_n)$, and

$$C = \begin{pmatrix} C^1 \\ \vdots \\ C^n \end{pmatrix}$$

Then

$$B = \begin{pmatrix} C^1AT_1 & C^1AT_2 & \cdots & C^1AT_n \\ C^2AT_1 & C^2AT_2 & \cdots & C^2AT_n \\ \vdots & \vdots & \ddots & \vdots \\ C^nAT_1 & C^nAT_2 & \cdots & C^nAT_n \end{pmatrix}$$

Consider element B_{im}, computed by

$$B_{im} = C^i A T_m = (C^{i1}, \cdots, C^{in}) \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} T_{1m} \\ \vdots \\ T_{nm} \end{pmatrix}$$

$$= (C^{i1}, \cdots, C^{in}) \begin{pmatrix} \sum_{j=1}^{n} A_{1j} T_{jm} \\ \vdots \\ \sum_{j=1}^{n} A_{nj} T_{jm} \end{pmatrix} = \sum_{l=1}^{n} \sum_{j=1}^{n} C^{il} A_{lj} T_{jm}$$

Now let $C = T = W$. Then

$$B_{im} = \sum_{l=1}^{n} \sum_{j=1}^{n} A_{lj} W_{il} W_{jm}$$

□

Lemma 26 *Let $B_{m \times n}$ and $Z_{n \times m}$ be complex matrices. Then*

$$\mathrm{tr}(BZ) = \sum_{j=1}^{m} \sum_{k=1}^{n} B_{jk} Z_{kj}$$

Proof.

$$\mathrm{tr}(BZ) = \mathrm{tr} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1m} \\ Z_{21} & Z_{22} & \cdots & Z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nm} \end{pmatrix}$$

$$= \sum_{k=1}^{n} B_{1k} Z_{k1} + \sum_{k=1}^{n} B_{2k} Z_{k2} + \cdots + \sum_{k=1}^{n} B_{mk} Z_{km} = \sum_{j=1}^{m} \sum_{k=1}^{n} B_{jk} Z_{kj}$$

□

Lemma 27 *Let $B_{m\times n}$ and $Z_{m\times n}$ be complex matrices. Then*

$$\operatorname{tr}(B^T Z) = \sum_{j=1}^{m}\sum_{k=1}^{n} B_{jk}Z_{jk} = \operatorname{tr}(ZB^T) = \operatorname{tr}(Z^T B) = \operatorname{tr}(BZ^T)$$

Proof.

$$\operatorname{tr}(B^T Z) = \operatorname{tr}\begin{pmatrix} B_{11} & B_{21} & \cdots & B_{m1} \\ B_{12} & B_{22} & \cdots & B_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1n} & B_{2n} & \cdots & B_{mn} \end{pmatrix}\begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{pmatrix}$$

$$= \sum_{j=1}^{m} B_{j1}Z_{j1} + \sum_{j=1}^{m} B_{j2}Z_{j2} + \cdots + \sum_{j=1}^{m} B_{jn}Z_{jn}$$

$$= \sum_{k=1}^{n}\sum_{j=1}^{m} B_{jk}Z_{jk} = \sum_{j=1}^{m}\sum_{k=1}^{n} B_{jk}Z_{jk}$$

$$= \sum_{k=1}^{n} B_{1k}Z_{1k} + \sum_{k=1}^{n} B_{2k}Z_{2k} + \cdots + \sum_{k=1}^{n} B_{mk}Z_{mk}$$

$$= \operatorname{tr}\begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{pmatrix}\begin{pmatrix} B_{11} & B_{21} & \cdots & B_{m1} \\ B_{12} & B_{22} & \cdots & B_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1n} & B_{2n} & \cdots & B_{mn} \end{pmatrix} = \operatorname{tr}(ZB^T)$$

Because $\operatorname{tr} A = \operatorname{tr} A^T$, we also have

$$\operatorname{tr}(B^T Z) = \operatorname{tr}(ZB^T) = \operatorname{tr}(Z^T B) = \operatorname{tr}(BZ^T)$$

□

Lemma 28 *Let $B, T \in \mathbf{C}^{n\times m}$ be complex matrices. Then*

$$\operatorname{tr}(BT^T) = \sum_{l=1}^{n}\sum_{k=1}^{m} B_{lk}T_{lk}$$

Proof.

$$\mathrm{tr}(BT^T) = \mathrm{tr}\left[\begin{pmatrix} B_{11} & \cdots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{nm} \end{pmatrix}\begin{pmatrix} T_{11} & \cdots & T_{n1} \\ \vdots & \ddots & \vdots \\ T_{1m} & \cdots & T_{nm} \end{pmatrix}\right] = \mathrm{tr}\, C$$

where $C = BT^T$. Element (i,j) of $\mathrm{tr}(BT^T)$ is

$$C_{ij} = \sum_{k=1}^{m} B_{ik}T_{jk}$$

Then

$$\mathrm{tr}(BT^T) = \sum_{l=1}^{n} C_{ll} = \sum_{l=1}^{n} B_{lk}T_{lk}$$

□

Lemma 29 *Let $B_{n\times m}$ and $Z_{n\times m}$ be complex matrices. Then*

$$\mathrm{tr}(B^H Z) = \sum_{j=1}^{m}\sum_{k=1}^{n} B^*_{kj}Z_{kj}$$

Proof.

$$\mathrm{tr}(B^H Z) = \mathrm{tr}\left[\begin{pmatrix} B^*_{11} & B^*_{21} & \cdots & B^*_{n1} \\ B^*_{12} & B^*_{22} & \cdots & B^*_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ B^*_{1m} & B^*_{2m} & \cdots & B^*_{nm} \end{pmatrix}\begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1m} \\ Z_{21} & Z_{22} & \cdots & Z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nm} \end{pmatrix}\right]$$

$$= \sum_{k=1}^{n} B^*_{k1}Z_{k1} + \sum_{k=1}^{n} B^*_{k2}Z_{k2} + \cdots + \sum_{k=1}^{n} B^*_{km}Z_{km} = \sum_{j=1}^{m}\sum_{k=1}^{n} B^*_{kj}Z_{kj}$$

□

Proposition 49 *Let $B = B^H$ and $Z = Z^H$ be $n \times n$ complex matrices. Then*

$$\mathrm{tr}(BZ) = \sum_{k=1}^{n} B_{kk} Z_{kk} + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} 2\,\mathrm{Re}\{B_{jk} Z_{kj}\}$$

Proof.

$$\mathrm{tr}(BZ) = \mathrm{tr}\left[\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{12}^* & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{1n}^* & B_{2n}^* & \cdots & B_{nn} \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{21}^* & \cdots & Z_{n1}^* \\ Z_{21} & Z_{22} & \cdots & Z_{n2}^* \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{pmatrix} \right]$$

$$= \sum_{k=1}^{n} B_{1k} Z_{k1} + B_{12}^* Z_{21}^* + \sum_{k=2}^{n} B_{2k} Z_{k2} + \cdots + \sum_{k=1}^{n-1} B_{kn}^* Z_{nk}^* + B_{nn} Z_{nn}$$

$$= \sum_{k=1}^{n} B_{kk} Z_{kk} + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} 2\,\mathrm{Re}(B_{jk} Z_{kj})$$

□

Lemma 30

$$\mathrm{tr}(Z^2) = \mathrm{tr}(ZZ) = \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ji} Z_{ij}$$

Proof.

$$\mathrm{tr}(ZZ) = \mathrm{tr}(Z^2) = \mathrm{tr}\left[\begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{pmatrix} \right]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ji} Z_{ij} = \sum_{j=1}^{n} \sum_{i=1}^{n} Z_{ji} Z_{ij}$$

□

Proposition 50

$$\mathrm{tr}(ZZ^T) = \mathrm{tr}(Z^TZ) = \sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}^2$$

Proof. Let

$$Z = (Z_1, \cdots, Z_n) = \begin{pmatrix} Z^1 \\ \vdots \\ Z^n \end{pmatrix}$$

Z_i are column vectors and the Z^j are row vectors. Then

$$\mathrm{tr}(Z^TZ) = \mathrm{tr}\left[\begin{pmatrix} Z_1^T \\ \vdots \\ Z_n^T \end{pmatrix}\begin{pmatrix} Z_1 & \cdots & Z_n \end{pmatrix}\right] = \mathrm{tr}\begin{pmatrix} Z_1^TZ_1 & \cdots & Z_1^TZ_n \\ \vdots & \ddots & \vdots \\ Z_n^TZ_1 & \cdots & Z_n^TZ_n \end{pmatrix}$$

$$= \sum_{j=1}^{n} Z_j^TZ_j = \sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}^2$$

□

Proposition 51 *Let* $Z \in \mathbf{C}^{p\times n}$. *Then*

$$\mathrm{tr}(Z^HZ) = \sum_{i=1}^{p}\sum_{j=1}^{n} |Z_{ij}|^2$$

Proof.

$$\mathrm{tr}(Z^HZ) = \mathrm{tr}\left[\begin{pmatrix} Z_{11}^* & Z_{21}^* & \cdots & Z_{p1}^* \\ Z_{12}^* & Z_{22}^* & \cdots & Z_{p2}^* \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1n}^* & Z_{2n}^* & \cdots & Z_{pn}^* \end{pmatrix}\begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{p1} & Z_{p2} & \cdots & Z_{pn} \end{pmatrix}\right]$$

Compute only the elements ending up on the diagonal. $\mathrm{tr}(Z^H Z) =$

$$\begin{bmatrix} Z_{11}^* Z_{11} + Z_{21}^* Z_{21} & & & \\ + \cdots + Z_{p1}^* Z_{p1} & & & \\ & Z_{12}^* Z_{12} + Z_{22}^* Z_{22} & & \\ & + \cdots + Z_{p2}^* Z_{p2} & & \\ & & \ddots & \\ & & & Z_{1n}^* Z_{1n} + Z_{2n}^* Z_{2n} \\ & & & + \cdots + Z_{pn}^* Z_{pn} \end{bmatrix}$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{n} Z_{ij}^* Z_{ij} = \sum_{i=1}^{p} \sum_{j=1}^{n} |Z_{ij}|^2 = \|Z\|_F^2$$

where $\|Z\|_F^2$ is the Frobenius norm for the matrix Z.

Alternate Proof. Let $Z = (Z_1, \cdots, Z_n)$ where $Z_i \in \mathbf{C}^p$. Then

$$\mathrm{tr}(Z^H Z) = \mathrm{tr}\left[\begin{pmatrix} Z_1^H \\ \vdots \\ Z_n^N \end{pmatrix} \begin{pmatrix} Z_1 & \cdots & Z_n \end{pmatrix} \right] = \mathrm{tr} \begin{pmatrix} Z_1^H Z_1 & \cdots & Z_1^H Z_n \\ \vdots & \ddots & \vdots \\ Z_n^H Z_1 & \cdots & Z_n^H Z_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} Z_i^H Z_i = \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ji}^* Z_{ji} = \sum_{i=1}^{n} \sum_{j=1}^{p} |Z_{ji}|^2$$

□

Proposition 52 *The two identities presented here are ones that occur in the complex matrix normal distribution. The derivation is so short that it will be*

included in the identity statement.

$$\Xi = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} A^*_{11} & A^*_{21} & \cdots & A^*_{n1} \\ A^*_{12} & A^*_{22} & \cdots & A^*_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A^*_{1n} & A^*_{2n} & \cdots & A^*_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{n} A_{1k}A^*_{1k} & \sum_{k=1}^{n} A_{1k}A^*_{2k} & \cdots & \sum_{k=1}^{n} A_{1k}A^*_{nk} \\ \sum_{k=1}^{n} A_{2k}A^*_{1k} & \sum_{k=1}^{n} A_{2k}A^*_{2k} & \cdots & \sum_{k=1}^{n} A_{2k}A^*_{nk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} A_{nk}A^*_{1k} & \sum_{k=1}^{n} A_{nk}A^*_{2k} & \cdots & \sum_{k=1}^{n} A_{nk}A^*_{nk} \end{bmatrix}$$

Therefore

$$\Xi_{ii'} = \sum_{k=1}^{n} A_{ik}A^*_{i'k}$$

$$\Sigma = \begin{bmatrix} B^*_{11} & B^*_{21} & \cdots & B^*_{n1} \\ B^*_{12} & B^*_{22} & \cdots & B^*_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ B^*_{1n} & B^*_{2n} & \cdots & B^*_{nn} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{s=1}^{p} B^*_{s1}B_{s1} & \sum_{s=1}^{p} B^*_{s1}B_{s2} & \cdots & \sum_{s=1}^{p} B^*_{s1}B_{sp} \\ \sum_{s=1}^{p} B^*_{s2}B_{s1} & \sum_{s=1}^{p} B^*_{s2}B_{s2} & \cdots & \sum_{s=1}^{p} B^*_{s2}B_{sp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{s=1}^{p} B^*_{sp}B_{s1} & \sum_{s=1}^{p} B^*_{sp}B_{s2} & \cdots & \sum_{s=1}^{p} B^*_{sp}B_{sp} \end{bmatrix}$$

Therefore

$$\Sigma_{jj'} = \sum_{s=1}^{p} B_{sj}^{*} B_{sj'}$$

□

Lemma 31 *Let* $A \in \mathbf{C}^{m\times p}$, $Z \in \mathbf{C}^{p\times q}$, $B \in \mathbf{C}^{q\times m}$. *Then*

$$\mathrm{tr}(AZB) = \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{q} A_{ij} Z_{jk} B_{ki}$$

Proof.

$$\mathrm{tr}(AZB) = \mathrm{tr}\left[\begin{pmatrix} A^1 \\ \vdots \\ A^m \end{pmatrix} Z \begin{pmatrix} B_1 & \cdots & B_m \end{pmatrix}\right]$$

where A^i is a row vector and B_j is a column vector. Then

$$\mathrm{tr}(AZB) = \mathrm{tr}\begin{bmatrix} A^1ZB_1 & A^1ZB_2 & \cdots & A^1ZB_m \\ A^2ZB_1 & A^2ZB_2 & \cdots & A^2ZB_m \\ \vdots & \vdots & \ddots & \vdots \\ A^mZB_1 & A^mZB_2 & \cdots & A^mZB_m \end{bmatrix}$$

$$= \sum_{i=1}^{m} A^i Z B_i = \sum_{i=1}^{m} \begin{pmatrix} A_{i1} & \cdots & A_{ip} \end{pmatrix} \begin{pmatrix} Z_{11} & \cdots & Z_{1q} \\ \vdots & \ddots & \vdots \\ Z_{p1} & \cdots & Z_{pq} \end{pmatrix} \begin{pmatrix} B_{1i} \\ \vdots \\ B_{qi} \end{pmatrix}$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{p} A_{ij}Z_{j1}, \sum_{j=1}^{p} A_{ij}Z_{j2}, \cdots, \sum_{j=1}^{p} A_{ij}Z_{jq}\right) \begin{pmatrix} B_{1i} \\ \vdots \\ B_{qi} \end{pmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{q} A_{ij}Z_{jk}B_{ki}$$

□

Lemma 32 *Let $A \in \mathbf{C}^{m\times p}$, $Z \in \mathbf{C}^{q\times p}$, $B \in \mathbf{C}^{q\times m}$. Then*

$$\mathrm{tr}(AZ^TB) = \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{q} A_{ij}Z_{kj}B_{ki}$$

Proof.

$$\mathrm{tr}(AZ^TB) = \mathrm{tr}\left[\begin{pmatrix} A^1 \\ \vdots \\ A^m \end{pmatrix} Z^T \begin{pmatrix} B_1 & \cdots & B_m \end{pmatrix}\right]$$

where A^i is a row vector and B_j is a column vector. Then

$$\mathrm{tr}(AZ^TB) = \mathrm{tr}\begin{bmatrix} A^1Z^TB_1 & A^1Z^TB_2 & \cdots & A^1Z^TB_m \\ A^2Z^TB_1 & A^2Z^TB_2 & \cdots & A^2Z^TB_m \\ \vdots & \vdots & \ddots & \vdots \\ A^mZ^TB_1 & A^mZ^TB_2 & \cdots & A^mZ^TB_m \end{bmatrix}$$

$$= \sum_{i=1}^{m} A^iZ^TB_i = \sum_{i=1}^{m} \begin{pmatrix} A_{i1} & \cdots & A_{ip} \end{pmatrix} \begin{pmatrix} Z_{11} & \cdots & Z_{q1} \\ \vdots & \ddots & \vdots \\ Z_{1p} & \cdots & Z_{qp} \end{pmatrix} \begin{pmatrix} B_{1i} \\ \vdots \\ B_{qi} \end{pmatrix}$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{p} A_{ij}Z_{1j}, \cdots, \sum_{j=1}^{p} A_{ij}Z_{qj}\right) \begin{pmatrix} B_{1i} \\ \vdots \\ B_{qi} \end{pmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{p}\sum_{k=1}^{q} A_{ij}Z_{kj}B_{ki}$$

□

Lemma 33 *Let A, B, C, D all be $p \times p$ (complex) matrices. Then*

$$\mathrm{tr}(AB^HCD) = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}B_j^HCD_i$$

where $B = [B_1, B_2, \cdots, B_p]$ *and* $D = [D_1, D_2, \cdots, D_p]$.

Proof.

$$\operatorname{tr}(AB^HCD) = \operatorname{tr}\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{pmatrix} \begin{pmatrix} B_1^H \\ B_2^H \\ \vdots \\ B_p^H \end{pmatrix} C \begin{pmatrix} D_1 & D_2 & \cdots & D_p \end{pmatrix}$$

$$= \operatorname{tr}\left[\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{pmatrix} \begin{pmatrix} B_1^HCD_1 & B_1^HCD_2 & \cdots & B_1^HCD_p \\ B_2^HCD_1 & B_2^HCD_2 & \cdots & B_2^HCD_p \\ \vdots & \vdots & \ddots & \vdots \\ B_p^HCD_1 & B_p^HCD_2 & \cdots & B_p^HCD_p \end{pmatrix} \right]$$

I do not have to do all the computations. I only need the sum of the diagonal elements of the product. therefore

$$\operatorname{tr}(AB^HCD) = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}B_j^HCD_i$$

Notice that the order of subscripts reverse. □

Proposition 53 *Let* A, F, C, D *all be* $p \times p$ *(complex) matrices. Then*

$$\operatorname{tr}(AFCD) = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}F^jCD_i$$

where

$$F = \begin{bmatrix} F^1 \\ F^2 \\ \vdots \\ F^p \end{bmatrix}$$

and $D = [D_1, D_2, \cdots, D_p]$.

Proof. This is merely (but a useful) corollary to lemma 33. □

Proposition 54 *Let* A, B, C, D *all be* $p \times p$ *(complex) matrices. Then*

$$\mathrm{tr}(A^H BCD) = \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ji}^* B_j C D_i$$

Proof.

$$\mathrm{tr}(A^H BCD) = \mathrm{tr} \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pp} \end{pmatrix}^H \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_p \end{pmatrix} C \begin{pmatrix} D_1 & D_2 & \cdots & D_p \end{pmatrix}$$

where

$$B = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_p \end{pmatrix}$$

and $D = \begin{pmatrix} D_1 & D_2 & \cdots & D_p \end{pmatrix}$. Then

$$\mathrm{tr}(A^H BCD) = \mathrm{tr}\left[\begin{pmatrix} A_{11}^* & A_{21}^* & \cdots & A_{p1}^* \\ A_{12}^* & A_{22}^* & \cdots & A_{p2}^* \\ \vdots & \vdots & \ddots & \vdots \\ A_{1p}^* & A_{2p}^* & \cdots & A_{pp}^* \end{pmatrix} \begin{pmatrix} B_1CD_1 & B_1CD_2 & \cdots & B_1CD_p \\ B_2CD_1 & B_2CD_2 & \cdots & B_2CD_p \\ \vdots & \vdots & \ddots & \vdots \\ B_pCD_1 & B_pCD_2 & \cdots & B_pCD_p \end{pmatrix} \right]$$

The sum of the diagonal elements is all that is needed.

$$\mathrm{tr}(A^H BCD) = \sum_{i=1}^{p} \sum_{j=1}^{p} A_{ji}^* B_j C D_i = \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} A_{ji}^* B_{jl} C_{lk} D_{ki}$$

□

K.3 Inverse

K.3.1 Partitioned Matrix Inverse

Lemma 34 *Partitioned Matrix Right Inverse. Let $Z \in M_n(\mathbf{C})$ be partitioned*

$$Z = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

Let $ZY = I_n$ where

$$Y = \begin{pmatrix} R & T \\ S & U \end{pmatrix}$$

Y is the Right Inverse *of Z. Then*

$$Z^{-1} = Y = \begin{bmatrix} (A - CD^{-1}B)^{-1} & -A^{-1}C(D - BA^{-1}C)^{-1} \\ -D^{-1}B(A - CD^{-1}B)^{-1} & (D - BA^{-1}C)^{-1} \end{bmatrix}$$

$$= \left[\begin{pmatrix} I_1 \\ -D^{-1}B \end{pmatrix} (A - CD^{-1}B)^{-1} \quad \begin{pmatrix} -A^{-1}C \\ I_2 \end{pmatrix} (D - BA^{-1}C)^{-1} \right]$$

A and D must be square matrices. This is Graybill theorem 8.2.1 [95].

Proof. Although I did the following proof, it is a common and easy proof that must have been done by many people.

$$ZY = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} R & T \\ S & U \end{pmatrix} = \begin{pmatrix} AR + CS & AT + CU \\ BR + DS & BT + DU \end{pmatrix}$$

$$= \begin{pmatrix} I_1 & 0_2 \\ 0_1 & I_2 \end{pmatrix} = I_n$$

This implies

$$DS = -BR \Rightarrow S = -D^{-1}BR$$

and

$$AT = -CU \Rightarrow T = -A^{-1}CU$$

Substituting into the main block diagonal terms,

$$AR + C(-D^{-1}BR) = (A - CD^{-1}B)R = I_1 \Rightarrow R = (A - CD^{-1}B)^{-1}$$

and

$$B(-A^{-1}CU) + DU = (D - BA^{-1}C)U = I_2 \Rightarrow U = (D - BA^{-1}C)^{-1}$$

Substituting back into Y, we obtain $Z^{-1} = Y =$

$$\left[\begin{pmatrix} I_1 \\ -D^{-1}B \end{pmatrix} (A - CD^{-1}B)^{-1} \quad \begin{pmatrix} -A^{-1}C \\ I_2 \end{pmatrix} (D - BA^{-1}C)^{-1} \right]$$

□

Lemma 35 *Partitioned Matrix Left Inverse. Let* $Z \in M_n(\mathbf{C})$ *be partitioned*

$$Z = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

Let $ZY = I_n$ *where*

$$Y = \begin{pmatrix} R & T \\ S & U \end{pmatrix}$$

Y *is the* Left Inverse *of* Z. *Then*

$$Z^{-1} = Y = \begin{bmatrix} (A - CD^{-1}B)^{-1}(I_1, -CD^{-1}) \\ (D - BA^{-1}C)^{-1}(-BA^{-1}, I_2) \end{bmatrix}$$

A and D must be square matrices. Although I did this theorem and its proof, it is so basic that it must have been done before.

Proof.

$$YZ = \begin{pmatrix} R & T \\ S & U \end{pmatrix} \begin{pmatrix} A & C \\ B & D \end{pmatrix} = \begin{pmatrix} RA + TB & RC + TD \\ SA + UB & SC + UD \end{pmatrix}$$

$$= \begin{pmatrix} I_1 & 0_2 \\ 0_1 & I_2 \end{pmatrix} = I_n$$

This implies

$$SA = -UB \Rightarrow S = -UBA^{-1}$$

and

$$TD = -RC \Rightarrow T = -RCD^{-1}$$

Substituting back into the main block diagonal terms,

$$RA - RCD^{-1}B = R(A - CD^{-1}B) = I_1 \Rightarrow R = (A - CD^{-1}B)^{-1}$$

and

$$-UBA^{-1}C + UD = U(D - BA^{-1}C) = I_2 \Rightarrow U = (D - BA^{-1}C)^{-1}$$

Substituting back into Y, we obtain

$$Z^{-1} = Y = \begin{bmatrix} (A - CD^{-1}B)^{-1} & -(A - CD^{-1}B)^{-1}CD^{-1} \\ -(D - BA^{-1}C)^{-1}BA^{-1} & (D - BA^{-1}C)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} (A - CD^{-1}B)^{-1}(I_1, -CD^{-1}) \\ (D - BA^{-1}C)^{-1}(-BA^{-1}, I_2) \end{bmatrix}$$

□

K.3.2 Complex Matrix Inversion Lemmas

Matrix inversion lemmas are frequently encountered in applied time series analysis. They are particularly useful when formulating Kalman filter algorithms. They are included here for the sake of completeness within the subject of complex matrix theory for acoustic signal processing.

Lemma 36 *Let A and B be $n \times n$ nonsingular complex matrices. Let Y be an $m \times m$ nonsingular complex matrix. Let X be an $m \times n$ complex matrix.*

Then the two following expressions imply each other. If one is true, then the other is true provided the necessary matrix inverses exist.

$$A^{-1} = B^{-1} + X^H Y^{-1} X \Leftrightarrow A = B - BX^H(XBX^H + Y)^{-1}XB$$

Proof. This proof follows Sinha and Kuszta's [246] proof for the case of real variables. Let

1 Let $A^{-1} = B^{-1} + X^H Y^{-1} X$

2 1 $\Rightarrow$ $A^{-1} - B^{-1} = X^H Y^{-1} X$

3 1 $\Rightarrow$ $AA^{-1}BX^H = A(B^{-1} + X^H Y^{-1} X)BX^H$

4 3 $\Rightarrow$ $BX^H = AX^H + AX^H Y^{-1} XBX^H$

5 4 $\Rightarrow$ $BX^H = AX^H Y^{-1}(Y + XBX^H)$

6 5 $\Rightarrow$ $BX^H(Y + XBX^H)^{-1}XB$

$$= AX^H Y^{-1}(Y + XBX^H)(Y + XBX^H)^{-1}XB = AX^H Y^{-1} XB$$

7 6 $\Rightarrow$ $B - BX^H(Y + XBX^H)^{-1}XB = B - A\underbrace{X^H Y^{-1} X}_{A^{-1}-B^{-1}} B$

$$= B - A(A^{-1} - B^{-1})B = B - B + A = A$$

8 $A = B - BX^H(XBX^H + Y)^{-1}XB$

□

Lemma 37 *Let A and B be $n \times n$ nonsingular complex matrices. Let Y be an $m \times m$ nonsingular complex matrix. Let X be an $m \times n$ complex matrix. Then the two following expressions imply each other. If one is true, then the other is true provided the necessary matrix inverses exist.*

$$A^{-1} = B^{-1} - X^H Y^{-1} X \Leftrightarrow A = B + BX^H(Y - XBX^H)^{-1}XB$$

Proof. This proof follows Sinha and Kuszta's [246] proof for the case of real variables. Let

1 Let $A^{-1} = B^{-1} - X^H Y^{-1} X$

2 $1 \Rightarrow$ $A^{-1} - B^{-1} = -X^H Y^{-1} X$

$B^{-1} - A^{-1} = X^H Y^{-1} X$

3 $1 \Rightarrow$ $AA^{-1}BX^H = A(B^{-1} - X^H Y^{-1} X)BX^H$

4 $3 \Rightarrow$ $BX^H = AX^H - AX^H Y^{-1} XBX^H$

5 $4 \Rightarrow$ $BX^H = AX^H Y^{-1}(Y - XBX^H)$

6 $5 \Rightarrow$ $BX^H(Y - XBX^H)^{-1}XB$

$= AX^H Y^{-1}(Y - XBX^H)(Y - XBX^H)^{-1}XB = AX^H Y^{-1} XB$

7 $6 \Rightarrow$ $B + BX^H(Y - XBX^H)^{-1}XB = B + A \underbrace{X^H Y^{-1} X}_{B^{-1} - A^{-1}} B$

$= B + A(B^{-1} - A^{-1})B = B + A - B = A$

8 $A = B + BX^H(Y - XBX^H)^{-1}XB$

□

Lemma 38 *Let A and B be $n \times n$ nonsingular complex matrices. Let X be an $n \times m$ complex matrix such that $(I - X^H BX)^{-1}$ exists. Then the two following expressions imply each other. If one is true, then the other is true provided the necessary matrix inverses exist.*

$$A^{-1} = B^{-1} - XX^H \Leftrightarrow A = B + BX(I - X^H BX)^{-1} X^H B$$

Proof. Let $A^{-1} = B^{-1} - XX^H$. Post-multiply by BX and premultiply by A to get

$$AA^{-1}BX = AB^{-1}BX - AXX^HBX$$

This implies

$$BX = AX - AXX^HBX = AX(I - X^HBX)$$

Post-multiply by $(I - X^HBX)^{-1}X^HB$ to get

$$BX(I - X^HBX)^{-1}X^HB = AXX^HB$$

Add B to both sides to get

$$B + BX(I - X^HBX)^{-1}X^HB = B + AXX^HB$$

Rearranging the initial equation, $XX^H = B^{-1} - A^{-1}$ implies

$$B+BX(I-X^HBX)^{-1}X^HB = B+A(B^{-1}-A^{-1})B = B+AB^{-1}B-AA^{-1}B = A$$

□

Lemma 39 *Let A and B be $n \times n$ nonsingular complex matrices. Let X be an $n \times m$ complex matrix such that $(I - XBX^H)^{-1}$ exists. Then the two following expressions imply each other. If one is true, then the other is true provided the necessary matrix inverses exist.*

$$A^{-1} = B^{-1} + XX^H \Leftrightarrow A = B - BX(I + X^HBX)^{-1}X^HB$$

Proof. Let $A^{-1} = B^{-1} + XX^H$. Post-multiply by BX and premultiply by A to get

$$AA^{-1}BX = AB^{-1}BX + AXX^HBX$$

This implies

$$BX = AX + AXX^HBX = AX(I + X^HBX)$$

Post-multiply by $(I + X^HBX)^{-1}X^HB$ to get

$$BX(I + X^HBX)^{-1}X^HB = AXX^HB$$

Add B to both sides to get

$$B + BX(I + X^HBX)^{-1}X^HB = B + AXX^HB$$

Rearranging the initial equation, $XX^H = B^{-1} - A^{-1}$ implies

$$B + BX(I - X^HBX)^{-1}X^HB = B + A(A^{-1} - B^{-1})B = B + B - A = 2B - A$$

Solve for A to get the final answer. □

Lemma 40 *Let $W^H = W > 0$ be partitioned in $\mathbf{C}^{p\times p}$ as*

$$W = \begin{pmatrix} \underbrace{Y^HY}_{l\times l} & \underbrace{Y^HZ}_{l\times(p-l)} \\ \underbrace{Z^HY}_{(p-l)\times l} & \underbrace{Z^HZ}_{(p-l)\times(p-l)} \end{pmatrix} = \underbrace{(Y,Z)}_{p\times k}^H \underbrace{(Y,Z)}_{k\times p}$$

where $Y \in \mathbf{C}^{k\times l}$. Z is a column vector. Then $W^{ii} \in W^{-1}$ is given by

$$W^{ii} = e_i^H(Y^HY)^{-1}e_i + \frac{\left|e_i^H(Y^HY)^{-1}Y^HZ\right|^2}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

where e_i is the standard basis vector of length l consisting of all zeros, except a 1 in position i. This is often used in deriving Kalman filters.

Proof. Let

$$W = \begin{pmatrix} A & B^H \\ B & D \end{pmatrix}$$

and

$$W^{-1} = \begin{pmatrix} R & S^H \\ S & U \end{pmatrix}$$

By lemma 34, $R^{-1} = A - B^H D^{-1} B$. By lemma 37,

$$R = A^{-1} + A^{-1}B^H(D - BA^{-1}B^H)^{-1}BA^{-1}$$

$$= (Y^HY)^{-1} + (Y^HY)^{-1}Y^HZ(Z^HZ - Z^HY(Y^HY)^{-1}Y^HZ)^{-1}Z^HY(Y^HY)^{-1}$$

$$= (Y^HY)^{-1} + (Y^HY)^{-1}Y^HZ[Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z]^{-1}Z^HY(Y^HY)^{-1}$$

Since Z is a column vector,

$$R = (Y^HY)^{-1} + \frac{(Y^HY)^{-1}Y^HZZ^HY(Y^HY)^{-1}}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

The $(i,i)^{th}$ element of R is $e_i^H R e_i$, given by

$$W^{ii} = e_i^H(Y^HY)^{-1}e_i + \frac{e_i^H(Y^HY)^{-1}Y^HZZ^HY(Y^HY)^{-1}e_i}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

$$= e_i^H(Y^HY)^{-1}e_i + \frac{\left|e_i^H(Y^HY)^{-1}Y^HZ\right|^2}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

□

Lemma 41 *Let $W^H = W > 0$ be partitioned in $\mathbf{C}^{p\times p}$ as*

$$W = \begin{pmatrix} Y^HY & Y^HZ \\ Z^HY & Z^HZ \end{pmatrix} = (Y,Z)^H(Y,Z)$$

where $Y \in \mathbf{C}^{k\times l}$. Z is a column vector. Then $W^{pp} \in W^{-1}$ is given by

$$W^{pp} = \frac{1}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

This is often used in deriving Kalman filters.

Proof. Let

$$W = \begin{pmatrix} A & B^H \\ B & D \end{pmatrix}$$

and

$$W^{-1} = \begin{pmatrix} R & S^H \\ S & U \end{pmatrix}$$

By lemma 34, $U^{-1} = D - BA^{-1}B^H$. Thus,

$$U = [Z^HZ - Z^HY(Y^HY)^{-1}Y^HZ]^{-1} = [Z^H(I - Y(Y^HY)^{-1}Y^H)Z]^{-1}$$

The $(j,j)^{th}$ element of U is given by

$$U^{jj} = e_j^HUe_j = e_j^H[Z^H(I - Y(Y^HY)^{-1}Y^H)Z]^{-1}e_j$$

$$W^{jj} = \frac{1}{Z^H\{I - Y(Y^HY)^{-1}Y^H\}Z}$$

with e_j as the standard basis vector of all zeros, except for a 1 is position j. □

K.4 Determinants

K.4.1 Basic Properties

Definition 78 *Let $X \in \mathbf{C}^{n\times n}$ have elements (X_{ij}). The matrix formed by deleting row i and column j from X is called the* minor *of X_{ij}.*

Definition 79 *Let $A \in \mathbf{C}^{n\times n}$ have elements (a_{ij}). Let X_{ij} be the minor of a_{ij}. Then*

$$A_{ij} = (-1)^{i+j} \det(X_{ij})$$

is the cofactor *of a_{ij}. Note that the cofactor is a scalar.*

Proposition 55 *Let $A \in \mathbf{C}^{n\times n}$ have elements (a_{ij}) and cofactors (α_{ij}). Then $\sum_{k=1}^{n} a_{jk}\alpha_{ik} = 0$. Similarly, $\sum_{k=1}^{n} a_{ki}\alpha_{kj} = 0$. This is Ayres problem 3.10 [35].*

Proof. This is essentially the proof given by Ayres, with statements made slightly more explicit. $\sum_{k=1}^{n} a_{ki}\alpha_{kj}$ is the determinant of some matrix. Call it B. For example, let $X^T \in \mathbf{C}^n$. Then

$$\det B = \sum_{k=1}^{n} X_k \alpha_{kj} \underset{=}{\text{def}} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,j-1} & x_1 & a_{1,j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,j-1} & x_2 & a_{2,j+2} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n1} & \cdots & a_{n,j-1} & x_n & a_{n,j+n} & \cdots & a_{nn} \end{pmatrix}$$

Substituting a_{ki} for X_k, we get

$$\det B = \sum_{k=1}^{n} a_{ki}\alpha_{kj}$$

$$= \det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} & \cdots & a_{1,j-1} & a_{1i} & a_{1,j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2i} & \cdots & a_{2,j-1} & a_{2i} & a_{2,j+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{ni} & \cdots & a_{n,j-1} & a_{ni} & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix}$$

B now has two identical columns, when $i \neq j$. We know because of this condition that its determinant is zero. [However, when X is not a linear combination of the columns of A, then $\det B = \sum_{k=1}^{n} X_k \alpha_{kj}$ is not necessarily zero, but rather is the determinant of a brand new matrix.] A similar proof applies for the row expansion case.□

Proposition 56 $A(\operatorname{adj} A) = \det(A) I_n$ *for* $A \in \mathbf{C}^{n \times n}$. *This is Ayres equation 6.2 [35]. When* A^{-1} *exists, then* $A^{-1} = \frac{\operatorname{adj} A}{\det A}$.

Proof. This proof is an expansion of Ayres' proof. This was motivated by wondering if

$$\operatorname{adj} A = [(-1)^{i+j} \det(X_{ij})]^T$$

or

$$\operatorname{adj} A = [(-1)^{i+j} \det(X_{ij})]^H$$

where X_{ij} is the minor of a_{ij} for $A = (a_{ij})$.

Let

$$\operatorname{adj} A = [(-1)^{i+j} \det(X_{ij})]^T = (\alpha_{ij})^T$$

where α_{ij} is the cofactor of a_{ij}. Then examine the product

$$A(\text{adj}\, A) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \cdots & \alpha_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{n} a_{1k}\alpha_{1k} & \sum_{k=1}^{n} a_{1k}\alpha_{2k} & \cdots & \sum_{k=1}^{n} a_{1k}\alpha_{nk} \\ \sum_{k=1}^{n} a_{2k}\alpha_{1k} & \sum_{k=1}^{n} a_{2k}\alpha_{2k} & \cdots & \sum_{k=1}^{n} a_{2k}\alpha_{nk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{nk}\alpha_{1k} & \sum_{k=1}^{n} a_{nk}\alpha_{2k} & \cdots & \sum_{k=1}^{n} a_{nk}\alpha_{nk} \end{pmatrix}$$

Recall that we proved earlier that $\sum_{k=1}^{n} a_{ik}\alpha_{jk} = 0$ for $i \neq j$. Thus

$$A(\text{adj}\, A) = \begin{pmatrix} \det A & & & \\ & \det A & & \\ & & \ddots & \\ & & & \det A \end{pmatrix} = I_n \det A$$

Further, when A^{-1} exists then

$$A^{-1} A \,\text{adj}(A) = A^{-1} I_n \det(A)$$

which implies

$$A^{-1} = \frac{\text{adj}\, A}{\det A}$$

Therefore

$$\text{adj}\, A = [(-1)^{i+j} \det(X_{ij})]^T$$

works! □

Proposition 57 $[\det(A)]^{-1} = \det(A^{-1})$.

Proof. $\det(AB) = \det(A)\det(B)$. Let $B = A^{-1}$. Therefore

$$\det(AA^{-1}) = \det(I) = 1 = \det(A)\det(A^{-1})$$

This implies $[\det(A)]^{-1} = \det(A^{-1})$. □

Lemma 42 $\det(A^*) = (\det A)^*$.

Proof. $A^* = (a_{ij}^*)$. By definition of the determinant,

$$\det(A^*) = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma_1}^* a_{2\sigma_2}^* \cdots a_{n\sigma_n}^*$$

where S_n is the set of all permutations of the ordered set $(1, 2, \cdots, n)$. Therefore σ takes on $n!$ different permutations from the form

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma_1 & \sigma_2 & \sigma_3 & \cdots & \sigma_n \end{pmatrix}$$

Then

$$\det(A^*) = \left(\sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma_1} a_{2\sigma_2} \cdots a_{n\sigma_n} \right)^* = (\det A)^*$$

Note that

$$\operatorname{sgn}\sigma = \begin{cases} +1, & \text{if } \sigma \text{ is an even permutation of } (1, 2, \cdots, n) \\ -1, & \text{if } \sigma \text{ is an odd permutation of } (1, 2, \cdots, n) \end{cases}$$

□

Proposition 58 $\det(A^H) = (\det A)^* = (\det A)^H$.

Proof.

$$\det(A^H) = \det[(A^*)^T] = \det(A^*) = (\det A)^* = (\det A)^H$$

since $\det(A)$ is a scalar and is thus equal to its transpose. □

Lemma 43 *Let A be a unitary complex $n \times n$ matrix. Then $\det(A) = e^{i\theta}$ for some $\theta \in \mathbf{R}$.*

Proof. If A is unitary, then $A^H = A^{-1}$. Thus $A^H A = I$.

$$\det(A^H A) = \det(I) = 1 = \det(A^H)\det(A) = |\det A|^2 = 1$$

Therefore $\det A = e^{i\theta}$ for some $\theta \in \mathbf{R}$. □

Lemma 44 *Let A be an orthonormal complex matrix where $A^T A = I$. Then $\det A = \pm 1$.*

Proof. If A is orthonormal, then $A^T A = I$ and hence $A^T = A^{-1}$. Then

$$\det(A^T A) = \det I = 1 = (\det A^T)(\det A) = (\det A)^2 = 1$$

Then $\det A = \pm 1$. □

Proposition 59 *If A is a skew-Hermitian matrix, then $[I + A][I - A]^{-1}$ is unitary. This comes from Littlewood (p. 19) [167].*

Proof. I have supplied the following proof. Let $A^H = -A$. Note that

$$(I - A)(I + A) = I - A^2 = (I + A)(I - A)$$

Then

$$[(I + A)(I - A)^{-1}]^H \cdot [(I + A)(I - A)^{-1}]$$

$$= (I - A)^{-H}(I + A)^H(I + A)(I - A)^{-1}$$

$$= (I - A)^{-H}(I - A)(I + A)(I - A)^{-1}$$

$$= (I - A)^{-H}(I + A)(I - A)(I - A)^{-1}$$

$$(I - A)^{-H}(I + A) = [(I + A)^H(I - A)^{-1}]^H$$

$$= [(I - A)(I - A)^{-1}]^H = I^H = I$$

Therefore $(I - A)(I - A)^{-1}$ is unitary. □

Proposition 60 *If B is unitary and -1 is not a characteristic root of B, then there exists skew-Hermitian matrix A such that*

$$B = [I + A][I - A]^{-1}$$

This comes from Littlewood (p. 19) [167].

Proof. I supplied the following proof. B is unitary implies $BB^H = I$. Let $B = (I + A)(I - A)^{-1}$. First, show that -1 is not reasonable as an eigenvalue of such a B.

$$\det[(I + A)(I - A)^{-1} - \lambda^2 I] = 0 = \det\{[(I + A) - \lambda^2(I - A)](I - A)^{-1}\}.$$

Assume $\det[(I-A)^{-1}] \neq 0$. Then

$$\det[(I+A)-\lambda^2(I-A)] = 0$$

$$= \det[(1+\lambda^2)A + (1-\lambda^2)I] \neq 0$$

Let $\lambda^2 = -1$. Then $\det[2I] = 2^m \neq 0$ where $\text{rank}(I) = m$.This contradicts $\det[B-\lambda^2 I] = 0$, so $\lambda^2 \neq -1$.

Now, consider $BB^H = I$, which implies $\det(BB^H) = 1$ or $\det B = e^{i\theta}$.

$$\det[(I+A)(I-A)^{-1}\{(I+A)(I-A)^{-1}\}^H] = 1$$

$$= \det[(I+A)(I-A)^{-1}(I-A)^{-H}(I+A)^H]$$

$$= [\det(I+A)][\det(I-A)]^{-1}[\det(I-A)^H]^{-1}[\det(I+A)^H]$$

This implies

$$[\det(I+A)][\det(I+A)^H] = [\det(I-A)][\det(I-A)^H]$$

which in turn implies

$$[\det(I+A)(I+A)^H] = [\det(I-A)(I-A)^H]$$

$$= \det(I+A+A^H+AA^H) = \det(I-A-A^H+AA^H)$$

This is true when $A^H = -A$. □

K.4.2 Partitioned Matrix Determinants

Lemma 45 *Let B be a complex square matrix partitioned as*

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

and let B_{11}^{-1} exist. Then

$$\det(B) = \det(B_{11}) \det(B_{22} - B_{21} B_{11}^{-1} B_{12})$$

This is a complexification of Graybill (p. 184) theorem 8.2.1(3) [95].

Proof. I provided the following derivation based on Graybill's derivation of lemma 46. Since $1 = \det(B_{11}) \det(B_{11}^{-1})$, we can write

$$\det(B) = \det(B_{11}^{-1}) \det(B) \det(B_{11})$$

Note for any matrix of the form

$$A = \begin{pmatrix} A_1 & 0 \\ A_2 & I \end{pmatrix}$$

and

$$C = \begin{pmatrix} I & C_2 \\ 0 & C_1 \end{pmatrix}$$

that by expanding the determinant of A along the right-most column repeatedly (or by the left columns of C) we get

$$\det(A) = \det(A_1) \text{ and } \det(C) = \det(C_1)$$

Thus

$$\det(B_{11}^{-1})\det(B) = \det\begin{bmatrix} B_{11}^{-1} & 0 \\ -B_{21}B_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \det(A)\det(B)$$

Since A and B are conformable, $\det(A)\det(B) = \det(AB)$. Therefore

$$\det(B_{11}^{-1})\det(B) = \det\left[\begin{pmatrix} B_{11}^{-1} & 0 \\ -B_{21}B_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\right]$$

$$= \det\begin{bmatrix} I & B_{11}^{-1}B_{12} \\ -B_{21} + B_{21} & B_{22} - B_{21}B_{11}^{-1}B_{12} \end{bmatrix}$$

$$= \begin{bmatrix} I & B_{11}^{-1}B_{12} \\ 0 & B_{22} - B_{21}B_{11}^{-1}B_{12} \end{bmatrix} = \det(B_{22} - B_{21}B_{11}^{-1}B_{12})$$

Finally,

$$\det(B) = \det(B_{11}^{-1})\det(B)\det(B_{11}) = \det(B_{11})\det(B_{22} - B_{21}B_{11}^{-1}B_{12})$$

Since determinants are polynomials, they commute. □

Lemma 46 *Let B be a complex square matrix partitioned as*

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

and let B_{22}^{-1} exist. Then

$$\det(B) = \det(B_{22})\det(B_{11} - B_{12}B_{22}^{-1}B_{21})$$

This is a complexification of Graybill (p. 184) theorem 8.2.1(2) [95].

Proof. This derivation is taken from Graybill.

$$\det(B) = \det(B_{22}) \det(B) \det(B_{22}^{-1})$$

Then

$$\det(B) \det(B_{22}^{-1}) = \det \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \det \begin{bmatrix} I & 0 \\ -B_{22}^{-1} B_{21} & B_{22}^{-1} \end{bmatrix}$$

$$= \det \left[\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -B_{22}^{-1} B_{21} & B_{22}^{-1} \end{pmatrix} \right]$$

$$= \det \begin{bmatrix} B_{11} - B_{12} B_{22}^{-1} B_{21} & B_{12} B_{22}^{-1} \\ B_{21} - B_{22} B_{22}^{-1} B_{21} & B_{22} B_{22}^{-1} \end{bmatrix}$$

$$= \det \begin{bmatrix} B_{11} - B_{12} B_{22}^{-1} B_{21} & B_{12} B_{22}^{-1} \\ 0 & I \end{bmatrix} = \det(B_{11} - B_{12} B_{22}^{-1} B_{21})$$

Therefore

$$\det(B) = \det(B_{22}) \det(B_{11} - B_{12} B_{22}^{-1} B_{21})$$

□

Proposition 61 *Let A, B, and I be $n \times n$ matrices. Then*

$$\det \begin{pmatrix} A & I \\ B & I \end{pmatrix} = \det(A - B)$$

Proof. This is a simple application of lemma 46.

$$\det \begin{pmatrix} A & I \\ B & I \end{pmatrix} = \det(I) \det(A - I I^{-1} B) = \det(A - B)$$

□

Proposition 62 *Let A, B, and I be $n \times n$ matrices. Then*

$$\det \begin{pmatrix} A & B \\ I & I \end{pmatrix} = \det(A - B)$$

Proof. By lemma 46,

$$\det \begin{pmatrix} A & B \\ I & I \end{pmatrix} = \det(I) \det(A - BI^{-1}I) = \det(A - B)$$

□

Proposition 63 *Let A and B be square matrices, not necessarily the same size. Then*

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \det(A) \det(B)$$

Proof. By lemma 46,

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \det(A) \det(B - 0A^{-1}0) = \det(A) \det(B)$$

□

K.4.3 Other Determinants

Lemma 47 *Let $A \in \mathbf{C}^{n \times m}$ and $B \in \mathbf{C}^{m \times n}$. Then*

$$\det(I_n + AB) = \det(I_m + BA)$$

This is Eaton's proposition 1.35 (p. 43) [74].

Proof. Apply lemma 45 and lemma 46 to

$$\begin{pmatrix} I_n & -A \\ B & I_m \end{pmatrix}$$

Then

$$\det \begin{pmatrix} I_n & -A \\ B & I_m \end{pmatrix} = \det(I_n)\det[I_m - BI_n^{-1}(-A)] = \det(I_m + BA)$$

$$= \det(I_m)\det[I_n - (-A)I_m^{-1}B] = \det(I_n + AB)$$

Therefore

$$\det(I_m + BA) = \det(I_n + AB)$$

□

Lemma 48 *Let $a, e \in \mathbf{C}$ and let matrices B, C, D be such that $BCD \in \mathbf{C}^{k\times k}$ and $CDB \in \mathbf{C}^{n\times n}$. Then*

$$\det\left(\frac{1}{a}BCD - eI_k\right) = (-1)^{n-k}\frac{e^k}{a^n}\det\left(\frac{1}{e}CDB - aI_n\right)$$

This is a corollary to Eaton's lemma 1.35 [74], supplied by me.

Proof. This is a simple application of lemma 45 and lemma 46.

$$\det \begin{pmatrix} aI_n & CD \\ B & eI_k \end{pmatrix} = \det(aI_n)\det(eI_k - B\frac{1}{a}I_n^{-1}CD)$$

$$= \det(eI_k)\det(aI_n - CD\frac{1}{e}I_k^{-1}B)$$

This implies

$$a^n \det(eI_k - \frac{1}{a}BCD) = e^k \det(aI_n - \frac{1}{e}CDB)$$

From this we get

$$(-1)^k a^n \det(\frac{1}{a}BCD - eI_k) = (-1)^n e^k \det(\frac{1}{e}CDB - aI_n)$$

We finally get

$$\det(\frac{1}{a}BCD - eI_k) = (-1)^{n-k}\frac{e^k}{a^n}\det(\frac{1}{e}CDB - aI_n)$$

□

Proposition 64 *Let* $\Sigma_{kk} > 0$, $X_{k\times n}$. *Then* $\det(XX^T - \lambda^2 I_n) = 0$ *implies that*

$$\det(X^T\Sigma^{-1}X - \lambda^2 I_n) = 0$$

This is a corollary to Eaton's lemma 1.35 [74], supplied by me.

Proof. This is a simple application of lemma 45 and lemma 46.

$$\det(XX^T - \lambda^2\Sigma_{kk}) = \det[(XX^T\Sigma^{-1} - \lambda^2 I_k)\Sigma]$$

$$= \det(XX^T\Sigma^{-1} - \lambda^2 I_k)\det(\Sigma)$$

$$= (-1)^{n-k}\lambda^{2k}\det(\lambda^{-2}X^T\Sigma^{-1}X - I_n)\det(\Sigma)$$

by lemma 48. This equals

$$(-1)^{n-k}\lambda^{2(k-n)}\det(X^T\Sigma^{-1}X - \lambda^2 I_n)\det(\Sigma)$$

By our hypothesis, this is zero. If we assume that $\lambda^2 \neq 0$ and $\Sigma > 0$, then we conclude that

$$\det(X^T\Sigma^{-1}X - \lambda^2 I_n) = 0$$

□

Lemma 49 *Let*

$$\begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

be a partitioned square matrix such that A^{-1} exists. Then

$$\det\begin{bmatrix} A \otimes I_p & C \otimes I_p \\ B \otimes I_p & D \otimes I_p \end{bmatrix} = \det\left[\begin{pmatrix} A & C \\ B & D \end{pmatrix} \otimes I_p\right] = \left[\det\begin{pmatrix} A & C \\ B & D \end{pmatrix}\right]^p$$

This was supplied by me.

Proof.

$$\det\begin{bmatrix} A \otimes I_p & C \otimes I_p \\ B \otimes I_p & D \otimes I_p \end{bmatrix}$$

$$= \det(A \otimes I_p)\det[D \otimes I_p - (B \otimes I_p)(A \otimes I_p)^{-1}(C \otimes I_p)]$$

$$= \det(A \otimes I_p)\det[D \otimes I_p - (B \otimes I_p)(A^{-1} \otimes I_p)(C \otimes I_p)]$$

$$= \det(A \otimes I_p)\det[D \otimes I_p - (BA^{-1} \otimes I_p)(C \otimes I_p)]$$

$$= \det(A \otimes I_p)\det[D \otimes I_p - (BA^{-1}C) \otimes I_p)]$$

$$= \det(A \otimes I_p)\det[(D - BA^{-1}C) \otimes I_p)]$$

$$= [\det A]^p[\det(D - BA^{-1}C)]^p = [\det(A)\det(D - BA^{-1}C)]^p$$

$$= \left[\det \begin{pmatrix} A & C \\ B & D \end{pmatrix}\right]^p = \det\left[\begin{pmatrix} A & C \\ B & D \end{pmatrix} \otimes I_p\right]$$

□

Note that even though the determinants are equal, the matrices are not equal.

$$\begin{pmatrix} A \otimes I_p & C \otimes I_p \\ B \otimes I_p & D \otimes I_p \end{pmatrix} \neq \begin{pmatrix} A & C \\ B & D \end{pmatrix} \otimes I_p$$

Proposition 65 *Let A be a square matrix. Then*

$$\det(I + A^2) = |\det(I + iA)|^2$$

Proof.

$$\det(I + A^2) = \det[(I + iA)(I - iA)]$$

$$= \det(I + iA)\det(I - iA) = |\det(I + iA)|^2$$

□

Appendix L

GRAM-SCHMIDT

We examine two Gram-Schmidt algorithms which give different results in the complex case. A discovery from this exercise is that it is possible via Gram-Schmidt to produce an orthonormal basis for a vector space without having the property of an inner product space. The first algorithm will examine the case where we use the bilinear operator $< x, y >= x^H y$ which is an inner product operator. The second algorithm will examine the bilinear operator $(x, y) = x^T y$ with does, indeed, produce an orthonormal basis, but this operator is not an inner product. The two orthonormal sets produced are generally not the same.

The basic algorithm is given as problems 5.1.10 and 5.1.11 of Stewart [259].

L.1 Algorithm Using $< x, y >= x^H y$

The following proof is set in general Hilbert space H, which is a complete inner product space. In our application, we define our inner product as $< x, y >= x^H y$ where x and y are vectors, which may be complex.

Let $\{x_n\}_{n=1}^N$ be a set of linearly independent elements (vectors) in Hilbert space H. Define the inner product using the engineering convention that the inner product is linear in the second argument. For example,

$$< x, \alpha y >= \alpha < x, y >$$

This is the reverse of the usual mathematician's convention, but is of greater practical use.

The following Gram-Schmidt orthonormalization process operates on $\{x_n\}_{n=1}^N$ to produce the orthonormal set $\{u_n\}_{n=1}^N$ which has the same span as $\{x_n\}_{n=1}^N$. The algorithm is as follows.

L.1.1 Inner Product Gram-Schmidt Algorithm

1 Let $v_1 = x_1$

2 Let $u_1 = \frac{v_1}{\|v_1\|}$ $\qquad \|v_1\| =< v_1, v_1 >^{1/2} \in \mathbf{R}^+$

3 Let $v_n = x_n - \sum_{i=1}^{n-1} < x_n, u_i > u_i$

4 Let $u_n = \frac{v_n}{\|v_n\|}$ $\qquad \|v_n\| =< v_n, v_n >^{1/2} \in \mathbf{R}^+$

Repeat steps 3 and 4 for $2 \leq n \leq N$.

L.1.2 Inner Product Gram-Schmidt Algorithm Proof

First, show

$$< u_2, u_1 >= 0 \tag{L.1}$$

$$< u_2, u_1 >= \left\langle \frac{v_2}{\|v_2\|}, \frac{v_1}{\|v_1\|} \right\rangle = \frac{1}{\|v_2\| \, \|v_1\|} \langle v_2, v_1 \rangle = c_{21} \langle v_2, v_1 \rangle$$

$$= c_{21} < x_2 - < x_2, u_1 > u_1, x_1 >$$

$$= c_{21} \{ \langle x_2, x_1 \rangle - \langle x_2, u_1 \rangle \langle u_1, x_1 \rangle \}$$

$$= c_{21} \left\{ \langle x_2, x_1 \rangle - \left\langle x_2, \frac{x_1}{\|v_1\|} \right\rangle \left\langle \frac{x_1}{\|v_1\|}, x_1 \right\rangle \right\}$$

Recall that

$$u_1 = \frac{v_1}{\|v_1\|} = \frac{x_1}{\|v_1\|}$$

$$= c_{21}\left\{\langle x_2, x_1\rangle - \frac{1}{\|v_1\|^2}\langle x_2, x_1\rangle\langle x_1, x_1\rangle\right\} = c_{21}\{\langle x_2, x_1\rangle - \langle x_2, x_1\rangle\} = 0$$

Therefore $< u_2, u_1 >= 0$.

Now, show

$$< u_1, u_1 >= 1 \tag{L.2}$$

$$< u_1, u_1 >= \left\langle \frac{v_1}{\|v_1\|}, \frac{v_1}{\|v_1\|} \right\rangle = \frac{1}{\|v_1\|^2} < v_1, v_1 >= 1$$

For $i, j \leq n-1$, assume $< u_i, u_j >= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. Then

$$< u_n, u_j >= \left\langle \frac{1}{\|v_n\|}\left(x_n - \sum_{i=1}^{n-1}\langle x_n, u_i\rangle u_i\right), u_j \right\rangle$$

$$= \frac{1}{\|v_n\|}\left\{< x_n, u_j > - \sum_{i=1}^{n-1}\langle x_n, u_i\rangle\langle u_i, u_j\rangle\right\}$$

$$\frac{1}{\|v_n\|}\{< x_n, u_j > - \langle x_n, u_j\rangle\} = 0$$

Thus $< u_n, u_j >= 0$.

Now, show

$$< u_n, u_n >= 1 \tag{L.3}$$

$$< u_n, u_n >= \left\langle \frac{v_n}{\|v_n\|}, \frac{v_n}{\|v_n\|} \right\rangle = 1$$

Therefore by induction, the algorithm produces an orthonormal set $\{u_n\}$ for all $n \in \mathbf{Z}$. *For the inner product, we see that the orthonormal set is unitary also.*

L.2 Algorithm Using $(x, y) = x^T y$.

The Gram-Schmidt process applied to a complex vector space $\mathbf{C}^n$ usually involves using the inner product. However, the inner product is not the only function of the two vectors that will produce a decomposition. For example, define $(x, y) = x^T y$ where $x, y \in \mathbf{C}^n$. (x, y) does not define an inner product. We cannot assert that $(x, x) \geq 0$ for all $x \in \mathbf{C}^n$. However, the function $(x, y) = x^T y$ can successfully be used in the Gram-Schmidt process to produce an orthonormal basis. In fact, let (x, y) be any operator such that

$$(\alpha x, y) = \alpha(x, y) = (x, \alpha y)$$

$$(x + y, z) = (x, z) + (y, z)$$

where $x, y, z \in \mathbf{C}^n$ and $\alpha \in \mathbf{C}$. The algorithm is the "same" as before.

L.2.1 Bilinear Gram-Schmidt Algorithm

Let $\{x_i\}_1^n$ be a set of linearly independent vectors in $\mathbf{C}^n$. The algorithm is as follows.

1	Let $v_1 = x_1$	
2	Let $u_1 = \frac{v_1}{(v_1, v_1)^{1/2}}$	$(v_1, v_1)^{1/2} \in \mathbf{C}$
3	Let $v_k = x_k - \sum_{i=1}^{k-1} < x_k, u_i > u_i$	
4	Let $u_k = \frac{v_k}{(v_k, v_k)^{1/2}}$	$(v_k, v_k)^{1/2} \in \mathbf{C}$

Repeat steps 3 and 4 for $2 \leq k \leq n$.

L.2.2 Proof

The proof closely follows the one using the inner product. First, show

$$(u_2, u_1) = 0 \tag{L.4}$$

$$(u_2, u_1) = \left(\frac{v_2}{(v_2, v_2)^{1/2}}, \frac{v_1}{(v_2, v_2)^{1/2}} \right)$$

$$= \frac{1}{(v_2, v_2)^{1/2}(v_1, v_1)^{1/2}} (v_2, v_1) = c_{21}(v_2, v_1)$$

$$= c_{21} (x_2 - (x_2, u_1)u_1, x_1) = c_{21} \{(x_2, x_1) - (x_2, u_1)(u_1, x_1)\}$$

$$= c_{21} \left\{ (x_2, x_1) - \left(x_2, \frac{x_1}{(v_1, v_1)^{1/2}} \right) \left(\frac{x_1}{(v_1, v_1)^{1/2}}, x_1 \right) \right\}$$

Recall that

$$u_1 = \frac{v_1}{(v_1, v_1)^{1/2}} = \frac{x_1}{(v_1, v_1)^{1/2}}$$

Then we obtain

$$(u_2, u_1) = c_{21} \left\{ (x_2, x_1) - \frac{1}{(v_1, v_1)} (x_2, x_1)(x_1, x_1) \right\}$$

$$= c_{21} \{(x_2, x_1) - (x_2, x_1)\} = 0$$

Therefore

$$(u_2, u_1) = 0 \tag{L.5}$$

Now, show $(u_1, u_1) = 1$.

$$(u_1, u_1) = \left(\frac{v_1}{(v_1, v_1)^{1/2}}, \frac{v_1}{(v_1, v_1)^{1/2}} \right) = \frac{1}{(v_1, v_1)} (v_1, v_1) = 1$$

For $i, j \leq n-1$, assume

$$(u_i, u_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Then

$$(u_n, u_j) = \left(\frac{1}{(v_n, v_n)^{1/2}} \left(x_n - \sum_{i=1}^{n-1} (x_n, u_i)\, u_i \right), u_j \right)$$

$$= \frac{1}{(v_n, v_n)^{1/2}} \left\{ (x_n, u_j) - \sum_{i=1}^{n-1} (x_n, u_i)\,(u_i, u_j) \right\}$$

$$\frac{1}{(v_n, v_n)^{1/2}} \{(x_n, u_j) - (x_n, u_j)\} = 0$$

Thus $(u_n, u_j) = 0$ when $j \neq n$.

Now, show

$$(u_n, u_n) = 1 \tag{L.6}$$

$$(u_n, u_n) = \left(\frac{v_n}{(v_n, v_n)^{1/2}}, \frac{v_n}{(v_n, v_n)^{1/2}} \right) = 1$$

Therefore by induction, the algorithm produces an orthonormal set $\{u_n\}$ for all $n \in \mathbf{Z}^+$. Compared to the previous Gram-Schmidt algorithm, note that

$$< v_k, v_k >^{1/2} \neq (v_k, v_k)^{1/2} \tag{L.7}$$

Thus the two orthonormal sets are not generally the same.

Appendix M

COMPLEX MATRIX DECOMPOSITIONS AND EIGENVALUES

This contains various decompositions of complex matrices. Most decompositions are related to the eigenvalue decomposition of an Hermitian positive definite matrix. Also included are some decompositions of triangular and rectangular matrices. Some of the proofs are given as algorithms or constructions. There are also a number of theorems that describe or exploit properties of eigenvalues.

Most of the theorems are straight-forward adaptations of similar theorems for the case of real matrices. In generalizing, special attention is required when the real case specifies uniqueness to ± 1. In the complex case, this sometimes will generalize to $e^{i\theta}$ for arbitrary $\theta \in \mathbf{R}$. Distinction is also required between symmetric and Hermitian complex matrices. It is shown, for example, that you cannot assume a symmetric complex matrix is a definite matrix, or even that its eigenvalues are all real. Recall that the present literature about zonal polynomials for complex matrices assume complex symmetric matrices.

The decompositions are needed to support the work on Jacobians and the development of distributional results.

M.1 Decomposition to a Product of a Triangular Matrix

Proposition 66 *Let T be an $n \times n$ complex upper triangular matrix with distinct diagonal elements. Let $\Lambda^2 = \mathrm{diag}(T_{11}, \cdots, T_{NN})$. Then there exists a nonsingular upper triangular matrix C satisfying $CT = \Lambda^2 C$. C is uniquely determined up to a (possibly different) multiplicative constant for each row. Row k of C is the left eigenvector of T corresponding to eigenvalue λ_k^2. This is a complexification of Takemura lemma 3.1.1 (p. 17)[265], stated without proof.*

Proof. The expansion of the identity is $CT =$

$$\begin{bmatrix} C_{11}T_{11} & C_{11}T_{12}+C_{12}T_{22} & C_{11}T_{13}+C_{12}T_{23}+C_{13}T_{33} & \cdots & \sum_{k=1}^{n} C_{1k}T_{kn} \\ 0 & C_{22}T_{22} & C_{22}T_{23}+C_{23}T_{33} & \cdots & \sum_{k=2}^{n} C_{1k}T_{kn} \\ 0 & 0 & C_{33}T_{33} & \cdots & \sum_{k=3}^{n} C_{1k}T_{kn} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & C_{nn}T_{nn} \end{bmatrix}$$

$$\Lambda^2 C = \begin{bmatrix} \lambda_1^2 C_{11} & \lambda_1^2 C_{12} & \lambda_1^2 C_{13} & \cdots & \lambda_1^2 C_{1n} \\ 0 & \lambda_2^2 C_{22} & \lambda_2^2 C_{23} & \cdots & \lambda_2^2 C_{2n} \\ 0 & 0 & \lambda_3^2 C_{33} & \cdots & \lambda_3^2 C_{3n} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & \lambda_n^2 C_{nn} \end{bmatrix}$$

Table M.1. $CT = D^2C$ Decomposition Pseudo-Code

C	Increment by rows
	DO α J=1,N-1
C	Increment by columns
	..DO β M=J+1,N
C	Compute C_{JM}
	....SUM=0.0
	....DO γ K=1,M-1
γ	......SUM=SUM+C(J,K)*T(K,M)
β	....C(J,M)=SUM/(T(J,J)-T(M,M))
α	..CONTINUE

By construction, $\lambda_k^2 = T_{kk}$ for $1 \leq k \leq n$. Each row of C may be determined independently from all other rows. The algorithm is given in table M.1

This computes

$$C_{JM} = \frac{\sum_{k=1}^{m-1} C_{JK} T_{KM}}{T_{JJ} - T_{MM}}$$

for $1 \leq J \leq N-1$ and $J+1 \leq M \leq N$, where we used $\lambda_J^2 = T_{JJ}$ in the derivation. The values of C_{JJ} are arbitrary, and C_{JJ} is independent of C_{KK} for $J \neq K$. For a fixed diagonal of C, the other entries of C are unique. The $\{C_{JJ}\}$ may be chosen to minimize numerical error in computations. Alternately, the computations may be slightly simplified by arbitrarily setting $C_{JJ} = 1$ for all

J. □

Lemma 50 *Let T be an $n \times n$ complex upper triangular matrix with distinct diagonal elements. Let*

$$\Lambda^2 = \mathrm{diag}(T_{11}, \cdots, T_{nn})$$

Then there exists a nonsingular upper triangular matrix C satisfying $CT = C\Lambda^2$. C is uniquely determined up to a (possibly different) multiplicative constant for each column. Column k of C is the right eigenvector of T corresponding to eigenvalue λ_k^2. This is a corollary to the complexification of Takemura lemma 3.1.1 [265].

Proof. Examine the structure of the following matrices.

$$\Lambda^2 C = \begin{bmatrix} C_{11}\lambda_1^2 & C_{12}\lambda_2^2 & C_{13}\lambda_3^2 & \cdots & C_{1n}\lambda_n^2 \\ 0 & C_{22}\lambda_2^2 & C_{23}\lambda_3^2 & \cdots & C_{2n}\lambda_n^2 \\ 0 & 0 & C_{33}\lambda_3^2 & \cdots & C_{3n}\lambda_n^2 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & C_{nn}\lambda_n^2 \end{bmatrix}$$

We also have $TC =$

$$\begin{bmatrix} T_{11}C_{11} & T_{11}C_{12} + T_{12}C_{22} & T_{11}C_{13} + T_{12}C_{23} + T_{13}C_{33} & \cdots & \sum_{k=1}^{n} T_{1k}C_{kn} \\ 0 & T_{22}C_{22} & T_{22}C_{23} + T_{23}C_{33} & \cdots & \sum_{k=2}^{n} T_{2k}C_{kn} \\ 0 & 0 & T_{33}C_{33} & \cdots & \sum_{k=3}^{n} T_{3k}C_{kn} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & & T_{nn}C_{nn} \end{bmatrix}$$

Table M.2. $CT = CD^2$ Decomposition Pseudo-Code

C	Increment by columns
	DO α M=2,N
C	Increment by rows
	..DO β J=M-1,1,-1
C	Compute C_{JM}
	....SUM=0.0
	....DO γ K=J+1,M
γ	.....SUM=SUM+T(J,K)*C(K,M)
β	....C(J,M)=SUM/(T(M,M)-T(J,J))
α	..CONTINUE

By construction, $\lambda_k^2 = T_{kk}$ for $1 \leq k \leq n$ and the values of C_{KK} are arbitrary. Construction of C_{JK} proceeds one column at a time from left to right, working from the diagonal to the top row. The algorithm is given in table M.2.

This computes

$$C_{JM} = \frac{\sum_{k=J+1}^{m} T_{JK} C_{KM}}{T_{MM} - T_{JJ}}$$

for $2 \leq M \leq N$ and $M - 1 \geq J \geq 1$. The values of C_{JJ} are arbitrary, and C_{JJ} is independent of C_{KK} for $J \neq K$. For a fixed diagonal of C, the other entries of C are unique.

Proposition 67 *Let A be an $m \times n$ complex matrix. Then there exists an*

upper $n \times n$ triangular matrix T having positive real elements on the diagonal, and a subunitary $m \times n$ matrix S such that $S^H S = I_n$, and $A = ST$. This is C. R. Rao equation 1b.2(ix) [213]. This is one version of a QR decomposition.

Proof. This proof is by C. R. Rao. Let $(\alpha_1, \cdots, \alpha_n)$ be the columns of A. Let $(\sigma_1, \cdots, \sigma_n)$ be the columns of S. Let $(t_{1i}, \cdots, t_{ii})$ be the nonzero elements of the i^{th} column of T. Then we have

$$(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\sigma_1, \sigma_2, \cdots, \sigma_n) \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{pmatrix}$$

$$= (\sigma_1 t_{11}, \sigma_1 t_{12} + \sigma_2 t_{22}, \cdots, \sigma_1 t_{1n} + \sigma_2 t_{2n} + \cdots + \sigma_n t_{nn})$$

From this, we see

$$\alpha_1 = \sigma_1 t_{11}, \alpha_2 = \sigma_1 t_{12} + \sigma_2 t_{22}, \cdots, \alpha_i = \sigma_1 t_{1i} + \sigma_2 t_{2i} + \cdots + \sigma_i t_{ii}$$

for $1 \leq i \leq n$. Let $\sigma_i^H \sigma_j = \delta_{ij}$. Then $\sigma_i^H \alpha_j = t_{ij}$ for $i \neq j$, and

$$\alpha_i^H \alpha_i = t_{1i}^2 + t_{2i}^2 + \cdots + t_{ii}^2$$

which implies

$$t_{ii} = [\alpha_i^H \alpha_i - t_{1i}^2 - t_{2i}^2 - \cdots - t_{i-1,i}^2]^{1/2}$$

beginning with α_1, then S and T are constructed as follows.

$$
\begin{aligned}
t_{11} &= [\alpha_1^H \alpha_1]^{1/2} \\
\sigma_1 &= \tfrac{1}{t_{11}} \alpha_1 \\
t_{12} &= \sigma_1^H \alpha_2 \\
t_{22} &= [\alpha_2^H \alpha_2 - t_{12}^2]^{1/2} \\
\sigma_2 &= \tfrac{1}{t_{22}} [\alpha_2 - \sigma_1 t_{12}] \\
&\vdots \\
&\left.\begin{aligned}
t_{ij} &= \sigma_i^H \alpha_j, i \in [1, j-1] \in \mathbf{N} \\
t_{jj} &= [\alpha_j^H \alpha_j - t_{1j}^2 - t_{2j}^2 - \cdots - t_{j-1,j}^2]^{1/2} \\
\sigma_j &= \tfrac{1}{t_{jj}} [\alpha_j - \sigma_1 t_{1j} - \sigma_2 t_{2j} - \cdots - \sigma_{j-1} t_{j-1,j}]
\end{aligned}\right\} j \in [2, n] \in \mathbf{N}
\end{aligned}
$$

Thus we have constructed the required S and T such that $A = ST$, T is $n \times n$ upper triangular with a positive real diagonal, and $S^H S = I_n$. □

Proposition 68 *Let A be an $m \times n$ complex matrix. Then there exists a subunitary matrix S of size $m \times n$, where $S^H S = I_n$, an $n \times n$ nonsingular* upper *triangular matrix C, and a diagonal matrix Λ^2 such that $AC = SC\Lambda^2$ or $A = SC\Lambda^2 C^{-1}$.*

Proof. By C. R. Rao equation 1b.2(ix) [213], we obtain an $m \times n$ subunitary matrix S where $S^H S = I_n$, and an upper $n \times n$ triangular matrix T with positive real diagonal elements such that $A = ST$. By lemma 50, we have a nonsingular upper triangular matrix C satisfying $TC = C\Lambda^2$, where column

k of C is the right eigenvector of T corresponding to eigenvalue λ_k^2. Thus $AC = STC = SC\Lambda^2$. □

Proposition 69 *Let A be an $m \times n$ complex matrix. Then there exists an* upper *triangular $m \times m$ matrix T having positive real elements on the diagonal, and a subunitary $m \times n$ matrix S such that $SS^H = I_m$ and $A = TS$. This is motivated by C. R. Rao equation 1b.2(ix) [213]. This is a version of a QR decomposition, except that the orthonormal basis matrix is now on the right side, and it is the set of rows of S that form the basis.*

Proof. I have followed the proof is almost exactly as C. R. Rao's proof in proposition 67. Let $(\alpha_1, \cdots, \alpha_m)$ be the rows of A. Let $(\sigma_1, \cdots, \sigma_m)$ be the rows of S. Let $(t_{1i}, \cdots, t_{im})$ be the i^{th} row of T. Then we have

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ & t_{22} & \cdots & t_{2m} \\ & & \ddots & \vdots \\ & & & t_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_m \end{pmatrix}$$

Thus, we see that the rows of A are

$$\alpha_m = t_{mm}\sigma_m$$

$$\alpha_{m-1} = t_{m-1,m-1}\sigma_{m-1} + t_{m-1,m}\sigma_m$$

$$\vdots$$

$$\alpha_i = t_{ii}\sigma_i + t_{i,i+1}\sigma_{i+1} + \cdots + t_{im}\sigma_m$$

$$\vdots$$

$$\alpha_1 = t_{11}\sigma_1 + t_{12}\sigma_2 + \cdots + t_{1m}\sigma_m$$

The proof is by construction. We constrain the construction by requiring

$$\sigma_i^H \sigma_j = \begin{cases} 1, \text{ for } i = j \\ 0, \text{ for } i \neq j \end{cases}$$

Given this constraint, we construct the T that satisfies $A = TS$, and in the process we also explicitly find S. To begin with, notice that from the expansion of α_i that if we post-multiply by σ_j^H, where $i \neq j$, we get

$$\alpha_i \sigma_j^H = t_{ii}\sigma_i\sigma_j^H + t_{i,i+1}\sigma_{i+1}\sigma_j^H + \cdots + t_{im}\sigma_m\sigma_j^H = t_{ij}$$

Also note that

$$\alpha_i \alpha_i^H = (t_{ii}\sigma_i + t_{i,i+1}\sigma_{i+1} + \cdots + t_{im}\sigma_m)(t_{ii}\sigma_i + t_{i,i+1}\sigma_{i+1} + \cdots + t_{im}\sigma_m)$$

$$= t_{ii}^2 + t_{i,i+1}^2 + \cdots + t_{im}^2$$

Solving the above for t_{ij} and t_{ii} we get $t_{ij} = \alpha_i \sigma_j^H$ for $i \neq j$ and

$$t_{ii} = [\alpha_i \alpha_i^H - t_{i,i+1}^2 - \cdots - t_{im}^2]^{1/2}$$

To see that T and S can actually be constructed by decomposing A, we solve iteratively, as follows. We begin by recalling $\alpha_m = t_{mm}\sigma_m$ which leads us to the first step in the iteration.

$$\alpha_m = t_{mm}\sigma_m$$

$$t_{mm} = [\alpha_m \alpha_m^H]^{1/2}$$

$$\sigma_m = \tfrac{1}{t_{mm}}\alpha_m$$

$$\alpha_{m-1} = t_{m-1,m-1}\sigma_{m-1} + t_{m-1,m}\sigma_m$$

$$t_{m-1,m} = \alpha_{m-1}\sigma_m^H$$

$$t_{m-1,m-1} = [\alpha_{m-1}\alpha_{m-1}^H - t_{m-1,m}^2]^{1/2}$$

$$\sigma_{m-1} = \tfrac{1}{t_{m-1,m-1}}[\alpha_{m-1} - t_{m-1,m}\sigma_m]$$

$$\alpha_{m-2} = t_{m-2,m-2}\sigma_{m-2} + t_{m-2,m-1}\sigma_{m-1} + t_{m-2,m}\sigma_m$$

$$t_{m-2,m} = \alpha_{m-2}\sigma_m^H$$

$$t_{m-2,m-1} = \alpha_{m-2}\sigma_{m-1}^H$$

$$t_{m-2,m-2} = [\alpha_{m-2}\alpha_{m-2}^H - t_{m-2,m-1}^2 - t_{m-2,m}^2]^{1/2}$$

$$\sigma_{m-2} = \tfrac{1}{t_{m-2,m-2}}[\alpha_{m-2} - t_{m-2,m-1}\sigma_{m-1} - t_{m-2,m}\sigma_m]$$

In general,

$$\left.\begin{aligned} \alpha_i &= t_{ii}\sigma_i + t_{i,i+1}\sigma_{i+1} + \cdots + t_{i,m}\sigma_m \\ t_{ik} &= \alpha_i \sigma_k^H, \ k = m, m-1, \cdots, i+1 \\ t_{ii} &= [\alpha_i \alpha_i^H - t_{i,i+1}^2 - \cdots - t_{i,m}^2]^{1/2} \\ \sigma_i &= \tfrac{1}{t_{ii}}[\alpha_i - t_{i,i+1}\sigma_{i+1} - \cdots - t_{i,m}\sigma_m] \end{aligned}\right\} \; i = m-1, m-2, \cdots, 1$$

Thus, we have constructed T and S where $SS^H = I_m$ and $A = TS$. □

Proposition 70 *Let A be an $m \times n$ complex matrix. Then there exists a* lower *triangular $m \times m$ matrix T having positive real elements on the diagonal, and a subunitary $m \times n$ matrix S such that $SS^H = I_m$ and $A = TS$. This is motivated by C. R. Rao equation 1b.2(ix) [213]. This is a form of a QR decomposition, except that the orthonormal basis is the set of rows of S, and it is on the right side.*

Proof. I have followed the proof almost exactly as C. R. Rao's proof in proposition 67. Let $(\alpha_1, \cdots, \alpha_m)$ be the rows of A. Let $(\sigma_1, \cdots, \sigma_m)$ be the rows of S. Let $(t_{1i}, \cdots, t_{ii})$ be the i^{th} row of T. Then we have

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} t_{11} & & & \\ t_{21} & t_{22} & & \\ \vdots & \vdots & \ddots & \\ t_{m1} & t_{m2} & \cdots & t_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_m \end{pmatrix}$$

Thus, we see that the rows of A are

$$\alpha_1 = t_{11}\sigma_1$$

$$\alpha_2 = t_{21}\sigma_1 + t_{22}\sigma_2$$

$$\vdots$$

$$\alpha_i = t_{i1}\sigma_1 + t_{i,2}\sigma_2 + \cdots + t_{ii}\sigma_i$$

$$\vdots$$

$$\alpha_m = t_{m1}\sigma_1 + t_{m2}\sigma_2 + \cdots + t_{mm}\sigma_m$$

Let $\sigma_i \sigma_j^H = \delta_{ij}$. Given this constraint, we get $\alpha_i \sigma_j^H = t_{ij}$ for $1 \leq j \leq i-1$ and $i \neq j$. Also note that

$$\alpha_i \alpha_i^H = t_{i1}^2 + t_{i,2}^2 + \cdots + t_{ii}^2$$

where $1 \leq i \leq m$. Now, solve for t_{ij} and t_{ii}.

$$t_{ij} = \alpha_i \sigma_j^H \text{ for } i \neq j$$

$$t_{ii} = [\alpha_i \alpha_i^H - t_{i1}^2 - t_{i2}^2 - \cdots - t_{i,i-1}^2]^{1/2}$$

Now, form the algorithm to construct T and S.

$$t_{11} = [\alpha_1 \alpha_1^H]^{1/2}$$

$$\sigma_1 = \tfrac{1}{t_{11}} \alpha_1$$

$$t_{21} = \alpha_2 \sigma_1^H$$

$$t_{22} = [\alpha_2 \alpha_2^H - t_{21}^2]^{1/2}$$

$$\sigma_2 = \tfrac{1}{t_{22}} [\alpha_2 - t_{21} \sigma_1]$$

$$\vdots$$

$$\left.\begin{aligned} t_{ij} &= \alpha_i \sigma_j^H, \ 1 \leq j \leq i-1 \\ t_{ii} &= [\alpha_i \alpha_i^H - t_{i1}^2 - t_{i2}^2 - \cdots - t_{i,i-1}^2]^{1/2} \\ \sigma_i &= \tfrac{1}{t_{ii}} [\alpha_i - t_{i1} \sigma_1 - t_{i2} \sigma_2 - \cdots - t_{i,i-1} \sigma_{i-1}] \end{aligned}\right\} \ i = 2, 3, \cdots, m$$

Thus, $A = TS$ where $SS^H = I_m$ and T is lower triangular with positive real elements on the diagonal. □

Proposition 71 *Let A be an $m \times n$ complex matrix. Then there exists a* lower *triangular $n \times n$ matrix T having positive real elements on the diagonal, and a*

subunitary $m \times n$ matrix S such that $S^H S = I_n$ and $A = ST$. This is motivated by C. R. Rao equation 1b.2(ix) [213]. This is a form of a QR decomposition.

Proof. I have followed the proof almost exactly as C. R. Rao's proof in proposition 67. Let $(\alpha_1, \cdots, \alpha_n)$ be the columns of A. Let $(\sigma_1, \cdots, \sigma_n)$ be the columns of S. Let $(t_{ii}, \cdots, t_{ni})$ be the nonzero elements of the i^{th} column of T. Then we have

$$(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\sigma_1, \sigma_2, \cdots, \sigma_n) \begin{pmatrix} t_{11} & & & \\ t_{21} & t_{22} & & \\ \vdots & \vdots & \ddots & \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix}$$

$$= (\sigma_1 t_{11} + \cdots + \sigma_n t_{n1}, \sigma_2 t_{22} + \cdots + \sigma_n t_{n2}, \cdots, \sigma_n t_{nn})$$

Let $\sigma_i^H \sigma_j = \delta_{ij}$. Then $\sigma_i^H \alpha_j = t_{ij}$ for $i \neq j$, and

$$\alpha_i^H \alpha_i = t_{ii}^2 + t_{i+1,i}^2 + \cdots + t_{ni}^2$$

Thus $t_{ij} = \sigma_i^H \alpha_j$ and

$$t_{ii} = [\alpha_i^H \alpha_i - t_{i+1,i}^2 - \cdots - t_{n,i}^2]^{1/2}$$

Now, write the algorithm to construct S and T.

$$t_{nn} = [\alpha_n^H \alpha_n]^{1/2}$$

$$\sigma_n = \tfrac{1}{t_{nn}} \alpha_n$$

$$t_{n,n-1} = \sigma_n^H \alpha_{n-1}$$

$$t_{n-1,n-1} = [\alpha_{n-1}^H \alpha_{n-1} - t_{n-1,n}^2]^{1/2}$$

$$\sigma_{n-1} = \tfrac{1}{t_{n-1,n-1}} [\alpha_{n-1} - \sigma_n t_{n,n-1}]$$

$$\vdots$$

$$\left.\begin{aligned} t_{ij} &= \sigma_i^H \alpha_j, \text{ for } i \neq j \\ t_{ii} &= [\alpha_i^H \alpha_i - t_{i+1,i}^2 - \cdots - t_{ni}^2]^{1/2} \\ \sigma_i &= \tfrac{1}{t_{ii}} [\alpha_i - \sigma_{i+1} t_{i+1,i} - \cdots - \sigma_n t_{ni}] \end{aligned}\right\} \; 1 \leq i \leq n-1 \text{ in reverse order}$$

Thus we have found lower triangular T and S subunitary such that $A = ST$.

□

M.2 Similarity Transformation

Lemma 51 *Let B be a nonsingular complex $n \times n$ matrix, and let A be any other complex $n \times n$ matrix. Then A and $B^{-1}AB$ have the same eigenvalues.*

Proof.

$$\det(B^{-1}AB - \lambda^2 I) = \det(B^{-1}AB - \lambda^2 B^{-1} I B)$$

$$= \det[B^{-1}(A - \lambda^2 I)B] = \det(B^{-1}) \det(A - \lambda^2 I) \det(B)$$

$$= \det(A - \lambda^2 I) \det(B^{-1}B) = \det(A - \lambda^2 I)$$

when B is nonsingular. □

Theorem 112 *Let $s, t \in \mathcal{F}^{n\times n}$ where $\mathcal{F}$ is* **R** *or* **C**. *Let $\lambda^2(\cdot)$ be the set of eigenvalues of its argument. Then*

$$\lambda^2(st) = \lambda^2(ts)$$

Proof. I do not have a record of the pedigree of this theorem or its proof. By similarity transformation, using lemma 51.

$$\lambda^2(st) = \lambda^2(s^{-1}sts) = \lambda^2(ts)$$

□

Corollary 32 *Let U be a unitary complex $n \times n$ matrix, and let A be any other complex $n \times n$ matrix. Then A and $U^H AU$ have the same eigenvalues.*

Proof. Although I provided this, it is also common knowledge. It is provided here for the sake of completeness. $U^{-1} = U^H$. Apply theorem 112 and the result follows immediately. A longer proof follows here.

$$\det(U^H AU - \lambda^2 I) = \det(U^H AU - \lambda^2 U^H IU)$$

$$= \det[U^H(A - \lambda^2 I)U] = \det(U^H)\det(A - \lambda^2 I)\det(U) = \det(A - \lambda^2 I)$$

Recall that $\det(U) = e^{i\theta}$ by lemma 43, and

$$\det(U^H)\det(U) = 1$$

□

Corollary 33 *Let V be an orthonormal complex matrix such that $V^TV = I$, and let A be any other $n \times n$ complex matrix. Then A and V^TAV have the same eigenvalues.*

Proof. Although I provided this, it is also common knowledge. It is provided here for the sake of completeness. $V^{-1} = V^T$. Apply theorem 112. The result follows immediately. An alternate proof follows.

$$\det(V^TAV - \lambda^2 I) = \det(V^TAV - \lambda^2 V^T I V)$$

$$= \det[V^T(A - \lambda^2 I)V] = \det(V^T)\det(A - \lambda^2 I)\det(V) = \det(A - \lambda^2 I)$$

□

M.3 Transformation to a Triangular Matrix with the Same Eigenvalues

Lemma 52 *Let $A \in \mathbf{C}^{m\times m}$. Then there exists a unitary matrix U such that U^HAU is an upper triangular matrix whose diagonal elements are the eigenvalues of A. This is a complexification of Muirhead's theorem A9.1 [187].*

Proof. This is a complexification of Muirhead's proof. Let $\lambda_1^2, \cdots, \lambda_m^2$ be the eigenvalues of A, and let x_1 be an eigenvector of A corresponding to λ_1^2. Let $x_2, \cdots, x_m$ be any other vectors such that $x_1, x_2, \cdots, x_m$ form a basis for $\mathbf{C}^m$. Using the inner-product Gram-Schmidt orthonormalization process given

in section L.1, construct from $x_1, x_2, \cdots, x_m$ an orthonormal basis given as the columns of the unitary matrix U_1, where the first column u_1 is proportional to x_1, so that u_1 is also an eigenvector of A corresponding to λ_1^2. Then the first column of AU_1 is $Au_1 = \lambda_1^2 u_1$, and hence the first column of $U_1^H AU_1$ is $\lambda_1^2 U_1^H u_1$. Since this is the first column of $\lambda_1^2 U_1^H U_1 = \lambda_1^2 I_m$, it is

$$\begin{pmatrix} \lambda_1^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Hence

$$U_1^H AU_1 = \begin{pmatrix} \lambda_1^2 & B_1 \\ 0 & A_2 \end{pmatrix}$$

where A_2 is $(m-1) \times (m-1)$. By lemma 43, $\det(U_1) = e^{i\theta}$. Thus

$$\det(U_1^H AU_1 - \lambda^2 I_m) = \det(U_1^H AU_1 - \lambda^2 U_1^H I_m U_1)$$

$$= \det(U_1^H) \det(A - \lambda^2 I_m) \det(U_1) = \det(A - \lambda^2 I_m)$$

$$= \det \begin{pmatrix} \lambda_1^2 - \lambda^2 & B_1 \\ 0 & A_2 - \lambda^2 I_{m-1} \end{pmatrix} = (\lambda_1^2 - \lambda^2) \det(A_2 - \lambda^2 I_{m-1})$$

Since A and $U_1^H AU_1$ have the same eigenvalues, then the eigenvalues of A_2 are $\lambda_2^2, \cdots, \lambda_m^2$.

Now, using a construction similar to that above, we want to find an orthonormal $(m-1) \times (m-1)$ matrix U_2 whose first column is an eigenvector

of A_2 corresponding to λ_2^2. Then

$$U_2^H A_2 U_2 = \begin{pmatrix} \lambda_2^2 & B_2 \\ 0 & A_3 \end{pmatrix}$$

where A_3 is $(m-2) \times (m-2)$ with eigenvalues $\lambda_3^2, \cdots, \lambda_m^2$.

Repeating this procedure an additional $m-3$ times we now define the orthonormal matrix

$$U = U_1 \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & U_3 \end{pmatrix} \cdots \begin{pmatrix} I_{m-2} & 0 \\ 0 & U_{m-1} \end{pmatrix}$$

Note that $U^H A U$ is upper triangular with diagonal elements equal to $\lambda_1^2, \cdots, \lambda_m^2$.

□

Lemma 53 *Let $A \in \mathbf{C}^{m \times m}$. Then there exists an orthonormal matrix V such that $V^T A V$ is an* upper *triangular matrix whose diagonal elements are the eigenvalues of A. This is a corollary to a complexification of Muirhead's theorem A9.1 [187].*

Proof. This is a complexification and adaptation of Muirhead's proof of his theorem A9.1. Note that even though a transpose is in the problem, this is still different from the real case.

Let $\lambda_1^2, \cdots, \lambda_m^2$ be the eigenvalues of A, and let x_1 be an eigenvector of A corresponding to λ_1^2. Let $x_2, \cdots, x_m$ be any other vectors such that $x_1, x_2, \cdots, x_m$ form a basis for $\mathbf{C}^m$. Using the bilinear Gram-Schmidt orthonormalization process given in section L.2, construct from $x_1, x_2, \cdots, x_m$ an orthonormal basis

given as the columns of the orthonormal complex matrix V_1, where the first column v_1 is proportional to x_1, so that v_1 is also an eigenvector of A corresponding to λ_1^2. Then the first column of AV_1 is $Av_1 = \lambda_1^2 v_1$, and hence the first column of $V_1^T AV_1$ is $\lambda_1^2 V_1^T v_1$. Since this is the first column of $\lambda_1^2 V_1^T V_1 = \lambda_1^2 I_m$, it is

$$\begin{pmatrix} \lambda_1^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Hence

$$V_1^T AV_1 = \begin{pmatrix} \lambda_1^2 & B_1 \\ 0 & A_2 \end{pmatrix}$$

where A_2 is $(m-1) \times (m-1)$. By lemma 44, $\det(V_1) = \pm 1$. Thus

$$\det(V_1^T AV_1 - \lambda^2 I_m) = \det(V_1^T AV_1 - \lambda^2 V_1^H I_m V_1)$$

$$= \det(V_1^T) \det(A - \lambda^2 I_m) \det(V_1) = \det(A - \lambda^2 I_m)$$

$$= \det \begin{pmatrix} \lambda_1^2 - \lambda^2 & B_1 \\ 0 & A_2 - \lambda^2 I_{m-1} \end{pmatrix} = (\lambda_1^2 - \lambda^2) \det(A_2 - \lambda^2 I_{m-1})$$

Since A and $V_1^T AV_1$ have the same eigenvalues, then the eigenvalues of A_2 are $\lambda_2^2, \cdots, \lambda_m^2$.

Now, using a construction similar to that above, we want to find an orthonormal $(m-1) \times (m-1)$ matrix V_2 whose first column is an eigenvector

of A_2 corresponding to λ_2^2. Then

$$V_2^T A_2 V_2 = \begin{pmatrix} \lambda_2^2 & B_2 \\ 0 & A_3 \end{pmatrix}$$

where A_3 is $(m-2) \times (m-2)$ with eigenvalues $\lambda_3^2, \cdots, \lambda_m^2$.

Repeating this procedure an additional $m-3$ times, we now define the orthonormal matrix

$$V = V_1 \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & V_3 \end{pmatrix} \cdots \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1} \end{pmatrix}$$

Note that $V^T A V$ is upper triangular with diagonal elements equal to $\lambda_1^2, \cdots, \lambda_m^2$.

$$V^T A V = \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1}^T \end{pmatrix} \cdots \begin{pmatrix} I_2 & 0 \\ 0 & V_3^T \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & V_2^T \end{pmatrix} V_1^T A V_1 \times$$

$$\times \begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & V_3 \end{pmatrix} \cdots \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1} \end{pmatrix}$$

$$= \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1}^T \end{pmatrix} \cdots \begin{pmatrix} I_2 & 0 \\ 0 & V_3^T \end{pmatrix} \begin{pmatrix} \lambda_1^2 & B_1 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & V_3 \end{pmatrix} \cdots \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1} \end{pmatrix}$$

$$= \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1}^T \end{pmatrix} \cdots \begin{pmatrix} \lambda_1^2 & & B_1 \\ 0 & \lambda_2^2 & B_2 \\ 0 & 0 & A_3 \end{pmatrix} \cdots \begin{pmatrix} I_{m-2} & 0 \\ 0 & V_{m-1} \end{pmatrix} \qquad \text{(M.1)}$$

$$= \begin{pmatrix} \lambda_1^2 & & B \\ & \lambda_2^2 & \\ & & \\ 0 & & \lambda_m^2 \end{pmatrix}$$

Note that the nature of the λ_i^2 was not at issue here. They are not necessarily real. Also, compared to lemma 52, $V \neq U$ in general, even though the eigenvalues are the same in both cases. □

M.4 Functions of Eigenvalues

Theorem 113 *Let A be an $n \times n$ complex matrix with eigenvalues $\lambda_1^2, \cdots, \lambda_n^2$. Then* $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i^2$. *This is a complexification of Graybill theorem 9.1.3 [95].*

Proof. By lemma 52, there exists a unitary matrix U such that $U^H AU$ is an upper triangular matrix whose diagonal elements are the eigenvalues of A. Call it T. By property of the trace function $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, we see that

$$\mathrm{tr}(A) = \mathrm{tr}(AI) = \mathrm{tr}(AUU^H) = \mathrm{tr}(U^H AU) = \mathrm{tr}(T) = \sum_{i=1}^{n} \lambda_i^2$$

□

Theorem 114 *Let A be an $n \times n$ complex matrix with eigenvalues $\lambda_1^2, \cdots, \lambda_n^2$. Then*

$$\det(A) = \prod_{i=1}^{n} \lambda_i^2$$

Proof. By lemma 52, there exists a unitary matrix U such that $U^H AU$ is an upper triangular matrix whose diagonal elements are the eigenvalues of A. Call it T. By lemma 43, $\det(U) = e^{i\theta}$. Thus

$$\det(I) = \det(U^H U) = 1$$

and

$$\det(A) = \det(U^H)\det(A)\det(U) = \det(U^H AU)$$

Since U and A are conformable square matrices. Thus

$$\det(A) = \det(T) = \prod_{i=1}^{n} \lambda_i^2$$

□

M.5 Eigenvalue Decomposition

Theorem 115 *(*Very Important*). If A is an Hermitian $m \times m$ matrix with eigenvalues $\lambda_1^2, \cdots, \lambda_m^2$, then there exists a unitary matrix U such that*

$$U^H AU = D = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2) = \Lambda^2$$

If $U = [U_1, \cdots, U_m]$, then U_i is an eigenvector of A corresponding to the eigenvalue λ_i^2. Moreover, if $\lambda_1^2, \cdots, \lambda_m^2$ are all distinct, then the representation $U^H AU = D$ is unique up to phase changes *in the first row of U. This is a complexification of Muirhead's theorem A9.2 [187].*

Proof. This is a complexification of Muirhead's proof. From lemma 52, there exists a unitary matrix U_1 such that

$$U_1^H A U_1 = \begin{pmatrix} \lambda_1^2 & B_1 \\ 0 & A_2 \end{pmatrix}$$

where $\lambda_2^2, \cdots, \lambda_m^2$ are the eigenvalues of A_2. $A^H = A$ implies

$$(U_1^H A U_1)^H = U_1^H A^H U_1 = U_1^H A U_1$$

is also Hermitian. thus, B_1 is a zero matrix, $B_1 = 0$. Similarly, each B_i in the proof of lemma 52 is zero $(i - 1, \cdots, m - 1)$. Thus, U given in lemma 52 satisfies

$$U^H A U = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2) = \Lambda^2$$

Observe that $UU^H AU = AU = U\Lambda^2$. Consequently $AU_i = U_i \lambda_i^2$ so that U_i is an eigenvector of A corresponding to the eigenvalue λ_i^2.

Now, suppose that we also have $Q^H AQ = D$ for a unitary matrix Q. Let $P = Q^H U$. Then

$$PD = (Q^H U)(U^H AU) = Q^H AU$$

and

$$DP = (Q^H AQ)(Q^H U) = Q^H AU$$

thus $PD = DP$. If $P = (p_{ij})$, it follows that $p_{ij}\lambda_j^2 = p_{ij}\lambda_i^2$. Since $\lambda_i^2 \neq \lambda_j^2$ by hypothesis, $p_{ij} = 0$ for all $i \neq j$. Note that

$$P^H P = U^H Q Q^H U = I$$

Since P is unitary and diagonal, it must have the form $P = \text{diag}(e^{i\theta_1}, \cdots, e^{i\theta_m})$. Thus

$$U = QP = Q \text{ diag}(e^{i\theta_1}, \cdots, e^{i\theta_m})$$

Note that when working with $A \in \mathbf{R}^{m \times m}$, this last property simplifies to saying that if $\lambda_1^2, \cdots, \lambda_m^2$ are all distinct, then the representation $U^H A U = D$ is unique up to sign changes in the first row of U. □

Caution. Not all eigenvalue decompositions can be written in the form of $A = Q\Lambda^2 Q^H$. The conditions on our theorem requiring A to be Hermitian give us the form we are familiar with. When A is not Hermitian, we get the form $A = Q\Lambda^2 Q^{-1}$.

Example. Let

$$A = \begin{pmatrix} 4 & -1 \\ 2 & 1 \end{pmatrix}$$

Then

$$\begin{pmatrix} 4 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and

$$\begin{pmatrix} 4 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Thus, the eigenvalues are $\{2, 3\}$ with associated nonnormalized eigenvectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The normalized eigenvectors are

$$\begin{pmatrix} \sqrt{1/5} \\ \sqrt{2/5} \end{pmatrix}$$

and

$$\begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \end{pmatrix}$$

Then we observe

$$\begin{pmatrix} \sqrt{1/5} & \sqrt{1/2} \\ \sqrt{2/5} & \sqrt{1/2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} \sqrt{1/5} & \sqrt{2/5} \\ \sqrt{1/2} & \sqrt{1/2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{5} + \frac{3}{2} & \frac{2\sqrt{2}}{5} + \frac{3}{2} \\ \frac{2\sqrt{2}}{5} + \frac{3}{2} & \frac{4}{5} + \frac{3}{2} \end{pmatrix} \neq \begin{pmatrix} 4 & -1 \\ 2 & 1 \end{pmatrix}$$

Thus we do not get back A when we compute $Q\Lambda^2 Q^H$. However, noting that

$$Q^{-1} = \begin{pmatrix} -\sqrt{5} & \sqrt{5} \\ 2\sqrt{2} & -\sqrt{2} \end{pmatrix}$$

we find that $A = Q\Lambda^2 Q^{-1}$. □

Theorem 116 *Let $A^T = A \in M_n(\mathbf{C})$ have eigenvalues $\lambda_1^2, \cdots, \lambda_n^2$. Then there exists a matrix V such that $V^T V = I$ and*

$$V^T A V = D = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2) = \Lambda^2$$

If $V = [V_1, \cdots, V_m]$, then V_i is an eigenvector of A corresponding to the eigenvalue λ_i^2. Moreover, if $\lambda_1^2, \cdots, \lambda_m^2$ are all distinct, then the representation $V^T AV = D$ is unique up to sign changes in the first row of V. This is a corollary to a complexification of Muirhead's very important theorem A9.2 [187].

Proof. This is a complexification of Muirhead's proof of his theorem A9.2. From lemma 53, there is an orthonormal $m \times m$ matrix V_1 such that

$$V_1^T AV_1 = \begin{pmatrix} \lambda_1^2 & B_1 \\ 0 & A_2 \end{pmatrix}$$

where $\lambda_2^2, \cdots, \lambda_n^2$ are the eigenvalues of A_2. $A^T = A$ implies

$$(V_1^T AV_1)^T = V_1^T A^T V_1 = V_1^T AV_1$$

Thus B_1 is a zero matrix. Similarly, each B_i in the proof of lemma 53 is zero for $1 \leq i \leq n-1$. Thus the V of lemma 53 satisfies

$$V^T AV = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2) = \Lambda^2$$

Now, $V^T V = I$ implies $V^T = V^{-1}$, which in turn implies $VV^T = I$. So,

$$VV^T AV = AV = V\Lambda^2$$

Consequently, $AV_i = V_i \lambda_i^2$, so that V_i is an eigenvector of A corresponding to the eigenvalue λ_i^2. Also, suppose $Q^T AQ = D$ for $Q^T Q = I$. Let $P = Q^T V$. Then

$$PD = (Q^T V)(V^T AV) = Q^T AV$$

and

$$DP = (Q^T AQ)(Q^T V) = Q^T AV$$

Thus $PD = DP$. Let $P = (p_{ij})$. Then $p_{ij}\lambda_j^2 = p_{ij}\lambda_i^2$. If $\lambda_i^2 \neq \lambda_j^2$, then $p_{ij} = 0$ for all $i \neq j$. Note that

$$P^T P = V^T Q Q^T V = I$$

Since P is orthonormal and diagonal,

$$P = \text{diag}(\pm 1, \pm 1, \cdots, \pm 1)$$

Thus

$$V = QP = Q\,\text{diag}(\pm 1, \pm 1, \cdots, \pm 1)$$

□

M.6 Hermitian Definiteness

Corollary 34 *If A is an Hermitian $m \times m$ matrix with eigenvalues $\lambda_1^2, \cdots, \lambda_m^2$. then $\lambda_i^2 \in \mathbf{R}$ for all $i \in [1, m]$. This is a corollary to a complexification of Muirhead's theorem A9.2 [187]. It is a widely known result.*

Proof. From theorem 115, we know there exists unitary U such that

$$U^H AU = \Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2)$$

Since $A = A^H$, we know

$$(U^H AU)^H = U^H AU$$

Therefore $(\Lambda^2)^H = \Lambda^2$. This can only be true if $\lambda_i^2 \in \mathbf{R}$ for all i.

Note that when $A^T = A \in M_m(\mathbf{C})$, we have orthonormal V such that

$$V^T A V = \Lambda^2$$

$A = A^T$ implies that $(V^T A V)^T = V^T A V$, and thus $(\Lambda^2)^T = \Lambda^2$. However, we cannot deduce from this condition that $\lambda^2 \in \mathbf{R}$ for any i. This is a fundamental difference between complex symmetric matrices and Hermitian matrices. This says you cannot automatically assume definiteness for a complex symmetric matrix. □

Theorem 117 *The $m \times m$ Hermitian matrix A is positive (negative) (semi)-definite if and only if the matrix*

$$\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2)$$

of eigenvalues also is. This is Stewart's corollary 6.5.3 [259].

Proof. Let U and Λ^2 be the matrices of eigenvectors and corresponding eigenvalues of A. Recall that $U^H A U = \Lambda^2$. Let $y = Ux$ for all $x \in \mathbf{C}^m$.

Λ^2 is positive (negative) (semi-)definite if and only if $U^H A U$ is. $x^H \Lambda^2 x > 0$ implies $x^H U^H A U x > 0$ for all nonzero x in $\mathbf{C}^m$. In turn, this implies $y^H A y > 0$ for all nonzero y in $\mathbf{C}^m$. Since U is nonsingular, $y = Ux$ is a one-to-one mapping. As x ranges over all $\mathbf{C}^m$, then y also ranges over all $\mathbf{C}^m$. The

inequality can be any of $>, \geq, \leq$, or $<$. Thus

$$x^H \Lambda^2 x > 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } y^H \Lambda^2 y > 0 \; \forall y \neq 0 \text{ in } \mathbf{C}^m$$

$$x^H \Lambda^2 x \geq 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } y^H \Lambda^2 y \geq 0 \; \forall y \neq 0 \text{ in } \mathbf{C}^m$$

$$x^H \Lambda^2 x < 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } y^H \Lambda^2 y < 0 \; \forall y \neq 0 \text{ in } \mathbf{C}^m$$

$$x^H \Lambda^2 x \leq 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } y^H \Lambda^2 y \leq 0 \; \forall y \neq 0 \text{ in } \mathbf{C}^m$$

Note that $X_i^* \lambda_i^2 x_i = |x_i|^2 \lambda^2$. This means

$$x^H \Lambda^2 x > 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } \lambda_i^2 > 0 \; \forall i$$

$$x^H \Lambda^2 x \geq 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } \lambda_i^2 \geq 0 \; \forall i$$

$$x^H \Lambda^2 x < 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } \lambda_i^2 < 0 \; \forall i$$

$$x^H \Lambda^2 x \leq 0 \; \forall x \neq 0 \text{ in } \mathbf{C}^m \text{ implies } \lambda_i^2 \leq 0 \; \forall i$$

□

Theorem 118 *The $m \times m$ matrix A^{-1} is Hermitian positive (negative) definite if and only if A is Hermitian positive (negative) definite.*

Proof. By theorem 115, let $A = U \Lambda^2 U^H$ where $\Lambda^2 = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2)$ is the matrix of eigenvalues with corresponding eigenvectors in matrix U. Let $\lambda_i^2 = 0$ for all i. Then

$$A^{-1} = (U \Lambda^2 U^H)^{-1} = U^{-H} \Lambda^{-2} U^{-1}$$

U is unitary. Thus $U^{-1} = U^H$ and $U = U^{-H}$, which implies $A^{-1} = U \Lambda^{-2} U^H$, where

$$\Lambda^{-2} = \text{diag}(\frac{1}{\lambda_1^2}, \cdots, \frac{1}{\lambda_m^2})$$

Note that $\frac{1}{\lambda_i^2} > 0$ if and only if $\lambda_i^2 > 0$, and $\frac{1}{\lambda_i^2} < 0$ if and only if $\lambda_i^2 < 0$. Therefore, if $\lambda_i^2 > 0$ for all i, then $\frac{1}{\lambda_i^2} > 0$ for all i. Thus A is Hermitian positive definite if and only if A^{-1} is Hermitian positive definite. Similarly, A is Hermitian negative definite if and only if A^{-1} is Hermitian negative definite.

□

M.7 Square Root Decomposition

Theorem 119 *(*Very Important!*) Let A be a non-negative definite complex $m \times m$ matrix. Then there exists a non-negative definite complex $m \times m$ matrix, written as $A^{1/2}$, such that $A = A^{1/2}(A^{1/2})^H$. There also exists a $B^{1/2}$ such that $(B^{1/2})^H(B^{1/2}) = A$. These are Hermitian Square Root matrices. Their existence provides a key to obtaining numerically robust methods such as in Kalman square root filtering. This is an important complexification of Muirhead's theorem A9.3 [187]. These are widely known results.*

Proof. This is a complexification of Muirhead's proof. Let H be a unitary matrix such that $H^H A H = D$, where

$$D = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2)$$

with $\lambda_1^2, \cdots, \lambda_m^2$ being the eigenvalues of A. Since A is nonnegative definite, $\lambda_i^2 \geq 0$ for $i = 1, \cdots, m$. Let

$$D^{1/2} = \text{diag}(\lambda_1, \cdots, \lambda_m) = (D^{1/2})^H$$

Then $D^{1/2}(D^{1/2})^H = D$. Let $A^{1/2} = HD^{1/2}H^H$. Then

$$(A^{1/2})(A^{1/2})^H = HD^{1/2}H^H(HD^{1/2}H^H)^H$$

$$= HD^{1/2}\underbrace{H^HH}_{I}(D^{1/2})^HH^H = HDH^H = A$$

Therefore $(A^{1/2})(A^{1/2})^H = A$.

Similarly, there exists a $B^{1/2}$ such that $(B^{1/2})^H(B^{1/2}) = A$. Let $(B^{1/2})^H = A^{1/2}$, and the proof is complete. If A is positive definite, then $A^{1/2}$ is also positive definite. □

Theorem 120 *Let A be a non-negative definite complex $m \times m$ matrix. Then there exists a non-negative definite complex $m \times m$ matrix, written as $A^{1/2}$, such that $A = A^{1/2}A^{1/2}$. This is another complexification of Muirhead's theorem A9.3 [187]. This is a widely known result.*

Proof. This is a complexification of Muirhead's proof. Let H be a unitary matrix such that $H^HAH = D$, where $D = \text{diag}(\lambda_1^2, \cdots, \lambda_m^2)$ with $\lambda_1^2, \cdots, \lambda_m^2$ being the eigenvalues of A. Since A is nonnegative definite, $\lambda_i^2 \geq 0$ for $i = 1, \cdots, m$. Let

$$D^{1/2} = \text{diag}(\lambda_1, \cdots, \lambda_m)$$

Then $D^{1/2}D^{1/2} = D$. Let $A^{1/2} = HD^{1/2}H^H$. Then

$$(A^{1/2})(A^{1/2}) = HD^{1/2}H^H(HD^{1/2}H^H)$$

$$= HD^{1/2}\underbrace{H^HH}_{I}D^{1/2}H^H = HDH^H = A$$

Therefore $(A^{1/2})(A^{1/2})^H = A$.

If A is positive definite, then $A^{1/2}$ is also positive definite.

Examining the proof, note that $A^{1/2} = (A^{1/2})^H$. □

Theorem 121 *Let A be an $m \times m$ non-negative definite complex matrix of rank r. Then (i) there is an $m \times r$ matrix B of rank r such that $A = BB^H$, and (ii) there is an $m \times m$ nonsingular matrix C such that*

$$A = C \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C^H$$

This is a complexification of Muirhead's theorem A9.4.

Proof. This is a complexification of Muirhead's proof. First, we prove (i). Let $D_1 = \text{diag}(\lambda_1^2, \cdots, \lambda_r^2)$ where $\lambda_1^2, \cdots, \lambda_r^2$ are the nonzero eigenvalues of A. Let H be an $m \times m$ unitary matrix such that

$$H^H A H = \text{diag}(\lambda_1^2, \cdots, \lambda_r^2, 0, \cdots, 0)$$

Partition H as $H = [H_1, H_2]$, where H_1 is $m \times r$ and H_2 is $m \times (m-r)$. Then

$$A = H \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} H^H = [H_1, H_2] \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_1^H \\ H_2^H \end{pmatrix} = H_1 D_1 H_1^H$$

Let $D_1^{1/2} = \text{diag}(\lambda_1, \cdots, \lambda_r)$. Then

$$A = H_1 D_1^{\frac{1}{2}} (D_1^{\frac{1}{2}})^H H_1^H = BB^H$$

where $B = H_1 D_1^{1/2}$ is $m \times r$ of rank r.

Now we prove (ii). Let C be an $m \times m$ nonsingular matrix whose first r columns are the columns of B in part (i). Then $C = [B, E]$.

$$C \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C^H = [B, E] \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B^H \\ E^H \end{pmatrix} = BB^H = A$$

□

Corollary 35 *Let A be an $m \times m$ non-negative definite complex matrix of rank r. Then (i) there is an $m \times r$ matrix B of rank r such that $A = B^H B$, and (ii) there is an $m \times m$ nonsingular matrix C such that*

$$A = C^H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C$$

This is another complexification of Muirhead's theorem A9.4 [187].

Proof. This is a complexification of Muirhead's proof. First, we prove (i). Let $D_1 = \text{diag}(\lambda_1^2, \cdots, \lambda_r^2)$ where $\lambda_1^2, \cdots, \lambda_r^2$ are the nonzero eigenvalues of A. Let H be an $m \times m$ unitary matrix such that

$$H^H A H = \text{diag}(\lambda_1^2, \cdots, \lambda_r^2, 0, \cdots, 0)$$

Partition H as $H = [H_1, H_2]$, where H_1 is $m \times r$ and H_2 is $m \times (m - r)$. Then

$$A = H \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} H^H = [H_1, H_2] \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} H_1^H \\ H_2^H \end{pmatrix} = H_1 D_1 H_1^H$$

Let $D_1^{1/2} = \text{diag}(\lambda_1, \cdots, \lambda_r)$. Then

$$A = H_1 D_1^{1/2} (D_1^{1/2})^H H_1^H = B^H B$$

where $B = (H_1 D_1^{1/2})^H$ is $r \times m$ of rank r.

Now we prove (ii). Let C be an $m \times m$ nonsingular matrix whose first r rows are the rows of B in part (i). Then

$$C = \begin{pmatrix} B \\ E \end{pmatrix}$$

$$C^H \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} C = [B^H, E^H] \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B \\ E \end{pmatrix}$$

$$= [B^H, 0] \begin{bmatrix} B \\ E \end{bmatrix} = B^H B = A$$

□

M.8 Unitary Transformations

Theorem 122 *Suppose that A and B are complex matrices where $A \in \mathbf{C}^{k \times m}$ and $B \in \mathbf{C}^{k \times n}$, with $m \leq n$. Then $AA^H = BB^H$ if and only if there is an $m \times n$ matrix H with $HH^H = I_m$ such that $AH = B$. This is a complexification of Muirhead's theorem A9.5 [187].*

Proof. This is a complexification of Muirhead's proof. Suppose there is an $m \times n$ matrix H with $HH^H = I_m$ such that $AH = B$. Then

$$BB^H = (AH)(AH)^H = AHH^H A^H = AA^H$$

Now, suppose that $AA^H = BB^H$. Let C be a $k \times k$ nonsingular matrix such that

$$AA^H = BB^H = C \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} C^H$$

where $\text{rank}(AA^H) = r$. Matrix C exists by theorem 121. Let

$$D = C^{-1}A, \quad E = C^{-1}B \tag{M.2}$$

and partition these as

$$D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}$$

$$E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}$$

where D_1 is $r \times m$, D_2 is $(k-r) \times m$, E_1 is $r \times n$, and E_2 is $(k-r) \times n$. Then

$$EE^H = \begin{bmatrix} E_1E_1^H & E_1E_2^H \\ E_2E_1^H & E_2E_2^H \end{bmatrix} = C^{-1}BB^H(C^{-1})^H = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Also

$$DD^H = \begin{bmatrix} D_1D_1^H & D_1D_2^H \\ D_2D_1^H & D_2D_2^H \end{bmatrix} = C^{-1}AA^H(C^{-1})^H = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

This implies $E_1E_1^H = I_r$, $D_1D_1^H = I_r$, $E_2 = 0$, and $D_2 = 0$. Thus

$$D = \begin{bmatrix} D_1 \\ 0 \end{bmatrix}$$

and

$$E = \begin{bmatrix} E_1 \\ 0 \end{bmatrix}$$

Let $\tilde{E}_2$ be an $(n-r) \times n$ matrix such that

$$\tilde{E} = \begin{bmatrix} \tilde{E}_1 \\ \tilde{E}_2 \end{bmatrix}$$

is an $n \times n$ unitary matrix. Let

$$\tilde{D} = \begin{bmatrix} D_1 & 0 \\ \tilde{D}_2 & \tilde{D}_3 \end{bmatrix}$$

where $\tilde{D}_2$ is $(n-r) \times m$, and $\tilde{D}_3$ is $(n-r) \times (n-m)$ such that $\tilde{D}$ is an $n \times n$ unitary matrix. Note that we use D_1, not $\hat{D}_1$. Then

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \tilde{E} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{E}_1 \\ \tilde{E}_2 \end{bmatrix} = \begin{bmatrix} \tilde{E}_1 \\ 0 \end{bmatrix} = E$$

and

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \tilde{D} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ \tilde{D}_2 & \tilde{D}_3 \end{bmatrix} = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D & 0 \end{bmatrix}$$

Notice that

$$E = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \tilde{E} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \tilde{D} \tilde{D}^H \tilde{E} \tag{M.3}$$

$$= \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^H \tilde{E} = \begin{bmatrix} \cap & \cap \end{bmatrix} Q \tag{M.4}$$

since $\tilde{D}$ is unitary. Define $Q = \tilde{D}^H \tilde{E}$.

Examine Q.

$$Q = \tilde{D}^H \tilde{E} = \begin{bmatrix} D_1^H & \tilde{D}_2^H \\ 0 & \tilde{D}_3^H \end{bmatrix} \begin{bmatrix} E_1 \\ \tilde{E}_2 \end{bmatrix} = \begin{bmatrix} D_1^H E_1 + \tilde{D}_2^H \tilde{E}_2 \\ \tilde{D}_3^H \tilde{E}_2 \end{bmatrix}$$

$\tilde{E}$ is unitary which implies $\tilde{E}\tilde{E}^H = I_n$. $\tilde{D}$ is unitary implies $\tilde{D}\tilde{D}^H = I_n = \tilde{D}^H \tilde{D}$.

Then

$$(\tilde{D}^H \tilde{E})(\tilde{D}^H \tilde{E})^H = \tilde{D}^H \tilde{E}\tilde{E}^H \tilde{D} = \tilde{D}^H \tilde{D} = I_n$$

Therefore Q is unitary.

Let Q be partitioned as

$$Q = \begin{bmatrix} H \\ P \end{bmatrix} \tag{M.5}$$

where H is $m \times n$ and P is $(n-m) \times m$. Then $HH^H = I_m$, since Q is unitary,

$$QQ^H = \begin{bmatrix} I_m & 0 \\ 0 & I_{n-m} \end{bmatrix}$$

Then

$$\begin{aligned} C^{-1}B &= E && \text{by equation M.2} \\ &= \begin{bmatrix} D & 0 \end{bmatrix} Q && \text{by equation M.4} \\ &= \begin{bmatrix} D & 0 \end{bmatrix} \begin{bmatrix} H \\ P \end{bmatrix} && \text{by equation M.5} \\ &= DH = C^{-1}AH && \text{by equation M.2} \end{aligned}$$

This implies $CC^{-1}B = CC^{-1}AH = B = AH$ which is the result we are seeking. □

Theorem 123 *Suppose that A and B are complex matrices where $A \in \mathbf{C}^{m\times k}$ and $B \in \mathbf{C}^{n\times k}$, with $m \leq n$. Then $A^H A = B^H B$ if and only if there exists some $H \in \mathbf{C}^{n\times m}$ such that $H^H H = I_m$ and $B = HA$. This is a corollary to a complexification of Muirhead's theorem A9.5 [187].*

Proof. This is a slight modification of a complexification to Muirhead's proof. Suppose there is some $H \in \mathbf{C}^{n\times m}$ such that $H^H H = I_m$ and $B = HA$. Then

$$B^H B = (HA)^H (HA) = A^H H^H HA = A^H A$$

Now, suppose $A^H A = B^H B$. Let C be a $k \times k$ non-singular matrix such that

$$A^H A = B^H B = C^H \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} C$$

where $\text{rank}(A^H A) = r$. C exists, by corollary 35. Let

$$D = AC^{-1},\ E = BC^{-1} \tag{M.6}$$

and partition these as $D = [D_1, D_2]$, $E = [E_1, E_2]$, where D_1 is $m \times r$, D_2 is $m \times (k-r)$, E_1 is $n \times r$, and E_2 is $n \times (k-r)$. Then

$$E^H E = \begin{bmatrix} E_1^H E_1 & E_1^H E_2 \\ E_2^H E_1 & E_2^H E_2 \end{bmatrix} = (BC^{-1})^H BC^{-1} = C^{-H} B^H BC^{-1}$$

$$= C^{-H} C^H \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} CC^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{k\times k}$$

Also

$$D^H D = \begin{bmatrix} D_1^H D_1 & D_1^H D_2 \\ D_2^H D_1 & D_2^H D_2 \end{bmatrix} = (AC^{-1})^H (AC^{-1})$$

$$= C^{-H} A^H A C^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{k \times k}$$

This implies $E_1^H E_1 = D_1^H D_1 = I_r$, $E_2 = 0$, and $D_2 = 0$. Thus $D = [D_1, 0]$ and $E = [E_1, 0]$.

Let $\tilde{E}_2$ be an $n \times (n-r)$ matrix such that $\tilde{E} = [E_1, \tilde{E}_2]$ is an $n \times n$ unitary matrix. Let

$$\tilde{D} = \begin{bmatrix} D_1 & \tilde{D}_2 \\ 0 & \tilde{D}_3 \end{bmatrix}$$

where $\tilde{D}_2$ is $m \times (n-r)$, and $\tilde{D}_3$ is $(n-m) \times (n-r)$ such that $\tilde{D}$ is an $n \times n$ unitary matrix. Then

$$\tilde{E} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = [E_1, \tilde{E}_2] \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = [E_1, 0] = E$$

and

$$\tilde{D} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = \begin{bmatrix} D_1 & \tilde{D}_2 \\ 0 & \tilde{D}_3 \end{bmatrix} \cdot \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D \\ 0 \end{bmatrix}$$

Notice that

$$E = \tilde{E} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = \tilde{E} \tilde{D}^H \tilde{D} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times k}^{*} = \tilde{E} \tilde{D}^H \begin{bmatrix} D \\ 0 \end{bmatrix} = Q \begin{bmatrix} D \\ 0 \end{bmatrix} \tag{M.7}$$

since $\tilde{D}$ is unitary and $Q = \tilde{E}\tilde{D}^H$. We see that

$$Q^H Q = (\tilde{E}\tilde{D}^H)^H(\tilde{E}\tilde{D}^H) = \tilde{D}\tilde{E}^H\tilde{E}\tilde{D}^H = \tilde{D}\tilde{D}^H = I_n$$

Therefore Q is unitary. Partition Q as

$$Q = [H, P] \tag{M.8}$$

where H is $n \times m$ and P is $n \times (n-m)$. Then $H^H H = I_m$ since Q is unitary and $Q^H Q = \begin{bmatrix} I_m & 0 \\ 0 & I_{n-m} \end{bmatrix}$. Then

$$\begin{aligned}
&BC^{-1} = E && \text{by equation M.6} \\
&= Q\begin{bmatrix} D \\ 0 \end{bmatrix} = [H, P]\begin{bmatrix} D \\ 0 \end{bmatrix} = HD && \left\{ \begin{array}{l} \text{by equations M.7} \\ \text{and M.8} \end{array} \right. \\
&= HAC^{-1} && \text{by equation M.6} \\
&\Rightarrow BCC^{-1} = HAC^{-1}C = B = HA && \text{which is the result we want}
\end{aligned}$$

□

Consider

$$\begin{bmatrix} \underbrace{I_r}_{r\times r} & \underset{r\times(k-r)}{0} \\ \underset{(n-r)\times r}{0} & \underset{(n-r)\times(k-r)}{0} \end{bmatrix}_{n\times k} \qquad \begin{bmatrix} \underset{m\times k}{D} \\ \underset{(n-m)\times k}{0} \end{bmatrix}_{n\times k}$$

Note that there are two different matrices of the form $\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. One has dimensions $k \times k$ and the other has dimensions $n \times k$. Dimensions of various matrices are $Q_{n\times n}$, $D_{m\times k}$, and $E_{n\times k}$.

Theorem 124 *Let A be an $n \times m$ complex matrix of* $\text{rank}(A) = m$ *where $n \geq m$. Then (i) A can be written as $A = H_1 B$, where H_1 is $n \times m$ with $H_1^H H_1 = I_m$ and B is $m \times m$ positive definite, and (ii) A can be written as*

$$A = H \begin{bmatrix} I_m \\ 0 \end{bmatrix} B$$

where H is $n \times n$ unitary and B is $m \times m$ positive definite. This is a complexification of Muirhead's theorem A9.6 [187].

Proof. This is a complexification of Muirhead's proof. (i) Let B be such that $B^H B = A^H A$. This B exists by theorem 119. B is the positive definite Hermitian square root of $(A^H A)$. By theorem 123, A can be written as $A = H_1 B$ where H_1 is $n \times m$ with $H_1^H H_1 = I_m$.

(ii) Let H_1 be the matrix in (i) such that $A = H_1 B$ and choose an $n \times (n-m)$ matrix H_2 so that $H = [H_1, H_2]$ is an $n \times n$ unitary matrix. Then

$$A = H_1 B = [H_1, H_2] \begin{bmatrix} I_m \\ 0 \end{bmatrix} B = H \begin{bmatrix} I_m \\ 0 \end{bmatrix} B$$

□

Theorem 125 *Let A be an $n \times m$ complex matrix of* $\text{rank}(A) = n$ *where $m \geq n$. Then (i) A can be written as $A = B H_1$, where H_1 is $n \times m$ with $H_1 H_1^H = I_n$ and B is $n \times n$ positive definite, and (ii) A can be written as $A = B[I_n, 0]H$ where H is $m \times m$ unitary and B is $n \times n$ positive definite. This is a corollary to a complexification of Muirhead's theorem A9.6 [187].*

Proof. This is a modified version of a complexification of Muirhead's proof. (i) Let B be such that $BB^H = AA^H$. This B exists by theorem 119. B is the positive definite Hermitian square root of (AA^H). By theorem 122, there exists an $n \times m$ matrix H_1 such that $H_1 H_1^H = I_n$ and $A = BH_1$.

(ii) Let H_1 be the matrix in (i) such that $A = BH_1$ and choose an $(m - n) \times m$ matrix H_2 so that

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

is an $m \times m$ unitary matrix. Then

$$A = BH_1 = B[I_n, 0] \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = B[I_n, 0]H$$

□

M.9 Cholesky or Bartlett Decomposition

Theorem 126 *(*Very Important*) If A is an $m \times m$ positive definite complex matrix, then there is a unique $m \times m$ upper triangular matrix T with positive diagonal elements such that $A = T^H T$. This is known as* Cholesky *or* Bartlett *decomposition (see p. 134)[259]. This is a complexification of Muirhead's theorem A9.7 [187].*

Proof. This is a complexification of Muirhead's proof. We prove it by induction. When $m = 1$ and A is positive definite, then $A > 0$ and there

exists a T such that $A = T^H T = \sqrt{A}\sqrt{A}$.

Suppose the result holds for positive definite matrices of size $m-1$. Partition the $m \times m$ matrix A as

$$A = \begin{bmatrix} A_{11} & A_{12}^H \\ A_{12} & A_{22} \end{bmatrix}$$

where A_{11} is $(m-1)\times(m-1)$. Assume there exists a unique $(m-1)\times(m-1)$ upper triangular matrix T_{11} with positive (therefore real) diagonal elements such that $A_{11} = T_{11}^H T_{11}$. Suppose

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix} = \begin{bmatrix} T_{11}^H & 0^H \\ x^H & y \end{bmatrix} \begin{bmatrix} T_{11} & x \\ 0 & y \end{bmatrix}$$

$$= \begin{bmatrix} T_{11}^H T_{11} & T_{11}^H x \\ x^H T_{11} & x^H x + y^2 \end{bmatrix}$$

where x is $(m-1)\times 1$ and $y \in \mathbf{R}$. Thus $a_{12}^H = x^H T_{11}$. Solving for x, $a_{12}^H T_{11}^{-1} = x^H$ which implies $x = (T_{11}^{-1})^H a_{12}$. Also, $a_{22} = x^H x + y^2$ implies

$$y^2 = a_{22} - x^H x = a_{22} - a_{12}^H T_{11}^{-1}(T_{11}^{-1})^H a_{12}$$

$$= a_{22} - a_{12}^H (T_{11}^H T_{11})^{-1} a_{12} = a_{22} - a_{12}^H A_{11}^{-1} a_{12}$$

where a_{12} is a column vector of dimension $(m-1)\times 1$. Since A is positive definite, then

$$a_{22} - a_{12}^H A_{11}^{-1} a_{12} > 0$$

The unique y satisfying this is

$$y = (a_{22} - a_{12}^H A_{11}^{-1} a_{12})^{1/2} > 0$$

We have thus found

$$T = \begin{bmatrix} T_{11} & x \\ 0 & y \end{bmatrix}$$

such that $A = T^H T$. □

Corollary 36 *If A is an $m \times m$ positive definite complex matrix, then there is a unique $m \times m$ lower triangular matrix L with positive diagonal elements such that $A = LL^H$. This is a corollary to a complexification of Muirhead's theorem A9.7 [187].*

Proof. By theorem 126, there is a unique $m \times m$ upper triangular matrix T with positive diagonal elements such that $A = T^H T$. Let $L = T^H$. Then $A = LL^H$. L is lower triangular. □

Theorem 127 *If A is a complex $n \times m$ matrix of rank m where $n \geq m$, then A can be uniquely written as $A = H_1 T$ where H_1 is $n \times m$ with $H_1^H H_1 = I_m$ and T is $m \times m$ upper triangular with positive real diagonal elements. This is a complexification of Muirhead's theorem A9.8.*

Proof. This is a complexification of Muirhead's proof. $A^H A$ is positive definite with dimensions $m \times m$. By theorem 126, there is a unique $m \times m$ upper triangular matrix with positive real diagonal elements such that $A^H A = T^H T$.

By theorem 123, there exists an $n \times m$ matrix H_1 such that $H_1^H H_1 = I_m$ and $A = H_1 T$. Because T is unique, we thus know H_1 is unique. Since $A^H A$ is $m \times m$ and positive definite, rank$(T) = m$. □

Corollary 37 *If A is a complex $m \times n$ matrix where $n \geq m$, then A can be uniquely written as $A = LH_1$ and* rank$(A) = m$ *where H_1 is $m \times n$ with $H_1 H_1^H = I_m$ and L is $m \times m$ lower triangular with positive real diagonal elements. This is a corollary to a complexification of Muirhead's theorem A9.8. This is also Srivastava lemma 1 [256].*

Proof. This is a modified complexification of Muirhead's proof. AA^H is positive definite with dimensions $m \times m$. By corollary 36, there is a unique lower triangular matrix L with positive real diagonal elements, such that $AA^H = LL^H$. Let L in this proof be A in theorem 122 and let A is this proof be B in theorem 122. Then by theorem 122, there exists an $m \times n$ matrix H_1 such that $H_1 H_1^H = I_m$ and $A = LH_1$. Because L is unique, we know H_1 is unique. Since AA^H is positive definite, rank$(L) = m$. □

M.10 Eigenvalues of Simply Modified Matrices

Lemma 54 *Let $\{t_i^2\}, i = 1, \cdots, p$ be the eigenvalues of real or complex square matrix X of dimension $p \times p$. Then the matrix $(I_p - X)$ has eigenvalues $\{1 -$*

$t_i^2\}, i = 1, \cdots, p$. *This lemma was motivated by a comment by Arnold (p. 418, bottom) [31].*

Proof. Let $\det(X - t^2 I) = 0$. Consider the matrix $(I - X)$.

$$\det[(I - X) - \tau^2 I] = 0$$

defines the eigenvalues of $(I - X)$. Then

$$\det[(I - X) - \tau^2 I] = \det(-X + I - \tau^2 I)$$

$$= \det[-X + (1 - \tau^2) I] = 0$$

Since this is zero, if I change the sign I still have zero. Thus

$$\det[X - (1 - \tau^2) I] = 0 = \det[-X + (1 - \tau^2) I]$$

Comparing with $\det[X - t^2 I] = 0$, we note that $1 - \tau^2 = t^2$, or $\tau^2 = 1 - t^2$. therefore if $\{t_i^2, i = 1, \cdots, p\}$ are the eigenvalues of X, then $\{1 - t_i^2, i = 1, \cdots, p\}$ are the eigenvalues of $I - X$. □

Proposition 72 *Let* $\{t_i^2\}, i = 1, \cdots, p$ *be the eigenvalues of real or complex square matrix* X *of dimension* $p \times p$. *Then the matrix* $(I_p + X)$ *has eigenvalues* $\{1 + t_i^2\}, i = 1, \cdots, p$. *This lemma was motivated by a comment by Arnold (p. 418, bottom) [31].*

Proof. Let $\det(X - t^2 I) = 0$. Consider the matrix $(I + X)$. $\det[(I + X) - \tau^2 I] = 0$ defines the eigenvalues of $(I + X)$. Then

$$\det[(I + X) - \tau^2 I] = \det(X + I - \tau^2 I) = \det[X - (\tau^2 - 1) I] = 0$$

We observe that $\tau^2 - 1 = t^2$ which implies $\tau^2 = t^2 + 1$. Therefore if $\{t_i^2 + 1\}_{i=1}^p$ are the eigenvalues of $I + X$. □

Proposition 73 *Let $\{t_i^2\}, i = 1, \cdots, p$ be the eigenvalues of real or complex square matrix X of dimension $p \times p$. Then the matrix $(aI_p + X)$ has eigenvalues $\{a + t_i^2\}, i = 1, \cdots, p$. This lemma was motivated by a comment by Arnold (p. 418, bottom) [31].*

Proof. Let $\det(X - t^2 I) = 0$. Consider the matrix $(aI + X)$.

$$\det[(aI + X) - \tau^2 I] = 0$$

defines the eigenvalues of $(aI + X)$. Then

$$\det[(aI + X) - \tau^2 I] = \det(X + aI - \tau^2 I) = \det[X - (\tau^2 - a)I] = 0$$

We observe that $\tau^2 - a = t^2$ which implies $\tau^2 = t^2 + a$. Therefore if $\{t_i^2 + a\}_{i=1}^p$ are the eigenvalues of $aI_p + X$. □

Proposition 74 *Let $\{t_i^2\}, i = 1, \cdots, p$ be the eigenvalues of real or complex square matrix X of dimension $p \times p$. Then the matrix $(aI_p + bX)$ has eigenvalues*

$$\{a + bt_i^2\}, i = 1, \cdots, p$$

This lemma was motivated by a comment by Arnold (p. 418, bottom) [31].

Proof. Let $\det(X - t^2 I) = 0$. Consider the matrix $(aI + bX)$.

$$\det[(aI + bX) - \tau^2 I] = 0$$

defines the eigenvalues of $(aI + bX)$. Then

$$\det[(aI + bX) - \tau^2 I] = \det(bX + aI - \tau^2 I)$$

$$= \det[bX - (\tau^2 - a)I] = b^p \det\left[X - \left(\frac{\tau^2 - a}{b}\right) I_p\right] = 0$$

defines the eigenvalues τ^2 of $(aI + bX)$ in terms of the eigenvalues $t^2 = \frac{\tau^2 - a}{b}$ of X. Thus $\tau_i^2 = a + bt_i^2$ are the eigenvalues of $(aI + bX)$. □

Proposition 75 *Let X have eigenvalue decomposition $\sum_{k=1}^{n} \lambda_k^2 P_k P_k^H$. Then the inverse of the matrix $(aI + bX)$ is given by*

$$X^{-1} = \sum_{k=1}^{n} \left(\frac{1}{a + b\lambda_k^2}\right) P_k P_k^H$$

Proof. By proposition 74, the eigenvalues of $(aI + bX)$ are $\{a + b\lambda_k^2\}_{k=1}^n$. Since

$$aI + bX = \sum_{k=1}^{n} (a + b\lambda_k^2) P_k P_k^H$$

then

$$(aI + bX)^{-1} = \sum_{k=1}^{n} \left(\frac{1}{a + b\lambda_k^2}\right) P_k P_k^H$$

□

Theorem 128 *Let X have eigenvalue decomposition $\sum_{k=1}^{n} \lambda_k^2 P_k P_k^H$. Then*

$$X = I - \sum_{k=1}^{n} (1 - \lambda_k^2) P_k P_k^H$$

Proof.

$$X = I - (I - X) = I - \left(\sum_{k=1}^{n} P_k P_k^H - \sum_{k=1}^{n} \lambda_k^2 P_k P_k^H\right)$$

$$= I - \sum_{k=1}^{n} (1 - \lambda_k^2) P_k P_k^H$$

This result is occasionally disguised where λ_k^2 has the form $\frac{1}{a_k}$. then

$$1 - \frac{1}{a_k} = \frac{a_k - 1}{a_k}$$

Also, for

$$\lambda_k^2 = \frac{1}{1 + b_k}$$

then

$$1 - \lambda_k^2 = \frac{b_k}{1 + b_k} = \frac{1}{\frac{1}{b_k} + 1}$$

is a form that appears in literature. This comes from looking at the inverse of a matrix

$$Y = I + \sum_{k=1}^{n} b_k P_k P_k^H = I + B$$

□

Proposition 76 *Let* $A = A^H$. *Then* $\mathrm{tr}(A^2) = \sum_{i=1}^{p} \lambda_i^4$ *where the* $\{\lambda_i^2\}_{i=1}^{p}$ *are the eigenvalues of* A.

Proof. Since $A = A^H$, A has an eigenvalue decomposition $A = \Gamma \Lambda^2 \Gamma^H$ where Λ^2 is the diagonal matrix of eigenvalues and $\Gamma \in U(p)$ is the matrix of eigenvectors.

$$A^2 = \Gamma \Lambda^2 \Gamma^H \Gamma \Lambda^2 \Gamma^H = \Gamma \Lambda^4 \Gamma^H$$

Therefore

$$\mathrm{tr}(A^2) = \mathrm{tr}\, \Lambda^4 = \sum_{i=1}^{p} \lambda_i^4$$

□

M.11 Singular Value Decomposition

Theorem 129 *(Singular Value Decomposition, SVD). Let* $A \in \mathbf{C}^{m\times n}$ *have* $\text{rank}(A) = r$. *Then* A *can be decomposed as* $A = P\Lambda Q^H$ *where* $P \in \mathbf{U}(m)$, $Q \in \mathbf{U}(n)$, *and* Λ *is an* $m \times n$ *matrix consisting of all zeros except for* r *positive elements* $(\lambda_1, \cdots, \lambda_r)$ *on the main diagonal. Without loss of generality, assume* $n \geq m$. *This is a widely known result.*

Proof. (Gratefully taken from C. R. Rao, pp. 42-43 [213]). Recall that the matrix $B^H = B = AA^H$ has the eigenvalue decomposition

$$B = \sum_{i=1}^{r} \lambda_i^2 P_i P_i^H$$

From C. R. Rao [213], let $Q_i = \lambda_i^{-1} A^H P_i$. Recall that $\{P_i\}$ are orthonormal, which means

$$P_i^H P_j = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

We then observe

$$Q_j^H Q_k = \lambda_j^{-1}\lambda_k^{-1} P_j^H AA^H P_k = \lambda_j^{-1}\lambda_k^{-1} P_j^H \left(\sum_{i=1}^{r} \lambda_i^2 P_i P_i^H\right) P_k$$

$$= \lambda_j^{-1}\lambda_k^{-1} \sum_{i=1}^{r} \lambda_i^2 \delta_{ij}\delta_{ik} = \lambda_j \lambda_k^{-1} \delta_{jk} = \delta_{jk}$$

Therefore the $\{Q_i\}$ are also orthonormal.

With a slight rearrangement of $Q_i^H = \lambda_i^{-1} P_i^H A$, we note that $\lambda_i Q_i^H = P_i^H A$. We also note that given any set of orthonormal vectors, we can complete that

set to form an orthonormal basis for its space. Thus we can find $\{P_i\}_{r+1}^m$ in $\mathbf{C}^m$ so that

$$\sum_{i=1}^{r} P_i P_i^H = I$$

This allows us to state

$$A = IA = (P_1 P_1^H + \cdots + P_m P_m^H)A$$

We substitute the relationship for $\{Q_i\}$, now extended to a set of size m, to obtain

$$A = P_1 P_1^H A + \cdots + P_m P_m^H A = \lambda_1 P_1 Q_1^H + \cdots + \lambda_r P_r Q_r^H$$

where $\lambda_{r+1} = \cdots = \lambda_m = 0$.

We know $\{Q_i\}_1^m$ are orthonormal. When $n > m$, we can extend this set to $\{Q_i\}_1^n$ to form an orthonormal basis for $\mathbf{C}^n$. Thus P and Q are unitary matrices where $P \in \mathbf{U}(m)$ and $Q \in \mathbf{U}(n)$. We can rewrite the expansion of A into

$$A = (P_1, \cdots, P_r, P_{r+1}, \cdots, P_{in}) \begin{bmatrix} \lambda_1 & & & 0 \\ & \ddots & & \vdots \\ & & \lambda_r & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^H \\ \vdots \\ Q_r^H \\ \vdots \\ Q_n^H \end{bmatrix} = P \Lambda Q^H$$

where $P \in \mathbf{C}^{m \times m}$, $\Lambda \in \mathbf{R}^{m \times m}$, $Q \in \mathbf{C}^{n \times n}$. $\square$

Appendix N

TRIGONOMETRY OF COMPLEX MATRICES

This was written just for the fun of it. It was motivated by looking at eigenvalue decompositions while playing with the zonal polynomial questions, and recalling the Cayley-Hamilton theorem that says a matrix satisfies its own characteristic equation. This led to looking at other functions of a matrix. The work presented here has potential application when the CS decomposition is used, such as using the matrices C and S in section 6.3.4 of Tague's thesis [263]. Much of the early part of this chapter is a complexification of material from the fine work by Curtis (pp. 45 ff) [64]. In this case, "Curtis" is the author's family name, not his Christian name.

N.1 Matrix Exponential and Logarithm Properties

Definition 80 *Exponential of a Matrix. Let A be a complex $n \times n$ matrix and define*

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

where $A^m = AA^{m-1}$. This series converges if each of the n^2 complex number series

$$(I)_{ij} + (A)_{ij} + \frac{1}{2!}(A^2)_{ij} + \cdots$$

converges. This defines a mapping

$$\exp : M_n(\mathbf{C}) \to M_n(\mathbf{C}) = GL(n, \mathbf{C})$$

This definition is from Curtis [64].

Proposition 77 *For any complex $n \times n$ matrix A, the series*

$$I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

converges. This is a complexification of Curtis' proposition 1 [64].

Proof. Let m be the largest $|a_{ij}|$ in A. Then the element of largest magnitude in the first term is 1. The element of largest magnitude in the second term is m. The element of largest magnitude in the third term is $\leq \frac{nm^2}{2!}$. The element of largest magnitude in the fourth term is $\leq \frac{n^2m^3}{3!}$, and so on. Any ij sequence is dominated by

$$1, m, \frac{nm^2}{2!}, \frac{n^2m^3}{3!}, \cdots, \frac{n^{k-2}m^{k-1}}{(k-1)!}, \cdots$$

Applying the ratio test to this sequence gives

$$\frac{n^{k-1}m^k}{k!}\frac{(k-1)!}{n^{k-2}m^{k-1}} = \frac{nm}{k}$$

Since m and n are fixed, the ratio goes to zero as $k \to \infty$, proving absolute convergence.□

Proposition 78 *If* 0 *is the zero matrix, then* $e^0 = I$.

Proposition 79 *If* A *is an* $n \times n$ *complex Hermitian matrix* ($A = A^H$), *then* A^n *is also an* $n \times n$ *complex Hermitian matrix.*

Proof. Since A is Hermitian, then $A = B^H B$ by theorem 119. From this we observe

$$A^n = (B^H B)^n = (B^H B)(B^H B)\cdots(B^H B) = (A^n)^H = (A^H)^n$$

□

Proposition 80 *For any* $n \times n$ *complex Hermitian matrix* A, *then* e^A *is also an* $n \times n$ *complex Hermitian matrix.*

Proof.

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

is a linear combination of $n \times n$ complex Hermitian matrices. Therefore e^A is an $n \times n$ complex Hermitian matrix. Therefore e^A is an $n \times n$ complex Hermitian matrix. □

Lemma 55 *If the matrices* A *and* B *commute, then*

$$e^{A+B} = e^A e^B$$

This is Curtis' proposition 2 [64].

Proof.

$$e^{A+B} = I + (A+B) + \frac{1}{2!}(A+B)^2 + \frac{1}{3!}(A+B)^3 + \cdots$$

$$= I + A + B + \frac{1}{2}A^2 + \frac{1}{2}AB + \frac{1}{2}BA + \frac{1}{2}B^2$$

$$+\frac{1}{6}A^3 + \frac{1}{6}A^2B + \frac{1}{6}ABA + \frac{1}{6}AB^2 + \frac{1}{6}BA^2 + \frac{1}{6}BAB + \frac{1}{6}B^2A + \frac{1}{6}B^3 + \cdots$$

Note that $e^{A+B} = e^{B+A}$ even if $AB \neq BA$. When $AB = BA$ then the above simplifies to

$$e^{A+B} = I + A + B + \frac{1}{2}A^2 + AB + \frac{1}{2}B^2$$

$$+\frac{1}{6}A^3 + \frac{1}{2}A^2B + \frac{1}{2}AB^2 + \frac{1}{6}B^3 + \cdots$$

Continuing,

$$e^A e^B = (I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots)(I + B + \frac{1}{2}B^2 + \frac{1}{6}B^3 + \cdots)$$

$$= I + A + B + \frac{1}{2}A^2 + AB + \frac{1}{2}B^2$$

$$+\frac{1}{6}A^3 + \frac{1}{2}A^2B + \frac{1}{2}AB^2 + \frac{1}{6}B^3 + \cdots$$

When $AB = BA$ then

$$e^A e^B = e^B e^A = e^{A+B} = e^{B+A}$$

□

Proposition 81 *e^A is nonsingular. This is a complexification of corollary 1 to Curtis' proposition 2 [64].*

Proof. Let $A \in M_n(\mathbf{C})$. A and $-A$ commute with respect to multiplication. By lemma 55,

$$I = e^0 = e^{A-A} = e^A e^{-A}$$

and thus

$$\det(I) = 1 = (\det e^A)(\det e^{-A})$$

which implies $\det(e^A) \neq 0$. Therefore e^A is nonsingular. □

Proposition 82 *If $A = -A^H$ (A is skew-Hermitian), then e^A is unitary. This is a complexification of Curtis' proposition 3 [64].*

Proof.

$$I = e^0 = e^{A-A} = e^{A+A^H} = e^A e^{A^H} (e^A)(e^{A^*})^T = (e^A)(e^A)^H$$

Thus e^A is unitary when A is skew-Hermitian. □

Theorem 130 *If A, B are $n \times n$ complex matrices and B is nonsingular, then*

$$e^{BAB^{-1}} = Be^A B^{-1}$$

and

$$\det e^{BAB^{-1}} = \det e^A$$

This is Curtis' proposition 4 [64].

Proof. This proof is by Curtis [64].

$$(BAB^{-1})^n = (BAB^{-1})(BAB^{-1}) \cdots (BAB^{-1}) = BA^n B^{-1}$$

$$e^{BAB^{-1}} = I + BAB^{-1} + \frac{1}{2!}(BAB^{-1})^2 + \frac{1}{3!}(BAB^{-1})^3 + \cdots$$

$$= I + BAB^{-1} + \frac{1}{2!}BA^2B^{-1} + \frac{1}{3!}BA^3B^{-1} + \cdots$$

$$= B(I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots)B^{-1} = Be^AB^{-1}$$

$$\det e^{BAB^{-1}} = \det(Be^AB^{-1}) = (\det B)(\det e^A)(\det B^{-1}) = \det e^A$$

□

Corollary 38 *If A, B are $n \times n$ complex matrices and B is unitary, then*

$$\det e^{BAB^H} = \det e^A$$

and

$$e^{BAB^H} = Be^AB^H$$

Proof. Substitute $B^H = B^{-1}$ into the proof of theorem 130. □

Definition 81 *Let $X \in M_n(\mathbf{C})$. Define*

$$\log(X) = (X - I) - \frac{1}{2}(X - I)^2 + \frac{1}{3}(X - I)^3 - \frac{1}{4}(X - I)^4 + \cdots$$

This is a complexification of a definition given by Curtis, p. 49 [64].

Proposition 83 *Let $X \in M_n(\mathbf{C})$. Then $\log(X)$ converges when the magnitude of the largest element of $X - I$ is less than $\frac{1}{n}$. Note: X can not be the zero matrix. This is a complexification of proposition 5 of Curtis [64].*

Proof. This is the proof by Curtis [64], except that I now am using the magnitude of complex quantities rather than just the absolute value. Let M be the magnitude of the largest element of $Y = X - I$. Then

$$|(Y)_{ij}| \leq M, \quad \left|\frac{1}{2}(Y^2)_{ij}\right| \leq \frac{nM^2}{2},$$

$$\left|\frac{1}{3}(Y^3)_{ij}\right| \leq \frac{n^2M^3}{3}, \quad \cdots \quad , \left|\frac{1}{k}(Y^k)_{ij}\right| \leq \frac{n^{k-1}M^k}{k}$$

By the ratio test of d'Alembert,

$$\lim_{k\to\infty}\left|\left(\frac{n^kM^{k+1}}{k+1}\right)\left(\frac{k}{n^{k-1}M^k}\right)\right| = \lim_{k\to\infty}\frac{k}{k+1}nM = nM$$

The series converges when $nM < 1$, or equivalently when $M < \frac{1}{n}$.□

Theorem 131 *In $M_n(\mathbf{C})$, let U be a neighborhood of I in which* log *is defined, and let V be a neighborhood of zero such that* $\exp(V)$ *is contained in U. Then (i) for $A \in V$, $\log e^A = A$, and (ii) for $X \in U$, $e^{\log X} = X$. This is a complexification of proposition 6 of Curtis [64].*

Proof. This proof is by Curtis [64], except that the matrices are now complex valued rather than real-valued.

(i) $A \in V$ implies $e^A \in U$ by hypothesis. This implies that $\log e^A$ exists. So,

$$\log e^A = (A + \frac{1}{2}A^2 + \cdots) - \frac{1}{2}(A + \frac{1}{2!}A^2 + \cdots)^2$$

$$+\frac{1}{3}(A + \frac{1}{2!}A^2 + \cdots)^3 - \cdots$$

$$= A + \left[\frac{1}{2!}A^2 - \frac{1}{2}A^2\right] + \left[\frac{1}{3!}A^3 - \frac{1}{2}A^3 + \frac{1}{3}A^3\right]$$

$$+\left[\frac{1}{4!}A^4-\frac{1}{2}\left\{\left(\frac{1}{2!}A^2\right)^2+2A\frac{1}{3!}A^3\right\}\frac{1}{3}(0)+\frac{1}{4}A^4\right]+\cdots$$

$$=A+0+0+0+\cdots=A$$

(ii)

$$\log(X)=(X-I)-\frac{1}{2}(X-I)^2+\frac{1}{3}(X-I)^3-\cdots$$

which implies

$$e^{\log X}=\left\{I+(X-I)-\frac{1}{2}(X-I)^2+\cdots\right\}+\frac{1}{2!}\left\{(X-I)-\frac{1}{2}(X-I)^2+\cdots\right\}^2$$

$$+\frac{1}{3!}\left\{(X-I)-\frac{1}{2}(X-I)^2+\cdots\right\}^3+\cdots$$

$$=X-\frac{1}{2}(X-I)^2+\frac{1}{2}(X-I)^2+\left\{\frac{1}{3}(X-I)^3-\frac{1}{2}(X-I)^3+\frac{1}{6}(X-I)^3\right\}+\cdots$$

$$=X+0+0+\cdots=X$$

□

Corollary 39 *Let U be a neighborhood of I in $M_n(\mathbf{C})$ in which* log *is defined. Let $X,Y\in U$. Let* $\log X$ *and* $\log Y$ *commute. Then*

$$\log(XY)=\log X+\log Y$$

This is a complexification of part 1 of Curtis' proposition 7 [64].

Proof. This is essentially the proof by Curtis [64] where the matrices are now understood to be complex. $e^{\log(XY)}=XY$ by theorem 131. $XY=e^{\log X}e^{\log Y}$, also by theorem 131. Since $\log X$ and $\log Y$ commute,

$$e^{\log X}e^{\log Y}=e^{\log X+\log Y}$$

by lemma 55. □

Corollary 40 *Let unitary X be in a neighborhood of I in $M_n(\mathbf{C})$ in which* log *is defined. Then* $\log X$ *is skew-Hermitian. This is a complexification of part 2 of Curtis' proposition 7 [64].*

Proof. This is a complexification of the proof by Curtis [64]. Since X is unitary, $X^H X = X X^H = I$. Thus X and X^H commute, which implies $\log X$ and $\log X^H$ commute. Then

$$0 = \log(I) = \log(XX^H) = \log X + \log X^H = \log X + (\log X)^H$$

This implies $\log X = -(\log X)^H$, showing that $\log X$ is skew-Hermitian.

Remark. This remark is supplied by me. The matrix functions exp and log are not simple generalizations of the univariate case. For example,

$$\frac{d}{dY} \log Y \neq Y^{-1}$$

even when Y^{-1} exists. This is easy to see since Y^{-1} is an $n \times n$ matrix, while $\frac{d}{dY} \log Y$ is an $n^2 \times n^2$ matrix. Let us see what $\frac{d}{dY} \log Y$ is.

$$\exp(\log Y) = Y \subset \text{Neighborhood}(I)$$

$$\log(\exp X) = X \subset \text{Neighborhood}(0)$$

$$\frac{d}{dX} \log(\exp X) = \frac{dX}{dX} = E_n E_n^T \neq I_{n^2}.$$

Treating $\log(\exp X)$ as a composition of functions $\log \circ \exp(X)$, we get

$$\frac{d}{dX}(\log(\exp X)) = [\frac{d}{dY} \log Y][\frac{d}{dX} \exp X] = E_n E_n^T$$

Thus

$$\frac{d}{dY}\log Y = E_n E_n^T [\frac{d}{dX}\exp X]^{-1} \tag{N.1}$$

when this inverse exists. □

N.2 Matrix COS and SIN Functions

Work in this section is supplied by me.

Definition 82 *Define the matrix cosine function as*

$$C(X) = \frac{1}{2}[\exp(iX) + \exp(-iX)] \tag{N.2}$$

and the matrix sine function as

$$S(X) = \frac{1}{2i}[\exp(iX) - \exp(-iX)] \tag{N.3}$$

Thus

$$\exp(iX) = C(X) + iS(X) \tag{N.4}$$

Note that $C(0) = I$ *and* $S(0) = 0$. *Unlike the univariate case,*

$$\frac{d}{dX}C(X) \neq -S(X) \tag{N.5}$$

and

$$\frac{d}{dX}S(X) \neq C(X) \tag{N.6}$$

The derivative matrices are $n^2 \times n^2$, *whereas* $C(X)$ *and* $S(X)$ *are* $n \times n$ *matrices.*

Proposition 84 *(a)*

$$S(A) = A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 - \frac{1}{7!}A^7 + \cdots$$

and (b)

$$C(A) = I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \cdots$$

Proof. For part (a):

$$S(A) = \frac{1}{2i}\left(e^{iA} - e^{-iA}\right)$$

$$= \frac{1}{2i}\sum_{k=0}^{\infty}\frac{1}{k!}[(iA)^k - (-iA)^k] = \frac{1}{2i}\sum_{k=0}^{\infty}\frac{1}{k!}[1-(-1)^k]\,i^kA^k$$

Note that

$$1-(-1)^k = \begin{cases} 0, \text{ when k s even} \\ 2, \text{ when k is odd} \end{cases}$$

Then

$$S(A) = \frac{1}{i}\left(iA - \frac{i}{3!}A^3 + \frac{i}{5!}A^5 + \cdots\right) = A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 + \cdots$$

$$= \sum_{k=1}^{\infty}\frac{1}{(2k-1)!}(-1)^{k+1}A^{2k-1}$$

For part (b):

$$C(A) = \frac{1}{2}(e^{iA} + e^{-iA}) = \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{k!}[(iA)^k + (-iA)^k]$$

$$= \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{k!}\underbrace{[1+(-1)^k]}_{\begin{matrix}0,\text{ k odd}\\ 2,\text{ k even}\end{matrix}}(iA)^k$$

$$= \frac{1}{0!}(iA)^0 + \frac{1}{2!}(iA)^2 + \frac{1}{4!}(iA)^4 + \frac{1}{6!}(iA)^6 + \cdots$$

$$= I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 - \frac{1}{6!}A^6 + \cdots$$

$$= \sum_{k=0}^{\infty} \frac{1}{(2k)!}(-1)^k A^{2k}$$

□

Proposition 85 *Let $X \in M_n(\mathbf{C})$ be an $n \times n$ complex matrix, and let $C(X)$ and $S(X)$ be defined as in definition 82. Then*

$$C^2(X) + S^2(X) = I_n$$

Proof. Note that X and $-X$ commute under matrix multiplication, which allows us to use lemma 55.

$$C^2(X) + S^2(X) = \frac{1}{2}\left[\exp(iX) + \exp(-iX)\right]\frac{1}{2}\left[\exp(iX) + \exp(-iX)\right]$$

$$+\frac{1}{2i}\left[\exp(iX) - \exp(-iX)\right]\frac{1}{2i}\left[\exp(iX) - \exp(-iX)\right]$$

$$= \frac{1}{4}\left[\exp(i2X) + 2\exp(iX - iX) + \exp(-i2x)\right]$$

$$-\frac{1}{4}\left[\exp(i2X) - 2\exp(iX - iX) + \exp(-i2x)\right]$$

$$= \exp(i0) = I_n$$

□

Lemma 56 *Let $D = \text{diag}(d_1, d_2, \cdots, d_n)$ be an $n \times n$ diagonal complex matrix. Then*

$$e^D = \begin{pmatrix} \exp(d_1) & & \\ & \ddots & \\ & & \exp(d_n) \end{pmatrix}$$

Proof.

$$e^D = I + D + \frac{1}{2!}D^2 + \frac{1}{3!}D^3 + \cdots$$

$$= \text{diag}\left\{\left(1 + d_i + \frac{1}{2!}d_i^2 + \frac{1}{3!}d_i^3 + \cdots\right)_i \quad \text{where } i = 1, \cdots, n\right\}$$

$$= \begin{pmatrix} e^{d_1} & & \\ & \ddots & \\ & & e^{d_n} \end{pmatrix}$$

□

Theorem 132 *Let a be a complex scalar. Then $e^{aI} = e^a I$.*

Proof. This follows directly from lemma 56. □

Theorem 133 *Let a be a complex scalar and $D = \text{diag}(d_1, \cdots, d_n)$ be an $n \times n$ diagonal complex matrix. Then*

$$e^{aD} = \begin{pmatrix} e^{d_1} & & \\ & \ddots & \\ & & e^{d_n} \end{pmatrix}^a$$

Proof. This follows directly from lemma 56. □

Theorem 134 *Let a be a complex scalar. Then*

$$C(aI) = [\cos(a)]I$$

and

$$S(aI) = [\sin(a)]I$$

where cos *and* sin *(with lower case c and s) are the usual scalar trigonometric functions.*

Proof.

$$C(aI) = \frac{1}{2}(e^{iaI} + e^{-iaI}) = \frac{1}{2}(e^{ia}I + e^{-ia}I)$$

$$= \frac{1}{2}(e^{ia} + e^{-ia})I = [\cos(a)]I$$

Similarly,

$$S(aI) = [\sin(a)]I$$

□

Corollary 41 $C(\frac{\pi}{2}I) = 0$ *and* $S(\frac{\pi}{2}I) = I$.

Proof. This follows directly from theorem 134. □

Theorem 135 *Let A and B commute under multiplication and $A, B \in M_n(\mathbf{C})$. Then*

$$S(A + B) = S(A)C(B) + C(A)S(B)$$

$$S(A - B) = S(A)C(B) - C(A)S(B)$$

$$C(A+B) = C(A)C(B) - S(A)S(B)$$

$$C(A-B) = C(A)C(B) + S(A)S(B)$$

Proof:

$$S(A)C(B) + C(A)S(B)$$

$$= \frac{1}{2i}\left(e^{iA} - e^{-iA}\right)\frac{1}{2}\left(e^{iB} + e^{-iB}\right) + \frac{1}{2}\left(e^{iA} + e^{-iA}\right)\frac{1}{2i}\left(e^{iB} - e^{-iB}\right)$$

$$= \frac{1}{4i}\left(e^{iA}e^{iB} + e^{iA}e^{-iB} - e^{-iA}e^{iB} - e^{-iA}e^{-iB}\right)$$

$$+\frac{1}{4i}\left(e^{iA}e^{iB} - e^{iA}e^{-iB} + e^{-iA}e^{iB} - e^{-iA}e^{-iB}\right)$$

$$= \frac{1}{4i}\left(2e^{iA}e^{iB} - 2e^{-iA}e^{-iB}\right) = \frac{1}{2i}\left(e^{iA}e^{iB} - e^{-iA}e^{-iB}\right)$$

Invoking lemma 55, we get

$$\frac{1}{2i}\left(e^{i(A+B)} - e^{-i(A+B)}\right) = S(A+B)$$

The other identities are proven in a similar fashion. □

Theorem 136 *If A and B commute, then $S(A)$ and $C(B)$ commute. That is, if $AB = BA$, then*

$$S(A)C(B) = C(B)S(A)$$

Proof. From theorem 135,

$$S(A)C(B) = \frac{1}{4i}(e^{iA}e^{iB} + e^{iA}e^{-iB} - e^{-iA}e^{iB} - e^{-iA}e^{-iB})$$

Now invoke $e^{iA}e^{iB} = e^{iB}e^{iA}$ from lemma 55. This gives us

$$\frac{1}{4i}\left(e^{iB}e^{iA} + e^{-iB}e^{iA} - e^{iB}e^{-iA} - e^{-iB}e^{-iA}\right)$$

$$= \frac{1}{2}\left(e^{iB} + e^{-iB}\right)\frac{1}{2i}\left(e^{iA} - e^{-iA}\right) = C(B)S(A)$$

□

Corollary 42 *$S(A)$ and $C(A)$ commute.*

Proof. A commutes with itself. Let $B = A$ in theorem 136. □

Theorem 137

(a) $S(2A) = 2S(A)C(A)$

(b) $C(2A) = [C(A)]^2 - [S(A)]^2 = I - 2[S(A)]^2 = 2[C(A)]^2 - I$

(c) $S(3A) = 3S(A) - 4[S(A)]^3$

(d) $C(3A) = 4[C(A)]^3 - 3C(A)$

(e) $[S(A)]^2 = \frac{1}{2}[I - C(2A)]$

(f) $[C(A)]^2 = \frac{1}{2}[I + C(2A)]$

Proof. The proof of these is strictly mundane, made possible by theorem 136. Note that A commutes with itself. For example, look at (c).

$$S(3A) = S(A + 2A) = S(A)C(2A) + C(A)S(2A)$$

$$= S(A)[C^2(A) - S^2(A)] + C(A)[S(A)C(A) + C(A)S(A)]$$

$$= S(A)C^2(A) - S^3(A) + S(A)C^2(A) + S(A)C^2(A)$$

$$= 3S(A)C^2(A) - S^3(A) = 3S(A)[I - S^2(A)] - S^3(A)$$

$$= 3S(A) - 3S^3(A) - S^3(A) = 3S(A) - 4S^3(A)$$

Proof of other parts follows similar mechanics. □

Theorem 138 *If A and B commute, then*

$$(a)\quad S(A)+S(B)=2S[\tfrac{1}{2}(A+B)]C[\tfrac{1}{2}(A-B)]$$

$$(b)\quad S(A)-S(B)=2C[\tfrac{1}{2}(A+B)]S[\tfrac{1}{2}(A-B)]$$

$$(c)\quad C(A)+C(B)=2C[\tfrac{1}{2}(A+B)]C[\tfrac{1}{2}(A-B)]$$

$$(d)\quad C(A)-C(B)=2S[\tfrac{1}{2}(A+B)]S[\tfrac{1}{2}(B-A)]$$

$$(e)\quad S(A)C(B)=\tfrac{1}{2}[S(A+B)+S(A-B)]$$

$$(f)\quad C(A)S(B)=\tfrac{1}{2}[S(A+B)-S(A-B)]$$

$$(g)\quad C(A)C(B)=\tfrac{1}{2}[C(A+B)+C(A-B)]$$

$$(h)\quad S(A)S(B)=\tfrac{1}{2}[C(A-B)-C(A+B)]$$

Proof. The proof is merely tedious algebra. Commutativity is required to invoke lemma 55.

(a) $2S[\frac{1}{2}(A+B)]C[\frac{1}{2}(A-B)]$

$= 2[S(\frac{1}{2}A)C(\frac{1}{2}B) + C(\frac{1}{2}A)S(\frac{1}{2}B)]$

$\times[C(\frac{1}{2}A)C(\frac{1}{2}B) + S(\frac{1}{2}A)S(\frac{1}{2}B)]$

$= 2[S(\frac{1}{2}A)C(\frac{1}{2}A)C^2(\frac{1}{2}B) + S^2(\frac{1}{2}A)S(\frac{1}{2}B)C(\frac{1}{2}B)$

$+C^2(\frac{1}{2}A)S(\frac{1}{2}B)C(\frac{1}{2}B) + S(\frac{1}{2}A)C(\frac{1}{2}A)S^2(\frac{1}{2}B)]$

$= 2[(S^2(\frac{1}{2}A) + C^2(\frac{1}{2}A))S(\frac{1}{2}B)C(\frac{1}{2}B)$

$+(S^2(\frac{1}{2}B) + C^2(\frac{1}{2}B))S(\frac{1}{2}A)C(\frac{1}{2}A)]$

$= 2[S(\frac{1}{2}B)C(\frac{1}{2}B) + S(\frac{1}{2}A)C(\frac{1}{2}A)]$

$= 2S(\frac{1}{2}A)C(\frac{1}{2}A) + 2S(\frac{1}{2}B)C(\frac{1}{2}B)$

$= S(A) + S(B)$

(b)-(d) Proof is similar to (a).

(e) $\frac{1}{2}[S(A+B) + S(A-B)]$

$\frac{1}{2}[S(A)C(B) + C(A)S(B) + S(A)C(B) - C(A)S(B)]$

$= \frac{1}{2}[2S(A)C(B)] = S(A)C(B)$

(f)-(h) Proof is similar to (e).

□

Theorem 139

$$[C(A) + iS(A)]^n = C(nA) + iS(nA)$$

Proof.

$$[C(A) + iS(A)]^n = [\frac{1}{2}e^{iA} + \frac{1}{2}e^{-iA} + \frac{1}{2}e^{iA} - \frac{1}{2}e^{-iA}]^n$$

$$= [e^{iA}]^n = e^{inA} = C(nA) + iS(nA)$$

□

Theorem 140 *Let $X, B \in M_n(\mathbf{C})$ and let B be nonsingular. Then*

$$BC(X)B^{-1} = C(BXB^{-1})$$

and

$$BS(X)B^{-1} = S(BXB^{-1})$$

Proof. Invoke theorem 130. To prove the first equality, we begin

$$\begin{aligned} BC(X)B^{-1} &= \tfrac{1}{2}[B\exp(iX)B^{-1} + B\exp(-iX)B^{-1}] \\ &= \tfrac{1}{2}[\exp(iBXB^{-1}) + \exp(-iBXB^{-1})] \\ &= c(BXB^{-1}) \end{aligned}$$

Similarly, the second equality is shown

$$\begin{aligned} BS(X)B^{-1} &= \tfrac{1}{2i}[B\exp(iX)B^{-1} - B\exp(-iX)B^{-1}] \\ &= \tfrac{1}{2i}[\exp(iBXB^{-1}) - \exp(-iBXB^{-1})] \\ &= S(BXB^{-1}) \end{aligned}$$

□

Corollary 43 *Let $X, B \in M_n(\mathbf{C})$ and let B be unitary. Then*

$$BC(X)B^H = C(BXB^H)$$

and

$$BS(X)B^H = S(BXB^H)$$

Proof. Since B is unitary, then $B^{-1} = B^H$. With this substitution, the proof follows that of theorem 140. □

N.3 Relating Trace and Determinant

This is a particularly nice result because the trace and determinant operators are functions of only the unordered eigenvalues of A.

Theorem 141 *Let $A \in M_n(\mathbf{C})$. Then*

$$\exp[\mathrm{tr}(A)] = \det[\exp(A)]$$

This is taken from Curtis (p. 55) [64].

Proof. Let A be diagonalized by the similarity transformation $D = BAB^{-1}$ where D is a diagonal matrix. Then

$$e^{BAB^{-1}} = Be^AB^{-1}$$

by theorem 130. We note

$$\det e^A = (\det B)(\det e^A)(\det B^{-1}) = \det(Be^AB^{-1})$$

$$= \det(e^{BAB^{-1}}) = \det e^D$$

Let

$$D = \begin{pmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_n \end{pmatrix}$$

and

$$d_i = \begin{pmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \delta_i & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

Thus $D = \sum_{i=1}^{n} d_i$ and

$$\det e^D = \prod_{i=1}^{n} \left(\det e^{d_i}\right)$$

since the d_i commute. Using the definition for matrix exponential,

$$e^{d_i} = I + d_i + \frac{1}{2!}d_i^2 + \cdots = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & e^{\delta_i} & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

which implies $\det e^{d_i} = e^{\delta_i}$. So,

$$\det e^D = \prod_{i=1}^{n} e^{\delta_i} = \exp\left(\sum_{i=1}^{n} \delta_i\right) = e^{\mathrm{tr}(D)} = e^{\mathrm{tr}(BAB^{-1})}$$

$$= \exp[\mathrm{tr}(AB^{-1}B)] = \exp(\mathrm{tr}\, A)$$

Therefore

$$\exp[\mathrm{tr}\, A] = \det[\exp A]$$

when A can be diagonalized by a similarity transformation. □

Appendix O

USEFUL IDENTITIES

Identities which have been useful in the development of this work are recorded here. Most of these are common identities recorded here for convenience's sake. There are, however, some nontrivial ones near the end of this short section. Lack of citing a reference on the simpler identities merely indicates they are very easy ones which I did myself and did not think important enough to find out who else has done them. I do not claim these as new contributions.

O.1 Sums

Proposition 86

$$\sum_{k=b}^{p} k = \frac{(p+b)(p-b+1)}{2}$$

Proof. This is a generalization of

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$$\sum_{k=b}^{p} k = (p-b+1)(b-1) + \sum_{k=1}^{p-b+1} k$$

$$= (p-b+1)(b-1) + \frac{1}{2}(p-b+1)(p-b+2)$$

$$= (p-b+1)[b-1+\frac{1}{2}(p-b+2)]$$

$$= \frac{1}{2}(p-b+1)(2b-2+p-b+2) = \frac{1}{2}(p+b)(p-b+1)$$

□

Proposition 87

$$\sum_{k=a}^{b} k = \sum_{i=1}^{b} i - \sum_{j=1}^{a-1} j = \frac{1}{2}[b(b+1) - (a-1)a]$$

Useful special cases:

$$\sum_{i=1}^{p-1} 2i = (p-1)p$$
$$\sum_{i=2}^{p} 2i = p(p+1) - 2$$
$$\sum_{i=2}^{p-1} 2i = (p-1)p - 2$$

□

O.2 Combinatorics

Proposition 88

$$(2m-1)(2m-3)(2m-5)\cdots 3 \cdot 1 = \frac{(2m)!}{2^m m!}$$

For EVEN m:
$$\begin{cases} (2m-1)(2m-3)(2m-5)\cdots(m+1) \\ = \frac{(2m)!\left(\frac{m}{2}\right)!}{2^{m/2}(m!)^2} = \frac{\left(\frac{m}{2}\right)!}{2^{m/2}}\binom{2m}{m} \end{cases}$$

For ODD m:
$$\begin{cases} (2m-1)(2m-3)(2m-5)\cdots m = \frac{(2m)!\left(\frac{m-1}{2}\right)!}{2^{(m+1)/2}m!(m-1)!} \\ = \frac{\left(\frac{m+1}{2}\right)!}{2^{(m-1)/2}}\binom{2m}{m-1} \end{cases}$$

□

O.3 Classical Distribution Properties

Theorem 142 *Chi-Square Distribution.*

The χ^2 distribution shows up in the evaluation of some important special cases in the properties of the Wishart distribution, both in the complex and real variables cases. The χ^2 distribution in this thesis refers to the usual real-variables case of random variables for χ^2. Many texts discuss the gamma distribution, and then point out that the χ^2 is merely a special case. Although true, it is an important enough special case to have a life of its own.

There are some properties we need in this work, and they are tabulated here. These are copied from Canavos (p. 149) [50]. They can be found in many texts. Let $x \sim \chi_n^2(0)$. Then the following results are true.

$$\mathcal{E}\{x\} = n$$

$$\text{var}(x) = 2n$$

$$\text{skewness}(x) = \frac{4}{\sqrt{2n}}$$

$$\text{kurtosis}(x) = 3\left(1 + \frac{4}{n}\right)$$

$$\text{mgf } m_x(t) = (1-2t)^{-n/2} \text{ for } 0 \leq t < \tfrac{1}{2}$$

$$\mathcal{E}\{x^m\} = 2^m \Gamma\left(\frac{n}{2} + m\right) / \Gamma\left(\frac{n}{2}\right)$$

$$\mathcal{E}\{x^2\} = 4\Gamma\left(\frac{n+4}{2}\right) / \Gamma\left(\frac{n}{2}\right)$$

$$\mathcal{E}\{x^2\} = 4k(k+1) \text{ for } n = 2k$$

$$\mathcal{E}\{y^2\} = k(k+1) \text{ for } x = 2y \text{ and } n = 2k$$

$$\mathcal{E}\left\{\frac{1}{x}\right\} = \frac{1}{n-2}, \quad n > 2$$

$$\mathcal{E}\left\{\frac{1}{x^2}\right\} = \frac{1}{(n-2)(n-4)}$$

$$\text{var}\left\{\frac{1}{x}\right\} = \frac{2}{(n-2)^2(n-4)}$$

□

O.4 Functions of a Hermitian Positive Definite Matrix

Lemma 57 *Let U be Hermitian positive definite. Let*

$$h(U) = \frac{1}{[\det U]^n} \exp[-\operatorname{tr} U^{-1}]$$

Then $h(U)$ is maximized when $U = \frac{1}{n} I$. This is a complexification of Arnold's lemma A.14 [31].

Proof. This is a complexification and expansion of Arnold's proof. Let U be a $p \times p$ Hermitian positive definite matrix. By theorem 118, U^{-1} is also Hermitian positive definite. By theorem 115, let U^{-1} have eigenvalues

$$\Lambda^2 = \text{diag}(t_1^2, \cdots, t_p^2)$$

with eigenvector matrix Γ. By Stewart corollary 6.5.3 [259] we know $t_i^2 > 0$ for all $i \in [1, p]$. Then by theorem 114 and theorem 113 we have

$$h(U) = \left(\prod_{i=1}^{p} t_i^2\right)^n \exp\left[-\sum_{i=1}^{p} t_i^2\right] \stackrel{\text{def}}{=} g(t_1^2, \cdots, t_p^2)$$

Now, find the $(t_1^2, \cdots, t_p^2)$ that maximizes $g(t_1^2, \cdots, t_p^2)$ over the set where $t_i^2 > 0$, and then find the matrix associated with those eigenvalues.

$$\frac{\partial}{\partial t_i^2} g(t_1^2, \cdots, t_p^2) = \frac{\partial}{\partial t_i^2} \prod_{i=1}^{p} t_i^{2n} e^{-t_j^2} = \left(n t_i^{2(n-1)} e^{-t_i^2} - t_i^{2n} e^{-t_i^2}\right) c$$

where

$$c = \prod_{\substack{j=1 \\ j \neq i}}^{p} t_j^{2n} e^{-t_j^2} > 0$$

because $t_i^2 > 0$ for all i. Continuing,

$$\left(n t_i^{2(n-1)} e^{-t_i^2} - t_i^{2n} e^{-t_i^2}\right) c = \left(n t_i^{-2} - 1\right) t_i^{2n} e^{-t_i^2} c$$

From this we see that

$$\frac{\partial}{\partial t_i^2} g(t_1^2, \cdots, t_p^2) = 0$$

if and only if $n t_i^{-2} - 1 = 0$, which implies $t_i^2 = n$. Then

$$\frac{\partial^2}{\partial (t_i^2)^2} g(t_1^2, \cdots, t_p^2) = \frac{\partial}{\partial t_i^2} (n t_i^{-2} - 1) t_i^{2n} e^{-t_i^2} c$$

$$= \left[\left(-n t_i^{-4}\right)\left(t_i^{2n} e^{-t_i^2}\right) + \left(n t_i^{-2} - 1\right)\left(n t_i^{2(n-1)} e^{-t_i^2} - t_i^{2n} e^{-t_i^2}\right)\right] c$$

$$= \left[(n^2 - n) t_i^{-4} - 2n t_i^{-2} + 1\right] t_i^{2n} e^{-t_i^2} c$$

Evaluating the second partial derivative at $t_i^2 = n$ gives us

$$\left[1 - \frac{1}{n} - 2 + 1\right] n^n e^{-n} c = -n^{n-1} e^{-n} c < 0$$

Thus $t_i^2 = n$ gives us a maximum. Therefore

$$U^{-1} = \Gamma(nI)\Gamma^H = nI$$

which implies $U = \frac{1}{n}I$. □

Theorem 143 *Let U and A be Hermitian positive definite $p \times p$ complex matrices. Define*

$$f(U) = [\det U]^{-n} \exp[-\operatorname{tr}(U^{-1}A)]$$

Then $f(U)$ is maximized when $U = \frac{1}{n}A$. This is a complexification of theorem A.15 of Arnold [31].

Proof. This is a complexification of Arnold's proof. By theorem 120, there exists $A^{1/2}$ such that $A = A^{1/2}A^{1/2}$. Then

$$f(U) = [\det U]^{-n} \exp[-\operatorname{tr}(U^{-1}A)] = [\det U]^{-n} \exp[-\operatorname{tr}(U^{-1}A^{1/2}A^{1/2})]$$

$$= [\det U]^{-n} \exp[-\operatorname{tr}(A^{\frac{H}{2}}U^{-1}A^{\frac{1}{2}})]$$

$$[\det A]^{-n}[\det(A^{-1/2}UA^{-1/2})]^{-n} \exp[-\operatorname{tr}(\{A^{-1/2}UA^{-1/2}\}^{-1})]$$

$$= [\det A]^{-n} h(A^{-1/2}UA^{-1/2})$$

where h is defined in lemma 57. Thus $f(U)$ is maximized when

$$A^{-1/2}UA^{-1/2} = \frac{1}{n}I$$

or equivalently, when $U = \frac{1}{n}A$. □

O.5 Properties of Unitarily Invariant Functions

Lemma 58 *Suppose that C is a fixed $p \times p$ complex matrix and Z is a $p \times p$ complex random matrix. If a function*

$$g(C) = \mathcal{E}\{\text{etr}(CZ)\}$$

satisfies

$$g(C) = \mathcal{E}\{\text{etr}(CZ)\} = \mathcal{E}\{\text{etr}(CU^H ZU)\} = g(UCU^H)$$

for all $p \times p$ unitary matrices U, then

$$\mathcal{E}\{Z_{ji} Z_{lk}\} = g_{ij,kl} \stackrel{\text{def}}{=} \frac{\partial^2}{\partial C_{ij} \partial C_{kl}} g(C)\Big|_{C=0} = b_1 \delta_{ij}\delta_{kl} + b_2 \delta_{il}\delta_{jk}$$

This is Tague's complexification [264] of Olkin and Rubin lemma 1 [199]. This lemma is used in the development of theory resulting in a beamforming example by Tague for computing the signal-to-noise ratio.

Proof. Let $U = [U_1, U_2, \cdots, U_p]$ be a unitary matrix. Then

$$\text{tr}(CU^H ZU) = \sum_{i=1}^{p} \sum_{j=1}^{p} C_{ij} U_j^H Z U_i$$

Note that the order of subscripts of C_{ij} are the opposite of $U_j^H Z U_i$. Expanding the assumed functional form and taking derivatives, we obtain

$$g_{ij,kl} = \frac{\partial^2}{\partial C_{ij} \partial C_{kl}} \int_{Z>0} \text{etr}(CU^H ZU) f(Z)(dZ)\Big|_{C=0}$$

$$= \int_{Z>0} U_j^H Z U_i U_l^H Z U_k \operatorname{etr}(CU^H ZU) f(Z)(dZ)\Big|_{C=0}$$

This shows that

$$g_{ij,kl} = \mathcal{E}\left\{U_j^H Z U_i U_l^H Z U_k\right\}$$

for all unitary matrices U.

Let $U = I$. Then

$$\mathcal{E}\left\{U_j^H Z U_j U_j^H Z U_j\right\} = \mathcal{E}\left\{e_j^H Z e_j e_j^H Z e_j\right\} = \mathcal{E}\left\{Z_{jj} Z_{jj}\right\} = \mathcal{E}\left\{Z_{jj}^2\right\} = g_{jj,jj}$$

By hypothesis, g is invariant for all unitary U. Therefore, g is unchanged if we let $U_i = e_j$ and $U_j = e_i$. Then

$$g_{ii,ii} = \mathcal{E}\left\{(Z_{ii})^2\right\} = \mathcal{E}\left\{U_i^H Z U_i U_i^H Z U_i\right\}$$

$$= \mathcal{E}\left\{e_j^H Z e_j e_j^H Z e_j\right\} = \mathcal{E}\left\{(Z_{jj})^2\right\}$$

where $1 \leq i, j \leq p$. If a column of unitary U is multiplied by a complex number with unit magnitude, then U remains unitary. If we exchange columns of unitary U, the new matrix is unitary. By picking special cases of U, we can show that most second order moments are zero.

Let $U = I$. Then

$$g_{ij,ij} = \mathcal{E}\left\{(Z_{ji})^2\right\}$$

Now, let $U_i = e_i$ and $U_j = e_j\sqrt{-1}$. Then

$$g_{ij,ij} = \mathcal{E}\left\{(Z_{ji})^2\right\} = \mathcal{E}\left\{(-e_j\sqrt{-1})^T Z e_i (-e_j\sqrt{-1})^T Z e_i\right\} = -\mathcal{E}\left\{(Z_{ji})^2\right\} = 0$$

Similarly,

$$g_{ii,ij} = g_{ii,jk} = g_{ij,ik} = g_{ij,kl} = 0$$

The only nonzero terms are $g_{ij,ji}, g_{ii,jj}$, and $g_{ii,ii}$, $1 \leq i, j \leq p$.

We relate these nonzero terms by applying Olkin and Rubin's trick to evaluate

$$g_{ii,ij} = 0 = \mathcal{E}\left\{U_i^H Z U_i U_j^H Z U_i\right\} = \sum_{\alpha,\beta,\gamma,\delta} U_{\alpha i} U_{\beta j}^* U_{\gamma k} U_{\delta l}^* g_{\alpha\beta,\gamma\delta}$$

where $U_{\beta j}^*$ is the complex conjugate of the complex scalar $U_{\beta j}$. Let $U = \exp(\epsilon F)$ where F is skew-Hermitian (which has purely imaginary diagonal terms). When $0 < \epsilon \ll 1$ then $U \approx I + \epsilon F$. Ignoring higher order terms, the main diagonal elements of U have the form $U_{ii} = 1 + ja_i$ for $a_i \in \mathbf{R}$. Also, $U_{ij} = f_{ij}$ and $U_{ji} = -f_{ij}^*$. The equation becomes

$$\epsilon f_{ij}^* g_{ii,ii} - \epsilon f_{ij}^* g_{ii,jj} - \epsilon f_{ij}^* g_{ij,ji} = 0$$

which implies

$$g_{ii,ii} = g_{ii,jj} + g_{ij,ji} = b_1 + b_2$$

and

$$g_{ij,kl} = b_1 \delta_{ij} \delta_{kl} + b_2 \delta_{il} \delta_{jk}$$

□

O.6 Some Special Definitions

Definition 83 *A function obeying the rule* $f(\delta Z) = \delta^a f(Z)$ *for all* δ *is called* homogeneous of degree a. *This is a complexification of a definition given in class by Krantz.*

The next two definitions should be blamed on me.

Definition 84 *A* generalized even function f *is a function* $f(Z)$ *that obeys the rule* $f(e^{i\theta}Z) = f(Z)$ *for all* $\theta \in \mathbf{R}$, *where* $Z \in \mathbf{C}^n$.

Definition 85 *A* generalized odd function f *is a function* $f(Z)$ *that obeys the rule* $f(e^{i\theta}Z) = e^{i\theta}f(Z)$ *for all* $\theta \in \mathbf{R}$, *where* $Z \in \mathbf{C}^n$.

Notice that when Z is restricted to $\mathbf{R}$ that $\theta \in \{n\pi \mid n \in \mathbf{Z}\}$, and these definitions specialize to the usual notions of $f(-x) = f(x)$ for even functions and $f(-x) = -f(x)$ for odd functions. Thus, odd functions are homogeneous of degree 1 in $e^{i\theta}$.

Definition 86 *Two functions* f, g *are called* algebraically independent *if for any polynomial function* $\sum a_{ij} f^i g^j = 0$ *with complex coefficients* a_{ij}, *we must have* $a_{ij} = 0$ *for all* i, j. *This definition was taken from Lang (p. 262) [160].*

O.7 Generalized Nested Operator

Definition 87 *Nested Operator. Let* $\Box$ *and* $\circ$ *be operators such that*

$$a \circ (b \Box c) = (a \circ b) \Box (a \circ c)$$

Then the nested operator is defined by

$$\bigwedge_{k=1}^{n}(a_k \square b_k)$$

$$\underset{=}{\text{def}}\ [a_1 \square b_1 \circ (a_2 \square b_2 \circ \{a_3 \square b_3 \circ \langle \cdots a_{n-1} \square b_{n-1} \circ (a_n \square b_n)\rangle\})]$$

$$= a_1 \square (b_1 \circ a_2) \square (b_1 \circ b_2 \circ a_3) \square (b_1 \circ b_2 \circ b_3 \circ a_4) \square \cdots$$

$$\square (b_1 \circ b_2 \circ \cdots \circ b_n \circ a_n) \square (b_1 \circ \cdots \circ b_n)$$

where $n \in \mathbf{N}$. *This is an extension of the definition given by Tuma (section 8.11) [268].*

Application: Polynomial. Let $\square$ be ordinary addition, let $\circ$ be ordinary multiplication, and let $b_k = x$ for all k. Then

$$\bigwedge_{k=1}^{n}(a_k + x) = a_1 + a_2 x + a_3 x^2 + \cdots + a_n x^{n-1} + x^n$$

Appendix P

INTEGRALS

The purpose of this portion is to make this thesis easier and quicker to read and understand, and for verifying those integrals which have been required or closely related to the thesis work. Many of these integrals can be done by most sophomores. However, some may require explanation. In many cases these integrals were not in Gradshteyn and Ryzhik [94] or in Abramowitz and Stegun [1]. The integrals are ordered according to their use of prior results. They fall into several categories. The most interesting category has to do with integration over groups, and those which involve zonal polynomials and hypergeometric functions. The next most interesting grouping consists of integrals over matrices. Finally, there are the routine tedious integrals which are uninteresting and should only be done once in a lifetime, and hence they are recorded so they will not have to be done again.

The integrals which are most important are the ones that define the multivariate Gamma function, the matrix Laplace transform, and those involving hypergeometric functions of one and two matrix arguments.

P.1 Easy Chain Rule Bookkeeping Method

These are the uninteresting integrals. I will try to make it more palatable by introducing a bookkeeping method to reduce the work involved in doing the chain rule.

Lemma 59 *Chain Rule Evaluation.* $\int udv = uv - \int vdu$.

Sometimes evaluating an integral by using the chain rule yields a long sequence of steps. To reduce the labor (and thus the opportunity for clerical error), a simple convention below permits efficient iteration. Given $\int udv$, write the uv term on the left half of a line. Follow the uv term by a vertical dashed line, which is then followed on the right half of that line by the $-\int(du)v$ term left in its integral form. This integral is now operated on by the chain rule with the result on the next line. The process continues repeatedly. The solution to the original integral is the sum of terms to the left of the dotted line.

An example of this technique is given in lemma 63, the evaluation of

$$\int (xt+y)^{-n} e^{-zxt} dt$$

$$
\begin{array}{lcl}
\int u^{(m)}v^{(n)} & \vdots & \\
= u^{(m)}v^{(n-1)} & \vdots & -\int u^{(m+1)}v^{(n-1)} \\
 & \swarrow & \\
-u^{(m+1)}v^{(n-2)} & \vdots & +\int u^{(m+2)}v^{(n-2)} \\
 & \swarrow & \\
+u^{(m+2)}v^{(n-3)} & \vdots & -\int u^{(m+3)}v^{(n-3)} \\
 & \swarrow & \vdots \\
-u^{(m+3)}v^{(n-4)} & \vdots & \\
\vdots & \vdots & \\
\overbrace{\cdots\cdots\cdots\cdots}^{\text{sum is solution}} & &
\end{array}
$$

P.2 Mundane Integrals

P.2.1 Exponential Integral Definition

Lemma 60 *Exponential Integral. This integral is listed here for reference purposes. This integral is used to express the results of one of the test distributions for examining a disjoint combination of sample eigenvalues.*

$$Ei(x) \underset{=}{\text{def}} -P.V. \int_{-x}^{\infty} t^{-1}e^{-t}dt, \; x > 0$$

This is found in Abramowitz and Stegun (p. 288) [1]. The path of integration excludes $t = 0$ and the path does not cross the negative real axis. A series

expansion is

$$Ei(x) = \gamma + \ln(x) + \sum_{n=1}^{\infty} \frac{x^n}{nn!}, \ x > 0$$

where $\gamma = 0.57721\ 56649$ *is Euler's constant.* $Ei(x)$ *is tabled in Abramowitz and Stegun [1] for* $\frac{1}{2} \leq x \leq 2$.

P.2.2 Integrals of Rational Functions

Proposition 89

$$\int t(xt+y)^{-1}dt = \frac{1}{x^2}[(xt+y) - y\ln(xt+y)]$$

Proof.

$$\int t(xt+y)^{-1}dt = \frac{1}{x}\int t\, d[\ln(xt+y)]$$

$$= \frac{1}{x}t\ln(xt+y) - \frac{1}{x}\int \ln(xt+y)dt = \frac{1}{x}t\ln(xt+y) - \frac{1}{x^2}\int \ln(xt+y)d(xt+y)$$

$$= \frac{1}{x}t\ln(xt+y) - \frac{1}{x^2}(xt+y)[\ln(xt+y) - 1]$$

$$= \left(\frac{xt}{x^2} - \frac{xt+y}{x^2}\right)\ln(xt+y) + \frac{xt+y}{x^2}$$

$$= \frac{1}{x^2}[(xt+y) - y\ln(xt+y)]$$

□

Proposition 90

$$\int \frac{t^n}{(xt+y)^m}dt = -\sum_{k=1}^{n+1} \frac{(m-1-k)!}{(m-1)!}\frac{n!}{(n+1-k)!}\frac{1}{x^k}t^{n+1-k}(xt+y)^{-(m-k)} + c$$

for $n < m$.

Proof.

$$\int t^n(xt+y)^{-m}dt$$

$$= -\frac{1}{(m-1)x}t^n(xt+y)^{-(m-1)} + \frac{1}{(m-1)x}n\int t^{n-1}(xt+y)^{-(m-1)}dt$$

$$= -\frac{1}{(m-1)}\frac{1}{x}t^n(xt+y)^{-(m-1)} - \frac{n}{(m-1)(m-2)}\frac{1}{x^2}t^{n-1}(xt+y)^{-(m-2)}$$

$$-\frac{n(n-1)}{(m-1)(m-2)(m-3)}\frac{1}{x^3}t^{n-2}(xt+y)^{-(m-3)} - \cdots$$

$$-\frac{n(n-1)\cdots(n-k+2)}{(m-1)(m-2)\cdots(m-k)}\frac{1}{x^k}t^{n-k+1}(xt+y)^{-(m-k)}$$

$$+\frac{n(n-1)\cdots(n-k+1)}{(m-1)(m-2)\cdots(m-k)}\frac{1}{x^k}\int t^{n-k}(xt+y)^{-(m-k)}dt$$

$$= -\sum_{k=1}^{n+1}\frac{(m-1-k)!}{(m-1)!}\frac{n!}{(n+1-k)!}\frac{1}{x^k}t^{n+1-k}(xt+y)^{-(m-k)}$$

□

Proposition 91

$$\int\left(\frac{t}{xt+y}\right)^n dt = \frac{ny}{x^{n+1}}(1-\ln(xt+y))$$

$$+\frac{nt}{x^n} - \sum_{k=1}^{n-1}\frac{n}{(n-k+1)(n-k)}\frac{1}{x^k}t^{n+1-k}(xt+y)^{-(n-k)} + c$$

Proof. In proposition 90, let $n = m$ and consider the last two terms when $k = m-1$. We get

$$-\frac{m(m-1)\cdots[m-(m-1)+2]}{(m-1)(m-2)\cdots[m-(m-1)]}\frac{1}{x^{m-1}}t^{m+1-(m-1)}(xt+y)^{-[m-(m-1)]}$$

$$+\frac{m(m-1)\cdots[m+1-(m-1)]}{(m-1)(m-2)\cdots[m-(m-1)]}\frac{1}{x^{m-1}}\int t^{m-(m-1)}(xt+y)^{-[m-(m-1)]}dt$$

$$= -\frac{m(m-1)\cdots(3)}{(m-1)(m-2)\cdots(1)}\frac{1}{x^{m-1}}t^2(xt+y)^{-1}$$

$$+\frac{m(m-1)\cdots(2)}{(m-1)(m-2)\cdots(1)}\frac{1}{x^{m-1}}\int t(xt+y)^{-1}dt$$

Recall proposition 89. We get

$$-\frac{m(m-1)\cdots(3)}{(m-1)(m-2)\cdots(1)}\frac{1}{x^{m-1}}t^2(xt+y)^{-1}$$

$$+\frac{m(m-1)\cdots(2)}{(m-1)(m-2)\cdots(1)}\frac{1}{x^{m+1}}\left[(xt+y)-y\ln(xt+y)\right]$$

$$=-\frac{m(m-1)\cdots(3)}{(m-1)(m-2)\cdots(1)}\frac{1}{x^{m-1}}t^2(xt+y)^{-1}+\frac{mt}{x^m}+\frac{my}{x^{m+1}}\left[1-\ln(xt+y)\right]$$

Also note:

$$\frac{(n-1-k)!}{(n-1)!}\frac{n!}{(n+1-k)!}=\frac{n}{(n-k+1)(n-k)}$$

Substitution into the last two terms of proposition 90 produces the result. □

Lemma 61

$$\int\frac{x^m}{ax+b}dx=\sum_{k=0}^{m-1}(-1)^k\frac{1}{a}\left(\frac{b}{a}\right)^k\frac{x^{m-k}}{m-k}+(-1)^m\frac{1}{a}\left(\frac{b}{a}\right)^m\ln(ax+b)$$

Proof. Solve by brute force algebraic division with a remainder term, and then integrate the result.

$$\frac{x^m}{ax+b}=\underbrace{\frac{1}{a}x^{m-1}}_{k=0}\underbrace{-\frac{1}{a}\left(\frac{b}{a}\right)x^{m-2}}_{k=1}+\cdots$$

The result of the k^{th} division (starting with $k=0$) produces the term

$$(-1)^k\left(\frac{1}{a}\right)\left(\frac{b}{a}\right)^k x^{m-k-1}$$

with a remainder for that division being

$$(-1)^{k+1}\left(\frac{b}{a}\right)^{k+1}x^{m-k-1}$$

Division number $k = m - 1$ produces a remainder term with a zero power of x and remainder $(-1)^m \left(\frac{b}{a}\right)^m$. Thus the result of the division is

$$\frac{x^m}{ax+b} = \left[\sum_{k=0}^{m-1} (-1)^k \frac{1}{a} \left(\frac{b}{a}\right)^k x^{m-k-1}\right] + (-1)^m \left(\frac{b}{a}\right)^m \frac{1}{ax+b}$$

Integrate this over all x to get the final result. □

Theorem 144 *Let k and p be positive integers. Then*

$$\int \frac{x^k}{(ax+b)^p} dx$$

$$= \begin{cases} \left(\frac{1}{a}\right)^{k+1} \sum_{m=0}^{k} \binom{k}{m} (-b)^m \left(\frac{1}{k-p-m+1}\right) (ax+b)^{k-p-m+1}, & k < p-1 \\ \left(\frac{1}{a}\right)^{k+1} \left\{ \binom{k}{k-p+1} (-b)^{k-p+1} \ln(ax+b) \right. & \\ \left. + \sum_{\substack{m=0 \\ m \neq k-p+1}}^{k} \binom{k}{m} (-b)^m \left(\frac{1}{k-p-m+1}\right) (ax+b)^{k-p-m+1} \right\}, & k \geq p-1 \end{cases}$$

Proof. Let $z = ax + b$. Then $x = \frac{1}{a}(z-b)$ and $dx = \frac{1}{a} dz$. Performing the change of variables, we get

$$\int \frac{x^k}{(ax+b)^p} dx = \frac{1}{a} \int \left[\frac{1}{a}(z-b)\right]^k z^{-p} dz$$

$$= \left(\frac{1}{a}\right)^{k+1} \int z^{-p} \sum_{m=0}^{k} \binom{k}{m} (-b)^m z^{k-m} dz$$

where the complicated exponent on z was chosen for the expansion to keep the dependence on k explicit. When $k \geq p-1$ we have a special case when

$m = k - p + 1$. We rearrange the problem to sharpen this dependence. The rearranged problem is

$$\int \frac{x^k}{(ax+b)^p} dx = \left(\frac{1}{a}\right)^{k+1} \sum_{m=0}^{k} \binom{k}{m} (-b)^m \int z^{k-p-m} dz$$

The result follows from the integration. □

P.2.3 Integrals Related to the Gamma Function

Definition 88 *The gamma function is defined by*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

where $\mathrm{Re}(z) > 0$ *and* z *is complex.*

Recall that $z\Gamma(z) = \Gamma(z+1)$ and when n is an integer we get $\Gamma(n+1) = n!$.

Lemma 62 *Let* $\mathrm{Re}(\alpha z + 1) > 0$. *Then*

$$\int_0^\infty t^{\alpha z} e^{-\beta t} dt = \beta^{-(\alpha z+1)} \Gamma(\alpha z + 1)$$

Proof. Perform the change of variables $x = \beta t$. Then $t = \frac{x}{\beta}$ and $dt = \frac{1}{\beta} dx$. The new limits of integration are $(0, \infty)$. Then

$$\int_0^\infty t^{\alpha z} e^{-\beta t} dt = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha z} e^{-\beta\left(\frac{x}{\beta}\right)} \frac{1}{\beta} dx = \beta^{-(\alpha z+1)} \int_0^\infty x^{\alpha z} e^{-x} dx$$

$$= \beta^{-(\alpha z+1)} \int_0^\infty x^{(\alpha z+1)-1} e^{-x} dx = \beta^{-(\alpha z+1)} \Gamma(\alpha z + 1)$$

when $\mathrm{Re}(\alpha z + 1) > 0$. □

Theorem 145 *Let n be a positive integer, and let a and b be real numbers. Then*

$$\int_a^b t^n e^{-t} dt = n! \sum_{k=0}^{n} \frac{1}{k!}(e^{-a}a^k - e^{-b}b^k)$$

$$= i2n! \exp\left(-\frac{a+b}{2}\right) \sum_{k=0}^{n} \frac{(ab)^{k/2}}{k!} \sinh[\omega(k)]$$

Proof. Apply the chain rule.

$$\int_a^b t^n e^{-t} dt = -t^n e^{-t}\Big|_a^b + n\int_a^b t^{n-1} e^{-t} dt$$

$$= -(t^n + nt^{n-1} + n(n-1)t^{n-2} + \cdots + n!)e^{-t}\Big|_a^b$$

$$= -e^{-t}\sum_{k=0}^{n} \frac{n!}{k!} t^k \Big|_a^b = -e^{-b}n!\sum_{k=0}^{n}\frac{1}{k!}b^k + e^{-a}n!\sum_{k=0}^{n}\frac{1}{k!}a^k$$

$$= n!\left(\sum_{k=0}^{n}\left[\frac{1}{k!}\right]\left[a^k e^{-a} - b^k e^{-b}\right]\right)$$

The last term in brackets looks inviting because of its symmetry. It can be a false oasis. If you go through the mathematics and let

$$\omega(k) = \frac{1}{2}\left[b - a + k\ln\left(\frac{a}{b}\right)\right] = \frac{1}{2}\left[(b - k\ln b) - (a - k\ln a)\right]$$

then you can manipulate the last term in brackets.

$$a^k e^{-a} - b^k e^{-b}$$

$$= \exp\left(-\frac{a+b}{2}\right)(ab)^{k/2}\left\{\left[\frac{a}{b}\right]^{k/2} e^{\left(\frac{b-a}{2}\right)} - \left[\frac{a}{b}\right]^{-k/2} e^{-\left(\frac{b-a}{2}\right)}\right\}$$

$$= \exp\left(-\tfrac{a+b}{2}\right)(ab)^{k/2}\left\{\exp\left[\tfrac{b-a}{2} + \tfrac{k}{2}\ln\left(\tfrac{a}{b}\right)\right]\right.$$

$$\left. - \exp\left[-\left(\tfrac{b-a}{2} + \tfrac{k}{2}\ln\left(\tfrac{a}{b}\right)\right)\right]\right\}$$

$$= 2i \exp\left(-\frac{a+b}{2}\right)(ab)^{k/2} \sinh[\omega(k)]$$

Then the result becomes

$$\int_a^b t^n e^{-t} dt = n! \sum_{k=0}^{n} (e^{-a} a^k - e^{-b} b^k) \frac{1}{k!}$$

$$= i2n! \exp\left(-\frac{a+b}{2}\right) \sum_{k=0}^{n} \frac{(ab)^{k/2}}{k!} \sinh[\omega(k)]$$

□

Corollary 44

$$\int_0^b t^n e^{-t} dt = n! \left(1 - e^{-b} \sum_{k=0}^{n} \frac{b^k}{k!}\right)$$

Proof. In theorem 145, let $a = 0$. Notice that the second form of the result is not as useful because $\ln(0)$ is undefined in the definition of $\omega(k)$. □

Corollary 45

$$\int_a^\infty t^n e^{-t} dt = n! e^{-a} \sum_{k=0}^{n} \frac{a^k}{k!}$$

Proof. In theorem 145, let $b = \infty$. □

Corollary 46

$$\int_0^\infty t^n e^{-t} dt = n! = \Gamma(n+1)$$

Remark. This merely follows the definition of $\Gamma(n+1)$, but we obtained it by letting $a = 0$ and $b = \infty$ in theorem 145. □

Corollary 47

$$\int_0^a w^b e^{-cw} dw = c^{-(b+1)} b! \left(1 - e^{-ac} \sum_{k=0}^{b} \frac{(ac)^k}{k!}\right)$$

Proof. Let $t = cw$ be a change of variables. Then $w = \frac{t}{c}$ and $dw = \frac{1}{c}dt$. The limits become $(0, ac)$.

$$\int_0^{ac} \left(\frac{t}{c}\right)^b e^{-t} \frac{1}{c} dt = c^{-(b+1)} \int_0^{ac} t^b e^{-t} dt = c^{-(b+1)} b! \left(1 - e^{-ac} \sum_{k=0}^{b} \frac{(ac)^k}{k!}\right)$$

□

Corollary 48

$$\int_0^\infty e^{-\alpha x} x^m dx = \alpha^{-(m+1)} m!$$

Proof. Let $t = \alpha x$ be a change of variables. Then

$$\int_0^\infty e^{-\alpha x} x^m dx = \int_0^\infty \left(\frac{t}{\alpha}\right)^m e^{-t} \frac{1}{\alpha} dt = \alpha^{-(m+1)} \int_0^\infty t^m e^{-t} dt$$

$$= \alpha^{-(m+1)} \int_0^\infty t^{(m+1)-1} e^{-t} dt$$

where we note that $\mathrm{Re}(m+1) = m + 1 > 0$ for $m \geq 0$. By the definition of the gamma function, we get

$$\alpha^{-(m+1)} \Gamma(m+1) = \alpha^{-(m+1)} m!$$

□

Lemma 63

$$\int (xt + y)^n e^{-zxt} dt = -\sum_{m=0}^{n} \left(\frac{1}{zx}\right)^{m+1} \frac{n!}{(n-m)!} (xt + y)^{n-m} e^{-zxt}$$

Proof. Apply the chain rule.

$$
\begin{array}{ll}
\int (xt+y)^n e^{-zxt}dt & \\
= -\left(\frac{1}{zx}\right)(xt+y)^n e^{-zxt} & \vdots \; +\left(\frac{1}{zx}\right) n \int (xt+y)^{n-1} e^{-zxt}dt \\
-\left(\frac{1}{zx}\right)^2 n(xt+y)^{n-1}e^{-zxt} & \vdots \; +\left(\frac{1}{zx}\right)^2 n(n-1)\int (xt+y)^{n-2}e^{-zxt}dt \\
-\left(\frac{1}{zx}\right)^3 n(n-1)(xt+y)^{n-2}e^{-zxt} & \vdots \; +\left(\frac{1}{zx}\right)^3 n(n-1)(n-2) \\
 & \quad \times \int (xt+y)^{n-3}e^{-zxt}dt \\
-\left(\frac{1}{zx}\right)^4 n(n-1)(n-2)(xt+y)^{n-3}e^{-zxt} & \vdots \; \vdots \\
\vdots & \vdots \; \vdots \\
\vdots & \vdots \; +\left(\frac{1}{zx}\right)^n n! \int e^{-zxt}dt \\
-\left(\frac{1}{zx}\right)^{n+1} n! e^{-zxt} & \vdots
\end{array}
$$

To get the answer, we sum the results in the left column. We get

$$\int (xt+y)^n e^{-zxt}dt = -\sum_{m=0}^{n}\left(\frac{1}{zx}\right)^{m+1}\frac{n!}{(n-m)!}(xt+y)^{n-m}e^{-zxt}$$

□

Proposition 92

$$\int x^n e^{-ax}dx = \sum_{k=0}^{n}(-1)^k\left(-\frac{1}{a}\right)^{k+1}\frac{n!}{(n-k)!}x^{n-k}e^{-ax}, n \geq 0$$

Proof.

$$
\begin{aligned}
&\int x^n e^{-ax} dx \\
&= \left(-\tfrac{1}{a}\right) x^n e^{-ax} && \vdots \; -\left(-\tfrac{1}{a}\right) n \int x^{n-1} e^{-ax} dx \\
&- \left(-\tfrac{1}{a}\right)^2 n x^{n-1} e^{-ax} && \vdots \; + \left(-\tfrac{1}{a}\right)^2 n(n-1) \int x^{n-2} e^{-ax} dx \\
&+ \left(-\tfrac{1}{a}\right)^3 n(n-1) x^{n-2} e^{-ax} && \vdots \; - \left(-\tfrac{1}{a}\right)^3 n(n-1)(n-2) \int x^{n-3} e^{-ax} dx \\
&\vdots && \vdots \; \vdots \\
&= \sum_{k=0}^{n} (-1)^k \left(-\tfrac{1}{a}\right)^{k+1} \tfrac{n!}{(n-k)!} x^{n-k} e^{-ax} && \text{for } n \geq 0
\end{aligned}
$$

□

P.2.4 Ratio of Exponential to Algebraic Term

Proposition 93

$$\int \frac{\exp(-zxt)}{xt+y} dt = \frac{1}{x} e^{zy} \operatorname{Ei}[-z(xt+y)]$$

Proof. Perform a change of variables. Let $u = z(xt+y)$. Then

$$t = \frac{1}{x}\left(\frac{u}{z} - y\right)$$

and

$$dt = \frac{1}{xz} du$$

Then

$$\int (xt+y)^{-1} e^{-zxt} dt = \int \frac{z}{u} e^{-u+zy} \left(\frac{1}{xz}\right) du = \frac{1}{x} e^{zy} \int u^{-1} e^{-u} du$$

$$= \frac{1}{x} e^{zy} \operatorname{Ei}(-u) = \frac{1}{x} e^{zy} \operatorname{Ei}[-z(xt+y)]$$

The proof is more easily seen to be consistent with other references regarding the exponential integral $\mathrm{Ei}(-u)$ by considering the definite integral

$$\int_a^b (xt+y)^{-1} \exp(-zxt)dt$$

The change of variables yields

$$\int_{z(xa+y)}^{z(xb+y)} \frac{z}{u} e^{yz-u} \frac{1}{xz} du = \frac{1}{x} e^{zy} \int_{z(xa+y)}^{z(xb+y)} u^{-1} e^{-u} du$$

$$= \frac{1}{x} e^{zy} \left[\int_{z(xa+y)}^{\infty} u^{-1} e^{-u} du - \int_{z(xb+y)}^{\infty} u^{-1} e^{-u} du \right]$$

$$= \frac{1}{x} e^{zy} \left[-\mathrm{Ei}[-z(xa+y)] + \mathrm{Ei}[-z(xb+y)] \right]$$

$$= \frac{1}{x} e^{zy} \left[\mathrm{Ei}[-z(xb+y)] - \mathrm{Ei}[-z(xa+y)] \right]$$

□

Proposition 94 *Let $n \geq 0$. Then*

$$\int (xt+y)^{-n} e^{-zxt} dt$$

$$= \left\{ \sum_{m=1}^{n-1} (-1)^m x^{-m} \frac{(n-1-m)!}{(n-1)!} (zx)^{m-1} e^{-zxt} (xt+y)^{-(n-m)} \right\}$$
$$+ (-1)^{n-1} \left(\frac{1}{x}\right)^n \frac{1}{(n-1)!} (zx)^{n-1} e^{zy} \mathrm{Ei}[-z(xt+y)]$$

Proof. Apply the chain rule.

$$\int (xt+y)^{-n} e^{-zxt}\,dt \quad \vdots$$

$$= -\left(\tfrac{1}{x}\right)\left(\tfrac{1}{n-1}\right)\int e^{-zxt}\,d(xt+y)^{-(n-1)} \quad \vdots$$

$$= -\left(\tfrac{1}{x}\right)\left(\tfrac{1}{n-1}\right) e^{-zxt}(xt+y)^{-(n-1)} \quad \vdots \quad -\left(\tfrac{1}{x}\right)\left(\tfrac{1}{n-1}\right)(zx)\int e^{-zxt}(xt+y)^{-(n-1)}\,dt$$

$$+\left(\tfrac{1}{x}\right)^2\left(\tfrac{1}{(n-1)(n-2)}\right)\times \quad \vdots \quad +\left(\tfrac{1}{x}\right)^2\left(\tfrac{1}{(n-1)(n-2)}\right)\times$$
$$\times(zx)e^{-zxt}(xt+y)^{-(n-2)} \qquad \times(zx)^2\int e^{-zxt}(xt+y)^{-(n-2)}\,dt$$

$$-\left(\tfrac{1}{x}\right)^3\left(\tfrac{1}{(n-1)(n-2)(n-3)}\right)\times \quad \vdots \ \vdots$$
$$\times(zx)^2 e^{-zxt}(xt+y)^{-(n-3)}$$

$$\vdots \qquad \vdots \ \vdots$$

$$+(-1)^{n-1}\left(\tfrac{1}{x}\right)^{n-1}\tfrac{1}{(n-1)!}\times \quad \vdots \quad +(-1)^{n-1}\left(\tfrac{1}{x}\right)^{n-1}\tfrac{1}{(n-1)!}(zx)^{n-1}\times$$
$$\times(zx)^{n-2}e^{-zxt}(xt+y)^{-1} \qquad \times\int e^{-zxt}(xt+y)^{-1}\,dt$$

$$+(-1)^{n-1}\left(\tfrac{1}{x}\right)^{n-1}\tfrac{1}{(n-1)!}\times \quad \vdots$$
$$\times(zx)^{n-1}\left(\tfrac{1}{x}\right)e^{-zy}\,\mathrm{Ei}[-z(xt+y)]$$

Add the contents of the left column to obtain the result. □

P.2.5 Product of Rational Term and Exponential

Theorem 146

$$\int \frac{x^n}{ax+b} e^{-cx} dx$$

$$= (-1)^n \frac{1}{a} \left(\frac{b}{a}\right)^n e^{bc/a} \ln(ax+b)$$

$$+ \sum_{m=0}^{\infty} \sum_{k=0}^{n+m-1} (-1)^{m+k} \frac{1}{a} \left(\frac{b}{a}\right)^k \frac{c^m}{m!} \frac{x^{n+m-k}}{(n+m-k)}$$

Proof.

$$I = \int \frac{x^n}{ax+b} e^{-cx} dx = \int \frac{x^n}{ax+b} \sum_{m=0}^{\infty} \frac{(-cx)^m}{m!} dx$$

$$= \sum_{m=0}^{\infty} \frac{(-c)^m}{m!} \int \frac{x^{n+m}}{ax+b} dx$$

Apply lemma 61.

$$I = \sum_{m=0}^{\infty} \frac{(-c)^m}{m!} \left[\sum_{k=0}^{n+m-1} (-1)^k \frac{1}{a} \left(\frac{b}{a}\right)^k \frac{x^{n+m-k}}{(n+m-k)} + (-1)^{n+m} \frac{1}{a} \left(\frac{b}{a}\right)^{n+m} \ln(ax+b) \right]$$

The $\frac{1}{n+m-k}$ factor in the first term prevents us from getting an exponential extracted. However, look at the second term.

$$\sum_{m=0}^{\infty} \frac{(-c)^m}{m!} (-1)^{n+m} \frac{1}{a} \left(\frac{b}{a}\right)^{n+m} \ln(ax+b)$$

$$= (-1)^n \frac{1}{a} \left(\frac{b}{a}\right)^n \ln(ax+b) \sum_{m=0}^{\infty} \frac{c^m}{m!} \left(\frac{b}{a}\right)^m$$

$$= (-1)^n \frac{1}{a} \left(\frac{b}{a}\right)^n \exp\left(\frac{bc}{a}\right) \ln(ax+b)$$

The final result is the sum of the terms. Exchanging the order of summation with corresponding adjustment of limits did not simplify the result. □

Theorem 147

$$\int \frac{x^k}{(ax+b)^p} e^{-cx} dx$$

$$= \left(\frac{1}{a}\right)^{k+1} e^{\frac{bc}{a}} \sum_{m=0}^{k} (-b)^m \left\{ \frac{1}{(p+m-k-1)!} \left(\frac{c}{a}\right)^{p+m-k-1} \mathrm{Ei}\left(-\frac{c}{a}(ax+b)\right) \right.$$

$$\left. - \sum_{n=1}^{p+m-k-1} \frac{(p+m-k-1-n)!}{(p+m-k-1)!} \left(\frac{c}{a}\right)^{n-1} e^{-\frac{c}{a}(ax+b)} (ax+b)^{-(p+m-k-n)} \right\}$$

Proof. Perform the change of variables $z = ax + b$. Then $x = \frac{1}{a}(z-b)$ and $dx = \frac{1}{a}dz$. The integral becomes

$$\int \left[\frac{1}{a}(z-b)\right]^k z^p e^{-\frac{c}{a}(z-b)} \frac{1}{a} dz = \left(\frac{1}{a}\right)^{k+1} e^{\frac{bc}{a}} \int e^{-\frac{cz}{a}} z^p (z-b)^k dz$$

$$= \left(\tfrac{1}{a}\right)^{k+1} e^{\frac{bc}{a}} \int e^{-\frac{cz}{a}} z^{-p} \sum_{m=0}^{k} (-b)^m z^{k-m} dz$$

$$= \left(\tfrac{1}{a}\right)^{k+1} e^{\frac{bc}{a}} \sum_{m=0}^{k} (-b)^m \int e^{-\frac{cz}{a}} z^{-(p+m-k)} dz$$

Concentrate on this integral.

$$\begin{array}{lll}
 & \int e^{-\frac{c}{a}z} z^{-(p+m-k)} dz = & \vdots \\
1 & -\left(\frac{1}{p+m-k-1}\right) e^{-\frac{c}{a}z} z^{-(p+m-k-1)} & \vdots \quad +\left(\frac{1}{p+m-k-1}\right)\left(\frac{c}{a}\right) \times \int e^{-\frac{c}{a}z} z^{-(p+m-k-1)} dz \\
2 & -\left(\frac{1}{(p+m-k-1)(p+m-k-2)}\right) \times \left(\frac{c}{a}\right) e^{-\frac{c}{a}z} z^{-(p+m-k-2)} & \vdots \quad +\left(\frac{1}{(p+m-k-1)(p+m-k-2)}\right) \times \left(\frac{c}{a}\right)^2 \int e^{-\frac{c}{a}z} z^{-(p+m-k-2)} dz \\
\vdots & \vdots & \vdots \quad \vdots \\
n & -\left(\frac{1}{(p+m-k-1)\cdots(p+m-k-n)}\right) \times \left(\frac{c}{a}\right)^{n-1} e^{-\frac{c}{a}z} z^{-(p+m-k-n)} & \vdots \quad \left(\frac{1}{(p+m-k-1)\cdots(p+m-k-n)}\right) \times \left(\frac{c}{a}\right)^{n} \int e^{-\frac{c}{a}z} z^{-(p+m-k-n)} dz \\
\vdots & \vdots & \vdots \quad \vdots \\
p+m-k-1 & -\left(\frac{1}{(p+m-k-1)!}\right) \times \left(\frac{c}{a}\right)^{p+m-k-2} e^{-\frac{c}{a}z} z^{-1} & \vdots \quad +\left(\frac{1}{(p+m-k-1)!}\right)\left(\frac{c}{a}\right)^{p+m-k-1} \times \int e^{-\frac{c}{a}z} z^{-1} dz
\end{array}$$

Concentrate on the integral $\int e^{-\frac{c}{a}z} z^{-1} dz$. Perform the change of variables $w = \frac{c}{a}z$. Thus $z = \frac{a}{c}w$ and $dz = \frac{a}{c}dw$. The integral is

$$\int e^{-\frac{c}{a}z} z^{-1} dz = \int e^{-w} \frac{c}{a} w^{-1} \frac{a}{c} dw = \int e^{-w} w^{-1} dw$$

Consider the definite integral $\int_u^v e^{-w} w^{-1} dw$. It equals

$$\int_u^\infty e^{-w} w^{-1} dw - \int_v^\infty e^{-w} w^{-1} dw$$

$$= -\operatorname{Ei}(-u) + \operatorname{Ei}(-v) = \operatorname{Ei}(-w)\Big|_v^u$$

Thus

$$\int e^{-\frac{c}{a}z} z^{-1} dz = \mathrm{Ei}\left(-\frac{c}{a}z\right)$$

Substitute into the expansion of

$$\int e^{-\frac{c}{a}z} z^{-(p+m-k)} dz$$

to get

$$\int e^{-\frac{c}{a}z} z^{-(p+m-k)} dz$$

$$= \frac{1}{(p+m-k-1)!}\left(\frac{c}{a}\right)^{p+m-k-1} \mathrm{Ei}\left(-\frac{c}{a}z\right)$$
$$- \sum_{n=1}^{p+m-k-1} \frac{(p+m-k-1-n)!}{(p+m-k-1)!}\left(\frac{c}{a}\right)^{n-1} e^{-\frac{c}{a}z} z^{-(p+m-k-n)}$$

The original integral is found by substituting $z = ax + b$ into

$$\left(\frac{1}{a}\right)^{k+1} e^{\frac{bc}{a}} \sum_{m=0}^{k} (-b)^m \times$$
$$\times \left\{ \frac{1}{(p+m-k-1)!}\left(\frac{c}{a}\right)^{p+m-k-1} \mathrm{Ei}\left(-\frac{c}{a}z\right) \right.$$
$$\left. - \sum_{n=1}^{p+m-k-1} \frac{(p+m-k-1-n)!}{(p+m-k-1)!}\left(\frac{c}{a}\right)^{n-1} e^{-\frac{c}{a}z} z^{-(p+m-k-n)} \right\}$$

□

P.2.6 Generalized Even and Odd Functions

Proposition 95 *Let $f(z)$ be a generalized even function for complex z. Then*

$$\int f(z)dz = 2\pi \int f(r)\, r\, dr$$

Proof.

$$\int f(z)dz = \int f(re^{i\theta})\, r\, dr\, d\theta$$

by changing variables to polar coordinates. This equals $\int f(r) r dr d\theta$ since $f(re^{i\theta}) = f(r)$ for all θ. Therefore

$$\int f(z)dz = 2\pi \int f(r)\, r\, dr$$

□

Proposition 96 *Let $f(z)$ be a generalized even functional for $z \in \mathbf{C}^n$. Then*

$$\int f(z)dz = (2\pi)^n \int f(\underline{r})\, r\, dr$$

where $\underline{r} \in \mathbf{R}^n$ and $r = r_1 r_2 \cdots r_n$.

Proof.

$$\int f(z)dz = \int f(r_1 e^{i\theta_1}, \cdots, r_n e^{i\theta_n})\, r_1 \cdots r_n dr_1 \cdots dr_n d\theta_1 \cdots d\theta_n$$

Note that each z_i is undergoing a change of variables to (r_i, θ_i) rather than the usual change to pure polar coordinates where there is a single true radial component. We can do this since each of the z_i are functionally independent. Since $f(re^{i\theta}) = f(r)$ for all θ, we get

$$\int f(r_1, \cdots, r_n)\, r_1 \cdots r_n dr_1 \cdots dr_n d\theta_1 \cdots d\theta_n$$

Since f is a generalized even function, this integral equals

$$(2\pi)^n \int f(r_1, \cdots, r_n)\, r_1 \cdots r_n dr_1 \cdots dr_n = (2\pi)^n \int f(\underline{r})\, r\, dr$$

□

Proposition 97 *Let $f(z)$ be a generalized odd function for complex z. Then* $\int f(z)dz = 0$.

Proof.

$$\int f(z)dz = \int f(re^{i\theta})rdrd\theta$$

by changing variables to polar coordinates where $-\pi \leq \theta \leq \pi$.

$$\int f(re^{i\theta})rdrd\theta = \int e^{i\theta} f(r)rdrd\theta = \left(\int e^{i\theta}d\theta\right)\left(\int f(r)rdr\right)$$

Since f is a generalized odd function, $\int e^{i\theta}d\theta = 0$. □

Proposition 98 *Let $f(z)$ be a generalized odd function for $z \in \mathbf{C}^n$. Then* $\int f(z)dz = 0$.

Proof.

$$\int f(z)dz = \int f(r_1 e^{i\theta_1}, \cdots, r_n e^{i\theta_n})r_1 \cdots r_n dr_1 \cdots dr_n d\theta_1 \cdots d\theta_n$$

$$= \left(\int e^{i\theta_1}d\theta_1\right) \cdots \left(\int e^{i\theta_n}d\theta_n\right) \int f(r_1, \cdots, r_n)r_1 \cdots r_n dr_1 \cdots dr_n$$

Note that $\int e^{i\theta_k}d\theta_k = 0$ for each $k \in [1, n]$. □

P.2.7 Exponentials

Lemma 64

$$\int e^{-|t|^2}dt = \pi$$

Proof.

$$\int e^{-|t|^2} dt = \int e^{-(t_R^2+t_I^2)} dt_R dt_I$$

$$= \left(\int_{-\infty}^{+\infty} e^{-t_R^2} dt_R\right)\left(\int_{-\infty}^{+\infty} e^{-t_I^2} dt_I\right) = \sqrt{\pi}\sqrt{\pi} = \pi$$

□

Lemma 65

$$\int_{-\infty}^{+\infty} t^{2(a-i)+1} e^{-t^2} dt = \frac{1}{2}\Gamma(a-i+1),\ \mathrm{Re}(a-i+1) > 0$$

Proof. Perform the change of variables $y = t^2$, which implies $dy = 2tdt$.

Then

$$\int_{-\infty}^{+\infty} t^{2(a-i)+1} e^{-t^2} dt = \frac{1}{2}\int_0^{\infty} t^{2(a-i)} e^{-t^2} 2t dt$$

$$= \frac{1}{2}\int_0^{\infty} [t^2]^{(a-i)} e^{-t^2} dt^2 = \frac{1}{2}\int_0^{\infty} y^{a-i} e^{-y} dy$$

From the definition of the gamma function, we know this is

$$\frac{1}{2}\int_0^{\infty} y^{(a-i+1)-1} e^{-y} dy = \frac{1}{2}\Gamma(a-i+1),\ \ \mathrm{Re}(a-i+1) > 0$$

□

Proposition 99

$$\int_{\mathbf{C}} e^{-a^{-1}|t|^2} dt = a\pi$$

Proof.

$$\int_{\mathbf{C}} e^{-a^{-1}|t|^2} dt = \int_{\mathbf{R}^2} e^{-a^{-1}(t_R^2+t_I^2)} dt_R dt_I$$

$$= \left(\int_{-\infty}^{+\infty} e^{-a^{-1}t_R^2} dt_R\right)\left(\int_{-\infty}^{+\infty} e^{-a^{-1}t_I^2} dt_I\right)$$

Let $y = a^{-1/2}t$, which implies $dy = a^{-1/2}dt$. Then

$$\int_{\mathbf{C}} e^{-a^{-1}|t|^2}dt = \left(a^{1/2}\int_{-\infty}^{+\infty} e^{-y_R^2}dy_R\right)\left(a^{1/2}\int_{-\infty}^{+\infty} e^{-y_I^2}dy_I\right)$$

$$= a^{1/2}\sqrt{\pi}a^{1/2}\sqrt{\pi} = a\pi$$

□

Proposition 100

$$\int_{-\infty}^{+\infty} t^{2a+1}e^{-b^{-1}t^2}dt = \frac{1}{2}b^{a+1}\Gamma(a+1), \mathrm{Re}(a+1) > 0$$

Proof. Let $u = b^{-1}t^2$, which implies $du = 2b^{-1}tdt$ and $t^2 = bu$. Then

$$\int_{-\infty}^{+\infty} t^{2a+1}e^{-b^{-1}t^2}dt = \frac{1}{2}b\int_{-\infty}^{+\infty} t^{2a}e^{-b^{-1}t^2}2b^{-1}tdt$$

$$= \frac{1}{2}b\int_0^{\infty}(bu)^a e^{-u}du = \frac{1}{2}b^{a+1}\int_0^{\infty} u^a e^{-u}du$$

We use the definition of the gamma function to lead us to

$$\int_{-\infty}^{+\infty} t^{2a+1}e^{-b^{-1}t^2}dt = \frac{1}{2}b^{a+1}\int_0^{\infty} u^{(a+1)-1}e^{-u}du$$

$$= \frac{1}{2}b^{a+1}\Gamma(a+1), \ \ \mathrm{Re}(a+1) > 0$$

□

Proposition 101

$$\int_0^{\infty} e^{-\alpha x^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}, \alpha \neq 0$$

Proof. Let $I = \int_0^\infty e^{-\alpha x^2} dx$ where $\alpha \in \mathbf{C}$ and $x \in \mathbf{R}$.. Then

$$I^2 = \left(\int_0^\infty e^{-\alpha x^2} dx\right)\left(\int_0^\infty e^{-\alpha y^2} dy\right) = \int_0^\infty \int_0^\infty e^{-\alpha(x^2+y^2)} dx dy$$

Let $x = r\cos\theta$ and $y = r\sin\theta$. Then $x^2 + y^2 = r^2$ and $dx\,dy = r\,dr\,d\theta$. This leads to

$$I^2 = \int_0^\infty \int_0^{\pi/2} e^{-\alpha r^2} r dr d\theta = \frac{\pi}{2}\int_0^\infty e^{-\alpha r^2} r dr$$

$$= -\frac{\pi}{4\alpha}\int_0^\infty e^{-\alpha r^2}(-2\alpha r) dr = -\frac{\pi}{4\alpha} e^{-\alpha r^2}\Big|_0^\infty$$

$$= \lim_{b\to\infty} -\frac{\pi}{4\alpha}\left(e^{-\alpha b} - 1\right) = \frac{\pi}{4\alpha}$$

Therefore

$$I = \int_0^\infty e^{-\alpha x^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}, \alpha \neq 0$$

□

This particular change of variables of solving for the square of the integral is one I have seen applied only to this particular example.

Proposition 102

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}}, \alpha \neq 0$$

This is a modification of integral 3.321.3 of Gradshteyn and Ryshik [94].

Proof. Consider the following integral. $I = \int_{-\infty}^{+\infty} e^{-\alpha x^2} dx$ for $\alpha \in \mathbf{C}$ and $x \in \mathbf{R}$. Then

$$I^2 = \left(\int_{-\infty}^{+\infty} e^{-\alpha x^2} dx\right)\left(\int_{-\infty}^{+\infty} e^{-\alpha y^2} dy\right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\alpha(x^2+y^2)} dx dy$$

Perform the change of variables $x = r\cos\theta$ and $y = r\sin\theta$. Then $x^2 + y^2 = r^2$ and $dxdy = rdrd\theta$. This implies

$$I^2 = 4\int_0^{\theta}\int_0^{\pi/2} e^{-\alpha r^2} r dr d\theta = 4\left(\frac{\pi}{2}\right)\int_0^{\infty} e^{-\alpha r^2} r dr$$

Note that

$$de^{-\alpha r^2} = e^{-\alpha r^2}(-2\alpha r)dr$$

Then

$$I^2 = -\frac{\pi}{\alpha}\int_0^{\infty} e^{-\alpha r^2}(-2\alpha r)dr = -\frac{\pi}{\alpha}e^{-\alpha r^2}\Big|_0^{\infty}$$

$$= \lim_{b\to\infty} -\frac{\pi}{\alpha}e^{-\alpha r^2}\Big|_0^{b} = \lim_{b\to\infty} -\frac{\pi}{\alpha}\left(e^{-\alpha b^2} - 1\right)$$

Since α is complex, let $\alpha = \beta + i\gamma$. Then

$$\lim_{b\to\infty}\left|e^{-\alpha b^2}\right| = \lim_{b\to\infty}\left|e^{-\beta b^2}e^{-i\gamma b^2}\right|$$

$$\leq \lim_{b\to\infty}\left|e^{-\beta b^2}\right|\cdot\left|e^{-i\gamma b^2}\right| = \lim_{b\to\infty}\left|e^{-\beta b^2}\right|\cdot 1 = 0$$

Therefore

$$I^2 = -\frac{\pi}{\alpha}(0 - 1) = \frac{\pi}{\alpha}$$

This implies

$$I = \int_{-\infty}^{+\infty} e^{-\alpha x^2}dx = \sqrt{\frac{\pi}{\alpha}}$$

for $\alpha \neq 0$. □

Proposition 103 *Let*

$$a! = a(a-1)(a-2)(a-3)\cdots$$

where a is not necessarily an integer. Then

$$\int u^m(1-u)^a du = -\sum_{k=0}^{m} \frac{a!}{(a+1+k)!}\frac{m!}{(m-k)!}u^{m-k}(1-u)^{a+1+k}$$

$$= -\frac{(1-u)^{a+1}}{(a+1)}\sum_{k=0}^{m}\frac{1}{\binom{a+1+k}{k}}\left[\binom{m}{k}u^{m-k}(1-u)^k\right]$$

Note that

$$\binom{m}{k}u^{m-k}(1-u)^k$$

is the probability of k failures in m trials when $0 < u < 1$, *where u is the probability of success on one trial.*

Proof.

$$\int u^m(1-u)^a du$$

$$= -\left(\frac{1}{a+1}\right) \times$$

$$\times \int u^m(-1)(a+1)(1-u)^a du \; =$$

$$= -\left(\frac{1}{a+1}\right)\int u^m d(1-u)^{a+1} \; =$$

index	summing column	$\cdots$	expansion column
{0}	$= -\left(\frac{1}{a+1}\right)u^m(1-u)^{a+1}$	$\vdots$	$+\frac{m}{(a+1)}\int u^{m-1}(1-u)^{a+1}du$
{1}	$-\frac{m}{(a+1)(a+2)}u^{m-1}(1-u)^{a+2}$	$\vdots$	$+\frac{m(m-1)}{(a+1)(a+2)}\times$ $\times \int u^{m-2}(1-u)^{a+2}du$
{2}	$-\left(\frac{m(m-1)}{(a+1)(a+2)(a+3)}\right)\times$ $\times (u^{m-2}(1-u)^{a+3})$	$\vdots$	$+\left(\frac{m(m-1)(m-2)}{(a+1)(a+2)(a+3)}\right)\times$ $\times \int u^{m-3}(1-u)^{a+3}du$

$$
\begin{array}{lll}
\vdots & \vdots & \vdots \quad \vdots \\
\{k\} & -\left(\frac{a!}{(a+1+k)!}\frac{m!}{(m-k)!}\right)\times & \vdots \quad +\left(\frac{a!}{(a+1+k)!}\frac{m!}{(m-k-1)!}\right)\times \\
 & \times\left(u^{m-k}(1-u)^{a+1+k}\right) & \times\int u^{m-k-1}(1-u)^{a+1+k}du \\
\vdots & \vdots & \vdots \quad \vdots \\
\{m-1\} & -\frac{a!m!}{(a+m)!}u(1-u)^{a+m} & \vdots \quad +\frac{a!m!}{(a+m)!}\int(1-u)^{a+m}du \\
\{m\} & -\frac{a!m!}{(a+m+1)!}(1-u)^{a+m+1} &
\end{array}
$$

The sum of terms on the left side of the table is

$$
\begin{aligned}
\int u^m(1-u)^a du &= -\sum_{k=0}^{m}\frac{a!}{(a+1+k)!}\frac{m!}{(m-k)!}u^{m-k}(1-u)^{a+1+k} \\
&= -\frac{(1-u)^{a+1}}{(a+1)}\sum_{k=0}^{m}\frac{(a+1)!k!}{(a+1+k)!}\frac{m!}{k!(m-k)!}u^{m-k}(1-u)^{a+1+k} \\
&= -\frac{(1-u)^{a+1}}{(a+1)}\sum_{k=0}^{m}\frac{\binom{m}{k}}{\binom{a+1+k}{k}}u^{m-k}(1-u)^{k}
\end{aligned}
$$

□

P.3 Integrals with Hypergeometric Function of Matrix Arguments and Zonal Polynomials

Definition 89 *Let z be a complex number, and let*

$$[a]_k \underset{=}{\text{def}} a(a+1)\cdots(a+k-1)$$

be Pockhammer's symbol. Then the generalized hypergeometric function (or series) of scalar argument z is

$$ {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; z) \underset{=}{\text{def}} \sum_{k=0}^{\infty} \frac{[a_1]_k \cdots [a_p]_k}{[b_1]_k \cdots [b_q]_k} \frac{z^k}{k!} $$

This is Muirhead's definition 1.3.1 [187]. This definition is an important building block for the material dealing with zonal polynomials.

Notes. The sets of numbers $\{a_i\}_1^p$ and $\{b_j\}_1^q$ are complex numbers. The $\{b_j\}_1^q$ cannot be zero or a negative integer. If any of the $\{a_i\}_1^p$ is zero or a negative integer, the series is finite. If $p \leq q$ and $|z| < \infty$, the series converges. If $p = q + 1$, the series converges if $|z| < 1$ and diverges if $|z| > 1$. If $p > q + 1$, the series diverges if $z \neq 0$.

Definition 90 *Let S be the set of all $n \times n$ nonsingular Hermitian matrices, $S = \{X = X^H, X \text{ nonsingular}\}$. Let $X \in S$. Let p and q be nonnegative integers. Let $\alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q$ be complex numbers such that $-\beta_j + (k-1)$ is not a nonnegative integer for $1 \leq j \leq q$ and $1 \leq k \leq n$. Let*

$$ [a]_k = a(a+1)\cdots(a+k-1) $$

which is Pockhammer's symbol. Let $Z_m(X)$ be the zonal polynomial of signature m. Define the hypergeometric function of single matrix argument X

by

$$ {}_pF_q(\alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q; X) $$

$$ = \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{Z_m(X)}{d!} $$

There are some important special cases worth mentioning.

$$ {}_0F_0(X) \quad = \text{etr}(X) = \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{Z_m(X)}{d!} = \sum_{d=0}^{\infty} \frac{(\text{tr}\, X)^d}{d!} $$

$$ {}_1F_0(\alpha; X) \quad = [\det(I - X)]^{-\alpha} $$

Note that ${}_pF_q(X) = {}_pF_q(\Lambda^2)$ *where* Λ^2 *is the diagonal matrix of eigenvalues of* X. *This is Gross and Richard's definition 6.1 [96]. It is very important in the work on zonal polynomials of matrix argument.*

Definition 91 *Let* S *be the set of all* $n \times n$ *nonsingular Hermitian matrices,* $S = \{X = X^H,\ X$ *nonsingular*$\}$. *Let* $X, Y \in S$. *Let* p *and* q *be nonnegative integers. Let* $\alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q$ *be complex numbers such that* $-\beta_j + (k-1)$ *is not a nonnegative integer for* $1 \leq j \leq q$ *and* $1 \leq k \leq n$. *Let*

$$ [a]_k = a(a+1)\cdots(a+k-1) $$

be Pockhammer's symbol. Let $Z_m(X)$ *be the zonal polynomial of signature* m. *Define the* hypergeometric function of two matrix arguments (X, Y) *by*

$$ {}_pF_q(\alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q; X, Y) $$

$$ = \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{Z_m(X) Z_m(Y)}{d! Z_m(I_n)} $$

This is a complexification of Muirhead's definition 7.3.2 [187], and it is Gross and Richard's equation 4.2 [97].

Theorem 148 *Let $X, Y \in S$ where S is the set of all nonsingular $n \times n$ Hermitian matrices. Let $U(n)$ be the set of all $n \times n$ unitary matrices. Let (dU) be the normalized Haar measure on $\mathbf{U}(n)$. Then*

$$\int_{\mathbf{U}(n)} {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; XU^HYU)(dU)$$

$$= {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; X, Y)$$

This is a complexification and slight modification of Muirhead's theorem 7.3.3 [187], and it is Gross and Richards' equation 4.3 [97].

Proof.

$$\int_{\mathbf{U}(n)} {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; XU^HYU)(dU)$$

$$= \int_{\mathbf{U}(n)} \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{Z_m(XU^HYU)}{d!}(dU)$$

$$= \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{1}{d!} \int_{\mathbf{U}(n)} Z_m(XU^HYU)(dU)$$

Applying Gross and Richards' proposition 5.5 [96], we get

$$= \sum_{d=0}^{\infty} \sum_{|m|=d} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{1}{d!} \frac{Z_m(X)Z_m(Y)}{d!Z_m(I_n)}$$

$$= {}_pF_q(\alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q; X, Y)$$

□

Proposition 104 . *Let $z \in \mathbf{C}$. Then*

$$\frac{\Gamma\left(\frac{1}{2}n\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left[\frac{1}{2}(n-1)\right]} \int_0^\pi e^{z\cos\theta}[\sin\theta]^{n-2}d\theta$$

$$= \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{\left[\frac{1}{2}n\right]_k k!} = {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}z^2\right)$$

This is a complexification of Muirhead's lemma 1.3.2 [187].

Proof. This is Muirhead's [187] proof with some steps filled in. By definition of the exponential function,

$$e^{z\cos\theta} = \sum_{k=0}^{\infty} \frac{1}{k!}(z\cos\theta)^k$$

This converges for all $|z| < \infty$. Then by Fubini [230], since the series converges, we can interchange the sum and integral. Thus

$$\int_0^\pi e^{z\cos\theta}[\sin\theta]^{n-2}d\theta = \sum_{k=0}^{\infty} \frac{z^k}{k!} \int_0^\pi [\cos\theta]^k[\sin\theta]^{n-2}d\theta$$

Observe that $[\sin\theta]^{n-2}$ is an *even* function about $\frac{\pi}{2}$ on the interval $(0, \pi)$. When k is *odd*, then $[\cos\theta]^k$ is an odd function about $\frac{\pi}{2}$ on the interval $(0, \pi)$. When k is even, then $[\cos\theta]^k$ is even. Thus

$$\int_0^\pi [\cos\theta]^k[\sin\theta]^{n-2}d\theta = \begin{cases} 2\int_0^{\pi/2}[\cos\theta]^k[\sin\theta]^{n-2}d\theta, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

Thus, if we let $2m = k$ for even k, then

$$\int_0^\pi e^{z\cos\theta}[\sin\theta]^{n-2}d\theta = \sum_{m=0}^{\infty} \frac{z^{2m}}{(2m)!} 2\int_0^{\pi/2} [\cos\theta]^{2m}[\sin\theta]^{n-2}d\theta$$

Perform a change of variables. Let $x = \sin^2\theta$. Then

$$\cos^2\theta = 1 - x$$

and

$$dx = 2\sin\theta\cos\theta\, d\theta$$

which implies

$$d\theta = \frac{1}{2}x^{-1/2}(1-x)^{-1/2}dx$$

The limits are changed from $\theta \in (0, \frac{\pi}{2})$ to $x \in (0,1)$. Then

$$2\int_0^{\pi/2}[\cos\theta]^{2m}[\sin\theta]^{n-2}d\theta$$

$$= 2\int_0^1(1-x)^m x^{(n-2)/2}\tfrac{1}{2}x^{-1/2}(1-x)^{-1/2}dx$$

Merely switching notation from m back to k (but not doing a change of variables), this integral is

$$\int_0^1 (1-x)^{k-\frac{1}{2}}x^{(n-3)/2}dx$$

We rewrite the exponents to place this integral into the form of the definition for the *beta function*, as given in Abramowitz and Stegun equation 6.2.1 [1].

$$\int_0^1 (1-x)^{(k+\frac{1}{2})-1}\, x^{(\frac{n-1}{2})-1}dx$$

$$\underset{=}{\text{def}}\ B\left(\tfrac{n-1}{2}, k+\tfrac{1}{2}\right) = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}+k\right)}$$

which is a commonly used identity for the beta function. The beta function is an important function in theoretical statistics.

With these changes, we now have

$$\int_0^\pi e^{z\cos\theta}[\sin\theta]^{n-2}d\theta = \sum_{k=0}^{\infty}\frac{z^{2k}}{(2k)!}\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}+k\right)}$$

Notes.

$$\frac{\Gamma\left(k+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{\left(k+\frac{n}{2}-1\right)\left(k+\frac{n}{2}-2\right)\cdots}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)\cdots}$$

$$= \left(k+\frac{n}{2}-1\right)\left(k+\frac{n}{2}-2\right)\cdots\left(k+\frac{n}{2}-k\right) = \left(\frac{n}{2}\right)_k$$

Also note that

$$\frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)(2k)!} = \frac{\left(k-\frac{1}{2}\right)\left(k-\frac{3}{2}\right)\left(k-\frac{5}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{(2k)(2k-1)(2k-2)\cdots 2\cdot 1\cdot\Gamma\left(\frac{1}{2}\right)}$$

$$= \frac{\left(\frac{2k-1}{2}\right)\left(\frac{2k-3}{2}\right)\left(\frac{2k-5}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{(2k)(2k-1)(2k-2)\cdots 2\cdot 1\cdot\Gamma\left(\frac{1}{2}\right)}$$

$$= \frac{\frac{1}{2^k}(2k-1)(2k-3)(2k-5)\cdots 1\cdot\Gamma\left(\frac{1}{2}\right)}{2^k(2k-1)(2k-3)(2k-5)\cdots 1\cdot\Gamma\left(\frac{1}{2}\right)k!} = \frac{1}{4^k k!}$$

Putting everything together,

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}\int_0^\pi e^{z\cos\theta}[\sin\theta]^{n-2}d\theta = \sum_{k=0}^{\infty} z^{2k}\frac{1}{\left(\frac{n}{2}\right)_k 4^k k!}$$

$$= \sum_{k=0}^{\infty}\frac{\left(\frac{1}{4}z^2\right)^k}{\left(\frac{n}{2}\right)_k k!} = {}_0F_1\left(\frac{n}{2};\frac{z^2}{4}\right)$$

Notice in the argument of ${}_pF_q(a_1,\cdots,a_p;b_1,\cdots,b_q;y)$ that since $p=0$ then there are no parameters $\{a_1,\cdots,a_p\}$. We have only $b_1=\frac{n}{2}$ and $y=\frac{1}{4}z^2$. We write ${}_0F_1(b_1;y)$ since we have no $\{a_i\}$. So, the lemma is proven.

For additional connections with the beta function, see Herz (p. 480, bottom) [106]. □

Theorem 149 *Let $a_i, b_j, c, z \in \mathbf{C}$. Then*

$$\int_0^\infty e^{-zt} t^{c-1} {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; kt) dt$$

$$= \Gamma(c) z^{-c} {}_{p+1}F_q(a_1, \cdots, a_p, c; b_1, \cdots, b_q; kz^{-1})$$

which holds for

$$\{p < q, \mathrm{Re}(c) > 0, \mathrm{Re}(z) > 0\}$$

and

$$\{p = q, \mathrm{Re}(c) > 0, \mathrm{Re}(z) > \mathrm{Re}(k)\}$$

This is Muirhead's lemma 1.3.3 [187], stated without proof.

Proof. The proof is straight forward with the following observations which come from the definition of the gamma function.

$$\int_0^\infty t^{(m+c)-1} e^{-zt} dt = z^{-(m+c)} z^{m+c} \int_0^\infty t^{(m+c)-1} e^{-zt} dt = z^{-(m+c)} \Gamma(m+c) \quad \text{(P.1)}$$

$$\Gamma(m+c) = (m+c-1)\Gamma(m+c-1) = \cdots \quad \text{(P.2)}$$

$$= (m+c-1)(m+c-2)\cdots(c+1)c\Gamma(c) = (c)_m \Gamma(c)$$

Now, substitute the definition of ${}_pF_q$ into our problem.

$$\int_0^\infty e^{-zt} t^{c-1} {}_pF_q(a_1, \cdots, a_p; b_1, \cdots, b_q; kt) dt$$

$$= \int_0^\infty e^{-zt} t^{c-1} \sum_{m=0}^{\infty} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{1}{m!} (kt)^m dt$$

Switching the order of summing and integration is allowed since the sum converges. This gives us

$$\sum_{m=0}^{\infty} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{k^m}{m!} \int_0^{\infty} e^{-zt} t^{(m+c)-1} dt$$

$$= \sum_{m=0}^{\infty} \frac{[a_1]_m \cdots [a_p]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{(kz^{-1})^m}{m!} z^{-c} \Gamma(m+c)$$

where we used our first observation, equation P.1, which leads us to

$$\Gamma(c) z^{-c} \sum_{m=0}^{\infty} \frac{[a_1]_m \cdots [a_p]_m [c]_m}{[\beta_1]_m \cdots [\beta_q]_m} \frac{(kz^{-1})^m}{m!}$$

$$= \Gamma(c) z^{-c} {}_{p+1}F_q(a_1, \cdots, a_p, c; b_1, \cdots, b_q; kz^{-1})$$

□

P.4 Integrals with Complex Multivariate Gamma Function

Definition 92 *Complex Multivariate Gamma Function,* $\mathbf{C}\Gamma_m(a)$.

$$\mathbf{C}\Gamma_m(a) \underset{=}{\text{def}} \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a-i+1)$$

$$= \pi^{m(m-1)/2} \Gamma(a) \Gamma(a-1) \cdots \Gamma(a-m+1)$$

$$= \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a-m+i)$$

where $\mathrm{Re}(a-m+1) > 0$. This is a complexification of Muirhead's theorem 2.1.12 [187]. It is James equation 83 [120]. This is $\tilde{\Gamma}_p(a)$ in Patil (p. 7) [205].

Discussion. The gamma function is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \ \mathrm{Re}(z) > 0$$

where z is complex. Recall the properties of the univariate $\Gamma(z)$.

$$\Gamma(n+1) = n!$$

$$\Gamma(z+1) = z\Gamma(z)$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Definition 93 *The real multivariate gamma function is defined to be*

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma[a - \frac{1}{2}(i-1)]$$

where $\mathrm{Re}[a - \frac{1}{2}(m-1)] > 0$. *This is Muirhead theorem 2.1.12.*

This function appears in the denominator of the real Wishart distribution density function, $W_p(n, \Sigma)$ where $m = p$ and $a = \frac{n}{2}$. $\Gamma_m(a)$ shows up in integrals that involve zonal polynomials. It also shows up in the cumulative distribution function and expected moments of the Type I Multivariate Beta distribution.

$\mathbf{C}\Gamma_m(a)$ appears in the denominator of the complex Wishart distribution $\mathbf{C}W_p(n, \Sigma)$ density function where $m = p$ and $a = n$.

Lemma 66

$$\int_{A^H=A>0} e^{-\operatorname{tr} A} |\det A|^{a-m} (dA)$$

$$= \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a-i+1) = \mathbf{C}\Gamma_m(a)$$

This is Muirhead's definition 2.1.10 [187] and James' equation (83) [120]. Herz [106] identifies $\mathbf{C}\Gamma_p(m)$ *as the generalization to matrix variables of the Eulerian integral of the second kind.*

Proof. This proof draws from Srivastava's derivation for the standard complex Wishart distribution $\mathbf{C}W_p(n, I)$. We begin with Srivastava's main result, his equation 4 (p. 314) [256],

$$P(B) = C_1 2^p |\det B|^{m-p} f(B)$$

where $B^H = B > 0$, and C_1 is a constant. Since $P(B)$ is a density, it integrates to 1.

$$1 = \int_{B^H=B>0} P(B)(dB) = \int_{B^H=B>0} C_1 2^{-p} |\det B|^{m-p} f(B)(dB)$$

Choose $f(B) = \pi^{-mp} e^{-\operatorname{tr} B}$ as we did in Srivastava's derivation of the density function for the complex Wishart distribution (see theorem 67). In that derivation, C_1 was evaluated to be

$$C_1 = \frac{2^p \pi^{pm}}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)}$$

Substituting into our integral, we obtain

$$1 = \frac{2^p \pi^{pm} 2^{-p} \pi^{-mp}}{\pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1)} \int_{B^H=B>0} e^{-\operatorname{tr} B} |\det B|^{m-p} (dB)$$

Dividing both sides by the constant in front of the integral, we get

$$\int_{B^H=B>0} e^{-\operatorname{tr} B} |\det B|^{m-p} (dB) = \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(m-i+1) = \mathbf{C}\Gamma_p(m)$$

□

Theorem 150 *Let $\Sigma = \Sigma^H$ and $A = A^H$ be $m \times m$ complex matrices where Σ and A are positive definite. Then the matrix Laplace transform of $(\det A)^{a-m}$ with respect to Σ is*

$$\int_{A>0} \operatorname{etr}(-\Sigma^{-1}A)(\det A)^{a-m}(dA)$$

$$= (\det \Sigma)^a \mathbf{C}\Gamma_m(a) = \mathcal{L}_{\Sigma^{-1}}\{(\det A)^{a-m}\}$$

This is a complexification of Muirhead theorem 2.1.11. This is also Herz equation (1.1) [106].

Proof. This is a complexification of Muirhead's proof. By theorem 119, we can decompose Σ into $\Sigma = BB^H$. By convention we use the symbol $\Sigma^{\frac{1}{2}}$ for B and call it the square root of Σ. Thus for $\Sigma = \Sigma^{\frac{1}{2}}\Sigma^{\frac{H}{2}}$ and $\Sigma^{-1} = \Sigma^{-\frac{H}{2}}\Sigma^{-\frac{1}{2}}$. Also, recall theorem 38 says that the Jacobian $J(A \to V)$ of the transformation $A = \Sigma^{\frac{1}{2}} V \Sigma^{\frac{H}{2}}$ where $V = V^H$ is given by

$$\left|\det \Sigma^{\frac{1}{2}}\right|^{2m} = (\det \Sigma)^m$$

With these preliminaries, perform the change of variables in our integral.

$$\int_{A>0} \operatorname{etr}(-\Sigma^{-1}A)(\det A)^{a-m}(dA)$$

$$= \int_{V>0} \operatorname{etr}\left(-\Sigma^{-1}\Sigma^{\frac{1}{2}}V\Sigma^{\frac{H}{2}}\right)\left(\det\left[\Sigma^{\frac{1}{2}}V\Sigma^{\frac{H}{2}}\right]\right)^{a-m}(\det\Sigma)^{m}(dV)$$

$$= \int_{V>0} \operatorname{etr}\left(-\Sigma^{\frac{H}{2}}\Sigma^{-1}\Sigma^{\frac{1}{2}}V\right)(\det\Sigma)^{a-m}(\det V)^{a-m}(\det\Sigma)^{m}(dV)$$

$$= \int_{V>0} \operatorname{etr}(-V)(\det V)^{a-m}(dV)(\det\Sigma)^{a}$$

From lemma 66 we recognize the integral as the complex multivariate gamma function, giving us

$$(\det\Sigma)^{a}\mathbf{C}\Gamma_m(a)$$

which completes the proof. □

Note. If you normalize the integral by $(\det\Sigma)^{a}\mathbf{C}\Gamma_m(a)$, then the integrand is a density function

$$\frac{(\det A)^{a-m}\operatorname{etr}(-\Sigma^{-1}A)}{(\det\Sigma)^{a}\mathbf{C}\Gamma_m(a)}$$

This is the density function for the complex Wishart distribution $\mathbf{C}W_m(a,\Sigma)$. □

Proposition 105

$$\int_{\mathbf{C}^{p\times p}}[\det(E^{H}E)]^{a-p}e^{-\operatorname{tr}(E^{H}E)}(dE) = \frac{\pi^{p^2}\mathbf{C}\Gamma_p(a)}{\mathbf{C}\Gamma_p(p)}$$

where $E \in M_p(\mathbf{C})$ and E is unstructured. This lemma was motivated by the proof of theorem 7.

Proof.

$$\int_{\mathbf{C}^{p\times p}}[\det(E^HE)]^{a-p}e^{-\operatorname{tr}(E^HE)}(dE)$$

$$=\pi^{p^2}\int_{\mathbf{C}^{p\times p}}[\det(E^HE)]^{a-p}\pi^{-p^2}e^{-\operatorname{tr}(E^HE)}(dE)$$

We recognize

$$\pi^{-p^2}e^{-\operatorname{tr}(E^HE)}(dE)$$

as a probability density function for the complex matrix normal distribution $\mathbf{C}N_{p,p}(0,I,I)$. Then the integral is the expected value of

$$[\det(E^HE)]^{a-p}$$

If $E\sim\mathbf{C}N_{p,p}(0,I,I)$, then $G=E^HE$ has the complex Wishart distribution $\mathbf{C}W_p(p,I)$. By theorem 79, we know

$$\mathcal{E}\{[\det(G)]^{a-p}\}=\frac{\mathbf{C}\Gamma_p(a)}{\mathbf{C}\Gamma_p(p)}$$

Thus

$$\int_{\mathbf{C}^{p\times p}}[\det(E^HE)]^{a-p}\operatorname{etr}[-E^HE](dE)$$

$$=\pi^{p^2}\mathcal{E}\{[\det(G)]^{a-p}\}=\pi^{p^2}\frac{\mathbf{C}\Gamma_p(a)}{\mathbf{C}\Gamma_p(p)}$$

Appendix Q

NOTATION

Q.1 Names of Variables

Sometimes the choice between Latin or Greek letters for a special variable is governed by broad consensus within a scientific community and this will override otherwise stated conventions. Such exceptions are noted in the next section.

Matrix: single upper case Latin or Greek letters such as

$$A, Z, \Gamma, \Theta, \Lambda, \Xi, \Sigma, \Phi$$

Vector: usually single lower case Latin or sometimes Greek letters such as

$$a, z, \beta, \theta, \omega, \mu$$

Scalar: usually lower case Greek or sometimes Latin letters such as

$$\alpha, \beta, \gamma, \sigma, a, b, c, d$$

Deterministic variables: usually chosen from early in the alphabet.

Random variables: usually chosen from late in the alphabet.

Distribution parameters: usually Greek letters.

Q.2 Special Notation

Special meanings are attached to the following symbols. This reservation is sometimes violated due to a paucity of available symbols.

$\mathcal{E}$ always is the expected value operator.

$i \quad = \sqrt{-1}$ when i is not an index. The letter j will not be used for $\sqrt{-1}$. I decided to use i since i is not current.

C identifies a distribution or function as the complex variables version of the appended symbol. By itself, or with a superscript, it refers to the set of complex numbers or the product space of complex numbers.

A^H is the Hermitian transpose of the matrix (or vector) A. The Hermitian transpose is the transpose of the complex conjugate of the matrix (or vector). Note that when applied to a scalar, this is merely the complex conjugate.

A^T is the transpose of the matrix (or vector) A.

σ^2 is reserved for the scalar variance parameter of a distribution.

δ is used as a noncentrality parameter for the complex Wishart distribution, which is a matrix.

Σ is reserved for the matrix covariance parameter of a distribution. For the complex matrix normal distribution, this is the covariance between column vectors. Some authors call this the variance-covariance matrix or the dispersion matrix.

Ξ is reserved for the row covariance matrix parameter for the complex

matrix normal distribution.

μ is reserved for the mean value parameter of a distribution.

λ always refers to a singular value.

λ^2 always refers to an eigenvalue.

Λ always refers to the rectangular matrix having the non-zero singular values on its main diagonal.

Λ^2 always refers to the square matrix having the non-zero eigenvalues on its main diagonal.

l almost always refers to a sample singular value. Sometimes l is an integer index.

l^2 always refers to a sample eigenvalue.

L almost always refers to the rectangular matrix having the non-zero sample singular values on its main diagonal.

L^2 always refers to the square matrix having the non-zero sample eigenvalues on its main diagonal.

W almost always refers to a Wishart matrix. Other matrices may also be Wishart matrices.

Φ is often reserved for use as a characteristic function of a distribution.

$\Delta(A)$ is a diagonal matrix consisting of the elements on the diagonal of matrix A.

$\{x_i\}_1^n$ means the sequence $x_1, x_2, \cdots, x_n$.

operator.

(dZ) is a differential form of the elements of Z, which possibly can be a matrix. This notation is used in connection with probability density functions and is a shorthand notation for the absolute value of the product of the element differentials, such as

$$dz_{11}dz_{12}\cdots dz_{1p}dz_{21}\cdots dz_{2p}\cdots dz_{np}$$

Equivalently, this is

$$\mid dz_{11} \bigwedge dz_{12} \bigwedge \cdots \bigwedge dz_{1p} \bigwedge dz_{21} \bigwedge \cdots \bigwedge dz_{2p} \bigwedge \cdots \bigwedge dz_{np} \mid$$

Q.3 Selected Abbreviations

IEEE	Institute of Electrical and Electronics Engineers
IRE	Institute of Radio Engineers, which later became IEEE (PGIT Vol.1, February 1953)
JASA	Journal of the Acoustical Society of America (Vol.1, October 1929)
JASA	Journal of the American Statistical Association (Vol.1, 1888/1889)
MLE	Maximum Likelihood Estimate
RKHS	Reproducing Kernel Hilbert Space
SIAM	Society of Industrial and Applied Mathematics *Note* : *Journal of the SIAM* later named SIAM Journal on Applied Mathematics
UMVU	Uniformly Minimum Variance Unbiased estimate

Q.4 Reference Author Names

The name of an author of a reference used in direct support of this research is *printed with this type style* in the bibliography to distinguish that reference from those used only for presentation of background and history. Prior to now, there was no canonical way of efficiently making this discrimination.

Q.5 Taxonomy of Logic

An attempt has been made to conform to Solow's taxonomy of statements of formal logic [252] (p. 37). His classifications are defined below. This historical taxonomy is not uniformly implemented. Exceptions were made where I judged a proposition in the context of other similar propositions. Any hierarchical taxonomy will fail because we have a multidimensional lattice of logic.

1. Proposition: A statement of interest that you are trying to prove.
2. Theorem: (Subjectively) extremely important propositions.
3. Lemma: Proposition used as a step in proving a theorems.
4. Corollary: Proposition whose veracity follows immediately from a theorem.
5. Axiom: Statement accepted without proof.

VITA

Curtis Irvin Caldwell was born on 04 March 1947 in Columbus, Ohio, United States of America. His father, Col. Elmer Irvin Caldwell, was a career U.S. Army soldier who served during World War II in North Africa, Italy, and France, and also in wars in Korea, and Viet Nam. As an Army dependent, Curtis lived for over a year in Japan, and in Germany for over three years beginning shortly after the Hungarian Revolution, and during the Czechoslovakian Uprising and the Second Berlin Crisis. It was during these years that he developed a deep sense of appreciation for the value of freedom that not everyone in the world enjoyed. From his mother, May Alice Wing Caldwell of Worthington, Ohio, Curtis learned that possession of knowledge and power incurs the obligation of its stewardship for the benefit of others. From his brother, Harold Earnest Caldwell, Curtis learned to love inquiry and analytical thought.

Curtis Caldwell attended grammar school in Germany, and Francis C. Hammond High School in Alexandria, Virginia, USA. He completed a B.S. in Computer Science at the University of South Carolina in 1972 under Dr. William Hines Linder, and an M.A. in Mathematical Sciences with a dual concentration in Statistics and Computer Science from University of North Florida under Drs. William J. Wilson and Yap Siong Chua. It was there that Curtis developed a love for seeing other people learn.

In addition to his interests in underwater acoustics and signal processing, Curtis has interests in Christian systematic theology and citizenship. He counts it a privilege to serve a nation of free people under God.

He is married to Susan Marion Belcher Caldwell, the daughter of Annie Lou Belcher and Jack Belcher, a liberator of the Nardheim Concentration Camp of World War II. His son is Joshua Benjamin Lee Caldwell, named for the spy sent in to the Promised Land who reported that God's promise is good, the favorite son, and the patriot-gentleman-soldier Robert E. Lee.

$|\cdot|$ will always refer to the magnitude. For $z \in C$ and $z = xe^{i\theta}$ for $x, \theta \in R$, then $|z| = x$. $|\cdot|$ will not be used for determinant.

det A will always refer to the determinant of a matrix.

tr A is the trace of a matrix, which is the sum of all the elements on the major diagonal.

etr $A = \exp(\mathrm{tr}(A))$

exp A is the exponential function, which usually refers to the scalar function e^z. It has also been defined for a matrix argument A, in which case e^A is a matrix.

0 is the zero matrix. When a matrix of all zero entries multiplies another matrix, the result is still a matrix of all zero entries, with appropriate dimensions. Rather than using a different notation for each null matrix, I have simply used 0 where the dimensions are assumed to be correct. Thus, I have also dispensed with the need for 0^T as the transpose of the null matrix.

diag(A) is an ordered $n-$tuple of elements of the main diagonal of the $n \times n$ matrix A. The context may determine if this $n-$tuple is a row vector or a column vector of n elements.

diag$(b_1, \cdots, b_n)$ is an $n \times n$ diagonal matrix with $(b_1, \cdots, b_n)$ as the elements on the diagonal.

$\bigwedge$ is (1) usually reserved for the exterior product operator (wedge product) used with differential forms and (2) sometimes reserved for the nested sum