GROUPS, LIE GROUPS, AND MULTIVARIATE STATISTICS

by

MARA LEE McLAREN
B. Math. (Hons.), M. Math. (Waterloo)

Thesis submitted for the Degree of
Doctor of Philosophy
in The University of Adelaide,
Department of Statistics,
November, 1975.

TABLE OF CONTENTS

SUMMARY

In this thesis an investigation of the fundamental relationships between Lie groups and multivariate statistics is presented. Examples are shown where group and Lie theory have been successfully applied to various problems in multivariate statistics, by utilizing the intrinsic symmetry of multivariate normal theory which can be described and analysed by group theory. On a more profound level, the part Lie groups play in the theory of non-central multivariate distributions is studied, in the hope that a new view of this topic will enable a wider audience to appreciate the substance of this difficult branch of statistics.

The first part of this work outlines the elementary theory of Lie groups and algebras, with emphasis on applying results directly to the groups of interest statistically, in an attempt to make this theory more accessible to statisticians. Some of these results are used to derive a new test concerning equality of eigenvalues and eigenvectors for two sample Wishart matrices, and to analyse other problems.

Next the manifold structures of the positive definite symmetric matrices and the Grassman manifold are studied, deriving their natural distance measures, and proving that the metric on the positive definite symmetric matrices obeys the triangle inequality. This requires an interesting lemma concerning bounds on the latent roots of the product of two such matrices.

Then the connection between the zonal polynomials of multivariate distribution theory, and the theory of symmetric spaces is explored. This is a non-trivial exercise, since it involves the complete representation and decomposition theory of semi-simple Lie groups and algebras. The symmetric space theory is currently only available in the literature in highly abstract and remote form (see for example the papers of Harish-Chandra). In an attempt to make this brilliant work more understandable, some down-to-earth examples, which are useful in statistics, are worked out in relative detail, so that some idea of the structures involved can be quickly appreciated.

These results are used to study some interesting new Koornwinder polynomials, to see where they overlap with the James' polynomials. Finally the generalized Hermite polynomials are found, and some possible directions for further research indicated.

An appendix contains a multi-purpose program that calculates the coefficients for the zonal polynomials, using a method derived by A.T. James which utilizes a basic symmetric space property of the polynomials. It is described in the chapter on symmetric spaces.

SIGNED STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, the thesis contains no material previously published or written by any other person, except where due reference is made in the text of the thesis.

M.L. McLaren

ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor A.T. James for his help and encouragement during the progress of this research, and to the National Research Council of Canada, and the University of Adelaide for their generous funding of my studies. Many thanks are due to Katherine Halsey for her intelligent typing of the manuscript.

"The universe is incredibly majestic and remarkable ... that to find out something about it, even about one tiny corner of it ... how it works ... how it operates, is one of the most rewarding experiences anyone can have."

Sir Mark Oliphant K.B.E., F.R.S., Ph.D.

CHAPTER 1

PRELIMINARIES

1.1 Introduction

This thesis embodies the results of a study undertaken into the possible uses of Lie group theory and related topics in multivariate statistics. That such a study was ever begun is not surprising for the essential structure of multivariate normal theory involves the two Lie groups, the general linear group, $G\ell(m,R)$, (as usual, R denotes the real numbers, C the complex) and the orthogonal group, $O(m,R)$. Further, the crucial space of positive definite real symmetric matrices, $M(m,R)$, can be realized as the coset space, $G\ell(m,R)/O(m,R)$, which inherits from its Lie group roots a symmetric space structure. It was thought that a very detailed study of these groups and coset spaces would help to develop a clear and natural understanding of the concepts and framework of multivariate statistics. A great deal is known about Lie groups, their algebras, and symmetric spaces, and one would imagine that contained in this large body of knowledge would be much information of value to statisticians, both theoretical and applied.

There are a few problems attached to such a study. For one thing, the literature on the subject is often presented in an inaccessible fashion, being geared generally for post-graduate students or academics in pure and applied mathematics. Another, in many ways more serious problem arises for the statistician. In other applications of Lie theory, such as nuclear physics and classical special function theory, one uses rather more difficult groups than we require, but of particular (and low!) dimensions. In multivariate statistics we use relatively elementary groups, but for the most part of unspecified dimension, being more interested in general structures than particular results.

These difficulties aside, a study of groups and Lie theory gives one real insight into multivariate normal statistics. It is the author's intention to <u>firmly</u> imbed multivariate statistics in Lie group theory, using as much geometry and intuition as possible. It is hoped that this new perspective on the subject may prove to be of some value in providing new approaches to different problems.

1.2 Outline

Apart from a short motivated section, the first part of this work follows in general the outline of a typical elementary text on Lie groups and algebras. Since much work must be covered, this is necessarily sketchy, with many references. It was thought that the reader would gain more from an understanding of the significance of theorems, and an ability to manipulate the various structure, than from rigorous proofs, which would in any case have made this far too long. Particular attention is paid to showing how any results apply to the groups in which we are interested. There are also occasional forays into statistical applications of various theorems and techniques. It is hoped that this approach will remove much of the tedium, besides clarifying immediately the central ideas.

The second part covers the more advanced topic of symmetric spaces, again with proofs mainly referenced, so that more time can be spent on the significance of the results. Similarly, much time is spent on applying results directly to the space of positive definite matrices, and to another coset space, the Grassman manifold, which occurs in statistical settings. This treatment is based to a large extent on the classic 'Differential Geometry and Symmetric Spaces' by S. Helgason. Although other statistical applications are explored, the main purpose of this section is to expose the underlying basis from which the non-central distribution theory of multivariate statistics has been derived. One hopes to provide an alternative approach to the basically algebraic path taken by Professor James and colleagues, and to provide a suitable framework for the study of a subject which appears, to someone without the necessary advanced mathematical background, to have been pulled out of a hat.

Notation is a non-trivial problem here, as anyone who has consulted the literature would agree. (Where did I see Lie theory ruefully called 'the study of invariance under change of notation'?) Any generally accepted conventions are used, with Helgason the arbiter, unless he is being inordinately unwieldy or idiosyncratic.

1.3 Groups and Coset Spaces

The multivariate normal density possesses many useful properties of invariance with respect to $G\ell(m,R)$ and $O(m,R)$, and these groups are well known to any student of multivariate statistics. It is of fundamental importance that if $\underset{\sim}{x}$ is an m-vector of observations from a multivariate normal population, with mean vector $\underset{\sim}{\mu}$, variance covariance matrix Σ, positive definite real symmetric, written

$$\underset{\sim}{x} \sim \text{MVN}\ (\underset{\sim}{\mu},\ \Sigma) \qquad , \quad \underset{\sim}{x} \in R^m$$

then $\qquad g\underset{\sim}{x} \sim \text{MVN}\ (g\underset{\sim}{\mu},\ g\Sigma g') \quad \forall\ g \in G\ell(m,R).$

That is, linear combinations of normal variates are again normal, an important property of the multivariate normal density. This invariance also holds under a more general group, the non-homogeneous affine group,

$$\underset{\sim}{x} \to g\underset{\sim}{x} + \underset{\sim}{b} \sim \text{MVN}\ (g\underset{\sim}{\mu} + \underset{\sim}{b},\ g\Sigma g') \quad \forall\ g \in G\ell(m,R) \quad \forall\ \underset{\sim}{b} \in R^m$$

which simply adds a mean shift. Moreover, if

$$\underset{\sim}{x} \sim \text{MVN}\ (\underset{\sim}{o},\ I)$$

where I is the m-by-m identity matrix, $\underset{\sim}{o}$ the o-vector, then

$$h\underset{\sim}{x} \sim \text{MVN}\ (\underset{\sim}{o},\ I) \qquad \forall\ h \in O(m,R)$$

since $hh' = h'h = I$ for every $h \in O(m,R)$. Thus any rotation or orthogonal transformation of a standard normal variate is again a standard normal variate. The density is completely invariant under orthogonal transformations.

These are basic tools of multivariate theory, as is the ability to rotate to canonical form. That is, if

$$\underset{\sim}{x} \sim \text{MVN}\ (\underset{\sim}{\mu},\ \Sigma)$$

then by the Principal Axis Theorem, there exists an $h \in O(m,R)$ such that

$$\underset{\sim}{y} = h\underset{\sim}{x} \sim MVN\ (h\underset{\sim}{\mu},\ \Lambda) \quad \text{where} \quad \Lambda = \begin{pmatrix} \lambda_1 & & o \\ & \lambda_2 \ddots & \\ o & & \lambda_m \end{pmatrix} \quad \text{diagonal.}$$

Thus $\underset{\sim}{y}$ is a vector of independent univariate normal variates, y_i, with variances λ_i. The λ_i are the eigenvalues or latent roots of Σ, and the rows of h are the eigenvectors or principal axes.

The multivariate normal density even enjoys a characterization as the only density whose standardized variates are invariate under $O(m,R)$, (and with finite cumulants). (e.g. Kendall and Stuart (1958) Vol. 2, 3rd ed. p. 365).

These properties are studied in depth in elementary courses in multivariate statistics. However the one-to-one relationship between the coset space $G\ell(m,R)/O(m,R)$ and the space of positive definite symmetric matrices, $M(m,R)$, of which Σ is a member is not often pointed out. By the Principal Axis Theorem, we know that each $p \in M(m,R)$ can be expanded as:

$$p = h\ell h' \quad ,\ h \in O(m,R),\quad \ell \text{ diagonal}$$

$$\therefore\ h'ph = \ell = \ell^{\frac{1}{2}}(\ell^{\frac{1}{2}})'$$

$$\therefore\ p = (h\ell^{\frac{1}{2}})(h\ell^{\frac{1}{2}})' = gg' \quad .$$

But this g is not unique since if $p = gg'$, then $p = (gh)(gh)'$ for any $h \in O(m,R)$ or

$$p = \big(gO(m,R)\big)\big(gO(m,R)\big)'.$$

The representation of $M(m,R)$ as $G\ell(m,R)/O(m,R)$ is one-to-one, since if

$$p = g_1 g_1{}' = g_2 g_2{}' \quad \text{for different} \quad g_1, g_2 \in G\ell(m,R)$$

then $\qquad g_2^{-1} g_1 = g_2'(g_1')^{-1}$

$$= g_2'(g_1^{-1})' = (g_1^{-1} g_2)'$$

but $\qquad (g_2^{-1} g_1)^{-1} = g_1^{-1} g_2$

$$\therefore \left((g_2^{-1} g_1)^{-1}\right)' = g_2^{-1} g_1$$

$$\therefore g_2^{-1} g_1 \in O(m,R)$$

$$\therefore g_2^{-1} g_1 = h \text{ for some } h \in O(m,R)$$

$$\therefore g_1 = g_2 h$$

$$g_2 = g_1 h'$$

$\therefore$ g_1 and g_2 must be in the same coset $g_1 O(m,R)$

$\therefore$ $M(m,R) = G\ell(m,R)/O(m,R)$.

Several tricks in multivariate statistics come from this relationship. For example, any $p \in M(m,R)$ can be decomposed as $p = tt'$ for t upper triangular, since $p = \left(g\ O(m,R)\right)\left(g\ O(m,R)\right)'$ for some g, and any $g \in G\ell(m,R)$ decomposes to $g_1 h$ for g_1 upper triangular, $h \in O(m,R)$.

As useful as these group structures are for the statistician, it is perhaps salutory for us to remember that they do not exist merely for our edification and enjoyment. They are important members of the so-called 'Classical Groups', which have been well studied in the past, but which have been by no means exhausted as sources of new and exciting mathematics. We require several basic definitions and notions, most of which will be well-known, but which are included here for completeness.

A group, G, is a collection of objects possessed of a rule of composition, $\cdot$, (usually omitted), such that:

(i) $g_1 \cdot g_2 \in G$ whenever $g_1, g_2 \in G$

(ii) $\exists\ e \in G$ s.t. $e \cdot g = g \cdot e \quad \forall\ g \in G$

(e is called the identity element of G)

(iii) $\forall\ g \in G$, $\exists\ g^{-1} \in G$ s.t. $g \cdot g^{-1} = g^{-1} \cdot g = e$

(g^{-1} is called the inverse of g)

We define $G\ell(m,R)$ as the set of non-singular real linear transformations of an m-dimensional vector space. Then $G\ell(m,R)$ can be represented as the set of m-by-m matrices with real entries, and non-zero determinant (which implies invertability), and these form a group under ordinary matrix multiplication with identity $I = \begin{pmatrix} 1 & & o \\ & 1 \ddots & \\ o & & 1 \end{pmatrix}$. That is

$$G\ell(m,R) = \{g \mid g = (g_{ij}), \underset{i,j=1,\dots m}{g_{ij}} \in R, \det g \neq o\} .$$

We define $O(m,R)$ as the set of orthogonal transformations of an m-dimensional vector space, or alternatively as the set of transformations leaving the quadratic form $\underset{\sim}{x}'\underset{\sim}{x}$ invariant, which gives:

$$O(m,R) = \{h \in G\ell(m,R) \mid h'h = I\} .$$

This forms a group since, for any h_1, $h_2 \in O(m,R)$,

$$(h_1 h_2)'(h_1 h_2) = h_2' h_1' \; h_1 h_2 = I$$

$$\therefore h_1 h_2 \in O(m,R) \text{ whenever } h_1, h_2 \in O(m,R).$$

$O(m,R)$ obviously has inverses, and the same identity, I, as $G\ell(m,R)$. The orthogonal group is sometimes called the rotation group, since any transformation of $\underset{\sim}{x}$ leaving $\underset{\sim}{x}'\underset{\sim}{x}$ invariant must be a simple rotation of $\underset{\sim}{x}$ about the origin. Since $h'h = I \;\; \forall \; h \in O(m,R)$, we have $\det h = \pm 1$. Thus $O(m,R)$, splits into two disjoint parts, the positive and negative rotations, which have positive and negative determinants respectively. That half with positive determinant is called the special orthogonal group, $SO(m,R)$, and becomes important later on because it is the connected component of the identity. Obviously $O(m,R)$ is a subgroup of $G\ell(m,R)$, as is $SO(m,R)$. Another important subgroup of $G\ell(m,R)$ is the special linear group, $S\ell(m,R)$, with

$$S\ell(m,R) = \{g \in G\ell(m,R) \mid \det g = 1\}.$$

Then $SO(m,R)$ is also a subgroup of $S\ell(m,R)$.

The complex extensions of these groups also come in useful. As in other branches of mathematics, the complex field, being of characteristic zero, adds considerable richness to certain results, particularly when finding roots of equations. It is often convenient to expand to the complex field to solve a problem, and then worry about bringing the result back to the reals. The complex general linear group is naturally enough

$$G\ell(m,C) = \{g = (g_{ij}) \mid \underset{i,j=1,\ldots m}{g_{ij} \in C}, \; \det g \neq o\}$$

and the complex orthogonal group is

$$O(m,C) = \{h \in G\ell m,C) \mid h'h = I\} \quad .$$

$S\ell(m,C)$ and $SO(m,C)$ are completely analogous. Another important group is the unitary group, $U(m)$, where

$$U(m) = \{u \mid \bar{u}'u = u\bar{u}' = I\}$$

and $\bar{u}$ is the complex conjugate of u. In many ways it is the more natural extension of $O(m,R)$. We can define it as the set of transformations leaving the hermitian form $\bar{\underset{\sim}{x}}'\underset{\sim}{x}$ invariant, and we have an isomorphism between the positive definite hermitian matrices $M(m,C) = \{p \mid \bar{p}' = p\}$ and the coset space $G\ell(m,C)/U(m)$. In the same way that $g'g$ is positive definite symmetric for any $g \in G\ell(m,R)$, we have $\bar{g}'g$ positive definite hermitian for any $g \in G\ell(m,C)$.

Occasionally one takes the view that these groups are quite abstract, defined purely as transformations, and possessing mere representations or realizations as matrix groups. This is sometimes useful, but often quite confusing. For example, $O(m,R)$ is sometimes defined as those transformations leaving an arbitrary quadratic form (e.g. $\underset{\sim}{x}'A\,\underset{\sim}{x}$, for $A \in M(m,R)$) invariant. This changes our familiar realization of O(m,R) drastically. We will avoid such abstractions.

The idea of a transitive transformation group may help in visualizing these groups. A group, G, is said to act transitively on a space X if each $g \in G$ is a one-to-one mapping on X (e.g. $\forall\, x \in X,\ g \cdot x \in X$, and is well-defined), and if for any two points $x_1, x_2 \in X$ there always exists a $g \in G$ which will carry x_1 into $x_2 = g \cdot x_1$. Thus X can be covered or generated by applying G to any one point of X. X is said to be homogeneous with respect to G. If we pick any point $x_0 \in X$, then the subgroup, G_0, of G which leaves x_0 fixed is called the isotropy group at x_0, or sometimes the stationary subgroup. Then X can be put in one-to-one correspondence with the coset space G/G_0, since for any $x \in X$ there exists a $g \in G$ such that $g \cdot x_0 = x$. But any member of $g\,G_0$ is another such map, so for each $x \in X$ we can associate the particular coset $g\,G_0$. The path traced out by x under transformations of G_0 is called the orbit of x.

The orthogonal group has been defined as rotations of vectors about their origin. In 3-dimensions, if we take a vector, say of unit length for convenience, and apply O(3,R) to it, it will trace out the sphere, S^2. If we then choose some point, say the 'north pole', $x_0 = (1,0,0)$ as origin, then the rotations leaving x_0 fixed are obviously the rotations about the axis through x_0, or $\left(\begin{array}{c:cc} 1 & 0 & 0 \\ \hdashline 0 & \multicolumn{2}{c}{\multirow{2}{*}{O(2,R)}} \\ 0 & & \end{array}\right)$. So we have O(3,R) acting transitively on S^2 which can be represented as the homogeneous space O(3,R)/O(2,R). Since S^2 is also homogeneous with respect to $SO(3,R)$, we get $S^2 = SO(3,R)/SO(2,R)$. The orbit of a point is a latitude line. Obviously this generalizes to arbitrary dimension, with $S^{m-1} = O(m,R)/O(m-1,R)$ as the hypersphere centred at $\underset{\sim}{o}$, of unit radius, in R^m. We can also consider $O(m,R)$ as acting transitively in the space of straight lines through $\underset{\sim}{o}$ in R^m, projective space P^{m-1}, by identifying the two points on an axis with each line. This concept generalizes in two ways. Consider the space of p-dimensional subspaces in R^m $(m \geqslant p)$, which is also homogeneous with respect to $O(m,R)$. But this time both rotations within a subspace and rotations in its orthogonal complement will leave it invariant, so that isotropy group is $O(m-p,R) \times O(p,R)$. This space is known as the Grassman manifold $G_{p,m-p} = O(m,R)/(O(m-p,R) \times O(p,R))$. The other is the space of p-forms, the sets of p orthonormal vectors in R^m, called the Stieffel manifold $V(p,m)$. The only transformations of $O(m,R)$ leaving p orthonormal vectors fixed are those in the complement of the space spanned by the p vectors. So $V(p,m) = O(m,R)/O(m-p,R)$.

We know $G\ell(m,R)$ acts transitively on $M(m,R)$ under congruence transformation

$$g : S \to g\,S\,g' \qquad g \in G\ell(m,R), \quad S \in M(m,R).$$

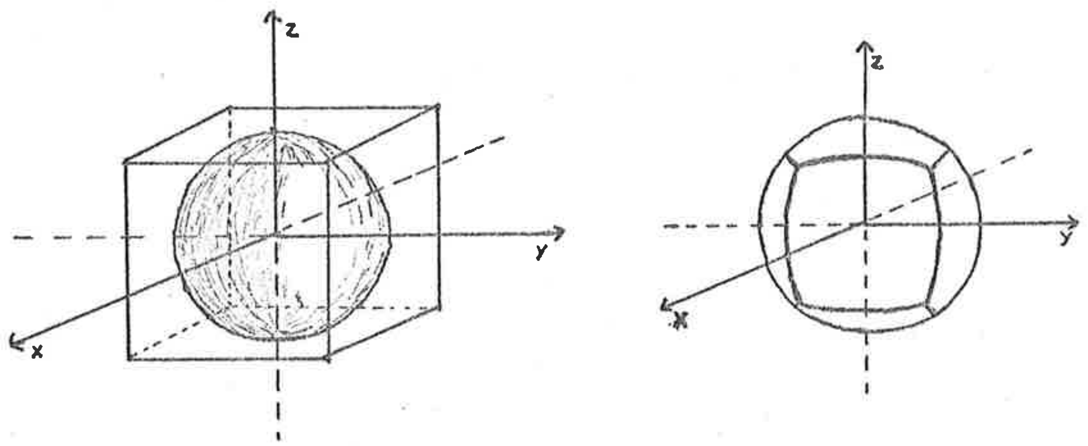
(In other words, the composition rule, $\cdot$, is a congruence transformation, $g \cdot S = g\,S\,g'$.) For, taking any $S_1, S_2 \in M(m,R)$ we can express each as $S_1 = g_1 g_1'$, $S_2 = g_2 g_2'$ for some $g_1, g_2 \in G\ell(m,R)$, and then $g_2 g_1^{-1}$ transforms S_1 into S_2 under congruence transformation. Also, the isotropy group at I in $M(m,R)$ is obviously $O(m,R)$, giving of course $M(m,R) = G\ell(m,R)/O(m,R)$.

1.4 A simple application

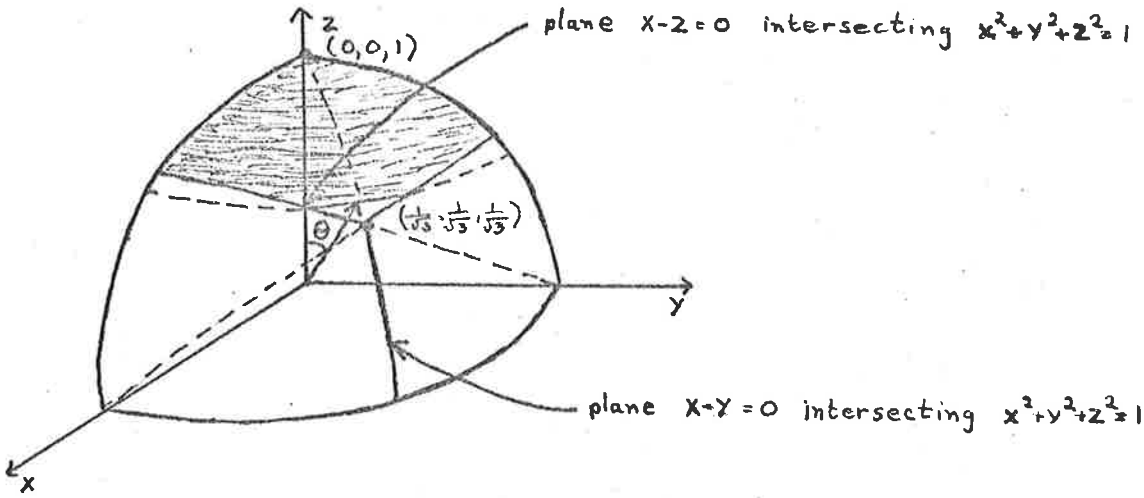
Sometimes just an appreciation of the group symmetries involved in a problem is enough to make the essential points stand out. For example, consider a crystal which is randomly orientated in space. One wishes to determine the probability density for the smallest angle this crystal must be turned to set it upright, or onto one of its faces. Such a problem might arise where crystals were being manufactured or found in different orientations, and one wished to determine whether their orientations were in fact random, or were perhaps subject to other forces.

First, for simplicity, take the crystal to be the unit cube, and randomly orientated in space, centred at $\underset{\sim}{o}$. So the orthogonal group is involved somewhere, since the cube is subject to orientation changing relations. But $O(3,R)$ we know acts transitively on spheres. What is really meant by a random orientation of a cube is that a sphere inside the cube is randomly rotated. Visualize the cube (upright) as containing S^2, with the centre of each face tangent to the sphere. If we pick a point on S^2 as origin, and then rotate S^2 randomly, this point will then be uniformly distributed on S^2. Once this is realized, the rest becomes simple. For obviously, the distance from our roving point to the nearest tangent point will give us the shortest angle the crystal must then be turned to make it upright.

It is a short step from picturing the boxed in sphere to projecting the edges of the cube onto the sphere, along the planes through the respective edges and $\underset{\sim}{o}$. This gives us a curiously marked sphere divided into six equally shaped and sized regions, by segments of great circles.



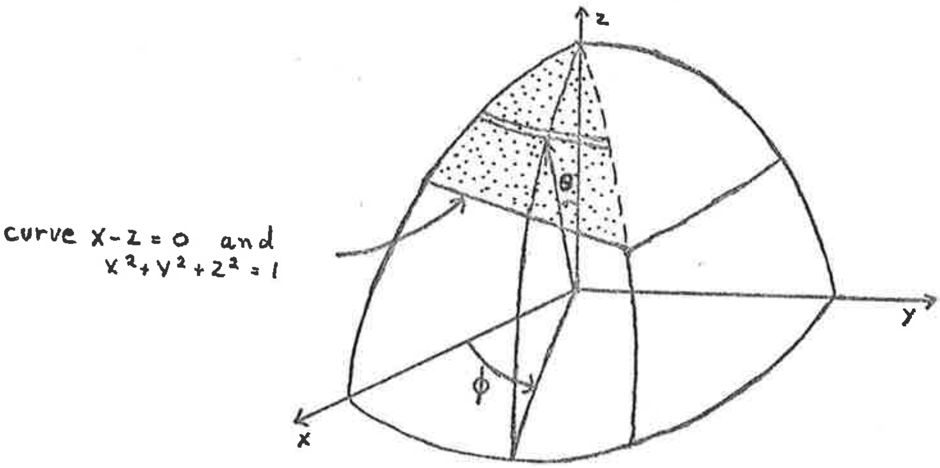
Each portion can be further divided neatly into 4 equal pieces by planes through the tangent point axes, and from various symmetries we need only consider the point as random on one of these. Taking the first quadrant, for example, and looking at the top section, the point has a $\frac{1}{24}$ chance of being on this piece, and if it does land there, then the shortest angle to a tangent face occurs along the path to (0,0,1).



The probability that that angle is between 0 and θ is then the surface area traced out by $0 \leqslant \theta^* \leqslant \theta$. This is obviously quite regular till θ reaches $\pi/4$, and then the rate decreases rapidly because of new edges. The mid-point of the first quadrant is $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ which makes an angle of

$\sin^{-1}\left(\sqrt{\frac{2}{3}}\right)$ or $\cot^{-1}\left(\frac{1}{\sqrt{2}}\right)$ (approximately $54° 44'$) with the three closest tangent faces, and is the maximum angle through which the cube might have to turn.

To calculate the density function, use the following spherical co-ordinate system (r set to one for convenience).



$$x = \operatorname{Sin} \theta \operatorname{Cos} \phi$$
$$y = \operatorname{Sin} \theta \operatorname{Sin} \phi$$
$$z = \operatorname{Cos} \theta$$

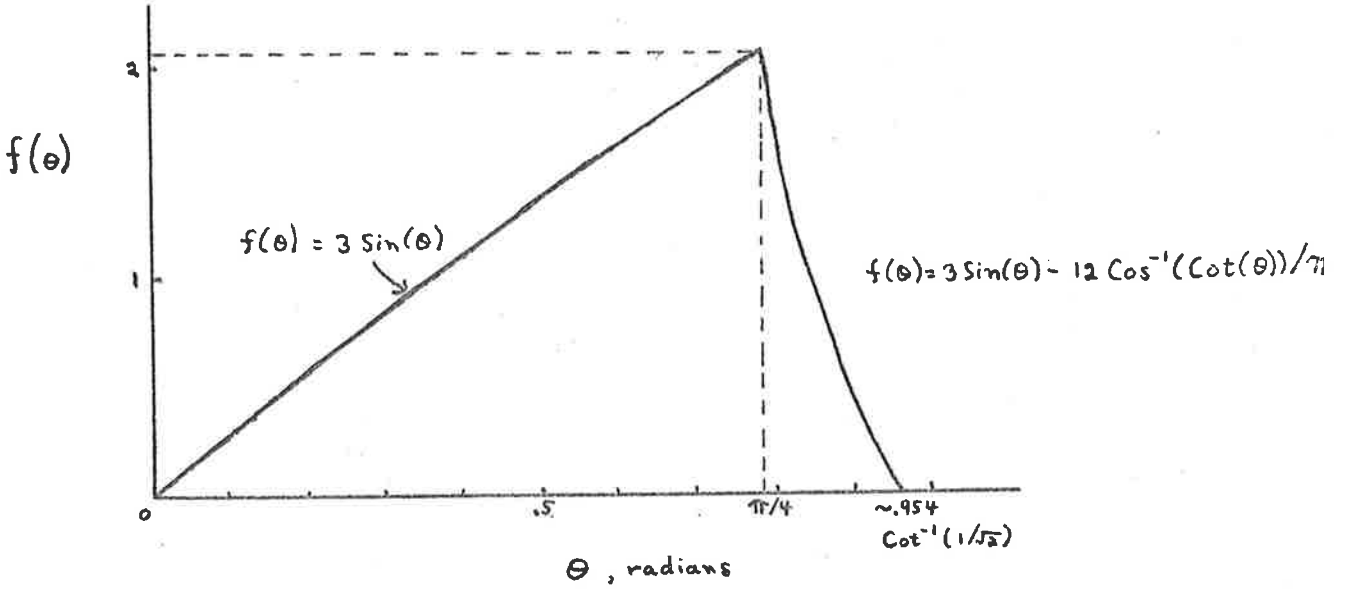
With this co-ordinate system, the infinitesimal area piece is just $\sin \theta \, d \theta \, d \phi$, and the area shaded is $\frac{1}{48}$ the total area of the sphere, (4π). So the density function for $0 \leqslant \theta \leqslant \pi/4$ is just

$$\frac{48}{4\pi} \int_{\phi=0}^{\pi/4} \operatorname{Sin} \theta \, d \theta \, d \phi = 3 \operatorname{Sin} \theta \, d \theta$$

and over the range $\frac{\pi}{4} \leqslant \theta \leqslant \operatorname{Cot}^{-1}\left(\frac{1}{\sqrt{2}}\right)$ is just

$$\frac{48}{4\pi} \int_{\phi = \operatorname{Cos}^{-1}(\operatorname{Cot} \theta)}^{\pi/4} \operatorname{Sin} \theta \, d \theta \, d \phi = (3 \operatorname{Sin} \theta - \frac{12}{\pi} \operatorname{Cos}^{-1}(\operatorname{Cot} \theta) \, d \theta$$

since the curve described by $x-z = 0$, $x^2+y^2+z^2 = 1$ becomes $\operatorname{Cot} \theta = \operatorname{Cos} \phi$, giving $\phi = \operatorname{Cos}^{-1}(\operatorname{Cot} \theta)$ along the one edge. A graph of this function shows its behaviour clearly.



Generalizing to other crystal shapes presents no conceptual problem, once the above is understood, since we will still have the basic spherical structure. The edges of an arbitrary crystal may, when projected onto S^2, give several different types of pieces. Each type will require a separate surface integral, and one will have to keep track of how many there are of each kind, but the basic idea remains the same. A similar technique was used, I believe, in Handscomb (1958) to solve a related problem concerning two randomly 'disorientated' cubes.

1.5 Lie groups and algebras

The most important properties of these matrix groups stem from their very rich topologies. As opposed to the ordinary concept of a group as a set of points which can be perhaps permuted, and which possess the discrete topology of artificial open and closed sets, the matrix groups are quite continuous, in fact analytic. It makes sense to discuss limits and neighbourhoods. One can construct sequences of group members which can approach arbitrarily close to another member, and ask perfectly valid questions concerning concepts such as closed-ness, open-ness, boundedness, compactness, and measurability.

For one thing, the group operations are analytic. If g_1, g_2 belong to a matrix group G, then $(g_1, g_2) \to g_1 \cdot g_2^{-1}$, from $G \times G \to G$, is an analytic map. (This neatly combines the properties $(g_1, g_2) \to g_1 \cdot g_2$ and $g_1 \to g_1^{-1}$ into one.) For another, a matrix group is an analytic manifold. Rigorous definitions of a manifold can be found in any differential geometry text (e.g. Helgason (1962), p.4). This is essentially the notion that if one can break the object up into small enough pieces such that each bit appears to be Euclidean or 'flat', then the object is a manifold. One can piece together a manifold from open subsets, and if u, say, is one of these pieces, then a homeomorphic map, ϕ, exists mapping u onto an open subset of R^m, for some fixed (throughout the manifold) m. Thus associated with each 'piece' is a co-ordinate system in R^m, usually denoted by $(x_1, x_2, \ldots x_m)$. The smoothness of the overlap between competing co-ordinate systems gives the type of manifold. In an analytic manifold, if u_1 and u_2 are intersecting open sets, and ϕ_1, ϕ_2 their respective homeomorphic maps, then $\phi_1 \cdot \phi_2^{-1}$ must be an analytic map from $\phi_2(u_1 \cap u_2)$ to $\phi_1(u_1 \cap u_2)$, which are open subsets of R^m.

Taking $G\ell(m,R)$ then, any $g \in G\ell(m,R)$ is an $m \times m$ matrix with real entries. If this matrix is strung out, row- or column-wise, it turns into a vector in R^{m^2}, giving the homeomorphism between $G\ell(m,R)$ and an open (since det $(g) \neq 0$) set in R^{m^2}. Similarly $G\ell(m,C)$ is an analytic manifold of dimension R^{2m^2}. We can then induce manifold structures on their various subgroups. Taking account of the added relationships among the matrix entries, this gives $O(m,R)$ of dimension $\frac{1}{2}m(m-1)$, $S\ell(m,R)$ of dimension m^2-1, $U(m)$ of dimension m^2, and so on (see e.g. Miller (1972), p.171).

It is this combination of group and topological structures which gives us the rich and rewarding subject of Lie groups - named after Sophus Lie who first studied them around the turn of the century. A Lie group is a group which is also an analytic manifold, and whose group operations are analytic. Thus our matrix groups are Lie groups, since matrix multiplication is obviously analytic. (Actually every Lie group is homeomorphic to a matrix group!) One can consider a Lie group as either a group with a smooth topology and group operations compatible with this topology, or as a nice manifold richly endowed with a compatible group structure. Whichever way, it is a most powerful concept, with far-reaching consequences. For example, it is a remarkable fact that if we have all the requirements for a Lie group, except merely continuous group operations, then this is sufficient to force them to be analytic. This was a long-standing conjecture of Hilbert's (the V problem), which was only proved comparatively recently (see Montgomery-Zippin, 1955). Even an isomorphism between two Lie groups, which is at least continuous, must necessarily be analytic. If their topologies coincide then they are identical. Also, any analytic subgroup (A subgroup and submanifold) of a Lie group must always be another Lie group, even if it is not closed, or even a subspace.

The special nature of Lie groups can be further appreciated by considering a Lie group near its identity, e. Restricting to such a region is no real bother, because we are dealing with an analytic group. A transformation from e to any other member carries the whole topology with it. Quoting Cohn (1957), 'No point is more important than any other point'. Whatever the group looks like near e is what it looks like anywhere else, and one can shift about as one pleases. For this reason one normally studies only the behaviour of a Lie group near its origin.

This brings us to that curious entity, the Lie algebra. It is essentially the tangent space to the manifold at e. But because a Lie group is more than just an ordinary manifold, this tangent space possesses many exceptional properties, not least of which is its well-defined multiplication, the Lie product. The Lie algebra of a group contains almost all the information required to identify and construct the group, and is also a vector space, so is in many ways easier to handle than the original group. Often Lie algebras are defined and studied independently of Lie groups, and only later is the tangent space at e of a Lie group shown to possess the required properties of this algebra. But we shall begin with tangent spaces, so first we need to know what they are.

The most usual approach to tangent spaces comes from a consideration of the space of real analytic functions, $C^w(M)$, on a manifold, M, and the set of differentiable curves on M, parametrized by $x_i(t)$, $i = 1,\ldots,m$. Take any curve, $\gamma(t)$, on M such that $\gamma(o) = p$, $p \in M$. Then for $\phi\cdot\gamma(t) = \{x_i(t)\}$,

$$\left.\frac{dx_i}{dt}\right|_{t=o} = \lambda^i$$

and the derivative of any function along $\gamma(t)$ at $t = o$ will be:

$$\left.\frac{d}{dt} f\,(x(t))\right|_{t=o} = \sum_{i=1}^{m} \lambda^i \left.\frac{\partial f}{\partial x_i}\right|_p$$

The operator $\sum_{i=1}^{m} \lambda^i \partial/\partial x_i$ is defined to be a tangent vector at p. The above is by no means rigorous. Care must be taken to keep straight when one is working on M or R^m, and numerous ϕ's and ϕ^{-1}'s are required for exactness. The $\partial/\partial x_i$'s are accepted shorthand for the map

$$f \to \frac{\partial(f\cdot\phi^{-1})}{\partial x_i}\cdot\phi \quad .$$

One then shows that the set of tangent vectors of differentiable curves through p form a vector space, M_p, with basis $\{\partial/\partial x_i,\ i = 1,\ldots,m\}$. Each tangent vector is an operator on $C^w(M)$, taking an analytic function defined on M into R. Every $p \in M$ has a tangent space. A vector field is a collection of tangent vectors, one from each tangent space on M, picked so that each collection is a derivation on $C^w(M)$. That is, X is a vector field if:

$$X(\alpha f+\beta g) = \alpha Xf + \beta Xg \qquad \alpha,\ \beta \in R$$

$$X(fg) = (Xf)g + f(Xg) \qquad f,\ g \in C^w(M).$$

Note that for $f \in C^w(M)$, $Xf \in C^w(M)$ also, for Xf is defined as $(Xf)(p) = X|_p\cdot f$.

The set of vector fields is denoted by $\mathcal{D}^1(M)$. Conversely, one could define vector fields as derivations on $C^w(M)$, and then produce tangent spaces by restricting $\mathcal{D}^1(M)$ to individual points in M. This whole process is sometimes called fibring, in modern literature. Vector fields were originally known as infinitesimal transformations.

Where one has vector spaces, one expects to find duals, and the dual of a tangent space is called the co-tangent space M_p^*, with a basis naturally defined by dx_i, $i = 1,\ldots,m$, such that $dx_i \frac{\partial}{\partial x_j} = \delta_j^i$ (Kronecher Delta), and for any $f \in C^w(M)$, df is a map from $x \in M_p$ into $x(f) \in R$, so df is a vector in M_p^*. The differential 1-forms are collections of co-tangent vectors, where now if w is a 1-form, and X a vector field, w X is a function on $C^w(M)$, where $wX|_p = w|_p\, X|_p$. These are denoted by $\mathcal{D}_1(M)$.

The notation of super- and sub-scripts is quite standard, having roots in the old differential geometry concepts of contra- and co-variant transformations and tensors. It is cross-products of the $\mathcal{D}^1(M)$'s which build up the tensor fields. One of these, the exterior differential forms, are important statistically. These are the set of alternate multilinear maps of $\mathcal{D}^1(M)\times\ldots\times\mathcal{D}^1(M)$ into R, for any number of crosses. Adding a multiplication to these, (the 'wedge' product), we can produce a maximal form which can be used as a volume measure on various groups and coset spaces of use in statistics (see James (1954)).

Returning to Lie groups, one might hope that the composition allowed on such manifolds would give an added richness to tangent spaces. If $\gamma_1(t)$ and $\gamma_2(t)$ are curves through e (at $t = o$), then $\gamma_1(t)\cdot\gamma_2(t)$ is another curve through

e, with tangent vector $X_1 + X_2$ for X_i the tangent at e of $\gamma_i(t)$. But we get much more than this. If X_1 and X_2 are tangent vectors at M_e, corresponding to $\gamma_1(t)$, $\gamma_2(t)$ as before, then $X_1X_2 - X_2X_1$ forms another tangent, denoted $[X_1, X_2]$, the commutation or Lie product, which corresponds to the curve $\gamma_1(\sqrt{t})\gamma_2(\sqrt{t})\gamma_1(\sqrt{t})^{-1}\gamma_2(\sqrt{t})^{-1}$. Alternatively, one can show a one-to-one relationship between M_e and the set of left-invariant vector fields (those vector fields invariant under dL_ρ where $L_\rho: g \to \rho g$ for $g, \rho \in G$). If $X \in M_e$ then there is only one left-invariant vector field $\tilde{X}$ with $\tilde{X}_e = X$. If $\tilde{X}_1$, $\tilde{X}_2$ are two left-invariant vector fields, then $[\tilde{X}_1, \tilde{X}_2]$ is another, with $[\tilde{X}_1, \tilde{X}_2]|_e = [X_1, X_2]$. Over a small enough region, we can express $\tilde{X}_1$ and $\tilde{X}_2$ as

$$\tilde{X}_1 = \sum_{i=1}^{m} \xi^i \frac{\partial}{\partial x_i} \quad \text{and} \quad \tilde{X}_2 = \sum_{i=1}^{m} \eta^i \frac{\partial}{\partial x_i}$$

for ξ^i, η^i analytic functions. Then

$$(\tilde{X}_1\tilde{X}_2)\, f = \sum_{i,j=1}^{m} \xi^i \frac{\partial \eta^j}{\partial x_i} \frac{\partial f}{\partial x_j} - \xi^i \eta^j \frac{\partial^2 f}{\partial x^i \partial x^j}$$

and the second term spoils $\tilde{X}_1\tilde{X}_2$ being another vector field. However $\tilde{X}_1\tilde{X}_2 - \tilde{X}_2\tilde{X}_1$ gets rid of this annoying factor. On M_e, obviously $[X,X] = 0$, and one can verify the Jacobi identity $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$ for all $X,Y,Z \in M_e$.

It does not really matter whether one considers a Lie algebra as the left-invariant vector fields, or as the tangent space at e of G, although the latter view is handier for our purposes.

Any vector space over a field, possessed of a bilinear product rule, $[\,,\,]$, such that $[X,X] = 0$, and the Jacobi identity, is called a Lie algebra. Often Lie algebras are defined in vacuo, and only later is the important connection with Lie groups brought out. This approach seems to obscure the fundamental and remarkable relationships between Lie groups and their algebras, which can just be touched on here. But first we need the exponential map, exp.

A Lie algebra is related to its group via the exponential map, which maps a neighbourhood of $\underset{\sim}{o}$ in the algebra, denoted as $\mathcal{G}$, onto a neighbourhood of e in G, in a one-to-one fashion. This process hinges on the so-called 1-parameter subgroups of a Lie group, which are curves $\gamma(t)$ $\{\gamma(o) = e\}$, such that $\gamma(t_1)\gamma(t_2) = \gamma(t_1+t_2)$. This gives an isomorphism between the real line and such curves through e in G. It is part of Lie theory to prove that such curves not only exist, but cover a neighbourhood of e, and that every $X \in \mathcal{G}$ is a tangent vector to one 1-parameter subgroup through e. From the definition of 1-parameter subgroups one can show that $\exp : X \to \gamma_X(t)|_{t=1}$ is the regular exponential map $\exp(X) = \sum_{n=o}^{\infty} \frac{1}{n!} X^n$. Basically, one shows that $\exp t X$ obeys

$$\frac{d}{dt}(\exp t X) = (\exp t X)X$$ giving the solution of the differential equation as $\exp t X = I + tX + \frac{1}{2}t^2X^2 + \ldots$ (for proof see Helgason (1962), p. 101).

It is a simple matter to prove that for any m-by-m matrix X, the series $\exp(X)$ is always well-defined and convergent. Some other useful properties are $\exp(A+B) = \exp A \cdot \exp B$ iff $AB = BA$, $\exp((s+t)A) = \exp(sA + tA)$, $\det(\exp A) = \exp(\text{trace } A)$ for det = determinant, and $\exp(X') = (\exp X)'$ for X' the transpose of X.

With exp we can now explicitly calculate the Lie algebras of our various groups and subgroups. By definition, $G\ell(m,R) = \{y | \det y \neq 0\}$, so if x is a member of the Lie algebra of $G\ell(m,R)$, denoted by $\mathcal{L}G\ell(m,R)$, then $\exp x \in G\ell(m,R$

So:

$$\det(\exp x) = \det(I + x + \tfrac{1}{2}x^2 + \ldots) \neq 0.$$

But this is true for any x at all. Therefore $\mathcal{L}G\ell(m,R)$ is composed of every $m\times m$ matrix, invertible or otherwise. (This is just a generalization of the mapping $x \to e^x$ of the whole real line into the positive reals.) Similarly, any $m\times m$ complex matrix is a member of the Lie algebra, $\mathcal{L}G\ell(m,C)$ of $G\ell(m,C)$.

If $x \in \mathcal{L}S\ell(m,R)$, the Lie algebra of the special linear group, then $\det(\exp(x)) = 1$. But $\det(\exp(x)) = \exp(\text{trace } x)$, and $\exp(\text{trace } x) = 1$ implies trace $x = 0$. So:

$$\mathcal{L}S\ell(m,R) = \{x \in \mathcal{L}G\ell(m,R) \mid \text{trace } x = 0\}.$$

The orthogonal group has a more interesting algebra, for if $x \in \mathcal{L}O(m,R)$, then

$$(\exp x)(\exp x)' = I$$

$$(\exp x)' = (\exp x)^{-1} = \exp(-x)$$

$$\exp(x') = \exp(-x)$$

$$x' = -x$$

$$x + x' = 0 \quad .$$

So x is skew-symmetric. We can easily verify this result for $\mathcal{L}O(2,R)$. By the above, all elements of the form $\begin{pmatrix} & -t \\ t & \end{pmatrix} \in \mathcal{L}O(2,R)$ for $t \in R$. Then:

$$\exp\begin{pmatrix} & -t \\ t & \end{pmatrix} = I + \begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} -t^2 & 0 \\ 0 & -t^2 \end{pmatrix} + \tfrac{1}{3!}\begin{pmatrix} 0 & t^3 \\ -t^3 & 0 \end{pmatrix} + \ldots$$

$$= \begin{pmatrix} 1-\tfrac{1}{2}t^2+\tfrac{1}{4!}t^4\ldots & -t+\tfrac{1}{3!}t^3\ldots \\ t-\tfrac{1}{3!}t^3+\ldots & 1-\tfrac{1}{2}t^2+\tfrac{1}{4!}t^4\ldots \end{pmatrix}$$

$$= \begin{pmatrix} \text{Cos } t & -\text{Sin } t \\ \text{Sin } t & \text{Cos } t \end{pmatrix} .$$

Similarly $\mathcal{L}O(m,C)$ is composed of complex skew-symmetric matrices, and $\mathcal{L}U(m)$ of skew-hermitian matrices (i.e. $x + \bar{x}' = 0$ for $\bar{x}$ complex conjugate of x).

The relationship between a group and its algebra is almost all one could wish for. I say 'its' algebra because every Lie group has an algebra, every Lie algebra a group, and there is only the one algebra associated with each group, and connected via exp. In fact, if two groups have the same algebra, then they are at least locally isomorphic. For example, the space of real numbers, R, is a group under addition, and R/Z or R mod the integers Z, is another additive group, but both are identical near e, and so have the same Lie algebra, even though R/Z is a wrapping of R. Also, $\mathcal{L}O(m,R)$ is identical to $\mathcal{L}SO(m,R)$, since SO(m,R) is actually just the connected component of e in O(2,R), so near e the two groups are indistinguishable. This property of Lie groups is often useful in showing local isomorphisms between seemingly unrelated Lie groups.

Another remarkable and important property is the relationship between Lie subgroups and subalgebras. B is a subalgebra of G if $\forall\ b_1, b_2 \in B$, $[b_1, b_2] \in B$. Corresponding to normal subgroup theory, if B is an ideal so $[b,g] \in B\ \forall\ g \in G$, then G/B is another subalgebra, and $G = B \oplus G/B$. If an algebra contains a subalgebra, then there is a Lie subgroup of the large group which corresponds to the subalgebra under exp. Also, any subgroup, B, of G will generate a Lie algebra B which will be a subalgebra of G. It is by no means obvious that this should happen, but the fact that it does helps to emphasize the special position of Lie group theory in ordinary group theory. As we would now expect, if B is an ideal, then exp B is a normal subgroup, exp B/G another group and $G = G_1 \times G_2 = \exp B \times \exp B/G$.

The above facts and results from Lie theory can be found in any text on Lie groups, such as Chevalley (1946), Pontrjagin (1958) or Cohn (1957) - a particularly fine little monograph.

The vector product of two Lie groups corresponds to the vector addition of their algebras. Thus if one could classify all Lie algebras, one could then essentially classify (up to local isomorphisms) all Lie groups (and vice-versa of course). E. Cartan set himself this task, and did succeed in classifying the special set of semi-simple groups this way, by decomposing their algebras into essential building blocks, the simple algebras (those with no proper ideals). We shall study this work later, since it is closely connected with symmetric space theory.

To recap then, we see that each group, G, possessed of a reasonable amount of topology, will have associated with it a flat vector space which possesses a rather curious composition rule, the Lie product, which is closely related to conjugation on the group $(g \to g_1 g g_1^{-1})$. The addition rule corresponds, to a first approximation, with multiplication on the group, and the Jacobi identity is just a reformation of the associative law on the group.

Some of this structure will now be used in separating various components in some likelihood ratio tests.

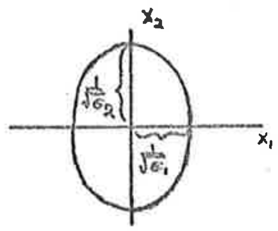
CHAPTER 2

STATISTICAL APPLICATION

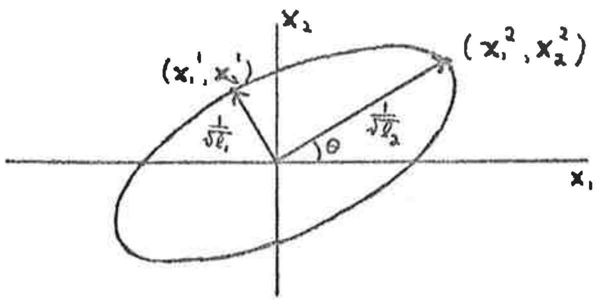
2.1 General

One can put a handle on the positive definite symmetric matrices by giving them a geometric representation, which is often of great help in understanding their structure. To each positive definite symmetric matrix, Σ, m-×-m, one associates the ellipsoid defined by the equation $\underset{\sim}{x}'\Sigma\underset{\sim}{x} = 1$, where $\underset{\sim}{x}' = (x_1, x_2, \ldots, x_m) \in R^m$. The lengths of the principal axes of the ellipsoid are the inverse square roots of the eigenvalues of Σ, and their directions give the eigenvectors. In two dimensions this becomes quite simple. For $\Sigma = \begin{pmatrix} \sigma_1 & o \\ o & \sigma_2 \end{pmatrix}$, say, then

$$(x_1, x_2)\begin{pmatrix} \sigma_1 & o \\ o & \sigma_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 1 \quad \text{gives}$$



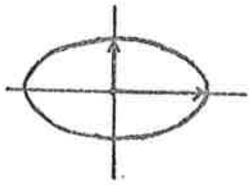
For Σ arbitrary, $\Sigma = \begin{pmatrix} \sigma_1 & \sigma_3 \\ \sigma_3 & \sigma_2 \end{pmatrix}$, $x'\Sigma\, x = 1$ gives



where ℓ_1, ℓ_2 are the eigenvalues of Σ, and $\sqrt{\ell_1}\begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix}$ and $\sqrt{\ell_2}\begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$ are the eigenvectors of Σ. In two dimensions then,

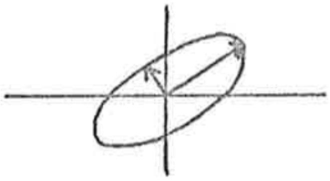
$$H = \begin{pmatrix} \sqrt{\ell_1}\; x_1^1 & \sqrt{\ell_2}\; x_1^2 \\ \sqrt{\ell_2}\; x_2^1 & \sqrt{\ell_2}\; x_2^2 \end{pmatrix} = \begin{pmatrix} \text{Cos}\ \theta & \text{Sin}\ \theta \\ -\text{Sin}\ \theta & \text{Cos}\ \theta \end{pmatrix} \in O(2,R)$$

for some θ, the angle the ellipsoid is rotated from standard form. Diagonal matrices give ellipses in standard form, but matrices with non-zero covariances are rotated about the origin. We can partition $M(m,R)$ into families where within each family the matrices all have the same eigenvalues or latent roots. In two dimensions then, if a family contained



it also contains all possible rotations of the ellipse about the origin.

Conversely, $M(m,R)$ could be partitioned into families where the members have the same eigenvectors or principal axes. A family of this type would then be all dilations and shrinkages; e.g.

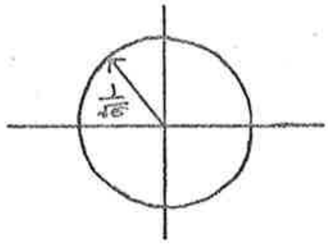


plus any other ellipse with the major and minor axes in the same directions.

A visualization of all possible ellipsoids in m-dimensions, centred at the origin, will then to a limited degree give one a mental image of $M(m,R)$.

Unfortunately the 2-dimensional picture is not always adequate. For example, $O(2,R)$ is a commutative (abelian) group. Rotating $\Sigma_{2\times 2}$ by θ_1, then θ_2 gives the same result as rotating by θ_2, then θ_1 (namely a rotation by the angle $\theta_1 + \theta_2$, which can be easily verified). In higher dimensions this does not happen, since $O(m,R)$ is not commutative for $m > 2$.

Still, this simple picture of $M(m,R)$ is often quite useful in multivariate hypothesis testing, and in dealing with $M(m,R)$ in general, since it often clarifies both questions and answers, besides pinpointing essential difficulties in certain problems. One of these concerns the behaviour of Σ if it has two or more equal roots. It is easy to forget that this has grave bearing on the principal axes, unless one has a very firm picture of $M(m,R)$. For, if $\Sigma = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}$, then $x'\Sigma x = 1$ gives



which is of course a circle, with principal axes in every direction. Using sampling statistics then, the closer two roots become the less information one has about the position of their respective eigenvectors. A zero root will mean that the ellipsoid takes up a smaller dimension, with the missing direction completely unknown.

One important question in multivariate analysis deals with deciding on equality of population variance-covariance matrices for different samples. Sometimes, especially in morphological work, it becomes important to compare not only whole variance-covariance matrices, but also various components of them. For example, in recent years [Blackith and Reyment, (1971)] it has become accepted practice to label eigenvectors of a sample variance-covariance matrix, when derived from biological measurement data, as various growth factors, such as size or shape. It is not the statistician's job to criticize the biologist's reasons for such labelling, but perhaps it is our job to point out the mathematical difficulties inherent in eigenvector/value techniques, so that such methods can be used with necessary care and caution.

Since the first eigenvector generally turns out to have positive entries, it has been labelled the size factor. This does seem reasonable in animal or plant data, since there is no a priori reason why the largest eigenvalue should be produced by a positive eigenvector, and it could be argued that it then must be measuring some large overall characteristic, most probably size. Of course this also occurs if the correlation terms are positive. One is, though, presuming that a size factor, if it exists, will necessarily have the largest variance of all. This may in fact be justifiable. However it does seem important to keep these assumptions in view.

There is also a certain arbitrariness about the choice of the diagonalizing matrix. For S a sample variance-covariance matrix, $S = HLH'$, $H \in O(m,R)$, L diagonal, elements $\ell_1, \ldots \ell_m$; $\ell_i \geqslant \ell_{i+1}$, then H has been chosen to give this ordering of the roots ℓ_i. Permuting the roots amounts to permuting the columns of H. Group theoretically, one can permute the elements of L by matrices of the form

$$k = \begin{bmatrix} 1_1 & & & & 0 \\ & \ddots & & & \\ & & \begin{matrix} 0 & -1 \\ 1 & 0 \end{matrix} & & \\ 0 & & & \ddots & \\ & & & & 1 \end{bmatrix},$$

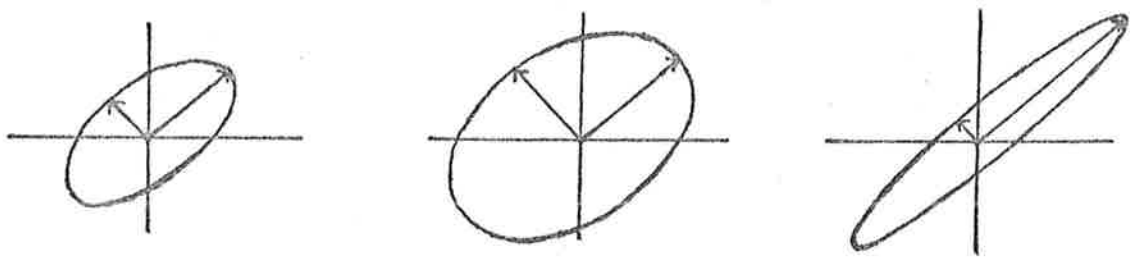
which form a subgroup K of $O(m,R)$. (Note this is a discrete subgroup of $O(m,R)$, with no smooth topology - it is not a Lie subgroup.) That is kLk' permutes the elements of L, $\forall\, k \in K$. Since $S = HLH' = Hk'kLk'kH' = (Hk')kLk'(Hk')'$ we see that Hk' also diagonalizes S. This is immaterial of course, since the same eigenvector remains related to the same latent root, but it is nice to be aware of what one is actually doing. Geometrically, if one imagines an hyperellipsoid in m-space, then this amounts to deciding to which standard form we wish to rotate. That is, which root goes with which axis. Choosing this by size, as is normally done, we then see that there may be different ways of rotating to this standard form. If $HLH' = S$, then for M diagonal, with ± 1 along the diagonal, we have

$H(MLM)H' = S = (HM)L(HM)'$. Thus M may reverse the signs of each entry in an eigenvector. Obviously if $h'Sh = \lambda$, then $(-h)'S(-h) = \lambda$ too. When H is calculated by a computer program, it has always had some arbitrary rule used to decide on the sign. Usually (and the one we use here) this is to set the ii^{th} elements positive, putting H close to I.

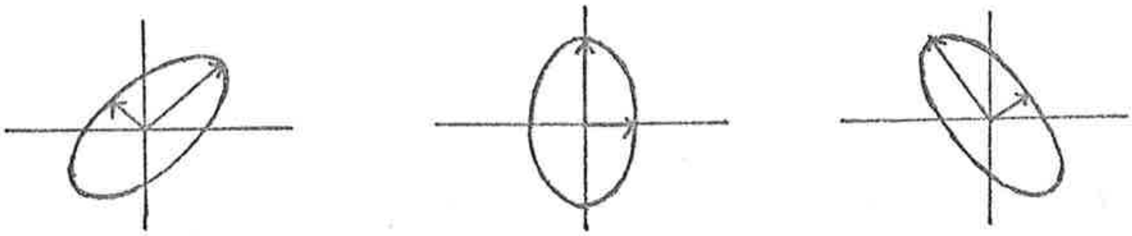
2.2 Eigenvalue-eigenvector tests of equality

In multivariate statistics, likelihood ratios tend to test several hypotheses at once, all embodied in the original null hypothesis, and discrepancies in any one of which can cause rejection of the whole. If the null hypothesis is rejected, one might wish to see just which factors are 'to blame', if this is at all possible. By utilizing the Lie group structure of $O(m,R)$, it was found possible to partition the likelihood ratio for testing equality of two population Wishart matrices into separate tests for equality of eigenvalues and eigenvectors. Such tests are of real value in morphological research in particular, where much work deals with comparisons of sexes or species.

Suppose we have S_1 and S_2 being two observations from the Wishart populations $W(n,m,\Sigma_1)$ and $W(n,m,\Sigma_2)$ respectively, with the same dimensions, m, and degrees of freedom, n. One wishes to test the null hypothesis, H_0: $\Sigma_1 = \Sigma_2$. This is equivalent to testing that both the eigenvalues and eigenvectors are identical. The ellipsoidal representation of the positive definite symmetric matrices, $M(m,R)$, is invaluable in understanding just how two covariance matrices can differ. They may have the same orientations or eigenvectors, but different latent roots or principle axes, i.e.

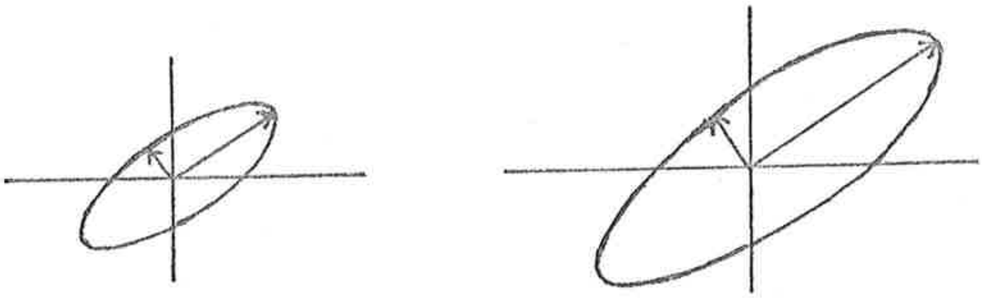


or have the same latent roots, but different axes, i.e.



or be quite different in both respects. Another perhaps interesting possibility is that the eigenvalues be in the same proportion, with one matrix being just a dilation of the other, and with equal eigenvectors,

i.e.



This then implies a simple scalar difference, $\Sigma_1 = k\Sigma_2$. One imagines this would be of interest in morphological work since it would seem to imply only a size difference in growth patterns. Even a simple scalar difference between the roots, without equal orientations, might have a physical interpretation in some problems.

One should appreciate why such questions pose real difficulty in multivariate work. For although we can draw a precise ellipsoid for an actual sample, the true population ellipsoid is simply somewhere around the sample, and we have no way of even drawing in percentile boundaries. One can imagine situations where likely population ellipsoids for both samples might have the same roots, but different eigenvectors, yet at the same time another pair of likely candidates might have equal eigenvectors but different roots, and where overall we reject $H_0 : \Sigma_1 = \Sigma_2$. One hopes this does not happen too often, since we have as yet no way of testing for this. The marginal density for the eigenvalues involves a hypergeometric function of matrix argument [James (1964)] which makes even calculating the maximum likelihood estimates (m.l.e.) for the population roots to all intents and purposes impossible, much less calculating joint m.l.e.'s.

The following test deals with this dilemma by compromise. We test both sub-hypotheses at once as best we can, by essentially looking at deviations of both factors from the m.l.e. for Σ, given $\Sigma_1 = \Sigma_2 = \Sigma$. This must be a rather crude test, no doubt, but it is at least a starting point, and is obviously better than the other known tests for approximately equal sample sizes, as we shall see later. If we start with $S_1 \sim W(n,m,\Sigma_1)$ and $S_2 \sim W(n,m,\Sigma_2)$, then the m.l.e. for Σ_1 is S_1 and for Σ_2 is S_2. But under $H_0: \Sigma_1 = \Sigma_2 = \Sigma$, the m.l.e. of Σ is $(S_1 + S_2)/2$, and the likelihood ratio criterion, λ, for testing H_0 is:

$$\lambda = \left(\frac{(\det S_1)^{\frac{1}{2}} \; (\det S_2)^{\frac{1}{2}}}{\operatorname{Det}\left(\frac{S_1 + S_2}{2}\right)} \right)^n$$

(see e.g. Anderson (1958)).

We also know that $S_1 = H_1 L_1 H_1'$ and $S_2 = H_2 L_2 H_2'$ for $H_i \in O(m,R)$ the eigenvectors, and L_i diagonal, the latent roots of S_i. By putting the roots in decreasing order of magnitude, and requiring the diagonal elements of H_i to be positive, we determine the H_i and L_i uniquely. By this choice, the H_i then have determinent 1. Using this fact, and the identity $\operatorname{Det}(A) \cdot \operatorname{Det}(B) = \operatorname{Det}(AB)$, we can derive a new expression for $\lambda^{1/n}$.

$$\lambda^{\frac{1}{n}} = \frac{(\det H_1 L_1 H_1')^{\frac{1}{2}} (\det H_2 L_2 H_2')^{\frac{1}{2}}}{\det \left(\frac{H_1 L_1 H_1' + H_2 L_2 H_2'}{2}\right)}$$

$$= \frac{(\det L_1)^{\frac{1}{2}} (\det L_2)^{\frac{1}{2}}}{\det (H_1') \det \left(\frac{H_1 L_1 H_1' + H_2 L_2 H_2'}{2}\right) \det H_2}$$

$$= \frac{(\det L_1)^{\frac{1}{2}} (\det L_2)^{\frac{1}{2}}}{\det \left(\frac{L_1 (H_1'H_2) + (H_1'H_2) L_2}{2}\right)} \quad . \tag{2}$$

We know that under the null hypothesis at least, H_1 is approximately equal to H_2, so that $H_1'H_2$ will be near the identity, I_m. But once in the neighbourhood of I_m, we can start to apply Lie algebra approximations to group members, via the exponential map. In this case, there exists an X such that

$$H_1'H_2 = \exp(X) = I + X + \tfrac{1}{2}X^2 + \ldots \quad .$$

But the Lie algebra of $O(m,R)$ is the space of skew-symmetric matrices, so

$$X = \begin{array}{c c} & \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{array}{c} i \\ \\ j \end{array} & \left[\begin{array}{ccccc} & \vdots & & \vdots & \\ \ldots & 0 & \ldots & x_{ij} & \ldots \\ & \vdots & & \vdots & \\ \ldots & -x_{ij} & \ldots & 0 & \ldots \\ & \vdots & & \vdots & \end{array}\right] \end{array}$$

with zeros along the diagonal, and $x_{ij} = -x_{ji}$.

For small x_{ij} then,

$$H_1'H_2 = \exp(X) \approx \left[\begin{array}{ccccc} & \vdots & & \vdots & \\ \ldots & 1-\frac{1}{2}\sum_{k=1}^{m} x_{ik}^2 & \ldots & x_{ij} & \ldots \\ & \vdots & & \vdots & \\ \ldots & -x_{ij} & \ldots & 1-\frac{1}{2}\sum x_{jk}^2 & \ldots \\ & \vdots & & \vdots & \end{array}\right] = H^*$$

We are going to substitute H^* for $H_1'H_2$ in (2).

The trouble with this approximation is that the further $H_1'H_2$ is from I, the worse the approximation may become, and this has nothing to do with the sample size n. However, over fairly broad regions of $O(m,R)$ this approximation should be reasonably good. In actual practice, when the x_{ij} are estimated, we can check the accuracy by seeing what effect this change has on λ. Appreciable changes would of course make one more cautious of drawing definite conclusions.

Also, put

$$L_1 = \begin{bmatrix} k_1 & & & \\ & k_2 & & \\ & & \ddots & \\ & & & k_m \end{bmatrix}, \quad L_2 = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_m \end{bmatrix}.$$

Making these substitutions in (2), one obtains

$$\lambda^{\frac{1}{n}} \approx \frac{\prod_{i=1}^{m} (k_i p_i)^{\frac{1}{2}}}{(\frac{1}{2})^m \det (L_1 H^* + H^* L_2)} \tag{3}$$

The denominator term, when expanded out, dropping all terms in the x_{ij}'s of degree greater than two, becomes:

$$\det (L_1 H^* + H^* L_2) \approx$$

$$2^m \sum_{i=1}^{m} \left(\frac{k_i+p_i}{2}\right) + 2^{m-2} \sum_{i<j} x_{ij}{}^2 (k_i-k_j)(p_i-p_j) \prod_{\substack{n=1 \\ n\neq i \\ n\neq j}}^{m} \left(\frac{k_n+p_n}{2}\right).$$

This can be seen by noting that the dominant diagonal terms are $(k_i + p_i)\left(1 - \frac{1}{2} \sum_{k=1}^{m} x_{ik}{}^2\right)$, and the off-diagonal terms are $x_{ij}(k_i + p_j)$ (i.e. the skew-opposite term is $-x_{ij}(k_j + p_i)$) so there are no terms of degree 1 left, the only degree 2 terms from the diagonal are of the form $-\ x_{ij}{}^2 \prod_{n=1}^{m} (k_n + p_n)$ and the squared off-diagonal terms are $x_{ij}{}^2 \left(\prod_{\substack{n=1 \\ n\neq i,j}}^{m} (k_n + p_n)\right)(k_i + p_j)(k_j + p_i)$. The third degree terms are almost non-existent, even if more terms are included. None arise from the diagonal, since X^3 has zeros along the diagonal. Off diagonal terms in X^3 add only higher degree terms in the determinant. The $x_{ij}x_{k\ell}$ terms in the off-diagonal, which we also omitted, only add terms of degrees 4 or more to the determinant. However, some cubic terms do arise, with terms

$$x_{i_1 j_1} x_{i_2 j_2} x_{i_3 j_3} \left((k_{i_1}+p_{j_1})(k_{i_2}+p_{j_2})(k_{i_3}+p_{j_3}) - (p_{i_1}+k_{j_1})(p_{i_2}+k_{j_2})(p_{i_3}+k_{j_3})\right)$$

which tend to cancel each other under H_0 at least.

If we let

$$L_1 = \begin{bmatrix} \ell_1+d_1 & & & \\ & \ell_2+d_2 & & \\ & & \ddots & \\ & & & \ell_m+d_m \end{bmatrix}, \quad L_2 = \begin{bmatrix} \ell_1-d_1 & & & \\ & \ell_2-d_2 & & \\ & & \ddots & \\ & & & \ell_m-d_m \end{bmatrix}$$

as a way of representing differences between the roots, then (3) becomes

$$\lambda^{\frac{1}{n}} \approx \frac{\prod_{i=1}^{m} (\ell_i^2 - d_i^2)^{\frac{1}{2}}}{\prod \ell_i + \frac{1}{4}\sum_{i<j} x_{ij}^2 (k_i-k_j)(p_i-p_j) \underbrace{\ell_1 \ldots \ell_m}_{\text{missing } \ell_i,\ \ell_j}} .$$

Dividing through by $\prod_{i=1}^{m} \ell_i$ gives:

$$\lambda^{\frac{1}{n}} \approx \frac{\prod_{i=1}^{m} \left(1 - \frac{d_i^2}{\ell_i^2}\right)^{\frac{1}{2}}}{1 + \frac{1}{4}\sum_{i<j} x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{\ell_i \ \ell_j}} .$$

Using the asymptotic results $-2 \log \lambda \sim \chi^2_{\frac{1}{2}m(m+1)}$, then for small d_i, x_{ij},

$$-2 \log \lambda \approx -n \sum_{i=1}^{m} \log\left(1-\left(\frac{d_i}{\ell_i}\right)^2\right) + 2n \log\left(1+\frac{1}{4}\sum_{i<j} x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{\ell_i \ \ell_j}\right)$$

$$\approx \sum_{i=1}^{m} n\left(\frac{d_i}{\ell_i}\right)^2 + \sum_{i<j} \frac{n \ x_{ij}^2 (k_i-k_j)(p_i-p_j)}{2\ell_i \ \ell_j} \sim \chi^2_{\frac{1}{2}m(m+1)}$$

Asymptotically then, we can use

$$\sum_{i=1}^{m} n\left(\frac{d_i}{\ell_i}\right)^2 \sim \chi^2_m \qquad (4)$$

as a test for equality of the latent roots, and

$$\sum_{i<j} n \ x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{2\ell_i \ \ell_j} \sim \chi^2_{\frac{1}{2}m(m-1)} \qquad (5)$$

as a test for equality of the eigenvectors.

This decomposition can be justified as follows. Asymptotically, under the hypothesis that the roots are identical, we have:

$$k_i \sim N(\lambda_i,\ 2\lambda_i^2/n), \quad p_i \sim N(\lambda_i,\ 2\lambda_i{}^2/n)$$

(see e.g. James (1966), p. 222).

Estimaging λ_i as $(k_i + p_i)/2 = \ell_i$, gives asymptotically,

$$\frac{k_i - p_i}{2} = d_i \mathrel{\dot\sim} N(o,\ \ell_i{}^2/n)$$

$$\text{or} \quad \frac{nd_i{}^2}{\ell_i{}^2} \mathrel{\dot\sim} \chi_1^2 .$$

Similarly, if the eigenvectors of Σ_1 and Σ_2 are the same, say the columns of H, then

$$H_1H' \approx \begin{pmatrix} \ddots & \vdots & & \\ \cdots & 1-\tfrac{1}{2}\sum_j y_{ij}^2 & \cdots\ y_{ik} & \cdots \\ & \vdots & \ddots & \\ & -y_{ik} & & \ddots \\ & \vdots & & \end{pmatrix}, \quad H_2H' = \begin{pmatrix} \ddots & \vdots & & \\ \cdots & 1-\tfrac{1}{2}\sum_j z_{ij}^2 & \cdots\ z_{ik} & \cdots \\ & \vdots & \ddots & \\ & -z_{ik} & & \ddots \\ & \vdots & & \end{pmatrix}$$

where $y_{ij} \mathrel{\dot\sim} N\left(o,\ \dfrac{\lambda_i\ \lambda_j}{n(k_i-k_j)(\lambda_i-\lambda_j)}\right)$, $z_{ij} \mathrel{\dot\sim} N\left(o,\ \dfrac{\sigma_i\ \sigma_j}{n(p_i-p_j)(\sigma_i-\sigma_j)}\right)$

for λ_i the roots of Σ_1, σ_i the roots of Σ_2 (see e.g. James (1969)). Also, $(H_1H')(H_2H')' = H_1H_2{}'$

$$\approx \begin{pmatrix} \ddots & \vdots & \\ \cdots & 1-\tfrac{1}{2}\sum_j x_{ij}^2 & \cdots +x_{ik} \\ & \vdots & \\ & -x_{ik} & \\ & \vdots & \end{pmatrix} \approx \begin{pmatrix} & \vdots & \\ \cdots & \cdots & \cdots y_{ik}-z_{ik} \cdots \\ & \vdots & \\ & -(y_{ik}-z_{ik}) & \\ & \vdots & \end{pmatrix}.$$

So under H_0, y_{ij} and z_{ij} are identically distributed, and our

$$x_{ij} \mathrel{\dot\sim} N\left(o,\ \frac{2(\ell_i\ell_j)}{n(p_i-p_j)(k_i-k_j)}\right)$$

is a very neat amalgam of the two variance estimates, giving asymptotically,

$$\frac{n}{2}\, x_{ij}{}^2\ (p_i-p_j)(k_i-k_j)/\ell_i\ell_j \mathrel{\dot\sim} \chi_1^2 .$$

The eigenvector terms are interesting, in that they show quite clearly the effect of close sample roots, which cause our information to be reduced drastically, till at worst we lose all ability to make any inferences about respective eigenvectors. This is precisely what one would expect to happen, since equal eigenvalues mean that the respective eigenvectors are any unit vectors within a certain subspace of dimension greater than one.

2.3 Comparisons

It is instructive to compare this test with other eigenvector tests. To the author's knowledge, the only other test for eigenvectors in the literature is the one developed by Anderson (1963), which tests the hypothesis that the population matrix for a particular sample has a fixed and given vector as an eigenvector. This is essentially a one-sample test which is often used in two-sample cases, treating the eigenvectors of the sample with largest degrees of freedom as fixed (see e.g. Blackith and Reyment (1971)). Another test has been developed by A.T. James, for testing that sets of population eigenvectors fill a certain fixed subspace in R^m, which for one eigenvector amounts to the same thing. This is outlined below, with his permission. This test again rests on the assumption that one is testing for a known vector (i.e. $n_2 = \infty$). For this reason they are both inappropriate in equal sample or approximately equal sample size situations.

We can express Anderson's test in our notation, quite simply. For, let $S_1 \sim W(n_1, m, \Sigma_1)$, where $S_1 = H_1 K H_1'$ for K diagonal, entries $k_1, \ldots, k_m$ along the diagonal, and $H_1 \in O(m,R)$. We wish to test whether a given vector $\underset{\sim}{h}$ is the i^{th} eigenvector of Σ_1. Then Anderson's test is:

$$\lambda = n_1 \; (k_i \underset{\sim}{h}' S_1^{-1} \underset{\sim}{h} + \frac{1}{k_i} \underset{\sim}{h}' S_1 \underset{\sim}{h} - 2) \rightsquigarrow \chi^2_{m-1} \quad .$$

It is easily verifiable that if $S_1 = H_1 K H_1'$, then $S_1^{-1} = H_1 K^{-1} H_1'$. So:

$$\lambda = n_1 \; (k_i \underset{\sim}{h}' H_1 K^{-1} H_1 \underset{\sim}{h} + \frac{1}{k_i} \underset{\sim}{h}' H_1 K H_1' \underset{\sim}{h} - 2).$$

Let $i = 1$, $m = 3$, in order to clarify the following calculations. Then,

$$\underset{\sim}{h}' H_1 = \left(\sqrt{1 - \hat{x}_{12}^2 - \hat{x}_{13}^2}\, ,\; \hat{x}_{12},\; \hat{x}_{13}\right)$$

and $H_1' \underset{\sim}{h} = (\underset{\sim}{h}' H_1)'$. We are simply picking up the first row of $H_2'H_1$, if $\underset{\sim}{h}$ were the first column of H_2. For arbitrary i,m, we would get the i^{th} row of $H_2'H_1$. So,

$$\lambda = n_1\left\{k_1\left(\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2}\;\; x_{12},x_{13}\right)\begin{pmatrix}\frac{1}{k_1} & o & o\\ o & \frac{1}{k_2} & o\\ o & o & \frac{1}{k_3}\end{pmatrix}\begin{pmatrix}\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2}\\ x_{12}\\ x_{13}\end{pmatrix}\right.$$

$$\left. + \frac{1}{k_1}\left(\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2}, x_{12}, x_{13}\right)\begin{pmatrix}k_1 & o & o\\ o & k_2 & o\\ o & o & k_3\end{pmatrix}\begin{pmatrix}\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2}\\ x_{12}\\ x_{13}\end{pmatrix} - 2\right\}$$

$$= n_1\left(1-\hat{x}_{12}^2-\hat{x}_{13}^2 + \frac{k_1}{k_2}\hat{x}_{12}^2 + \frac{k_1}{k_3}\hat{x}_{13}^2\right.$$
$$\left. + 1-\hat{x}_{12}^2-\hat{x}_{13}^2 + \frac{k_2}{k_1}\hat{x}_{12}^2 + \frac{k_3}{k_1}\hat{x}_{13}^2 - 2\right)$$

$$= n_1\ \hat{x}_{12}^2\ \frac{(k_1-k_2)^2}{k_1\ k_2} + n_1\ \hat{x}_{13}^2\ \frac{(k_1-k_3)^2}{k_1\ k_3} \sim \chi_2^2 .$$

For any i,m then, this obviously turns out as

$$\sum_{\substack{j=1\\ j\neq i}}^{m} n_1\ \hat{x}_{ij}^2\ \frac{(k_i-k_j)^2}{k_i\ k_j} \sim \chi_{m-1}^2 .$$

Note that these $\hat{x}_{ij}$ are one row of $H_2'H_1$, whereas the x_{ij} in my test are averages of skew-members, $\hat{x}_{ij} - \hat{x}_{ji}$, of $H_2'H_1$. Since in equal sample size situations one is just as interested in the behaviour of the second set of eigenvectors with respect to arbitrary ones of the first, as in the other way round, this test is obviously inappropriate in such cases. My test would be:

$$\sum_{\substack{j=1\\ j\neq i}}^{m} n\ x_{ij}^2\ \frac{(k_i-k_j)(p_i-p_j)}{\ell_i\ \ell_j} \sim \chi_{m-1}^2 .$$

Professor A.T. James has developed an exact test for population eigenvectors generating a prescribed subspace. Again, $S_1 \sim W(n_1, m, \Sigma)$, and let $H_{m\times q}$ be composed of q orthonormal column vectors spanning the prescribed subspace. Choose H_2 $(m\times(m-q))$ such that $[H \vdots H_2] \in O(m,R)$. Then:

$$[H \vdots H_2]'\ S_1\ [H \vdots H_2] = \left[\begin{array}{c:c} S_{11} & S_{12}\\ \hdashline S_{21} & S_{22}\end{array}\right] .$$

But under H_0,

$$[H_1 \vdots H_2]' \; \Sigma \; [H_1 \vdots H_2] = \begin{bmatrix} \Sigma_{11} & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \Sigma_{22} \end{bmatrix}$$

which puts us squarely into the classical sample canonical correlation situation for the null case $(\Sigma_{12} = 0)$. That is, the eigenvalues, $w_1, \ldots, w_q$, of $\mathrm{Det}(S_{12} S_{22}^{-1} S_{12} - wS_{11}) = 0$ are distributed as:

$$\frac{\Gamma_q(\frac{1}{2}n_1) \; \pi^{\frac{1}{2}q^2}}{\Gamma_q(\frac{1}{2}(n_1-m+q)) \; \Gamma_q(\frac{1}{2}(m-q))} \prod_{i=1}^{q} w_i^{\frac{1}{2}(m-2q-1)} (1-w_i)^{\frac{1}{2}(n_1-m-1)} \prod_{i<j} (w_i - w_j) \wedge dw_i$$

(For $m = 1$ this is simply the density of the sample multiple correlation coefficient R^2.)

It is well-known that if the w_i are distributed as such, then the terms $f_i = w_i/(1-w_i)$, where the f_i are the roots of $\mathrm{Det}(S_{12} S_{22}^{-1} S_{21} - f(S_{11} - S_{12} S_{22}^{-1} S_{21})) = 0$ are distributed as:

$$\frac{\Pi^{\frac{1}{2}q^2} \; \Gamma_q(\frac{1}{2}(n_1+q))}{\Gamma_q(\frac{1}{2}(n_1+m+q)) \; \Gamma_q(\frac{1}{2}m) \; \Gamma_q(\frac{1}{2}q)} \prod_i \frac{f_i^{\frac{1}{2}(m-q-1)}}{(1+f_i)^{\frac{1}{2}(n+q)}} \prod_{i<j} (f_i - f_j) \wedge df_i$$

which is the multivariate analysis of variance case, $\mathrm{Det}(W - f_i B) = 0$ (see e.g. Anderson (1958)).

For $q = 1$ this then becomes the density of $f = R^2/(1-R^2)$ and from Anderson (1958) p. 89 we have

$$\frac{R^2}{1-R^2} \frac{(n+1-m)}{(m-1)} \sim F_{m-1, n+1-m} \quad .$$

We have further that the roots of

$$\mathrm{Det}\big((H'SH \; H'S^{-1}H - I) - \lambda I\big) = 0$$

are the same as the roots of

$$\mathrm{Det}(S_{12} S_{22}^{-1} S_{21} - f(S_{11} - S_{12} S_{22}^{-1} S_{21})) = 0 \qquad (5a)$$

since $\quad [H \vdots H_2]' S [H \vdots H_2] = \begin{bmatrix} S_{11} & \vdots & S_{12} \\ \cdots & \cdots & \cdots \\ S_{21} & \vdots & S_{22} \end{bmatrix}$.

$\therefore \; H'SH = S_{11}$.

Also, let $[[H \vdots H_2]' \, S [H \vdots H_2]]^{-1} = \begin{bmatrix} S^{11} & \vdots & S^{12} \\ \cdots & \vdots & \cdots \\ S^{21} & \vdots & S^{22} \end{bmatrix}$.

$$\text{So} \quad S^{11} = \left[\left[\begin{bmatrix} H' \\ \cdots \\ H_2 \end{bmatrix} S \, [H \vdots H_2] \right]^{-1} \right]_{11}$$

$$= \left[\begin{bmatrix} H' \\ \cdots \\ H_2{}' \end{bmatrix} S^{-1} \, [H \vdots H_2] \right]_{11}$$

$$= H'S^{-1}H \quad .$$

But $(S^{11})^{-1} = S_{11} - S_{12} S_{22}^{-1} S_{21}$.

So (5a) can be written

$$\mathrm{Det}(S_{11} - (S^{11})^{-1} - f(S^{11})^{-1}) = 0$$

$$\mathrm{Det}(S_{11} S^{11} - I_q - fI_q) = 0$$

$$\mathrm{Det}((H'SHH'S^{-1}H - I_q) - fI_q) = 0 \; .$$

So the eigenvalues of $(H'SHH'S^{-1}H - I_q)$ can be used to test the hypothesis that the eigenvectors of Σ span a subspace with basis H, using the same criteria used in testing the general linear hypothesis in MANOVA (see e.g. Anderson (1958), p. 221-223).

For $q = 1$, the root of $\mathrm{Det}(\underset{\sim}{h}'S\underset{\sim}{h}\underset{\sim}{h}'S^{-1}\underset{\sim}{h} - 1 - f) = 0$ is of course just $\underset{\sim}{h}'S\underset{\sim}{h}\underset{\sim}{h}'S^{-1}\underset{\sim}{h} - 1$ and so

$$(\underset{\sim}{h}'S\underset{\sim}{h}\underset{\sim}{h}'S^{-1}\underset{\sim}{h}-1) \, \frac{(n+1-m)}{(m-1)} \qquad F_{m-1,n+1-m} \, ,$$

an exact statistic for testing if $\underset{\sim}{h}$ is an eigenvector of Σ.

We can compare this test with Anderson's by expressing this criterion in our notation. As before, let $S_1 \sim W(n_1, m\ \Sigma_1)$ with $S_1 = H_1 K H_1{}'$, K diagonal, entries k_i, H orthogonal. Then $S_1^{-1} = H_1 K^{-1} H_1{}'$. Again choose $m = 3$ for simplicity, giving:

$$\underset{\sim}{h}'S_1\underset{\sim}{h}\underset{\sim}{h}'S_1^{-1}\underset{\sim}{h} = \underset{\sim}{h}'H_1 K H_1{}'\underset{\sim}{h}\underset{\sim}{h}'H_1 K^{-1} H_1{}'\underset{\sim}{h}$$

$$= \left(\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2},\ \hat{x}_{12},\ \hat{x}_{13}\right) \begin{pmatrix} k_1 & & \\ & k_2 & \\ & & k_3 \end{pmatrix} \begin{pmatrix} \sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2} \\ \hat{x}_{12} \\ \hat{x}_{13} \end{pmatrix} \times$$

$$\times \left(\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2},\ \hat{x}_{12},\ \hat{x}_{13}\right) \begin{pmatrix} \frac{1}{k_1} & & \\ & \frac{1}{k_2} & \\ & & \frac{1}{k_3} \end{pmatrix} \begin{pmatrix} \sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2} \\ x_{12} \\ x_{13} \end{pmatrix}$$

$$= \left(k_1 (1-\hat{x}_{12}^2-\hat{x}_{13}^2) + k_2 \hat{x}_{12}^2 + k_3 \hat{x}_{13}^2\right) \times$$

$$\times \left(\frac{1}{k_1} (1-\hat{x}_{12}^2-\hat{x}_{13}^2) + \frac{1}{k_2} \hat{x}_{12}^2 + \frac{1}{k_3} \hat{x}_{13}^2\right) .$$

$\therefore \quad \underset{\sim}{h}' S_1 \underset{\sim}{h}\underset{\sim}{h}' S_1^{-1} \underset{\sim}{h} - 1 =$

$$\frac{(k_1-k_2)^2}{k_1 k_2} \hat{x}_{12}^2 + \frac{(k_1-k_2)^2}{k_1 k_3} \hat{x}_{13}^2$$

$$- \frac{(k_1-k_2)^2}{k_1 k_2} \hat{x}_{12}^4 - \frac{(k_1-k_3)^2}{k_1 k_3} \hat{x}_{13}^4 + \left(\sum_{i<j} \frac{(k_i-k_j)^2}{k_i k_j}\right) \hat{x}_{12}^2\ \hat{x}_{13}^2 .$$

So for small $\hat{x}_{ij}$ the quartic terms are negligible and

$$\left(\frac{(k_1-k_2)^2}{k_1 k_2} \hat{x}_{12}^2 + \frac{(k_1-k_3)^2}{k_1 k_3} \hat{x}_{13}^2\right) \frac{n_1-2}{2} \sim F(2, n_1-2)$$

when Anderson's test is

$$n_1 \frac{(k_1-k_2)^2}{k_1 k_2} \hat{x}_{12}^2 + n_1 \frac{(k_1-k_3)^2}{k_1 k_3} \hat{x}_{13}^2 \sim \chi_2^2 .$$

This obviously generalizes to arbitrary m. But recall that James' criterion tests for $\underset{\sim}{h}$ being <u>any</u> eigenvector of Σ_1, and not a particular one, as in Anderson's test. The fourth degree terms are only insignificant for small $\hat{x}_{ij}$, that is, when $\underset{\sim}{h}$ is near the i^{th} eigenvector. If it were actually closer to another one, the criterion could be expressed more like Anderson's test for the other eigenvector. In our example, the vector $(\sqrt{1-\hat{x}_{12}^2-\hat{x}_{13}^2}, \hat{x}_{12}, \hat{x}_{13})$ would have small $\hat{x}_{12}, \hat{x}_{13}$, if $\underset{\sim}{h}$ were near the first eigenvector, but if $\underset{\sim}{h}$ were actually near the second, say, we could then express the same vector as $(\hat{x}_{21}, \sqrt{1-\hat{x}_{21}^2-\hat{x}_{23}^2}, \hat{x}_{23})$ for small $\hat{x}_{21}, \hat{x}_{23}$. Following through the above calculations would then give:

$$\underset{\sim}{h}'S_1\underset{\sim}{h}\underset{\sim}{h}'S_1^{-1}\underset{\sim}{h} - 1 = \frac{(k_1-k_2)^2}{k_1k_2}\,\hat{x}_{21}^2 + \frac{(k_2-k_3)^2}{k_2k_3}\,\hat{x}_{23}^2 + \text{terms of fourth degree}$$

which is exactly the <u>value</u> of the previous expansion, but looks much more like Anderson's test for $\underset{\sim}{h}$ being the second eigen-vector. Of course the particular situation of any problem would decide which test was the more appropriate.

Both tests are designed essentially for one-sample problems, and quite fail us in two-sample, equal sample size situations. But what about the unequal, but still comparable sample size case? The likelihood ratio test for $H_0: \Sigma_1 = \Sigma_2$, where $S_1 \sim W(n_1, \Sigma_1)$, $S_2 \sim W(n_2, \Sigma_2)$ is:

$$\lambda = \frac{\operatorname{Det} S_1^{\frac{1}{2}n_1}\ \operatorname{Det} S_2^{\frac{1}{2}n_2}}{\operatorname{Det}\left(\dfrac{n_1S_1 + n_2S_2}{n_1 + n_2}\right)^{\frac{1}{2}(n_1+n_2)}} .$$

Again, let $S_1 = H_1L_1H_1'$, $S_2 = H_2L_2H_2'$, and

$L_1 = \begin{pmatrix} k_1 & & o \\ & \ddots & \\ o & & k_m \end{pmatrix}$, $L_2 = \begin{pmatrix} p_1 & & o \\ & \ddots & \\ o & & p_m \end{pmatrix}$. Also, let $\ell_i = \dfrac{n_1k_i + n_2p_i}{n_1 + n_2}$,

where before, $\ell_i = \dfrac{k_i + p_i}{2}$, for $n_1 = n_2$.

Expanding as before, with $(H_1H_2') \approx H^*$ gives :

$$\lambda = \frac{(k_1\ldots k_m)^{\frac{1}{2}n_1}\ (p_1\ldots p_m)^{\frac{1}{2}n_2}}{\left(\ell_1\ell_2\ldots\ell_m - \sum\limits_{i<j}\dfrac{n_1n_2}{(n_1+n_2)^2}(k_i-k_j)(p_i-p_j)\underbrace{\ell_1\ldots\ell_m}_{\text{missing } \ell_i,\ell_j} + \ldots\right)^{\frac{1}{2}(n_1+n_2)}}$$

$$= \frac{\prod\limits_{i=1}^{m}\left(\dfrac{k_i}{\ell_i}\right)^{\frac{1}{2}n_1}\left(\dfrac{p_i}{\ell_i}\right)^{\frac{1}{2}n_2}}{\left(1 - \sum\limits_{i<j}\dfrac{n_1n_2}{(n_1+n_2)^2}\,x_{ij}^2\,\dfrac{(k_i-k_j)(p_i-p_j)}{\ell_i\ell_j}\right)^{\frac{1}{2}(n_1+n_2)}} .$$

Let

$$d_i = \ell_i - k_i = \frac{n_2(p_i-k_i)}{n_1+n_2} .$$

Geometrically, stretching k_i, ℓ_i, p_i on R^1, d_i is the length from k_i to ℓ_i.

$$\underbrace{k_i \qquad\qquad\qquad\qquad}_{d_i = \frac{n_2(p_i-k_i)}{n_1+n_2}}\ \ell_i = \frac{n_1k_1+n_2p_i}{n_1+n_2} \qquad\qquad p_i$$

Then, $p_i = \ell_i + \frac{n_1}{n_2} d_i$.

So, $$\left(\frac{k_i}{\ell_i}\right)^{\frac{1}{2}n_1} \left(\frac{p_i}{\ell_i}\right)^{\frac{1}{2}n_2} = \left(\frac{\ell_i - d_i}{\ell_i}\right)^{\frac{1}{2}n_1} \left(\frac{\ell_i + \frac{n_1}{n_2} d_i}{\ell_i}\right)^{\frac{1}{2}n_2}$$

$$= \left(1 - \frac{d_i}{\ell_i}\right)^{\frac{n_1}{2}} \left(1 + \frac{n_1}{n_2} \frac{d_i}{\ell_i}\right)^{\frac{n_2}{2}}$$

$$\approx \left(1 - \frac{n_1}{2} \left(\frac{n_1 + n_2}{2n_2}\right) \frac{d_i^2}{\ell_i^2}\right) .$$

This gives,

$$-2 \log \lambda \approx \sum_{i=1}^{m} n_1 \left(\frac{n_1 + n_2}{2n_2}\right) \frac{d_i^2}{\ell_i^2} + \sum_{i<j} \frac{n_1 n_2}{n_1 + n_2} x_{ij}^2 \frac{(k_i - k_j)(p_i - p_j)}{\ell_i \ell_j}$$

$$\sim \chi^2_{\frac{1}{2}m(m+1)} .$$

Note, the $n_1 \left(\frac{n_1 + n_2}{2n_2}\right) \frac{d_i^2}{\ell_i^2}$ terms are symmetric with d_i^*, where $k_i + d_i + d_i^* = p_i$, or $p_i - \ell_i = d_i^*$.

When $n_1 = n_2 = n$, we get $\sum_{i=1}^{m} n \frac{d_i^2}{\ell_i^2} + \sum_{i<j} \frac{n}{2} x_{ij}^2 \frac{(k_i - k_j)(p_i - p_j)}{\ell_i \ell_j}$

as before, with $\ell_i = \frac{k_i + p_i}{2}$, $d_i = \frac{k_i - p_i}{2}$.

Now, let $n_2 \to \infty$.

This gives $\sum_{i=1}^{m} \frac{n_1}{2} \frac{d_i^2}{\ell_i^2} + \sum_{i<j} n_1 x_{ij}^2 \frac{(k_i - k_j)(p_i - p_j)}{\ell_i \ell_j}$

$$\sim \chi^2_{\frac{1}{2}m(m+1)}$$

where now $\ell_i = p_i$, $d_i = p_i - k_i$.

As a check, let us take the limit of λ as $n_2 \to \infty$.

We have:

$$\lambda = \frac{\text{Det } S_1^{\frac{1}{2}n_1} \ \text{Det } S_2^{\frac{1}{2}n_2}}{\text{Det}\left(\frac{n_1 S_1 + n_2 S_2}{2}\right)^{\frac{1}{2}(n_1 + n_2)}}$$

$$= \frac{(\text{Det } S_1 S_2^{-1})^{\frac{1}{2}n_1}}{\text{Det}\left(\frac{n_1 S_1 S_2^{-1} + n_2 I}{n_1 + n_2}\right)^{\frac{1}{2}(n_1 + n_2)}}$$

on dividing through by $(\text{Det } S_2)^{\frac{1}{2}(n_1 + n_2)}$.

Let $m = n_1 + n_2$, so λ becomes:

$$\lambda = \frac{(\mathrm{Det}(S_1 S_2^{-1}))^{\frac{1}{2}n_1}}{\mathrm{Det}\left(\left(\frac{n_1 (S_1 S_2^{-1} - I)}{m} + I\right)^{\frac{1}{2}m}\right)} \quad \text{since} \quad (\mathrm{Det}\ A)^n = \mathrm{Det}(A^n)$$

$$\text{then} \lim_{m\to\infty} \lambda = \frac{(\mathrm{Det}(S_1 S_2^{-1}))^{\frac{1}{2}n_1}}{\mathrm{Det}\ \left(e^{\frac{1}{2}n_1 (S_1 S_2^{-1} - I)}\right)}$$

$$= \frac{\mathrm{Det}\ (S_1 S_2^{-1})^{\frac{1}{2}n_1}}{e^{\mathrm{tr}(\frac{1}{2}n_1 (S_1 S_2^{-1} - I))}}$$

$$= \left(\mathrm{Det}\ (S_1 S_2^{-1})\ e^{\mathrm{tr}(I - S_1 S_2^{-1})}\right)^{\frac{1}{2}n_1} .$$

This should be the likelihood ratio test for $\Sigma_1 = \Sigma_0$ specified, which from Anderson (1958), p. 264 is:

$$\lambda_1 = \left(\frac{e}{n_1}\right)^{\frac{1}{2}mn_1} \det(n_1 S_1 \Sigma_0^{-1})^{\frac{1}{2}n_1}\ e^{-\frac{1}{2}\mathrm{tr}\ n_1 S_1 \Sigma_0^{-1}}$$

$$= e^{\frac{1}{2}mn_1 - \frac{1}{2}\mathrm{tr}\ n_1 S_1 \Sigma_0^{-1}}\ \det\ (S_1 \Sigma_0^{-1})^{\frac{1}{2}n_1}$$

$$= \det\ (S_1 \Sigma_0^{-1})^{\frac{1}{2}n_1}\ e^{\frac{1}{2}n_1 \mathrm{tr}\ (I - S_1 \Sigma_0^{-1})}$$

$$= \left(\det\ S_1 \Sigma_0^{-1}\ e^{\mathrm{tr}(I - S_1 \Sigma_0^{-1})}\right)^{\frac{1}{2}n_1}$$

which is exactly $\lambda\ (n_2 = \infty)$ with $\Sigma_0 = S_2$, as required.

Now, decomposing $\lambda\ (n_2 = \infty)$, $S_2^{-1} = H_2 \begin{pmatrix} \frac{1}{p} & & \\ & \ddots & \\ & & \frac{1}{p_m} \end{pmatrix} H_2{}'$, and $S_1 = H_1 \begin{pmatrix} k_1 & & \\ & \ddots & \\ & & k_m \end{pmatrix} H_1{}'$ as before. So $\det(S_1 S_2^{-1}) = \prod_{i=1}^{m} \left(\frac{k_i}{p_i}\right)$

and $\mathrm{tr}(S_1 S_2^{-1}) = \mathrm{tr}(H_1 L_1 H_1{}' H_2 L_2^{-1} H_2{}') = \mathrm{tr}(L_1 (H_1{}' H_2) L_2^{-1} (H_1{}' H_2)')$

since $\mathrm{tr}\ AB = \mathrm{tr}\ BA$.

$\therefore\ \mathrm{tr}(L_1 (H_1{}' H_2) L_2^{-1} (H_1{}' H_2)') \approx \mathrm{tr}(L_1 H^* L_2^{-1} H^{*\prime})$

and $\lambda(n_2 = \infty) \approx \left(\Pi\left(\frac{k_i}{p_i}\right)\ e^{\mathrm{tr}(-L_1 H^* L_2^{-1} H^{*\prime} + I)}\right)^{\frac{1}{2}n_1}$.

Expanding out $\mathrm{tr}(L_1 H^* L_2^{-1} H^{*\prime})$ gives:

$$\mathrm{tr}(L_1 H^* L_2^{-1} H^{*\prime}) = \sum_{i=1}^{m} \frac{k_i}{p_i} + \sum_{i<j} x_{ij}{}^2\ \frac{(k_i - k_j)(p_i - p_j)}{p_i p_j} .$$

$$\text{So}\ {}_{n_2=\infty}\lambda \approx \left(\prod_{i=1}^{m}\left(\frac{k_i}{p_i}\right)\ e^{m - \Sigma\frac{k_i}{p_i} - \sum_{i<j} x_{ij}{}^2\ \frac{(k_i - k_j)(p_i - p_j)}{p_i p_j}}\right)^{\frac{1}{2}n_1}$$

$$\approx \left(\prod_{i=1}^{m} \left(\frac{k_i}{p_i} e^{1-\frac{k_i}{p_i}} \right) \prod_{i<j} e^{-x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{p_i p_j}} \right)^{\frac{1}{2}n_1} .$$

But:

$$\frac{k_i}{p_i} e^{1-\frac{k_i}{p_i}} = \frac{k_i}{p_i} e^{\frac{p_i-k_i}{p_i}}$$

$$= \frac{p_i-d_i}{p_i} e^{\frac{d_i}{p_i}} \quad , \text{ if } d_i=p_i-k_i$$

$$= \left(1 - \frac{d_i}{p_i}\right) e^{\frac{d_i}{p_i}}$$

$$= \left(1 - \frac{d_i}{p_i}\right)\left(1 + \frac{d_i}{p_i} + \frac{d_i^2}{2p_i^2} + \ldots\right)$$

$$\approx \left(1 - \frac{d_i^2}{2p_i^2}\right)$$

which gives:

$$-2 \log_{n_2 = \infty} \lambda \approx \sum_{i=1}^{m} \frac{n_1 d_i^2}{2p_i^2} + \sum_{i<j} n_1 x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{p_i p_j}$$

which are the separate limits of $-2 \log \lambda$ as found previously.

These are asymptotically correct, since if S_1 is actually an observation from the population S_2 (i.e. $n_2=\infty$), then

$$k_i \sim N\left(p_i, \frac{2p_i^2}{n_1}\right) \Rightarrow \frac{n_1}{2} \frac{d_i^2}{p_i^2} \sim \chi_1^2$$

and

$$x_{ij} \sim N\left(o, \frac{n_1 \; p_i p_j}{(k_i-k_j)(p_i-p_j)}\right) \Rightarrow \frac{n_1 (k_i-k_j)(p_i-p_j)}{n_1 p_i p_j} x_{ij}^2 \sim \chi_1^2 .$$

However, the results for the eigenvector test depend on the eigenvalues for both sample matrices. Our x_{ij} is the skew-average $\left(\frac{x_{ij}^*-x_{ji}^*}{2}\right)$ of the inner product of the i^{th} eigenvector of S_1 with the j^{th} eigenvector of S_2, and the j^{th} eigenvector of S_1 with the i^{th} eigenvector of S_2. The one-sample tests involve only one sample S_1, and an arbitrary unit vector, $\underset{\sim}{h}$, without regard to the possible population $\underset{\sim}{h}$ may have come from, and ignoring this information if it is present. I think the one-sample tests are more appropriate when $n_2 >> n_1$, and one can reject the hypothesis that the roots are equal. But if the population roots are equal, then my test is probably better, since it weights favourably the sample roots about which we know most. My test would also be better for comparable sample sizes, regardless of the eigenvalue behaviour, since it weights evenly the variation from both samples.

Suppose one now wishes to test that discrepancies between the roots take the form of a simple scalar difference. That is, $\Gamma_1 = K\Gamma_2$ for some unidentified scalar K, where Γ_i is the diagonal matrix of eigenvalues of Σ_i. This would mean that in the representation of Σ_1 and Σ_2 as hyperellipsoids, namely $\underset{\sim}{x}'\Sigma_1\underset{\sim}{x} = 1$ and $\underset{\sim}{x}'\Sigma_2\underset{\sim}{x} = 1$, these two objects would be similar in the geometric sense, although perhaps with different orientations. The same orientation or eigenvectors would then mean $\Sigma_1 = K\Sigma_2$.

A further decomposition of the latent roots statistics yields such a test. Form new variables, y_i, from the old, d_i/ℓ_i, via:

$$d_i/\ell_i = a_{i1}\, y_1 + a_{i2}\, y_2 + \ldots + a_{im}\, y_m \qquad (6)$$

where $A = (a_{ij})$ is orthogonal. Then

$$\sum_{i=1}^{m} n_1\left(\frac{n_1+n_2}{2n_2}\right)\frac{d_i^{\,2}}{\ell_i^{\,2}} = \sum_{i=1}^{m}\frac{n_1(n_1+n_2)}{2n_2}\left(\sum_{j=1}^{m}(a_{ji})^2\right) y_i^{\,2} \mathrel{\dot\sim} \chi^2_m \,.$$

Any A will give a new partitioning of $\sum_{i=1}^{m} n_1\frac{(n_1+n_2)}{2n_2}\frac{d_i^{\,2}}{\ell_i^{\,2}}$.
But requiring $a_{11} = a_{21} = \ldots = a_{m1} = \frac{1}{\sqrt{m}}$ makes y_1 a measure of scalar differences in the root structures, with the other y_i's recording other (orthogonal) deviations.

For, if

$$\frac{k_i}{p_i} = \frac{\ell_i - d_i}{\ell_i + \frac{n_1}{n_2} d_i} = K \quad , \forall\, i \; ,$$

then $$\frac{d_i}{\ell_i} = \frac{1-K}{1+\frac{n_1}{n_2}K} \quad \forall\, i \; .$$

Asymptotically then, $\sum\limits_{i=2}^{m} n\left(\sum\limits_{j=1}^{m} (a_{ji})^2\right) y_i^{\,2} \sim \chi^2_{m-1}$

if the only difference between the roots is a scalar size one, and:

$$\frac{n_1(n_1+n_2)}{2n_2} \sum_{j=1}^{m} (a_{j1})^2 y_1^{\,2} = \frac{n_1(n_1+n_2)}{2n_2} y_1^{\,2} \sim \chi^2_1 \quad \text{if}$$

$K = 1$.

Solving (6), $y_1 = \frac{1}{\sqrt{m}} \left(\sum\limits_{i=1}^{m} \left(\frac{d_i}{\ell_i}\right)\right)$ giving

$$\frac{n_1(n_1+n_2)}{2n_2\, m} \left(\sum_{i=1}^{m} \frac{d_i}{\ell_i}\right)^2 \sim \chi^2_1$$

and

$$\frac{n_1(n_1+n_2)}{2n_2} \Sigma\left(\frac{d_i}{\ell_i}\right)^2 - \frac{n_1(n_1+n_2)}{2n_2\, m} \left(\Sigma\, \frac{d_i}{\ell_i}\right)^2 \sim \chi^2_{m-1} \; .$$

I suggest this test be used cautiously, and only where the latent roots of both samples are well spread out, unless the sample sizes be very large, for the asymptotic densities used here take no account of linkage factors between roots (see James (1966).

2.4 Data Analysis

Jolicoeur and Mosimann (1960) measured three variables on 24 male and 24 female species of Chrysemys Picta Marginata. The resulting variance-covariance estimates were:

$$S_1 = \begin{bmatrix} 451.39 & 271.17 & 168.70 \\ * & 171.73 & 103.29 \\ * & * & 66.65 \end{bmatrix} \quad \text{(female)}$$

$$S_2 = \begin{bmatrix} 138.77 & 79.15 & 37.38 \\ * & 50.04 & 21.65 \\ * & & 11.26 \end{bmatrix} \quad \text{(male)}$$

with $n = 23$, $m = 3$.

Testing $H_0 : \Sigma_1 = \Sigma_2$, the likelihood ratio test is:

$$\lambda = \left(\frac{\det S_1^{\frac{1}{2}} \det S_2^{\frac{1}{2}}}{\det \left(\frac{S_1 + S_2}{2} \right)} \right)^n = (.5699)^{23} .$$

Using the asymptotic χ^2 test, we obtain

$$-2 \log \lambda = 25.8 \sim \chi^2_6$$

so H_0 can be rejected at the 0.1% level. However, it was thought that this result was perhaps caused basically by unequal root structures, and in an attempt to test this, the previous decomposition was found.

We have:

$$S_1 = H_1 \begin{bmatrix} 680.41 & 0 & 0 \\ 0 & 6.502 & 0 \\ 0 & 0 & 2.857 \end{bmatrix} H_1' = H_1 \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} H_1'$$

$$S_2 = H_2 \begin{bmatrix} 195.28 & 0 & 0 \\ 0 & 3.686 & 0 \\ 0 & 0 & 1.103 \end{bmatrix} H_2' = H_2 \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix} H_2'$$

with

$$H_1 = \begin{bmatrix} .8126 & -.5454 & -.2053 \\ .4955 & .8321 & -.2493 \\ .3068 & .1009 & .9464 \end{bmatrix}, \qquad H_2 = \begin{bmatrix} .8401 & -.4876 & -.2376 \\ .4919 & .8695 & -.0449 \\ .2285 & -.0792 & .9703 \end{bmatrix}$$

and

$$H_1'H_2 = \begin{bmatrix} .9966 & .01033 & .08227 \\ -.02587 & .9814 & .1901 \\ -.07878 & -.1916 & .9783 \end{bmatrix}.$$

The matrix X, such that $\exp(X) \approx H_1'H_2$ was estimated by simply taking skew-averages of the ij,ji elements of $H_1'H_2$ to get estimates of the ij element of X. (e.g. $x_{12} = (.01033 + .02587)/2 = .01810$). This gives

$$X = \begin{bmatrix} 0 & .01810 & .08053 \\ -.01810 & 0 & .19081 \\ -.08053 & -.19081 & 0 \end{bmatrix}.$$

This was checked against the matrix log expansion:

$$\log H_1'H_2 = (H_1'H_2-I)-\tfrac{1}{2}(H_1'H_2-I)^2+\tfrac{1}{3}(H_1'H_2-I)^3-\tfrac{1}{4}(H_1'H_2-I)^4+\ldots$$

where less than 1% difference was recorded between corresponding terms of X and $\log H_1'H_2$, using the first four terms of the expansion. So the extra work involved in evaluating $\log H_1'H_2$ seems unnecessary.

Substituting

$$H^* = \begin{bmatrix} .9966 & .01810 & .08053 \\ -.0181 & .9814 & .19081 \\ -.08053 & -.19081 & .9783 \end{bmatrix}$$

for $H_1'H_2$ in the likelihood ratio causes negligible change to λ. Even less change is caused by neglecting x_{ij} terms of degree greater than two in the determinant expansion of $\det\left((S_1+S_2)/2\right)$. For, the latent root values give:

$$\ell_1 = 437.85 \qquad d_1 = 242.55$$

$$\ell_2 = 5.094 \qquad d_2 = 1.408$$

$$\ell_3 = 1.980 \qquad d_3 = .877$$

and with these terms we can follow the approximations.

First:

$$\det\left((S_1+S_2)/2\right)$$

$$= \det\left((L_1 H_1' H_2 + H_1' H_2 L_2)/2\right) = 5{,}561 \qquad \lambda = (.5699)^{23}$$

Then: $$\det\left((L_1 H^* + H^* L_2)/2\right) = 5{,}574 \qquad \lambda^* = (.5686)^{23}$$

Finally: $$\ell_1 \ell_2 \ell_3 + \sum_{\substack{i<j \\ s \neq i,j}} \frac{x_{ij}^2}{4} (k_i - k_j)(p_i - p_j) \ell_s = 5{,}577 \qquad \lambda^{**} = (.5683)^{23} .$$

The major sources of error are then the crude log approximations,

$$-2 \log \left(1 - \frac{d_i^2}{\ell_i^2}\right)^{\frac{n}{2}} \approx n \frac{d_i^2}{\ell_i^2}$$

and $$-2 \log \left(1 + \sum_{i<j} \frac{x_{ij}^2}{4} \frac{(k_i-k_j)(p_i-p_j)}{\ell_i \ell_j}\right)^{-n} \approx \frac{n}{2} \sum_{i<j} x_{ij}^2 \frac{(k_i-k_j)(p_i-p_j)}{\ell_i \ell_j}$$

However, since these results are asymptotically correct, it suggests perhaps that some of this 'error' is in fact a correction for the $-2 \log \lambda$ approximation itself, although I have not yet attempted to prove this. It might also be possible to find correction factors for the log approximations.

Testing for equality of roots, using (4) we obtain $\sum \frac{n d_i^2}{\ell_i^2} = 23(.3069 + .0764 + .1962) = 13.33 \sim \chi_3^2$, which can be rejected at the 0.5% level. Then using (5) to test for equality of the eigenvectors gives:

$$\sum_{i<j} \frac{n}{2} x_{ij}^2 \frac{((k_i-k_j)(p_i-p_j))}{\ell_i \ell_j} =$$

$$= \frac{23}{2} (.01897 + .98416 + .03399)$$

$$= 11.9 \sim \chi_3^2 \quad .$$

Even this can be rejected at the 1% level, so we must reject the idea that the eigenvectors are the same. There seem to be minor but significant differences between them, which turn up as loss of orthogonality between the first and third eigenvectors of the two matrices.

A simple preliminary calculation would have indicated what was happening, for the latent root part of this decomposition is just the numerator of $\lambda^{\frac{1}{n}}$, divided by $\prod_{i=1}^{m} \ell_i$, and the eigenvector part is just the denominator divided by the same, both parts being then 1, if $S_1 = S_2$.

In this example:

$$\frac{\text{Det } S_1^{\frac{1}{2}} \text{ Det } S_2^{\frac{1}{2}}}{\ell_1 \ell_2 \ell_3} = .7177$$

and
$$\frac{\text{Det}\left(\frac{S_1 + S_2}{2}\right)}{\ell_1 \ell_2 \ell_3} = 1.2593 \quad .$$

Both deviate from one by comparable amounts, so both contribute roughly the same amount to the rejection of H_0. This seems in itself to give a rather quick, though crude, look at what is happening.

At this point we thought perhaps the root differences could be accounted for by a scalar size factor, and devised the previously mentioned test. In this case the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{bmatrix}$$

will orthogonally repartition $\sum_{i=1}^{3} n\left(\frac{d_i}{\ell_i}\right)^2$ into $\sum_{i=1}^{3} n\, y_i^2$ where $\frac{d_i}{\ell_i} = \sum_{j=1}^{m} a_{ij}\, y_j$, and y_2, y_3 measure deviations from a simple scalar size factor between the two sets of roots, and y_1 measures how much that scalar size factor differs from unity.

Solving the equations $d_i/\ell_i = \Sigma\, a_{ij}\, y_j$ gives

$y_1 = \frac{1}{\sqrt{3}} \Sigma \frac{d_i}{\ell_i}$. Then:

$$n\, y_1^2 = \frac{n}{3}\left(\sum_{i=1}^{3} \frac{d_i}{\ell_i}\right)^2 = 12.43 \sim \chi_1^2$$

which can be rejected at the .5% level, and

$$ny_2^2 + ny_3^2 = 13.33 - 12.43 = .90 \sim \chi_2^2$$

$$\left(\text{found by subtraction} \quad n\Sigma\left(\frac{d_i}{\ell_i}\right)^2 - ny_1^2\right)$$

which is quite acceptable. Thus we have good support for our hypothesis that the differences between the roots of S_1 and S_2 can be accounted for by a non-unity scalar size factor.

To see what effect this result had on the eigenvector test, we used the asymptotic density for the latent roots to find asymptotic maximum likelihood estimates for $\lambda_1, \lambda_2, \lambda_3$, under $H_0: \Sigma_1 = \mathcal{H}_1 \Lambda \mathcal{H}_1$, and $\Sigma_2 = \mathcal{H}_2 K \Lambda \mathcal{H}_2'$. From James (1969), the asymptotic likelihood function for the roots under H_0 is:

$$\mathcal{L} = \text{const.} \times \prod_i \alpha_i^{\frac{1}{2}n} e^{-\frac{1}{2}n\sum_i k_i\alpha_i} \prod_{i<j} (k_i-k_j)^{-\frac{1}{2}}(\alpha_j-\alpha_i)^{-\frac{1}{2}} \Pi(\alpha_i/K)^{\frac{1}{2}n}.$$

$$\cdot\ e^{-\frac{1}{2}n\Sigma p_i\alpha_i/K} \prod_i (p_i-p_j)(\alpha_j-\alpha_i)/K$$

$$= \text{const.} \times \Pi\ \alpha_i^n\ e^{-\frac{1}{2}n(\sum_i k_i\alpha_i + \frac{1}{K}\sum_i p_i\alpha_i} \prod_{i<j} (\alpha_i-\alpha_i) K^{\frac{1}{2}m\left(\frac{m-1}{2}-n\right)}$$

$$\times \text{ a function of } k_i\text{'s}, p_i\text{'s}$$

where the k_i's are roots of S_1, p_i's are roots of S_2, and $\alpha_i = 1/\lambda_i$. Taking derivatives of the log likelihood:

$$\frac{\partial}{\partial k} (\log \mathcal{L}) = \frac{1}{2} \frac{\Sigma p_i\alpha_i}{K} + \frac{\frac{1}{2}m\left(\frac{m-1}{2}\right)}{n} - \frac{1}{2}m = 0$$

$$\frac{\partial}{\partial \alpha_i} (\log \mathcal{L}) = \frac{1}{\alpha_i} - \frac{1}{2}(k_i+p_i/K) + \frac{1}{n} \sum_{\substack{j\neq i \\ i<j}} 1/(\alpha_j-\alpha_i) = 0.$$

Taking the limit as $n \to \infty$ gives two relations

$$K = \sum_i p_i\alpha_i/m \quad , \quad \lambda_i = \frac{1}{2}(k_i+p_i/K)$$

by which the maximum likelihood estimates can be calculated recursively. Numerical evaluation gave:

$$\hat{K} = 2.51634 \qquad \hat{\lambda}_1 = 232.839$$
$$\hat{\lambda}_2 = 3.135$$
$$\hat{\lambda}_3 = 1.1192 \quad .$$

Substituting in the test for the eigenvectors, using $\hat{\lambda}_i$ for k_i, $\hat{K}\ \hat{\lambda}_i$ for p_i gave:

$$\Sigma\ \frac{n}{2}\ x_{ij}{}^2\ \frac{(\hat{\lambda}_i-\hat{\lambda}_j)^2 4\hat{K}}{\hat{\lambda}_i\hat{\lambda}_j\ (1+\hat{K})^2} = \tfrac{23}{2}(.0193 + 1.21 + .0385) = 14.579$$
$$\sim \chi_3^2$$

which can be rejected at the 0.5% level. This even further strengthens the previous result that the eigenvectors are different for the two samples.

The following is another asymptotic test for the eigenvalues of the two populations, Σ_1, Σ_2, being proportional, which serves as a check on our previous analysis. As before, let S_1 have latent roots k_1, k_2, k_3, and S_2 have roots p_1, p_2, p_3. Then asymptotically, under $H_0: k_i = K\,p_i$, for K some constant, we have

$$k_i \sim N(K\lambda_i, 2(K\lambda_i)^2/n), \quad p_i \sim N(\lambda_i, 2\lambda_i{}^2/n)$$

$$\text{or} \quad X = \begin{bmatrix} k_1 & p_1 \\ k_2 & p_2 \\ k_3 & p_3 \end{bmatrix} \sim N\left(M = \begin{bmatrix} K\lambda_1 & \lambda_1 \\ K\lambda_2 & \lambda_2 \\ K\lambda_3 & \lambda_3 \end{bmatrix},\ \Sigma_1 \otimes \Sigma_2 \right)$$

$$\text{for } \Sigma_1 = \begin{bmatrix} K^2 & o \\ o & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2\lambda_1{}^2/n & o & o \\ o & 2\lambda_2{}^2/n & o \\ o & o & 2\lambda_3{}^2/n \end{bmatrix} .$$

But $M'M$ is of rank 1, since the two column vectors are proportional, so fill a one-dimensional subspace, and if $X \sim N(M, \Sigma_1 \otimes \Sigma_2)$ then $\Sigma_2^{-\frac{1}{2}}\ X'\Sigma_1^{-1}X\Sigma_2^{-\frac{1}{2}} = (\Sigma_1^{-\frac{1}{2}}\ X\Sigma_2^{-\frac{1}{2}})'(\Sigma_1^{-\frac{1}{2}}\ X\Sigma_2^{-\frac{1}{2}})$ has latent roots whose distribution is derived from the non-central Wishart. Testing for the second latent root being zero then tests our original hypothesis that the roots are proportional. Using the asymptotic m.l.e.'s for $\lambda_1, \lambda_2, \lambda_3,$

and K which were found above, gives

$$\hat{\Sigma}_2^{-\frac{1}{2}} X' \hat{\Sigma}_1^{-\frac{1}{2}} X \hat{\Sigma}_2^{-\frac{1}{2}} = \begin{pmatrix} 35.16 & 33.84 \\ 33.84 & 35.16 \end{pmatrix}$$

with latent roots 69.003 and 1.314. But these are asymptotically χ^2 on 4 and 2 degrees of freedom respectively, which makes the first root extremely significant ($\geqslant$.001% level) (i.e. $K \neq 1$), and the second is at the 50% level, which makes H_0 quite acceptable.

2.5 A Test for Complexity

Such decompositions into recognizable components are unfortunately not universal, as the following example demonstrates. This example is probably more interesting for the inter-related group and symmetric space structures that arise, and as a concrete example of the injection process for complex and real spaces.

Consider a set of observations from a quadravariate normal population, so for each vector of observations we have:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix} , \quad \Sigma_{4\times 4} \right) .$$

Now suppose one thought that these observations had actually come from a bivariate complex normal population, and one wished to test this hypothesis. That is:

$$H_0 \; : \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} x_1 + ix_2 \\ x_3 + ix_3 \end{bmatrix} \sim \text{BVCN}\left(\begin{bmatrix} M_1 + iM_2 \\ M_3 + iM_4 \end{bmatrix} , \quad \Sigma^*_{2\times 2} \right)$$

with Σ^* Hermitian. Such situations could occur with electronic experiments and with time-series.

Remember that Σ^* Hermitian means $\bar{\Sigma}^{*\prime} = \Sigma^*$, thus it can easily be seen that

$$\Sigma^* = \begin{pmatrix} a^* & c^* - id^* \\ c^* + id^* & b^* \end{pmatrix} , \quad a^*, b^*, c^*, d^* \in R .$$

Now, to see what restrictions this places on Σ (or how to inject the 2×2 Hermitian matrices into the 4×4 real symmetric ones) we note that if Σ^* is invariant under some transformation, say G^*, of C_2, then Σ must be invariant under the equivalent transformation in R_4. With this in mind, let us inspect $G^* = \begin{pmatrix} i & o \\ o & i \end{pmatrix}$. Thus

$$G^* \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \text{BVCN} \left(G^* \begin{bmatrix} \mu_1 + i\mu_2 \\ \mu_3 + i\mu_4 \end{bmatrix}, \quad G^* \Sigma^* \bar{G}^{*\prime} \right)$$

so

$$\begin{bmatrix} ix_1 - x_2 \\ ix_3 - x_4 \end{bmatrix} \sim \text{BVCN} \left(\begin{bmatrix} i\mu_1 - \mu_2 \\ i\mu_3 - \mu_4 \end{bmatrix}, \begin{pmatrix} i & o \\ o & i \end{pmatrix} \begin{pmatrix} a & c-id \\ c+id & b \end{pmatrix} \begin{pmatrix} -i & o \\ o & -i \end{pmatrix} \right)$$

$$\sim \text{BVCN} \left(\begin{bmatrix} i\mu_1 - \mu_2 \\ i\mu_3 - \mu_4 \end{bmatrix}, \begin{pmatrix} a & c-id \\ c+id & b \end{pmatrix} \right).$$

Thus Σ^* is invariant under the action of G^*. But from the action of G^* on a vector in C_2, we see that the equivalent transformation, G, on R_4 will be

$$G = \left[\begin{array}{cc:cc} o & -1 & \multicolumn{2}{c}{o} \\ 1 & o & & \\ \hdashline & & o & -1 \\ \multicolumn{2}{c}{o} & 1 & o \end{array}\right]$$

since

$$G \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \\ -x_4 \\ x_3 \end{bmatrix}$$

which corresponds to

$$G^* \begin{bmatrix} x_1 + ix_2 \\ x_3 + ix_4 \end{bmatrix} = \begin{bmatrix} ix_1 - x_2 \\ ix_3 - x_4 \end{bmatrix}$$

So we must have $\Sigma = G\Sigma G'$. But it is simple to verify that, for

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix}$$

this requires $\sigma_{11} = \sigma_{22}$, $\sigma_{33} = \sigma_{44}$, $\sigma_{12} = \sigma_{34} = 0$, $\sigma_{13} = \sigma_{24}$, and $\sigma_{14} = -\sigma_{23}$, or

$$\Sigma = \begin{bmatrix} a & o & c & d \\ o & a & -d & c \\ c & -d & b & o \\ d & c & o & b \end{bmatrix} .$$

Also, let $\underset{\sim}{y}' = (y_1, y_2, y_3, y_4)$ be an observed vector, for the means. Then under H_0

$$E(\underset{\sim}{y}\ \underset{\sim}{y}') = E\begin{bmatrix} y_1^2 & y_1y_2 & y_1y_3 & y_1y_4 \\ y_1y_2 & y_2^2 & y_2y_3 & y_2y_4 \\ y_1y_3 & y_2y_3 & y_3^2 & y_3y_4 \\ y_1y_4 & y_2y_4 & y_3y_4 & y_4^2 \end{bmatrix} = \begin{bmatrix} a & o & c & d \\ o & a & -d & c \\ c & -d & b & o \\ d & c & o & b \end{bmatrix} = \Sigma .$$

But for $\underset{\sim}{Z}' = (y_1+iy_2\,,\ y_3+iy_3)$, then also under H_0,

$$E(\underset{\sim}{Z}\ \underset{\sim}{\bar{Z}}') = E\begin{bmatrix} y_1^2+y_2^2 & y_1y_3+y_2y_4 - i(y_1y_4-y_2y_3) \\ y_1y_3+y_2y_4 + i(y_1y_4-y_2y_3) & y_3^2+y_4^2 \end{bmatrix} = \begin{bmatrix} a^* & c^*-id^* \\ c^*+id^* & b^* \end{bmatrix}.$$

So, $a^* = 2a$, $b^* = 2b$, $c^* = 2c$, $d^* = 2d$. Since we started with 4 parameters in $M(2,C)$ the space of 2×2 Hermitian matrices, and we now have only 4 parameters for $M(4,R)$ left, thus we have found all the restrictions required for the injection process from $M(2,C) \to M(4,R)$ which is:

$$2\begin{pmatrix} a & c-id \\ c+id & b \end{pmatrix} \quad \to \quad \begin{pmatrix} a & o & c & d \\ o & a & -d & c \\ c & -d & b & o \\ d & c & o & b \end{pmatrix}$$

$$\Sigma^* \quad \to \quad \Sigma .$$

There are obviously no restrictions required on the means under H_0, so the whole problem reduces to testing for Σ taking the above form.

One can easily derive, in the usual way, the likelihood ratio test for H_0, as

$$\lambda \quad \frac{|S|^{\frac{1}{2}n}}{|S^*|^{\frac{1}{2}n}}$$

where S is the ordinary estimate for Σ, for unfixed alternative, say

$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{12} & s_{22} & s_{23} & s_{24} \\ s_{13} & s_{23} & s_{33} & s_{34} \\ s_{14} & s_{24} & s_{34} & s_{44} \end{pmatrix}$$

$$\text{and } S^* = \begin{pmatrix} \frac{s_{11}+s_{22}}{2} & 0 & \frac{s_{13}+s_{24}}{2} & \frac{s_{14}-s_{23}}{2} \\ 0 & \frac{s_{11}+s_{22}}{2} & \frac{s_{23}-s_{14}}{2} & \frac{s_{13}+s_{24}}{2} \\ \frac{s_{13}+s_{24}}{2} & \frac{s_{23}-s_{14}}{2} & \frac{s_{33}+s_{44}}{2} & 0 \\ \frac{s_{14}-s_{23}}{2} & \frac{s_{13}+s_{24}}{2} & 0 & \frac{s_{33}+s_{44}}{2} \end{pmatrix} = \begin{pmatrix} \hat{a} & 0 & \hat{c} & \hat{d} \\ 0 & \hat{a} & -\hat{d} & \hat{c} \\ \hat{c} & -\hat{d} & \hat{b} & 0 \\ \hat{d} & \hat{c} & 0 & \hat{b} \end{pmatrix} .$$

Since the test for a bivariate normal sample actually being a complex sample is just the regular sphericity test (i.e. equal variances, zero correlation), it seemed natural to expect $-2 \log \lambda$ for the 4-D case to split into two sphericity tests, for $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$, and a test for the 'complexity' of the intercorrelations. That is, we <u>expected</u>

$$\lambda^{2/n} = \frac{\det \begin{bmatrix} \hat{a}+\varepsilon_1 & \varepsilon_2 & \hat{c}+\varepsilon_5 & \hat{d}+\varepsilon_6 \\ \varepsilon_2 & \hat{a}-\varepsilon_2 & -\hat{d}+\varepsilon_6 & \hat{c}-\varepsilon_5 \\ \hat{c}+\varepsilon_5 & -\hat{d}+\varepsilon_6 & \hat{b}+\varepsilon_3 & \varepsilon_4 \\ \hat{d}+\varepsilon_6 & \hat{c}-\varepsilon_5 & \varepsilon_4 & \hat{b}-\varepsilon_3 \end{bmatrix}}{\det \begin{bmatrix} \hat{a} & 0 & \hat{c} & \hat{d} \\ 0 & \hat{a} & -\hat{d} & \hat{c} \\ \hat{c} & -\hat{d} & \hat{b} & 0 \\ \hat{d} & \hat{c} & 0 & \hat{b} \end{bmatrix}} \approx 1 - \sum_i k_i \varepsilon_i^2$$

for some constants k_i, where ε_1, ε_2 would test sphericity for the first two variables, ε_3, ε_4 sphericity for the next two, and ε_5, ε_6 the intercorrelations. Unfortunately this did <u>not</u> happen. The structure seems much more complicated. However, since $S^* \in M(4,R)$ we have $S^* = HLH'$ for L diagonal, $H \in O(4,R)$. In fact, it turns out that

$$L = \begin{bmatrix} \ell_1 & & & 0 \\ & \ell_1 & & \\ & & \ell_2 & \\ 0 & & & \ell_2 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} v & -w & -x & -y \\ w & v & y & -x \\ x & -y & v & w \\ y & x & -w & v \end{bmatrix}$$

with H a member of the quaternian group, no less. Also, if $H'S^*H = L$, then

$$\bar{H}^{*\prime}\begin{pmatrix}2a & 2c-i2d\\ 2c+i2d & 2b\end{pmatrix}H^* = \begin{pmatrix}2\ell_1 & \\ & 2\ell_2\end{pmatrix}$$

for $H^* = \begin{pmatrix}v+iw & -x+iy\\ x+iy & v-iw\end{pmatrix}, \in U(2)$. Alternatively we can view this as an injection of $U(2) \to O(4,R)$ in the obvious way.

Then:

$$\lambda^{\frac{2}{n}} = \det S/\det S^*$$

$$= \frac{\det H'SH'}{\det L}$$

$$= \det\begin{bmatrix}\ell_1+\delta_1 & \delta_2 & \delta_5 & \delta_6\\ \delta_2 & \ell_1-\delta_1 & \delta_6 & -\delta_6\\ \delta_5 & \delta_6 & \ell_2+\delta_3 & \delta_4\\ \delta_6 & -\delta_5 & \delta_4 & \ell_2-\delta_3\end{bmatrix}/(\ell_1\ell_2)^2$$

$$\approx 1-\frac{1}{\ell_1^2}(\delta_1^2+\delta_2^2)-\frac{1}{\ell_2^2}(\delta_3^2+\delta_4^2)-\frac{2}{\ell_1\ell_2}(\delta_5^2+\delta_6^2) .$$

So for small δ_i,

$$-2\log\lambda \approx \frac{n}{\ell_1^2}(\delta_1^2+\delta_2^2)+\frac{n}{\ell_2^2}(\delta_3^2+\delta_4^2)+\frac{2n}{\ell_1\ell_2}(\delta_5^2+\delta_6^2) \sim \chi_6^2 .$$

There is not much use in justifying the various component distributions unless one could also attach labels to the components, and ask reasonable questions concerning deviations from complexity, and so far I have been unable to name these δ_i's. However I thought the structures which arise in this problem were interesting enough on their own to warrant their inclusion.

CHAPTER 3

DISTANCE MEASURES ON MANIFOLDS - PROPERTIES AND APPLICATIONS

3.1 Introduction

In this chapter we shall investigate more closely the manifold structures of $M(m,R)$ and $G_{m,p}$. This involves an understanding of what is meant by the distance between two points for non-Euclidean but Riemannian spaces. The concept of distance is generalized via a metric form $(ds)^2 = \sum_{i,j} g_{ij}\, dx_i dx_j$ for $g = (g_{ij})$ a positive definite symmetric matrix, the Riemann structure, and dx_i local coordinates. For Euclidean spaces g is just the identity. A curve joining two points is called geodesic if it is locally the path of shortest length between the points. That is, along short enough segments of the curve, each segment is of minimum length. A geodesic curve may fail to be the shortest route globally however. For example, as a sphere, segments of the great circles are geodesic curves, but a geodesic joining two points could be the long way around, instead of the short way. For arbitrary Riemannian manifolds, calculating geodesic curves and distances can get complicated, but symmetric spaces prove rather easier to handle, since so much of their structure derives from their underlying group structures.

Classically, a symmetric space, S, is a Riemann manifold whose curvature tensor is invariant under parallel translation (i.e. along a geodesic path) - a very differential geometric statement which essentially generalizes the concept of a sphere (of constant curvature). Cartan (1927) proved that this condition was exactly equivalent to S being a factor space $S = K \times G/H$, for G a semi-simple Lie group,

H an isotropy subgroup of G with respect to S, and K a motion group of Euclidean space. The concept of semi-simplicity, though important, is not required here, and will be covered later. Suffice to say that a simple Lie algebra is non-abelian (non-commutative), with no proper ideals, that any semi-simple Lie algebra is composed of simple ones, and that a Lie group is (semi-)simple if its Lie algebra is (semi-)simple. It is curious that such an essentially manifold condition could be so closely related to such strong group concepts. Cartan (1914) had already catalogued the simple Lie groups, so it was then immediately possible to effectively list the irreducible symmetric spaces. We shall study this theory more fully in the next chapter on zonal polynomials, since it is an integral part of the theory of these 'special functions'.

Both $M(m,R) = G\ell(m,R)/O(m,R)$, and $G_{m,p} = O(m+p,R)/O(m,R) \times O(p,R)$ are symmetric spaces, for $G\ell(m,R) = R \times S\ell(m,R)$ and $O(m,R) = Z^2 \times SO(m,R)$, and both $S\ell(m,R)$ and $SO(m,R)$ are semi-simple. Note that any symmetric space S can also be written as $S = G/H$ for G a transitive transformation group and Lie group of S, and H an isotropy subgroup of G which leaves some element $p \in S$ fixed. Using this notation, any reasonable measure put on S must be invariant under G, so that if the distance between $p_1, p_2 \in S$ is $D(p_1,p_2)$, then $D(p_1,p_2) = D(g\cdot p_2, g\cdot p_2)\ \forall\ g \in G$. Also, on a symmetric space, S, the geodesics through any $p \in S$ are of the form $\exp(tX)\cdot p$ for X a member of the Lie algebra of G (see e.g. Helgason (1962) p. 173). These two conditions are just about all we need to find geodesic curves and distances on $M(m,R)$ and $G_{m,p}$.

3.2 The Positive Definite Symmetric Matrices.

Consider p_1, p_2 as two elements of $M(m,R)$. We wish to find a geodesic curve, say $\gamma(t)$, joining p_1 to p_2 with $0 \leqslant t \leqslant 1$ and $\gamma(o) = p_1, \gamma(1) = p_2$. But from the group invariance property, if $\gamma(t)$ is a geodesic joining p_1 to p_2, then $g \cdot \gamma(t) = g\, \gamma(t) g'$ must be a geodesic curve joining $g\, p_1 g'$ to $g\, p_2 g'$, for every $g \in G\ell(m,R)$. But we can always choose g such that $g\, p_1 g' = I_m$, $g\, p_2 g' = L$, for L diagonal, elements $\ell_1, \ell_2, \ldots, \ell_m$ along the diagonal. For we know $p_1 = H_1 L_1 H_1'$, $p_2 = H_2 L_2 H_2'$ for $H_i \in O(m,R)$, L_i diagonal, and if $H_3 \in O(m,R)$ diagonalizes $L_1^{-\frac{1}{2}} H_1 H_2 L_2 H_2' H_1' L_1^{-\frac{1}{2}}$, then $g = H_3 L_1^{-\frac{1}{2}} H_1$ will be such a transformation matrix. So we have shifted the problem to finding a geodesic curve from I to L, where L is now diagonal. But:

$$\left(\exp t(\log L^{\frac{1}{2}})\right) \cdot I = (\exp t \log L/2)\, I (\exp t \log L/2)'$$
$$= \exp t \log L$$
$$= (L)^t$$

is a curve joining I to L, which by our second condition must be a geodesic. So $\gamma(t) = g^{-1}(L)^t(g^{-1})'$ is a geodesic curve joining p_1 to p_2. Even nicer, the distance along $\gamma(t)$ from I to L is simply the length of the straight line, $t \log L$, from $\underset{\sim}{O}$ to $\log L$ in the Lie algebra of $G\ell(m,R)$, which is

$$\sqrt{(\log \ell_1)^2 + (\log \ell_2)^2 + \ldots + (\log \ell_m)^2}\,,$$

since L is diagonal. But this is precisely $D(p_1, p_2)$. (Note that the ℓ_i are the latent roots of $p_1 p_2^{-1}$ or $p_2^{-\frac{1}{2}} p_1 p_2^{-\frac{1}{2}}$.) We can verify this by using the metric form $(ds)^2 = \operatorname{trace}(S^{-1} dS\, S^{-1} dS)$ developed by Maass (1955), which is invariant under $S \to g\, S\, g'$, $g \in G\ell(m,R)$. For S diagonal, elements s_{ii} along diagonal, this becomes $(ds)^2 = \Sigma\left(\frac{1}{s_{ii}} ds_{ii}\right)^2$. So along the curve $(L)^t$, the metric tensor g is diagonal,

$g_{ii} = 1/\ell_i{}^{2t}$ and

$$D(I,\ L) = \int_I^L \sqrt{\sum_{ij}\sum g_{ij} \frac{d}{dt} (\ell_j)^t dt}$$

$$= \int_0^1 \sqrt{\sum_i (\log\ \ell_i)^2 dt} = \sqrt{\sum_i (\log\ \ell_i)^2}\ .$$

3.3 <u>The triangle inequality on $M(m,R)$</u>

The distance measure, D, in $M(m,R)$ is a local metric, but is it global? It is obvious that $D(\Sigma_1, \Sigma_2) \geqslant 0$, $\forall\ \Sigma_1, \Sigma_2 \in M(m,R)$, and that $D(\Sigma_1, \Sigma_2) = 0$ implies $\Sigma_1 = \Sigma_2$. But does D obey the triangle inequality globally? In other words, is a geodesic curve through two points on $M(m,R)$ unique? The answer is yes. We shall prove this explicitly for $m = 2$, and outline the proof for higher dimensions. The proof requires an interesting lemma concerning bounds on the roots of $\Sigma_1 \cdot \Sigma_2^{-1}$, given the roots of Σ_1 and Σ_2.

<u>Theorem</u>: The distance measure D on $M(2,R)$ obeys the triangle inequality.

<u>Proof</u>: Let Σ_1, Σ_2, Σ_3 be any 3 points on $M(2,R)$. Then there exists a $g \in G\ell(2,R)$ such that

$$g\ \Sigma_1\ g' = I$$

$$g\ \Sigma_2\ g' = L = \begin{pmatrix} \ell_1 & 0 \\ 0 & \ell_2 \end{pmatrix}, \quad \ell_1 \geqslant \ell_2$$

$$g\ \Sigma_3\ g' = \Sigma, \quad \text{arbitrary}$$

where

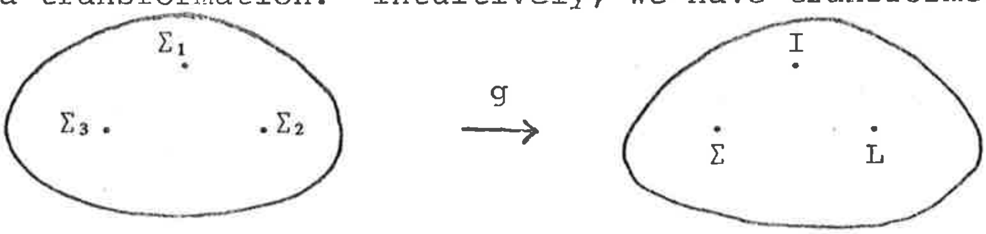
$$D(\Sigma_1,\ \Sigma_2) = D(I,\ \Sigma)$$

$$D(\Sigma_1,\ \Sigma_3) = D(I,\ L)$$

$$D(\Sigma_2,\ \Sigma_3) = D(\Sigma,\ L),$$

from the group invariance of D.

For, if $\Sigma_1 = H_1 L_1 H_1'$, let $g_1 = L_1^{-\frac{1}{2}} H_1'$. Then $g_1 \Sigma_2 g_1' = H_2 L H_2$, for some $H_2 \in O(2,R)$, L diagonal, and $g = H_2 g_1$ will be such a transformation. Intuitively, we have transformed:

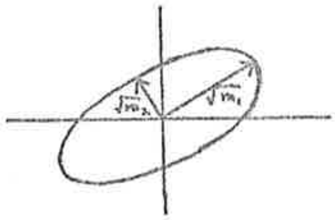
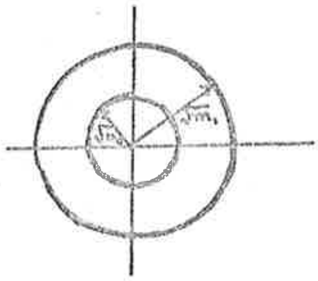


Let m_1, m_2 be the latent roots of Σ, $m_1 \geqslant m_2$. So, $D(\Sigma,I) = \left((\log m_1)^2 + (\log m_2)^2\right)^{\frac{1}{2}}$. Also, $D(I,L) = \left((\log \ell_1)^2 + (\log \ell_2)^2\right)^{\frac{1}{2}}$ and $D(\Sigma,L) = D(\Sigma L^{-1},I) = D(L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}},I) = D(L^{-1}\Sigma,I)$. We need the following lemma.

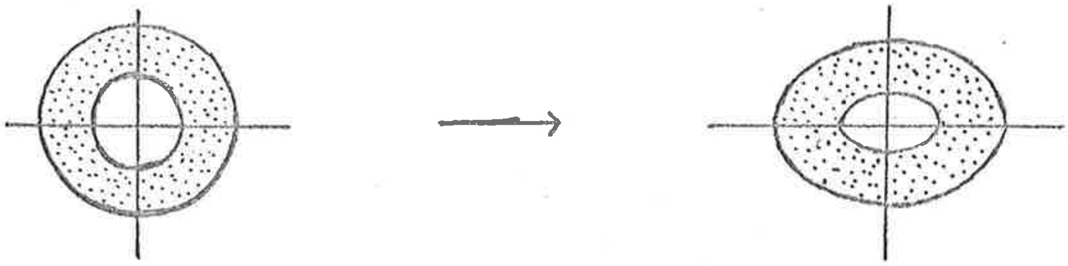
Lemma: If $\Sigma \in M(m,R)$ has roots m_1, m_2, $m_1 \geqslant m_2$, and $L = \begin{pmatrix} \ell_1 & 0 \\ 0 & \ell_2 \end{pmatrix}$, $\ell_1 \geqslant \ell_2$, then the largest possible root of $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ is m_1/ℓ_2, at which time the smallest root, m_2/ℓ_1 also occurs, and Σ is in standard form $\begin{pmatrix} m_2 & 0 \\ 0 & m_1 \end{pmatrix}$.

Proof: Identify Σ with the ellipse $x\Sigma^{-1}x' = 1$, (so that Σ is identified with an ellipse where principal axes are $\sqrt{m_1}$, $\sqrt{m_2}$). Rotating this ellipse through 2π then fills an anullus, radii $\sqrt{m_1}$, $\sqrt{m_2}$,

e.g.

Σ :  , $H\Sigma H', \forall H \in O(2,R)$: 

Under the transformation $\Sigma \to L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$, the anullus becomes deformed as



e.g. $x\Sigma^{-1}x' = 1 \to x(L^{\frac{1}{2}}\Sigma^{-1}L^{\frac{1}{2}})x' = 1$

where the outer circle, radius $\sqrt{m_1}$ becomes the ellipse

$$x\begin{pmatrix} \sqrt{\frac{m_1}{\ell_2}} & \\ & \sqrt{\frac{m_1}{\ell_1}} \end{pmatrix}x' = 1,$$ the inner one, radius $\sqrt{m_2}$ becomes

$$x\begin{pmatrix} \sqrt{\frac{m_2}{\ell_2}} & \\ & \sqrt{\frac{m_2}{\ell_1}} \end{pmatrix}x' = 1.$$ Thus every ellipse, $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ is inside the deformed annullus. It is thus obvious that the

very largest possible root for any $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ is then $\frac{m_1}{\ell_2}$, at which time the second root is $\frac{m_2}{\ell_1}$, which is also the smallest possible root, for $\mathrm{Det}(L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}) = m_1 m_2/\ell_1\ell_2$, a constant. Also these values only occur when Σ is in standard form $\Sigma = \begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix} = \begin{pmatrix} o & -1 \\ 1 & o \end{pmatrix}\begin{pmatrix} m_1 & o \\ o & m_2 \end{pmatrix}\begin{pmatrix} o & 1 \\ -1 & o \end{pmatrix}$. Q.E.D.

So any pair of roots of $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ will take the form $\frac{m}{a\ell_2}$ and $\frac{am_2}{\ell_1}$, for $1 \leqslant a \leqslant \frac{m_1/\ell_2}{m_2/\ell_1}$. We first show that the triangle inequality holds for the extreme root values, then that $D(\Sigma,L)$ is a maximum at these values.

Suppose the roots of $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ are $\frac{m_1}{\ell_2}$, $\frac{m_2}{\ell_1}$ so $\Sigma = \begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix}$.

Then

$$\left(D\left(\begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix},L\right)\right)^2 = \left(D\left(\begin{pmatrix} \frac{m_1}{\ell_2}, & o \\ o & \frac{m_2}{\ell_1} \end{pmatrix},I\right)\right)^2$$

$$= (\log m_1 - \log \ell_2)^2 + (\log m_2 - \log \ell_1)^2$$

$$= (\log m_1)^2 + (\log m_2)^2 + (\log \ell_1)^2 + (\log \ell_2)^2$$

$$- 2(\log m_1 \log \ell_2 + \log m_2 \log \ell_1).$$

Also $\left(D\left(\begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix},I\right) + D\left(I,\begin{pmatrix} \ell_1 & o \\ o & \ell_2 \end{pmatrix}\right)\right)^2$

$$= (\log m_1)^2 + (\log m_2)^2 + (\log \ell_1)^2 + (\log \ell_2)^2$$

$$+ 2\sqrt{((\log m_1)^2 + (\log m_2)^2)((\log \ell_1)^2 + (\log \ell_2)^2)}$$

But $(\log m_1)^2(\log \ell_1)^2 + (\log m_2)^2(\log \ell_2)^2 \geqslant 2 \log m_1 \cdot \log m_2 \cdot \log \ell_1 \cdot \log \ell_2$ (simply the well-known inequality $x^2+y^2 \geqslant 2xy$). Adding $(\log m_1)^2(\log \ell_2)^2 + (\log m_2)^2(\log \ell_1)^2$ to both sides of this inequality, and taking square roots, gives

$$+(((\log m_1)^2+(\log m_2)^2)((\log \ell_1)^2+(\log \ell_2)^2))^{\frac{1}{2}} \geqslant \pm(\log m_1 \log \ell_2 + \log m_2 \log \ell_1).$$

So by inspection, $\left(D\left(\begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix},I\right) + D\left(I,\begin{pmatrix} \ell_1 & o \\ o & \ell_2 \end{pmatrix}\right)\right)^2$

$$\geqslant \left(D\left(\begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix},\begin{pmatrix} \ell_1 & \\ & \ell_2 \end{pmatrix}\right)\right)^2$$

giving $D\left(\begin{pmatrix} m_2 & \\ & m_1 \end{pmatrix}, I\right) + D\left(I, \begin{pmatrix} \ell_1 & o \\ o & \ell_2 \end{pmatrix}\right) \geq D\left(\begin{pmatrix} m_2 & \\ & m_1 \end{pmatrix}, \begin{matrix} \ell_1 & \\ & \ell_2 \end{matrix}\right) = D\left(\begin{matrix} \frac{m_1}{\ell_2}, & \\ & \frac{m_2}{\ell_1} \end{matrix}, I\right)$

So for extreme values, the inequality holds.

Let $\frac{m_1}{\ell_2} = x_1$, $\frac{m_2}{\ell_1} = x_2$.

Then any pair of roots of $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ will be $\frac{x_1}{a}$, ax_2 with $1 \leq a \leq x_1/x_2$.

And: $D(\Sigma, L) = \left(\left(\log \frac{x_1}{a}\right)^2 + (\log ax_2)^2\right)^{\frac{1}{2}}$

$$= ((\log x_1)^2 + (\log x_2)^2 + 2(\log a)^2 - 2 \log a(\log x_1 - \log x_2))^{\frac{1}{2}}$$

But, since $1 \leq a \leq x_1/x_2$, we have

$$o \leq \log a \leq \log x_1 - \log x_2$$

$$o \leq (\log a)^2 \leq \log a (\log x_1 - \log x_2)$$

$$o \leq 2(\log a)^2 \leq 2 \log a (\log x_1 - \log x_2) .$$

So $2(\log a)^2 - 2 \log a (\log x_1 - \log x_2) \leq 0$

$\therefore D(\Sigma, L) \leq \left((\log x_1)^2 + (\log x_2)^2\right)^{\frac{1}{2}} = D\left(\begin{pmatrix} m_2 & o \\ o & m_1 \end{pmatrix}, \begin{pmatrix} \ell_1 & o \\ o & \ell_2 \end{pmatrix}\right)$

$\therefore D(\Sigma, L) \leq D(\Sigma, I) + D(I, L)$

$\therefore D(\Sigma_2, \Sigma_3) \leq D(\Sigma_2, \Sigma_1) + D(\Sigma_1, \Sigma_3)$

for any 3 arbitrary points on $M(2,R)$.

Thus the triangle inequality holds an all of $M(2,R)$ with metric D, and any two points on $M(2,R)$ can be joined by a unique curve whose distance is a global minimum.

Generalizing this proof becomes an exercise in combinatorics. Let $\Sigma = M(p,R)$ have roots $m_1, \ldots, m_p$ with $m_i \geq m_{i+1}$, and let L be diagonal, with diagonal elements $\ell_1, \ldots, \ell_p$; $\ell_i \geq \ell_{i+1}$. Then the largest possible root of $L^{-\frac{1}{2}}\Sigma L^{-\frac{1}{2}}$ is m_1/ℓ_p, again found by deforming the sphere, radius $\sqrt{m_1}$, traced out by all possible rotations of $\tilde{x}\,\Sigma^{-1}\tilde{x} = 1$. In this case, Σ has part standard form

$$\left(\begin{array}{ccc:c} & & & o \\ & * & & \vdots \\ & & & o \\ \hdashline o & \cdots & o & m_1 \end{array}\right) .$$

Repeating the process, conditioning on this root, gives the next largest root as m_2/ℓ_{p-1}, and Σ as

$$\left(\begin{array}{c:cc} & 0 & 0 \\ * & \vdots & \vdots \\ & 0 & 0 \\ \hdashline 0\ldots0 & m_2 & 0 \\ 0\ldots0 & 0 & m_1 \end{array}\right).$$

The next largest root, conditional on these two, is then m_3/ℓ_{p-2}. But this remains true, even if we only require Σ to have form

$$\left(\begin{array}{c:cc} & 0 & 0 \\ * & \vdots & \vdots \\ & 0 & 0 \\ \hdashline 0\ldots0 & \multicolumn{2}{c}{\Sigma^*_{2\times 2}} \\ 0\ldots0 & & \end{array}\right)$$

where Σ^*_{22} need only have roots m_1, m_2. In other words, the product of the two top roots need only remain constant in this subspace, (i.e. $m_1/a\ell_p$ and am_2/ℓ_{m-1} for $1 \leqslant a \leqslant \frac{m_1/\ell_p}{m_2/\ell_{p-1}}$) for the next largest root to take extreme value m_3/ℓ_{p-2}.

Eventually we get the extreme values of all the roots as m_1/ℓ_p, $m_2/\ell_{p-1}, \ldots, m_p/\ell_1$, where each root is as large as possible, given the product of the roots above it. For these roots one can easily show that triangle inequality holds. But any other root system can be reached by progressive dilation-shrinkages of the form $\frac{x_1}{a}$, $x_2 a$ used in the 2-dimensional case, because of our conditioning requirements. For, consider the extreme roots strung on a line, and the positions to which we wish to shift also marked. Choose any new position. Then by shrinking the closest larger extreme root, dilating the closest smaller one, till the new value is reached by one of them, which we then fix permanently (reduce dimension by 1), the unfixed root is then at the largest value it can take, given the fixed root, and the ones above it. It cannot get bigger, without moving a larger extreme root down, or get smaller except by moving a smaller extreme root up. But the smaller roots are also still at extreme values since they are only conditional on the product of all the roots above each of them, which has not changed. Obviously we can repeat this procedure as many times as necessary. But progressive changes of the form $\frac{x_i}{a}$, $x_{i+1}\, a$, $1 \leqslant a \leqslant x_i/x_{i+1}$ can be shown to keep the $D(\Sigma, L) \leqslant D\left(\begin{pmatrix} m_p \ldots m_1 \end{pmatrix} \begin{pmatrix} \ell_1 \ldots \ell_p \end{pmatrix}\right)$. So the triangle inequality holds for general dimension p.

Notice that having L diagonal in the lemma is no real restriction. For, let $\Sigma_1, \Sigma_2 \in M(p,R)$, with roots $(\ell_1, \ldots, \ell_p), (m_1, \ldots, m_p)$ as usual. Then the roots of $\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}} = HL^{-\frac{1}{2}}H'\Sigma_2 HL^{-\frac{1}{2}}H'$ are the same as the roots of $L^{-\frac{1}{2}}H'\Sigma_2 HL^{-\frac{1}{2}}$, where $\Sigma_1 = HLH'$, $H \in O(p,R)$, and $H'\Sigma_2 H$ has the same roots as Σ_2. So the lemma holds for arbitrary Σ_1, Σ_2, with known roots. This lemma appears to be in some sense a generalization of the Poincaré separation theorem (Bellman (1960), p. 117) which deals with bounds on the roots of one matrix, as the dimension increases. It may prove useful for finding bounds in multivariate work, where functions of $H'\Sigma_1 H\Sigma_2$; $\Sigma_1, \Sigma_2 \in M(m,R)$, $H \in O(m,R)$ often occur.

3.4 Statistical Uses

This structure on $M(m,R)$ can be used to test various hypotheses concerning Wishart matrices. For example, suppose S is a sample Wishart matrix, $S \sim W(n,m,\Sigma)$, and one wishes to make inferences about Σ. From James (1973) (where there is another derivation of the distance measure), if $\ell_1, \ell_2, \ldots, \ell_m$ are the latent roots of $S\Sigma^{-1}$, then

$$d^2 = \tfrac{1}{2} n \sum_{i=1}^{m} (\log \ell_i)^2$$

is asymptotically distributed as χ^2 on $\tfrac{1}{2}m(m+1)$ degrees of freedom. So d^2 can be used to test for membership of S in various possible populations. The distance metric could also be used to investigate relationships among sets of variance-covariance matrices, perhaps as a spatial aid in such fields as discriminant analysis or cluster analysis. Sometimes this metric aids in understanding seemingly peculiar results in multivariate work. For example, we had originally used the Jolicoeur-Mosimann data to numerically try out a test for the existence of a simple scalar difference between two population Wishart matrices. That is, for $S_1 \sim W(n_1,m,\Sigma_1)$ and $S_2 \sim W(n_2,m,\Sigma_2)$, we wished to test that $\Sigma_1 = K\Sigma_2$ for undefined K. Under such a null hypothesis, the derivatives of the likelihood function yield two transcendental equations:

$$\Sigma_2 = \frac{1}{n_1+n_2}(n_2 S_2 + n_1 S_1/K)$$

$$K = \frac{1}{m}\operatorname{tr}(S_1\Sigma_2^{-1}).$$

Evaluating these iteratively gave estimates

$$\hat{\Sigma}_2 = \begin{bmatrix} 162.67 & 95.62 & 53.55 \\ * & 60.51 & 32.17 \\ * & * & 19.40 \end{bmatrix} \qquad \hat{K} = 2.41938$$

But the original data was:

$$S_1 = \begin{bmatrix} 451.39 & 271.17 & 168.70 \\ * & 171.73 & 103.29 \\ * & * & 66.65 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 138.75 & 79.15 & 37.38 \\ * & 50.04 & 21.65 \\ * & * & 11.26 \end{bmatrix}$$

and our estimates of Σ_2 is uniformly 'larger' than S_2, while our estimate of K is smaller than we would expect, since the individual scalar differences are all greater than 3. Also $\hat{K}\,\hat{\Sigma}_2$ is 'smaller' than S_1. These results do not appear so strange when we use the natural metric. For, if $\Sigma_2 = K\Sigma_1$, then $S_2 S_1^{-1} \approx KI$. Let $x_1, \ldots x_m$ be the latent roots of $S_2 S_1^{-1}$. Then the value of K for which $D(S_2 S_1^{-1}, KI) = D\left(\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_m \end{pmatrix}, KI\right)$ is a minimum is just at $K = \sqrt[m]{x_1 \ldots x_m} = \sqrt[m]{\dfrac{k_1 \ldots k_m}{p_1 \ldots p_m}}$. For,

$$\frac{\partial}{\partial K}\left(D\left(\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_m \end{pmatrix}, KI\right)\right)^2 = \frac{\partial}{\partial K} \sum_{i=1}^{m} \left(\log \frac{x_i}{K}\right)^2$$

$$= \sum_{i=1}^{m} -2\frac{\left(\log \frac{x_i}{K}\right)}{K} = 0$$

giving:

$$\log \; x_1 \ldots x_m = \log \; (K^m)$$

$$K = \sqrt[m]{x_1 \ldots x_m} = \sqrt[m]{\frac{k_1 \ldots k_m}{p_1 \ldots p_m}}.$$

For the Jolicoeur-Mosimann data, this is 2.5156, which is much closer to the m.l.e. of K than the value of ~ 3 which our intuitive Euclidean minds had expected.

Once again, as in the previous analyses, the smaller roots play as important a role in any calculations as the largest one, which emphasizes the importance of removing random noise variables before doing tests such as these.

3.5 The Grassman Manifold

The Grassman manifold is fundamentally different from the space of positive definite symmetric matrices, being compact. So we cannot expect a global metric, obeying a triangle inequality. But with some care, the geodesic curves and distances on $G_{m,p}$ can still be useful. $G_{m,p}$ was constructed as the coset space $O(m+p,R)/O(m,R) \times O(p,R)$ where each point x is a hyperplane, represented as $[H_1 \vdots H_2]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}$, where H_1 is any set of m orthogonal column vectors spanning x, and H_2 similarly spans the orthogonal compliment of x. Again we make use of canonical forms. If $x_1, x_2 \in G_{m,p}$ then

$$x_1 \leftrightarrow [H_{11} \vdots H_{21}]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}$$

$$x_2 \leftrightarrow [H_{12} \vdots H_{22}]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}.$$

But $D(x_1, x_2)$ must be invariant under orthogonal transformations, so

$$\begin{aligned} D(x_1, x_2) &= D([H_{11}, H_{21}]' \cdot x_1, \quad [H_{11}, H_{21}] \cdot x_2) \\ &= D\left(I\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}, \ [H_{13}, H_{23}]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}\right). \end{aligned}$$

Now,

$$I\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}$$

is invariant under any transformation of the form

$$\begin{bmatrix} H_1 & \\ & H_2 \end{bmatrix} \in \begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}$$

so

$$D\left(I\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}, \ [H_{13}, H_{23}]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}\right)$$

$$= D\left(I\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}, \ \begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}[H_{13}, H_{23}]\begin{bmatrix} O(m,R) & 0 \\ 0 & O(p,R) \end{bmatrix}\right)$$

and one can always find members of $O(m,R) \times O(p,R)$ which will turn $[H_{13}, H_{23}]$ to canonical form. That is (assuming $m \leqslant p$):

$$x = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & & & & & & & & \\ -\sin\theta_1 & \cos\theta_1 & & & & & & & & \\ & & \cos\theta_2 & \sin\theta_2 & & & & & & \\ & & -\sin\theta_2 & \cos\theta_2 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & \cos\theta_m & \sin\theta_m & & & \\ & & & & & -\sin\theta_m & \cos\theta_m & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix},$$

with block diagonal entries $\begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}$ or 1 along the diagonal, zeros elsewhere. But from Roy (1947), these θ_i are simply the critical angles between the two planes, x_1 and x_2. One curve joining I to x is obviously defined by

$$\begin{bmatrix} \cos t\theta_1 & \sin t\theta_1 & & & & & & & & \\ -\sin t\theta_1 & \cos t\theta_1 & & & & & & & & \\ & & \cos t\theta_2 & \sin t\theta_2 & & & & & & \\ & & -\sin t\theta_2 & \cos t\theta_2 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & \cos t\theta_m & \sin t\theta_m & & & \\ & & & & & -\sin t\theta_m & \cos t\theta_m & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix},$$

with zeros everywhere else.

But this is just

$$\exp t \begin{pmatrix} 0 & \theta_1 & & & & & & & \\ -\theta_1 & 0 & & & & & & 0 & \\ & & 0 & \theta_2 & & & & & \\ & & -\theta_2 & 0 & & & & & \\ & & & & \ddots & & & & \\ & 0 & & & & 0 & \theta_m & & \\ & & & & & -\theta_m & 0 & & \\ & & & & & & & 0 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 0 \end{pmatrix},$$

which means that path is geodesic.

So the geodesic joining x_1 to x_2 can be found by transforming back again. As before, one can verify that the arc length from x_1 to x_2 is again the absolute length of the line from $\underset{\sim}{o}$ to

$$\begin{pmatrix} 0 & \theta_1 & & & & & & \\ -\theta_1 & 0 & & & & 0 & & \\ & & \ddots & & & & & \\ & & & 0 & \theta_m & & & \\ & 0 & & -\theta_m & 0 & & & \\ & & & & & 0 & & \\ & & & & & & \ddots & \\ & & & & & & & 0 \end{pmatrix}$$

in the Lie algebra of $O(m+p,R)$ which is just

$$\sqrt{\theta_1^2 + \theta_2^2 + \ldots + \theta_m^2} = D(x_1, x_2).$$

Notice that although x_1 and x_2 were both m-dimensional hyperplanes in R^{m+p}, $m \leqslant p$, that this restriction is actually unnecessary. For different dimension planes, there are only as many critical angles as the smaller of the two planes. For this reason it is often just as easy to work out generalized results on the Grassmann manifolds. This point is important later on, when we study generalized functions on symmetric spaces.

3.6 Applications

Several statistical tests suggested themselves, based on the critical angles. For from Hotelling (1936), the sample canonical correlations, $\mathrm{Cos}^2\,\theta_1$, $\mathrm{Cos}^2\,\theta_2,\ldots,\mathrm{Cos}^2\,\theta_m$, between two sets of observed vectors, X_1 and X_2, are simply the squares of the cosines of the critical angles between the planes spanned by X_1 and X_2. A test for independence of the two sets is then a test that the population canonical correlations are zero, or that the population critical angles are $\pi/2$. As correlations, we have the (admittedly) asymptotic result for the canonical correlations, $\cos^2\,\theta_i$,

$$\mathrm{Cos}^2\,\theta_i \sim N\left(\mathrm{Cos}^2\,\psi_i,\ (1-\mathrm{Cos}^2\,\psi_i)^2/n\right)$$

where the $\mathrm{Cos}^2\,\psi_i$ are the population canonical correlations. As functions of the canonical correlations then:

$$\theta_i \sim N(\psi_i,\ \tan^2\psi_i/4n).$$

As a test for given ψ_i being the population critical angles, one could then use:

$$4\,n\sum_i (\theta_i-\psi_i)^2/\tan^2\theta_i \sim \chi^2_m.$$

Given two sets of sample canonical correlations, $\mathrm{Cos}^2\,\theta_i$, $\mathrm{Cos}^2\,\rho_i$, one might wish to test whether they in fact came from identical populations, in which case:

$$2\,n\,\Sigma\,(\theta_i-\rho_i)^2/\tan^2\left((\theta_i+\rho_i)/2\right) \sim \chi^2_m$$

yields such a test.

CHAPTER 4

4.1 Introduction

Most people who have studied the theory of multivariate distributions, within the last fifteen years, are at least aware of the existence of the generalized hypergeometric functions of matrix argument,

$${}_pF_q\left((a_1),\ldots,(a_p);\quad (b_1),\ldots(b_q);\; X\right) \quad ,\; X \in M(m,R)$$

which occur in the non-central parts of the non-null densities derived from the multivariate normal. [see e.g. James (1961), (1964)]. However, few statisticians have any real understanding of what these functions are all about, and a great deal of mystery appears to surround those illusive creatures, the zonal polynomials, $C_k(X)$, which generate the ${}_pF_q$ functions,

$${}_pF_q\;(\ldots;\;\ldots;\;X) \;=\; \sum_{k=0}^{\infty}\;\sum_{k}\;\frac{(a_1)\ldots(a_p)}{(b_1)\ldots(b_q)}\;\;C_k(X). \tag{7}$$

The aim of this chapter is to provide a firm background for these functions, by placing them in their natural symmetric space context. Hopefully this will give the reader an intuitive grasp of what a zonal polynomial is, what it can do, and why. This approach has as far as I know, been rather neglected in the statistical literature, and it is precisely this side which provides a perspective on the subject.

It seems only fair to mention that this topic is difficult. It involves much high-powered mathematics which in such a short essay can only be dealt with superficially. But I hope that this outline, the result of several years study of the subject, will provide enough landmarks and sign-posts to make the same task easier for others.

If one point has become clear in the previous work, I hope it is the idea that the various matrix groups and spaces encountered in statistics have a most definite structure and shape attached to them, by which their behaviour is controlled. Their statistical importance is only incidental to their existence. With this in mind, we turn to more advanced Lie group and symmetric space theory.

One view of symmetric spaces is that of a simple extension part Euclidean space, where one relaxes the requirement of 'flatness' according to certain rules. The resulting symmetry of these spaces has meant that such topics as integration theory, fourier analysis and differential operators can be rewardingly studied on them. Basically one can study global functions on symmetric spaces using the same techniques as in Euclidean functional analysis. It is clear that such things as compactness versus non-compactness will play a vital role in this theory, since this is analogous to the problems encountered in function theory on making a simple change from R^1 to S^1, the simplest symmetric space, $S^1 \approx SO(2,R)/SO(1,R) = SO(2,R)$ (except R^1 itself!).

Fundamentally, there are certain unique functions attached to every symmetric space, $S = G/H$, the zonal spherical functions. These are the only functions, $f(g)$, on G which are invariant under transformations $f(g) \to f(h_1 g h_2)$, $\forall\, h_1, h_2 \in H$, and with the added property that

$$\int_H f(g_1 h g_2)\, dh = f(g_1) \cdot f(g_2).$$

From the first property we see that these two-sided invariant (under H) functions on G can be considered as one-sided invariant functions on S, by associating cosets gH with

$m \in S$. The second property allows integration of bi-invariant functions over H if they can be expressed in terms of these zonal spherical functions. The theory of fourier analysis for such bi-invariant has been much developed in recent years [e.g. Harish-Chandra (1958) I and II, Helgason (1962)]. A.T. James' zonal polynomials are the zonal spherical functions of $M(m,R)$, the space of positive definite symmetric matrices. Much of their value in multivariate analysis derives from their defining properties. Both their existence and form require deep results from the theory of group representations on semi-simple Lie groups, and this is what we are going to investigate. General reference texts are listed and described at the end of section 4.3.

The special functions of mathematical physics, which are also known as the higher transcendental functions, have been around for a long time, as we can see from their names - Bessel, Laguerre, Jacobi, Legendre, etc. These analytic functions of one variable possess such amazing and varied properties that many mathematicians made them their lives' work. They obey characteristic differential equations, addition, product, duplicating and generating formulae, and appear time and time again as solutions to physical problems [e.g. see Erdelyi et al. (1953) V.1,2,3]. Cartan (1929) first noticed a connection between special functions and Lie group theory, and since that time this intimate relationship has been much studied, especially since it proved fundamental to the theory of quantum physics (see e.g. Boerner (1963)). It was found that most of the special functions were related to particular Lie transformation groups via group representation theory. They turn up as matrix elements in certain group representations, and most properties arise as a result of the group behaviour. Much work has been done lately in filling in gaps in the representation theory, so that all properties of the special functions could be derived from the Lie group approach, and I believe this work has been substantially completed [see e.g. Miller (1968), Vilenkin (1968), Talman (1968)].

Unfortunately, many of these derivations appear to have been worked out only with the aid of the solutions. Rigorous theoretical results do not appear practical enough, and practical examples do not appear logical enough. It is quite pleasant then to find that the statisticians have got in ahead of the applied mathematicians, and used these techniques to derive completely new, and as yet little understood functions, which are matrix generalizations of the well-known univariate special functions.

The concept of a group representation is fundamental. Let G be an arbitrary group (finite or otherwise), V an arbitrary vector space of dimension n (possibly ∞), and $G\ell(n,V)$ the group of linear invertible transformations on V. Then, for $g \in G$, a transformation $T: g \to T(g) \in G\ell(n,V)$ is called a group representation if it takes the group operations with it. That is,

$$T(g_1 \cdot g_2) = T(g_1) \cdot T(g_2)$$

$$T(g^{-1}) = (T(g))^{-1}$$

$\big(\Rightarrow T(e) = I$, for e identity of G, I the identity of $G\ell(n,V)\big)$.

Notice that $T(G)$ by no means has to fill $G\ell(n,V)$, just be a subgroup of it. It is difficult to appreciate the importance of this structure at first glance, but consider the following questions. Suppose we choose a basis for V, so the $T(g)$'s take matrix form. Will there be subspaces of V invariant under all $T(g)$'s? That is, can a $V_i \in V$ exist, such that

$$T(g)v \in V_i \quad \forall\ v \in V_i, \quad g \in G\ ?$$

If V_i is such a subspace, is $V_i^{\perp}$ also invariant? Can we repeat this process till no further decompositions exist? Is so, will any two such complete decompositions be necessarily equivalent? Naturally, if the answers were always no, our representation structure would not be very interesting, but much time and space in textbooks (see e.g. Boerner (1963) and Miller (1972)) is spent in proving when and how these questions are answered affirmatively. Essentially, any vector space with an invariant inner product will decompose into mutually orthogonal subspaces, each invariant under T, and this decomposition can be continued till the smallest blocks are reached (the irreducible spaces). If $V = \bigoplus_{i=1}^{p} V_i$ each V_i irreducible variant, then we can construct $T_i(g)$'s such that $T(g) = \bigoplus_{i=1}^{p} T_i(g)$. Further, if v_j, $j = 1,\ldots,m$ is a basis for V_i, and w_j, $j = 1,\ldots,m$ is another, (or a basis for V_k, say) and

$$\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = B \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}, \ B \in G\ell(m,V_i) \ \text{(or } Aut(Vi,Vk))$$

then under $v_j \to w_j$, $T_i(g) \to B^{-1}T_i(g)B = T_k(g)$ and the two representations are said to be equivalent. So one can further collect the $T_i(g)$'s into equivalence classes, and by judicious choice of basis have

$$T(g) = \begin{bmatrix} \bigoplus_{m_2} T_1(g) & & & O \\ & \bigoplus_{m_2} T_2(g) & & \\ & & \ddots & \\ O & & \bigoplus_{m_i} T_i(g) & \\ & & & \ddots \end{bmatrix}$$

where each $T_i(g)$ is irreducible, and inequivalent to $T_j(g)$,

$i \neq j$, and repeated m_i times in the representation of G in V. One of the beautiful results of representation theory is that any two complete decompositions of G in V will be identical up to equivalence.

The study of group representations is basically to study all possible representations of G in every possible vector space, to find out when one has equivalent representations, and to be able to label or categorize the irreducible inequivalent ones in some way. Quite a formidable task, one might say. That any of this is possible is due to the fact that it is the group itself which determines what irreducible representations are allowable. Any particular vector space, V, is only picking up what ones it can from the set of all possible representations. On the other hand, if one knew, for example, how many finite-dimensional irreducible representations there were, and had turned up that number of explicit inequivalent representations in some space, then one would know that it was the complete finite-dimensional set.

For a Lie group, G, one can extend the notion of a group representation, T, to a representation, $\mathcal{T}$, of its Lie algebra, $\mathcal{G}$, with the product rule replaced by

$$\mathcal{T}([g_1, g_2]) = \mathcal{T}(g_1)\cdot\mathcal{T}(g_2) - \mathcal{T}(g_2)\cdot\mathcal{T}(g_1) \quad . \qquad \forall\, g_1, g_2 \in \mathcal{G} .$$

This proves a fruitful extension, connecting T and $\mathcal{T}$ by the exponential map, even though in general one lacks a complete one-to-one relationship between the irreducible group and algebra representations (from G being only locally defined for a given $\mathcal{G}$).

4.2 Root space decompositions of semi-simple Lie groups.

The representation theory of semi-simple Lie groups and their blood-relations ($G\ell(m,R)$, $O(m,R)$) becomes particularly clear-cut and illuminating owing to the very considerable structure of their algebras, which allows all finite dimensional representations to be found. It is this structure which we are going to investigate, in gory detail. The literature on this, the root space decomposition theory of semi-simple Lie algebras, is invariably done in such generality as to be incomprehensible, losing all intuition and form. We hope to avoid this by explicitly dissecting $\mathcal{L}SO(m,R)$, the Lie algebra of the special orthogonal group, and to a lesser degree, the Lie algebra of the special linear group, $\mathcal{L}S\ell(m,R)$.

If one can compare this process to a dissection, then the Killing form, K, takes the place of a scalpel. K is a symmetric bilinear form on a Lie algebra, G, defined as $K(X,Y) = \text{trace}$ (adjoint $X \cdot$ adjoint Y), where (adjoint $X)\cdot Z = [X,Z] = XZ - ZX$, for $X, Y, Z \in G$. One considers adjoint $X \cdot$ adjoint Y as a linear transformation on G, and K becomes an inner product (possibly degenerate: $K(X,Y) = 0$ $\forall X \in G \not\Rightarrow Y = 0$).

A Lie algebra is called semi-simple if it has no solvable ideals. An ideal is the algebra generalization of a normal subgroup, where a subalgebra, $\mathcal{H}$, of G is an ideal if $[g,h] \in \mathcal{H}$ $\forall$ $g \in G$, $h \in \mathcal{H}$. With this definition, the factor space $G/\mathcal{H}$ becomes a subalgebra, in the same way that G/H is a subgroup for H a normal subgroup of G. Solvability means that the derived series,

$$\mathcal{H}, [\mathcal{H},\mathcal{H}], [[\mathcal{H},\mathcal{H}],[\mathcal{H},\mathcal{H}]], \ldots$$

goes to zero sooner or later, for $\mathcal{H}$ a Lie (sub-) algebra. Since the last ideal before 0 must be abelian (i.e.

$[h_1, h_2] = 0 \ \forall\ h_1, h_2 \in \mathcal{H} \Rightarrow \mathcal{H}$ is abelian) we have semi-simplicity being equivalent to G containing no abelian ideals at all.

The special nature of semi-simple Lie algebras becomes more apparent when it can be shown that this restriction is equivalent to the Killing form being non-degenerate, thus making it an important tool for semi-simple algebras. Roughly, one shows that if G is not semi-simple, then it contains at least one ideal, the abelian one defined above, on which K is degenerate. For, if A is an abelian ideal, then for $a_1, a_2 \in A$, adjoint $a_1 \cdot g = [a_1, g] \in A \ \forall\ g \in G$, so adjoint $a_2 \cdot [a_1, g] = 0$ and $K(a_1, a_2) = 0 \ \forall\ a_1, a_2 \in A$. Alternatively, if K is degenerate, then the space $A = \{g_i \in G | K(g_i, g) = 0 \ \forall\ g \in G\}$ can be shown to be solvable (see e.g. Samelson (1969) §1, sections 6-10). Notice that the restriction of the Killing form to an ideal is the same as the Killing form of the restriction, although this is not true for arbitrary subalgebras, since the ideal structure drags everything into itself the first go. A Lie group is called semi-simple if its Lie algebra is semi-simple.

Our strengthened inner product can now be used to show that any semi-simple algebra is the direct sum of simple ones (those with no non-trivial ideals at all). This is accomplished by factoring out ideals which by definition are not solvable, using the Killing form to find orthogonal complements. For, if G_1 is any ideal in G, then $G_1^{\perp} = \{X \in G | K(X,Y) = 0 \ \forall\ Y \in G_1\}$ and from the non-degeneracy of K, $G_1 \cap G_1^{\perp} = 0$, so $G = G_1 \oplus G_1^{\perp}$. Repeating this process finally gives a direct sum decomposition of G in terms of simple algebras.

It is a quite remarkable fact that there are essentially only 9 simple Lie algebras, each corresponding to a precise connected tree graph. It then follows that the semi-simple algebras (and thus their respective groups) correspond to combinations of these trees. From a particular graph (the Dynkin diagram), one can recover (up to a local isomorphism) the particular semi-simple group associated with it.

The first step in deriving these diagrams is to extend to the complex field. Each semi-simple real lie algebra is a real form of a semi-simple complex one and each semi-simple complex algebra is the complexification of a real one. In practice this is quite easy. If G is real, then the complexification G^c is only $G^c = \{X + iY | X, Y \in G\}$. Thus a basis for G is also a basis for G^c, where one now allows multiplication by complex numbers. In essence, one deals with the complexified algebra exactly as if it were real, except that one can now found eigenvalues to certain determinental equations which occur in the root space decompositions of the algebras, which is what this whole theory is called. A complex semi-simple Lie algebra may have several different real forms, however (though only one compact one). Much of the detailed exposition of Lie groups and symmetric spaces involves switching back and forth, real to complex, compact to non-compact, proving results in whichever demain is easiest, then carrying them back to others if possible.

Of the nine simple complex Lie groups, 4 are 'natural', the rest existing on theoretical grounds, but not appearing in any natural historic manner. The natural ones are $SL(n+1,C)$, $n \geqslant 1$; $SO(2n+1,C)$, $n \geqslant 2$; $S_p(n,C)$, $n \geqslant 3$, the symplectic group; and $SO(2n,C)$, $n \geqslant 4$. $SL(n+1,C)$ is the complexification of $SU(n+1)$, the rest including the rare ones are complexifications of themselves with C replaced by R.

The maximal abelian subalgebras of a semi-simple Lie algebra, called the Cartan subalgebras (after the father of this subject, Cartan (1894), (1914)), are the key to the whole dissection - the backbone, one might say. This fact is hard to appreciate ex loco. It is even more difficult to understand when treated most generally as a maximal solvable subalgebra, or even worse, as the subalgebra generated by the zero root eigenvectors of the adjoint representation of a vaguely defined 'regular' element of G (see e.g. Samelson (1969), p.36). (A semi-simple Lie algebra has no solvable ideals, but that does not stop it having solvable subalgebras.) The different approaches are consistent, for a maximal solvable subalgebra does turn out abelian, and is spanned by eigenvectors, h, belonging to zero roots of the adjoint representation, adjoint $g \cdot h = [g,h] = 0$ for g regular. Also, every maximal abelian subalgebra is isomorphic to every other one, so the same decomposition will occur, regardless of choice. (For proofs, see Samelson (1969) or Helgason (1962)). But in each case there is one particular abelian subalgebra which is of obvious importance. One supposes it is the existence of these subalgebras which provided Cartan with the right clues.

The diagonal matrices with determinent one, which form an abelian group, are an important subgroup of $S\ell(n,C)$. Its corresponding algebra, the space of diagonal matrices with trace zero (e.g. $\det(\exp L) = I \Rightarrow \exp \operatorname{tr} L = I \Rightarrow \operatorname{tr} L = 0$ for $L \in \mathcal{L}S\ell(n,C)$) is a cartan subalgebra of $\mathcal{L}S\ell(n,C)$. The canonical form for a member of $SO(n,C)$ is

$$\begin{bmatrix} \cos\theta_1 & \sin\theta_1 & & & & \\ -\sin\theta_1 & \cos\theta_1 & & & & \\ & & \cos\theta_2 & \sin\theta_2 & & \\ & & -\sin\theta_2 & \cos\theta_2 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

where the bottom right-hand 1 is present only for n odd. These elements form a maximal abelian subgroup of $SO(n,R)$, with the corresponding algebra:

$$\begin{bmatrix} 0 & \theta_1 & & & & \\ -\theta_1 & 0 & & & & \\ & & 0 & \theta_2 & & \\ & & -\theta_2 & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

again with the bottom zero present only for n odd. The complexification of these elements form a Cartan subalgebra of $\mathcal{L}SO(n,C)$. I have not seen this to my mind natural Cartan subalgebra used in the literature. Those curious redefinitions of $SO(n,R)$ mentioned in chapter 1 are invariably substituted here, to give a diagonal Cartan subalgebra, which may simplify the arithmetic, but wreaks havoc with one's familiar view of $SO(n,R)$. We stick to the above subalgebra, which gives a very clear decomposition, at the expense of just slightly more difficult mechanics.

The following is a brief outline of the decomposition process. Let G be a semi-simple Lie algebra over C, and Γ a Cartan subalgebra, also sometimes called a trunk of G. Then the solutions λ_i, g_i of

$$[h_0, g] = \lambda(h_0)g \ ,$$

for $g \in G$, h_0 a non-extraordinary (regular) member of Γ, and λ a linear function defined on Γ, describe the algebra completely. Each λ is the same function over all of Γ (i.e. $[h,g] = \lambda(h)g \ \forall\ h \in \Gamma$ if it is true for some h_0). For any $g \in \Gamma$, $\lambda(h_0) = 0$, a zero root, and the rank of both Γ and G is defined to be the number of zero roots. The eigenvectors of h_0 corresponding to non-zero roots form a basis for the rest of G not taken care of by the ℓ vectors in Γ which generate itself. If α is a root of adjoint h_0, then so is $-\alpha$, and there are precisely ℓ linearly independent non-zero roots. If Γ_α, Γ_β are eigenvectors for the roots α, β, then $\Gamma_{\alpha+\beta} = [\Gamma_\alpha, \Gamma_\beta]$ is the eigenspace for $\alpha + \beta$ if $\alpha + \beta$ is a root too, otherwise $[\Gamma_\alpha, \Gamma_\beta] = 0$. The whole structure then resolves into looking for the ladders, the roots of the form $\alpha + t\,\beta$, for t a string of integers. For each non-zero root, α, one can find a vector X_α in Γ defined by

$$K(X_\alpha, h_0) = \lambda_\alpha(h_0),$$

and the Dynkin diagram is deduced from the angles between the X_α.

None of the above makes very much sense without a few explicit examples. Instead of proving these results in general, we demonstrate they certainly hold for SO(n,C). SO(4,C) is a good place to begin, since it is non-trivial, and a knowledge of this decomposition helps tremendously with decomposing all the rest, even the odd dimensional ones.

The Lie algebra, $\mathcal{L}$SO(4,C), is the set of skew-symmetric matrices with arbitrary complex entries, and from above, the Cartan subalgebra is composed of elements of the form

$$\left[\begin{array}{cc:cc} 0 & -\theta_1 & 0 & 0 \\ +\theta_1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & -\theta_2 \\ 0 & 0 & +\theta_2 & 0 \end{array}\right], \qquad \theta_1, \theta_2 \in C.$$

So choose $h_0 = \begin{bmatrix} 0 & -\theta_1 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\theta_2 \\ 0 & 0 & \theta_2 & 0 \end{bmatrix}$, where it remains essentially variable, so that we will show explicitly that the roots are the same functions for every element of Γ. A non-regular element would have some linear restriction among the parameters, such as $\theta_1 = \theta_2$, or $\theta_i = 0$. Choosing an arbitrary element $g \in G$ gives

$$g = \begin{bmatrix} 0 & -g_{21} & -g_{31} & -g_{41} \\ g_{21} & 0 & -g_{32} & -g_{42} \\ g_{31} & g_{32} & 0 & -g_{43} \\ g_{41} & g_{42} & g_{43} & 0 \end{bmatrix}, \quad g_{ij} \in C.$$

Then one can easily calculate

$$\text{adjoint } h_0 \cdot g = (h_0 g - g h_0)$$

$$= \begin{bmatrix} 0 & 0 & \theta_2 g_{41}+\theta_1 g_{32} & -\theta_2 g_{31}+\theta_1 g_{42} \\ 0 & 0 & \theta_2 g_{42}-\theta_1 g_{31} & -\theta_2 g_{32}+\theta_1 g_{41} \\ -\theta_2 g_{41}-\theta_1 g_{32} & -\theta_2 g_{42}+\theta_1 g_{31} & 0 & 0 \\ \theta_2 g_{31}-\theta_1 g_{42} & \theta_2 g_{32}+\theta_1 g_{41} & 0 & 0 \end{bmatrix}.$$

So the effect of adjoint h_0, considered as a linear transformation on G is:

$$\text{adjoint } h_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\theta_2 & -\theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & -\theta_1 & 0 \\ 0 & \theta_1 & 0 & 0 & -\theta_2 & 0 \\ 0 & 0 & \theta_1 & \theta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ for } g \text{ strung out, } g = \begin{bmatrix} g_{21} \\ g_{31} \\ g_{41} \\ g_{32} \\ g_{42} \\ g_{43} \end{bmatrix}.$$

Solving $\det(\text{adjoint } h_0 - \lambda I_{6\times 6}) = 0$ gives roots:

$$\lambda = 0,\ 0,\ i(\theta_1+\theta_2),\ i(\theta_1-\theta_2),\ -i(\theta_1+\theta_2),\ -i(\theta_1-\theta_2)$$

and eigenvectors:

$$e_1 = \begin{bmatrix} 0 & 0 & -i & -1 \\ 0 & 0 & -1 & i \\ i & 1 & 0 & 0 \\ 1 & -i & 0 & 0 \end{bmatrix} \text{ such that } [h_0, e_1] = i(\theta_1+\theta_2)e_1$$

$$e_2 = \begin{bmatrix} o & o & i & -1 \\ o & o & 1 & i \\ -i & +1 & o & o \\ -1 & -i & o & o \end{bmatrix} \text{ such that } [h_0, e_2] = i(\theta_1 - \theta_2)e_2$$

$$e_3 = \begin{bmatrix} o & o & i & -1 \\ o & o & -1 & -i \\ -i & 1 & o & o \\ 1 & i & o & o \end{bmatrix} \text{ such that } [h_0, e_3] = -i(\theta_1 + \theta_2)e_3$$

$$\text{and } e_4 = \begin{bmatrix} o & o & i & 1 \\ o & o & -1 & i \\ -i & -1 & o & o \\ 1 & -i & o & o \end{bmatrix} \text{ such that } [h_0, e_4] = -i(\theta_1 - \theta_2)e_4 .$$

Any member of Γ is an eigenvector of the zero roots, and since there are two, the rank of G is 2. And there are also only and exactly two linearly independent non-zero roots. For the general case, the next step is to choose ℓ (rank) linearly independent roots under some order imposed on the parameters, so as to get the ℓ smallest positive roots (the fundamental system) under the ordering. In any particular case the ordering is obvious - here choose $\theta_1 > \theta_2 > 0$, so the two smallest positive roots are $\lambda_1 = i(\theta_1 + \theta_2)$ and $\lambda_2 = i(\theta_1 - \theta_2)$ - but expressing this concept for arbitrary algebras is messy.

So we see that there are two ladders: $(-\lambda_1, o, \lambda_1)$ and $(-\lambda_2, o, \lambda_2)$, but the two systems are disconnected, for $[e_1, e_2] = [e_1, e_4] = [e_2, e_3] = [e_3, e_4] = 0$, so e_1 and e_3 from one part, e_2, e_4 the other, with

$$[e_1, e_3] = \begin{bmatrix} o & -4i & o & o \\ 4i & o & o & o \\ o & o & o & -4i \\ o & o & 4i & o \end{bmatrix} = h_1, \quad [e_2, e_4] = \begin{bmatrix} o & 4i & o & o \\ -4i & o & o & o \\ o & o & o & -4i \\ o & o & 4i & o \end{bmatrix} = h_2$$

The root vectors, X_i, are those elements of Γ satisfying $K(X_i, h_0) = \lambda_i(h_0)$ which in this case are:

$$X_1 = \begin{bmatrix} o & i/4 & o & o \\ -i/4 & o & o & o \\ o & o & o & i/4 \\ o & o & -i/4 & o \end{bmatrix} \text{ and } X_2 = \begin{bmatrix} o & i/4 & o & o \\ -i/4 & o & o & o \\ o & o & o & -i/4 \\ o & o & i/4 & o \end{bmatrix} .$$

Then the Cartan integers, a_{ij}, are defined as

$$a_{ij} = \frac{2K(X_i,\ X_j)}{K(X_j,\ X_j)}$$

Here $A = (a_{ij}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

The Dynkin diagram is graph of the fundamental system, a vertex, i, for each fundamental root, λ_i, two vertices, i, j joined by $a_{ij} \cdot a_{ji}$ edges. These a_{ij} are a prescription for the ladders or strings, as will be made clearer later. For now, the Dynkin diagram of SO(4,C) is obviously

$$\boxed{\begin{array}{ccc} \cdot & \qquad\qquad & \cdot \\ \lambda_1 & & \lambda_2 \end{array}}\ .$$

Each component of SO(4,C), (e_1, e_3, h_1) and (e_2, e_4, h_2) corresponding to the strings $(-\lambda_1, o, \lambda_1)$ and $(-\lambda_2, o, \lambda_2)$ is a rank one simple algebra, the same as $\mathcal{L}S\ell(2,C)$, as one can easily verify - choose Cartan subalgebra $\begin{pmatrix} \theta_1 & o \\ o & \theta_1 \end{pmatrix}$.

So SO(4,C) is not simple, but can be decomposed into the direct sum of two simple rank 1 algebras, each isomorphic to $\mathcal{L}S\ell(2,C)$.

$\mathcal{L}SO(6,C)$ has a slightly more interesting structure.

Let $h_0 = \begin{bmatrix} o & -\theta_1 & & & & \\ \theta_1 & o & & & & \\ & & o & -\theta_2 & & \\ & & +\theta_2 & o & & \\ & & & & o & -\theta_3 \\ & & & & \theta_3 & o \end{bmatrix}$, analogously. Then

$$[h_0, g] = \left[\begin{array}{cc:cc:cc} o & o & \multicolumn{2}{c:}{\text{Skew elements}} & \multicolumn{2}{c}{\text{Skew elements}} \\ o & o & & & & \\ \hdashline -\theta_2 g_{41} - \theta_1 g_{32} & -\theta_2 g_{42} + \theta_1 g_{31} & o & o & \multicolumn{2}{c}{\text{Skew elements}} \\ \theta_2 g_{31} - \theta_1 g_{42} & \theta_2 g_{32} + \theta_1 g_{41} & o & o & & \\ \hdashline -\theta_3 g_{61} - \theta_1 g_{52} & -\theta_3 g_{62} + \theta_1 g_{51} & -\theta_3 g_{63} - \theta_2 g_{54} & -\theta_3 g_{64} + \theta_2 g_{53} & o & o \\ \theta_3 g_{51} - \theta_1 g_{62} & \theta_3 g_{52} + \theta_1 g_{61} & \theta_3 g_{53} - \theta_2 g_{64} & \theta_3 g_{54} + \theta_2 g_{63} & o & o \end{array}\right]$$

which is just forming repeats of $\mathcal{L}SO(4,C)$, for all pairs of parameters in h_0. So by inspection almost, the roots of adjoint h_0 are

$$0,0,0,\ i(\theta_1+\theta_2),\quad i(\theta_2+\theta_3),\quad i(\theta_1+\theta_3)$$
$$i(\theta_1-\theta_2),\quad i(\theta_2-\theta_3),\quad i(\theta_1-\theta_3)$$
$$-i(\theta_1+\theta_2),\ -i(\theta_2+\theta_3),\ -i(\theta_1+\theta_3)$$
$$-i(\theta_1-\theta_2),\ -i(\theta_2-\theta_3),\ -i(\theta_1-\theta_3)\ .$$

So there are 15 roots of g with respect to h_0, 3 zero roots, and many dependent ones. Again choosing the ordering $\theta_1 > \theta_2 > \theta_3 > \theta$ gives the three smallest linearly independent positive roots $\lambda_1 = i(\theta_1-\theta_2)$, $\lambda_2 = i(\theta_2-\theta_3)$, $\lambda_3 = i(\theta_2+\theta_3)$, which one can easily verify will generate all the rest. The fundamental system also has the property that every other positive root is an integral linear combination of these, for non-negative integers. (λ_1 and λ_2 for $SO(4,C)$ obey this degeneratively.)

Solving for the fundamental root vectors gives:

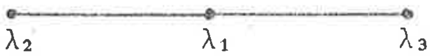
$$X_1 = \begin{bmatrix} 0 & i/4 & 0 & 0 & 0 & 0 \\ -i/4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i/4 & 0 & 0 \\ 0 & 0 & +i/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},\quad X_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i/4 & 0 & 0 \\ 0 & 0 & -i/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -i/4 \\ 0 & 0 & 0 & 0 & i/4 & 0 \end{bmatrix},$$

$$X_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i/4 & 0 & 0 \\ 0 & 0 & -i/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & i/4 \\ 0 & 0 & 0 & 0 & -i/4 & 0 \end{bmatrix}$$

with $K(X_1,h_0) = i(\theta_1-\theta_2)$, $K(X_2,h_0) = i(\theta_2-\theta_3)$, $K(X_3,h_0) = i(\theta_2+\theta_3)$

$$(a_{ij}) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

and Dynkin diagram



.

One can show this is the same diagram as for $\mathcal{L}S\ell(4,C)$, so $SO(6,C)$ is locally isomorphic to $S\ell(4,C)$, although there are much easier ways of showing this. The strings or ladders can be read off from (a_{ij}) as follows. For λ_i, λ_j, a_{ij}, then the j^{th} string of i is

$$\lambda_i,\ \lambda_i-\lambda_j,\ \lambda_i-2\lambda_j,\ldots,\lambda_i-a_{ij}\ \lambda_j$$

(see e.g. Samelson (1969), p.42). So the strings of $\mathcal{L}SO(6,C)$ are $(\lambda_1,\lambda_1+\lambda_2)$, $(\lambda_1,\lambda_1+\lambda_3)$ and $(-\lambda_i,0,\lambda_i)$, $i = 1,\ldots,3$.

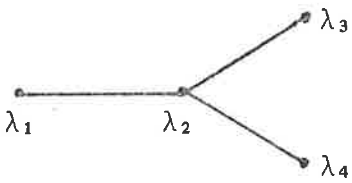
The class pattern for $\mathcal{L}SO(2n,C)$ starts to appear with $\mathcal{L}SO(8,C)$. Choosing h_0 as usual, with 4 parameters, the adjoint representation again copies $\mathcal{L}SO(4,C)$ for every combination of pairs, giving roots

$$\begin{array}{lllllll}
0 & i(\theta_1+\theta_2) & i(\theta_2+\theta_3) & i(\theta_3+\theta_4) & i(\theta_1+\theta_4) & i(\theta_1+\theta_3) & i(\theta_2+\theta_4) \\
0 & i(\theta_1-\theta_2) & i(\theta_2-\theta_3) & i(\theta_3-\theta_4) & i(\theta_1-\theta_4) & i(\theta_1-\theta_3) & i(\theta_2-\theta_4) \\
0 & -i(\theta_1+\theta_2) & -i(\theta_2+\theta_3) & -i(\theta_1+\theta_4) & -i(\theta_1+\theta_4) & -i(\theta_1+\theta_3) & -i(\theta_2+\theta_4) \\
0 & -i(\theta_1-\theta_2) & -i(\theta_2-\theta_3) & -i(\theta_3-\theta_4) & -i(\theta_1-\theta_4) & -i(\theta_1-\theta_3) & -i(\theta_2-\theta_4)
\end{array}$$

of which $\lambda_1 = i(\theta_1-\theta_2)$, $\lambda_2 = i(\theta_2-\theta_3)$, $\lambda_3 = i(\theta_3-\theta_4)$, $\lambda_4 = i(\theta_3+\theta_4)$ form a fundamental root system, with

$$(a_{ij}) = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & -1 \\ 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{pmatrix}$$

and Dynkin diagram

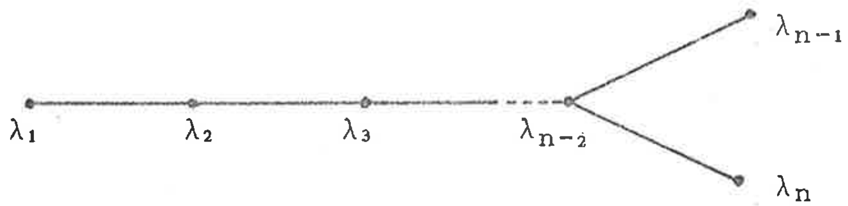


.

For higher degree n, one keeps getting $\frac{1}{2}n(n-1)$ copies of $\mathcal{L}SO(4,C)$, with the n fundamental roots

$\lambda_1 = i(\theta_1-\theta_2), \lambda_2 = i(\theta_2-\theta_3), \ldots, \lambda_{n-1} = i(\theta_{n-1}-\theta_n), \lambda_n = i(\theta_{n-1}+\theta_n)$

and Dynkin diagram:



The odd dimensional algebras, $\mathcal{L}SO(2n+1,C)$, are in some ways more interesting, possessing some 'double' bonds. The extra row and column of zeros in the Cartan subalgebra makes a deal of difference.

For $\mathcal{L}SO(5,C)$, take h_0 as

$$h_0 = \begin{bmatrix} 0 & -\theta_1 & 0 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\theta_2 & 0 \\ 0 & 0 & +\theta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and g as usual.

Then:

$$\text{adjoint } h_0 \cdot g = \begin{bmatrix} 0 & 0 & +\theta_2 g_{41}+\theta_1 g_{32} & -\theta_2 g_{31}+\theta_1 g_{42} & \theta_1 g_{52} \\ 0 & 0 & +\theta_2 g_{42}-\theta_1 g_{31} & -\theta_2 g_{32}-\theta_1 g_{41} & -\theta_1 g_{51} \\ -\theta_2 g_{41}-\theta_1 g_{32} & -\theta_2 g_{42}+\theta_1 g_{31} & 0 & 0 & \theta_2 g_{54} \\ +\theta_2 g_{31}-\theta_1 g_{42} & +\theta_2 \ {}_{32}+\theta_1 \ {}_{41} & 0 & 0 & -\theta_2 g_{53} \\ -\theta_1 g_{52} & +\theta_1 g_{51} & -\theta_2 g_{54} & \theta_2 g_{53} & 0 \end{bmatrix}.$$

Adjoint h_0 has representation:

$$\text{adjoint } h_0 = \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0 \\ 0&0&-\theta_2&-\theta_1&0&0&0&0&0&0 \\ 0&\theta_2&0&0&-\theta_1&0&0&0&0&0 \\ 0&\theta_1&0&0&-\theta_2&0&0&0&0&0 \\ 0&0&\theta_1&\theta_2&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&-\theta_1&0&0 \\ 0&0&0&0&0&0&\theta_1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&-\theta_2 \\ 0&0&0&0&0&0&0&0&\theta_2&0 \end{bmatrix} \text{ for } g = \begin{bmatrix} g_{21} \\ g_{31} \\ g_{41} \\ g_{32} \\ g_{42} \\ g_{43} \\ g_{51} \\ g_{52} \\ g_{53} \\ g_{54} \end{bmatrix}.$$

This is just a repeat of $\mathcal{L}SO(4,C)$ plus two extra pieces, which we can solve by inspection. The roots are:

$$0, 0, \; i(\theta_1+\theta_2), \; i(\theta_1-\theta_2), \quad i\theta_1, \quad i\theta_2$$
$$-i(\theta_1+\theta_2), -i(\theta_1-\theta_2), \quad -i\theta_1, \; -i\theta_2 \quad .$$

The Cartan subalgebra, and thus $\mathcal{L}SO(5,C)$, is of rank 2 (same as $\mathcal{L}SO(4,C)$, but with fundamental root system

$$\lambda_1 = i(\theta_1-\theta_2), \; \lambda_2 = i\theta_2 \quad .$$

The fundamental root vectors are then

$$X_1 = \begin{bmatrix} 0 & i/6 & 0 & 0 & 0 \\ -1/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i/6 & 0 \\ 0 & 0 & i/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } X_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i/6 & 0 \\ 0 & 0 & -i/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$(a_{ij}) = \begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix} \quad .$$

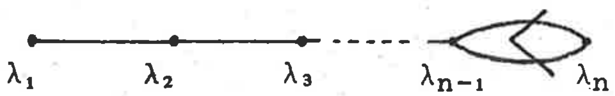
So X_1 and X_2 are joined by <u>2</u> edges, with Dynkin diagram

.

But the ladders are unsymmetric. There is a $(\lambda_1, \lambda_1 + \lambda_2, \lambda_1 + 2\lambda_2)$, but no string of length 3 for λ_2, only $(\lambda_2, \lambda_2+\lambda_1)$. So an arrow is attached, pointing to λ_1, to show which one can be added twice. It is not difficult to see what will happen for higher degree odd-dimensional $\mathcal{L}SO(2n+1,C)$ algebras. These will be the replications of $\mathcal{L}SO(4,C)$, the same as for $\mathcal{L}SO(2n,C)$, plus n extra sets of roots, the simple $\pm i\theta_j$'s, and a fundamental system

$$i(\theta_1-\theta_2), \; i(\theta_2-\theta_3), \ldots \; i(\theta_{n-1}-\theta_n), \quad i\theta_n$$

with $i\theta_n$ connected to $i(\theta_{n-1}-\theta_n)$ doubly, giving the diagram

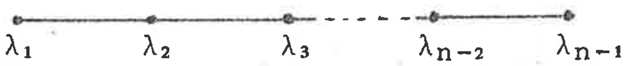


.

The root space decomposition for $\mathcal{L}S\ell(n,C)$ follows quite analogously. Choose h_0 diagonal, entries $\theta_1 \ldots \theta_n$, where $\sum_{i=1}^{n} \theta_i = 0$ (one linear constraint). Then it is easy to show that the roots of (adjoint h_0) are $\theta_i - \theta_j$ $\forall$ i, j (including the n-1 zero roots, - $\mathcal{L}S\ell(n,C)$ is of rank n-1), and each $\theta_i - \theta_j$ has eigenvector the matrix e_{ij} with entries 1 in the ijth position, zeros elsewhere. Ordering $\theta_1 > \theta_2 > \ldots > \theta_n$ gives a fundamental basis $\theta_1 - \theta_2$, $\theta_2 - \theta_3, \ldots,$ $\theta_{n-1} - \theta_n$, with

$$(a_{ij}) = \begin{pmatrix} 2 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & \ddots & & \\ & & \ddots & \ddots & & \\ & & & & 2 & -1 \\ 0 & & & & -1 & 2 \end{pmatrix}$$

and Dynkin diagram



The only major class we have left out is that for the symplectic algebras. On top of these four big classes, there are five fixed-rank exceptional algebras. From certain constraints on the Cartan integers, it is possible to tediously prove that only the diagrams corresponding to these algebras can exist, so that these 9 classes are the complete set of simple complex Lie algebras (see e.g. Varadarajan (1974), chapter 4 or Samelson (1969), chapter 2).

4.3 Zonal Polynomials

The representation theory of a complex semi-simple Lie algebra is related to its root space decomposition in a particularly elegant way. One can associate with each type of root structure a vector space (mutually orthogonal pieces for each simple component) where each root is considered a vector, and the fundamental system generates the vector diagram. The set of linear transformations which leave the whole diagram fixed (in one sense, all reversals of 'ladders') is the Weyl group, W, and it can be shown that any irreducible representation of the algebra is invariant under W. Thus one can parametrize the irreducible representations within the subspace, V/W. On top of this, the finite dimensional irreducible representations are parametrized by certain integral linear combinations of the roots in that subspace, which form a lattice. These results follow from a study of the weights of a representation: if T is a representation, then λ is a weight, and x a weight vector if $T(h)x = \lambda(h)x \ \forall\, h \in$ Cartan subalgebra. Each primitive string $(-\alpha, o, \alpha)$ in the root space decomposition, with vectors $(e_{-\alpha}, X_\alpha, e_\alpha)$ suitably normalized, gives an isomorphism with $\mathcal{L}S\ell(2,C)$ whose representation theory is well-known. From the commutator relations of a basis $(e_{-\alpha}, X_\alpha, e_\alpha)$ of $\mathcal{L}S\ell(2,C)$, it is not difficult to show that if x is a weight vector for $\lambda(H)$, then $T(e_{\pm\alpha})x$ is a weight vector for $\lambda(H) \pm \alpha(H)$, and it is known that all weights of $T(X)$ are integral for T an irreducible representation of $\mathcal{L}S\ell(2,C)$. By moving up and down these representation ladders (which must end for finite dimensional representations) it was found possible to label each irreducible finite dimensional representation by its highest weight in V/W, and since these will be in a lattice, there are a countable number of them. A reasonably clear and concise account of these calculations can be found in Samelson (1969), chapter 3.

The imbedding process mentioned previously, and Weyl's famous 'Unitary Trick' can now be used to connect the representation theory of complex groups to that of compact real groups. Each complex semi-simple Lie group, G^c, has several real forms (i.e. groups G such that $G^c = \{G+iG\}$), but only one of these is compact - or any two compact real forms are conjugate (see e.g. Helgason (1962), chapter 3, or Varadarajan (1974), chapter 4). Weyl's Unitary Trick amounts to showing that if a function on G^c is zero on G compact, then it is zero everywhere. This gives the mechanism to carry over representations on the complex group to representations on the real compact group. That is, if T^c is an irreducible finite dimensional representation of G^c, then the restriction T of T^c to G is again an irreducible finite dimensional representation of G, and the relationship between the T^c's and T's in 1-1. But it can be shown in other ways that every representation, finite or otherwise, of a compact group decomposes into the countable sum of finite dimensional irreducible representations (see e.g. Talman (1968), chapter 7). So the complete set of finite dimensional irreducible algebra representations $\mathcal{T}^c$ define the set of finite-dimensional irreducible group representations T^c (1-1 for the connected component of the identity, with modifications for other locally isomorphic groups) which in turn define the finite dimensional irreducible representations T of G, for G compact. Then relationships between the various real forms allow this set to be carried over to all the real forms. Herman Weyl then, in a brilliant tour-de-force, used these root structures to calculate explicit formulae in terms of the roots for both the dimension and character of each and every finite dimensional irreducible representation, for all the classical groups (see Weyl (1925), (1926 a and b)). The character χ of a representation is just the trace of the representation, important for being constant on each equivalence class. That is trace $(AT(g)A^{-1})$ = trace $(T(g))$ for any linear transformation A.

One might ask what this has to do with zonal polynomials. Remember that every symmetric space takes the form $K \times G/H$, for G semi-simple over R, H a maximal compact subgroup of G, and K a transformation group of Euclidean space. It can be shown (see Helgason (1962), chapter 5-8) that G/H decomposes into compact and non-compact components, with each component further decomposing into irreducible parts G_i/H_i for G_i simple. There is also a remarkable duality between the compact and non-compact spaces, each compact space being the dual of a non-compact space (see e.g. Helgason (1962), chapter 5). Consequently one need only catalogue all G_i/H_i for G_i's compact, and all possible H_i's, to have a complete list of the symmetric space building blocks. But getting back to zonal polynomials, the class of zonal spherical functions, including the zonal polynomials, arise in the following fashion. Suppose T_λ is an irreducible representation of G, in some space V_λ, and V_λ contains a vector v which is invariant under $T_\lambda(h)$ for every $h \in H$, the maximal compact subgroup of G, so:

$$T_\lambda(h)v \;=\; v \quad \forall\, h \in H \,.$$

Then, choosing v as the first basis element of V_λ,

$$T_\lambda(h) = \begin{bmatrix} 1 & o \ldots o \\ o & \\ \vdots & *(\text{arbitrary}) \\ o & \end{bmatrix} \quad \forall\, h \in H$$

we have for arbitrary $g \in G$

$$T_\lambda(g) = \begin{bmatrix} t_{11}(g) & * \ldots * \\ * & \\ * & \\ \vdots & * \\ * & \end{bmatrix}$$

where the first row is invariant under left transformation $(T_\lambda(hg))$, the first column is invariant under right transformation, $T_\lambda(gh)$, and $t_{11}(g) = t_{11}(h_1 g h_2)\ \forall\, h_1, h_2 \in H$. Such an animal is called a zonal spherical function. Notice we purposely allowed infinite dimensional representations here. It is probably worthwhile repeating this fact - that each $t_{11}(g)$ can be considered as a function on M, invariant with respect to H, because of its double invariance. The main job now is to work out just which representations of G possess such an invariant vector (the representations of class 1). This is non-trivial.

Applying these results to $G\ell(m,R)$, from the work of Weyl (1939) it turns out that there is one finite dimensional irreducible representation of $G\ell(m,R)$ for every set of integers $(k_1, k_2, \ldots, k_m)$, $k_i \geqslant k_j$, $i \geqslant j$. But the only representations with a spherical function (with respect to $O(m,R)$) are those with label $2k = (2k_1, 2k_2, \ldots, 2k_m)$, $k_i \geqslant 0$. This was shown explicitly in James (1961), by considering the representations of $G\ell(m,R)$ in the space of homogeneous polynomials in the elements of some $g_0 \in G\ell(m,R)$. Each spherical function on $G\ell(m,R)$ then becomes a polynomial of degree $2k = \sum_i 2k_i$. Further, from the invariance of each $t_{11}^{2k}(g)$ under left and right translation by $O(m)$, we have

$$t_{11}^{2k}(g) = t_{11}^{2k}(h_1 g h_2) = t_{11}^{2k}(L)$$

$$\text{for } h_1, h_2 \in O(m,R), \quad L \text{ diagonal.}$$

This can be seen easily from the isomorphism $M(m,R) = G\ell(m,R)/O(m,R)$ and the Principal Axis Theorem which show that every $g \in G\ell(m,R)$ can be expressed as $h_1'Lh_2'$ for some $h_1, h_2 \in O(m,R)$, L diagonal,

$$\begin{aligned} \text{i.e. } \forall\, m \in M(m,R), \quad m &= h_1'L^2h_1, \quad L \text{ diagonal} \\ &= h_1'LL'h_1 \\ &= h_1'Lh_2'h_2\,L'h_1 \quad \forall\, h_2 \in O(m,R) \\ &= (h_1'Lh_2')(h_1'Lh_2')' \\ &= g\,g', \quad g \in G\ell(m,R). \end{aligned}$$

This is a particular form of the general result that $G = HAH$ for H maximally compact in G, A a Cartan subgroup (whose algebra is a Cartan subalgebra of $\mathcal{G}$).

Considering these zonal spherical functions as functions on $M(m,R)$ then gives a polynomial in $Y = gg'$, a function of the latent roots $y_i = \ell_i^2$, for $L = \begin{pmatrix} \ell_1 & & \\ & \ddots & \\ & & \ell_m \end{pmatrix}$.

But more importantly, it was shown that each zonal polynomial takes the form

$$C_k(Y) = c_k(y_1^{k_1}\, y_2^{k_2} \cdots y_m^{k_m} + \text{symmetric terms})$$

$$+ \sum_{\rho \leqslant k} c_\rho\, (y_1^{r_1}\, y_2^{r_2} \cdots y_m^{r_m} + \text{symmetric terms})$$

for some constants c_k, where symmetric terms means all permutations of the k_i's, and a partition $\rho = (r_1, \ldots r_m)$ of degree k still, is said to be less than $k = (k_1, \ldots, k_m)$ if for the first r_i different from k_i, $r_i < k_i$.

I have ignored any problems of normalization. Obviously if $t_{11}(gg')$ is a zonal spherical function, so is any multiple of it (if v is an invariant vector, so is αv, α scalar). A.G. Constantine chose to normalize these things so that

$$(\mathrm{tr}\ Y)^{k} = \sum_{k} C_{k}(Y).$$

If we normalize $t_{11}^{k}(Y)$ to be 1 at $Y = I$, then

$$\frac{C_{k}(Y)}{C_{k}(I)} = t_{11}^{k}(Y) \quad (= C_{k}^{*}(Y) \text{ in the literature.})$$

Another normalization, $Z_{k}(Y)$, chosen by A.T. James (1961) to give integer coefficients in the polynomial expansions, is related to $C_{k}(Y)$ via

$$C_{k}(Y) = [\chi_{[2k]}(1)\ 2^{k}\ k!/(2k)!]\ Z_{k}(Y)$$

where $\chi_{[2k]}(1)$ is the dimension of the representation $[2k]$ of the symmetric group on $2k$ elements (see James (1964)). The symmetric groups are closely related to the representation theory of $G\ell(m,R)$, as can be seen from Weyl (1939), Boerner (1963) or Miller (1972).

The significance of having a semi-explicit form for these spherical functions cannot be over-emphasized. Most of the abstract theory is only concerned with proving existence of these functions. The non-statistical experts in this field do not know or care what forms these functions actually take, and are for the most part unaware of both of their uses and of their concrete existence in multivariate distribution theory.

The above zonal polynomials are only the finite dimensional ones, although they are all the ones of this class. There are spherical functions attached to the infinite dimensional representations of $G\ell(m,R)$, but we as yet have no idea what they will be like.

Harish-Chandra (1953 I)(1954 II and III)(1958 I and II) has derived general integral formulae for the zonal spherical functions (see Helgason (1962), chapter 10) in terms of the restricted root structures (see §4.5) for symmetric spaces. These include the spherical functions of the infinite-dimensional representations, when the integer restrictions of the finite dimensional representations are relaxed. Bhanu Murti (1960) calculates the general integral formula for $S\ell(m,R)/SO(m,R)$, which may sooner or later provide a method of deriving explicit expressions for them, although this has not yet been possible.

Here we collect a few facts about zonal polynomials. For one thing, there is only the one invariant vector in each irreducible subspace, giving the relationship

$$\int_{O(m,R)} \text{trace } T^{2k}(gh)\,dh = \frac{C_k(gg')}{C_k(I)}, \qquad h \in O(m,R).$$

The fundamental property for spherical functions ϕ_k of G,

$$\int_H \phi_k(g_1 h g_2)\,dh = \phi(g_1)\,\phi(g_2), \quad h \in H$$

(see Helgason (1962), chapter 10)

translates into

$$\int_{O(m,R)} C_k(Y\ h\ Y\ h')\,dh = \frac{C_k(Y_1)\ C_k(Y_2)}{C_k(I)}, \quad h \in O(m,R),$$

for the spherical functions considered on $M(m,R)$.
Another basic property of zonal spherical functions in general is that they are eigenfunctions of every differential operator, D, on $M = G/K$, which is invariant under G (Helgason (1962), chapter 10). This fact is what finally provided a useable method of calculating the coefficients c_ρ - this and knowing the general form of these functions! For there is one important

operator, the Laplace-Beltrami operator, Δ, which always exists, and which is reasonably easy to calculate. From Helgason (1962) p. 387,

$$\Delta = \frac{1}{|\det g|^{\frac{1}{2}}} \sum_{i=1}^{m} \partial_i \left(\sum_j g^{ji} |\det g|^{\frac{1}{2}} \partial_j\right) \tag{8}$$

for $g = g_{ij}$ the Riemann structure on M, $(g_{ij})^{-1} = (g_{ij})$ It has become important in the abstract theory of symmetric spaces to have an explicit form for the Laplace-Beltrami operator on each space, especially for the restriction of Δ to Cartan subgroups, the radial part of the operator. This can be calculated quite simply from some deep results of Harish-Chandra (1958) on root space decompositions. We shall study this later. Actually the radial part, Δ', on $M(m,R)$ was calculated in James (1968) from (8), by factoring out that part of the operator which dealt with the latent roots (the Cartan subgroup of diagonal matrices). So each $C_k\begin{pmatrix} y_1 \cdots & \\ & y_m \end{pmatrix}$ is an eigenfunction of Δ'

$$\text{i.e.}\quad \Delta' C_k\begin{pmatrix} y_1 \cdots & \\ & y_m \end{pmatrix} = \lambda_k C_k\begin{pmatrix} y_1 \cdots & \\ & y_m \end{pmatrix}$$

for λ_k the eigenvalue of Δ', y_i the latent roots of Y.

Explicit calculations in James (1968) yielded recurrence formulae by which the constant terms, c_ρ, in the expansion of $C_k(Y)$ in terms of the $M_\rho = y_1^{r_1} y_2^{r_2} \ldots y_m^{r_m}$ + symmetric terms, the monomial symmetric functions, could be evaluated (apart from a scale factor). The necessary calculations for the c_ρ are messy for hand computation, but do lend themselves quite well to computer evaluation, being basically sets of nested loops. I have developed such a program which is printed in full in the appendix, together with user instructions. This program, together with a discussion of possible used for the coefficients and polynomials (aside from the aesthetic pleasure of just being able to write them down) is to appear in Applied Statistics, JRSS, series C (see McLaren (1976)). It calculates the coefficients for the zonal polynomials, $Z_k(Y)$, where the first coefficient, c_k, that fixes the normalization, is evaluated from James (1968), 5.14, a combination of the dimension formula $\chi_{2k}(1)$ from James (1964), 19, and other relations found in James (1964), §9.

This normalization causes the last coefficient c_{1k} corresponding to the partition $(1,1,\ldots,1)$ (k times) to be $k!$, which serves as a handy check on the output.

Since the program can be easily extended to double precision, much higher degree coefficients can be calculated than had been possible previously (up to order 20 or 21, instead of 12 or 13). Certain intrinsic difficulties became apparent when these higher degree coefficients started to appear. There is a coefficient c_ρ for each $\rho \leqslant k$ using the previously mentioned ordering, and for a high degree k there can be hundreds of these coefficients. Also the final coefficient, being $k!$, is a very large number, (the first coefficient is of the order $10 \times k!$) even using double precision, and the highest orders calculable are around $k = 20$, depending on the machine. These are important considerations if one were considering the task of actually evaluating a hypergeometric function. For the hypergeometric functions according to their definition (7) are power sums of the $C_k(Y)$'s.

Historically, these generalized hypergeometric functions were first introduced by Carl Herz (1955), via integral (Laplace) transforms. Hypergeometric functions of matrix (positive definite symmetric) argument can be defined via their integral transforms in the same way that the univariate special functions can be defined (see Erdelyi *et al* (1953)). However such definitions do not give explicit formulae for these functions, but merely show their theoretical existence, and give properties of them. At the same time their power sums of zonal polynomials were appearing in multivariate non-null densities (see e.g. James (1960), (1961)). A real breakthrough in the whole theory was achieved when A.G. Constantine (1963) showed that these power sums of zonal polynomials were in fact the same hypergeometric functions defined by Herz, by proving they had the same transforms. At one go then, one had these functions fitting into the general scheme of special functions, and a concrete form for these functions. The two-matrix functions

$$ {}_pF_q\,(a_1,\ldots a_p;\ b_1,\ldots b_q;\ S,T) = \sum_{k=0}^{\infty} \sum_{k} \frac{(a_1)_k\ldots(a_p)_k}{(b_1)_k\ldots(b_q)_k} \frac{C_k(S)\quad C_k(T)}{C_k(I_m)\,k!} \quad \text{for} \quad S,T \in M(m,R) $$

which arise from $\displaystyle\int_{O(m,R)} C_k(S\ h\ T\ h')\,dh = \frac{C_k(S)\quad C_k(T)}{C_k(I_m)}$

in certain densities, were quite unexpected however. No one-dimensional analogue seems to exist. Even the so-called Appell functions (Erdelyi *et al* (1953) Vol. 1, chapter 5) see to be more closely related to the 2-by-2 matrix functions considered as functions of their 2 latent roots.

I hope that this section has given some idea of the scope and complexity of the whole subject of zonal polynomials in multivariate statistics. An admittedly rather personal guide to the mathematical literature may be of some value. For the statistical theory, of course, we have the series of papers by A.T. James, in particular the 1961, 1964 and 1968 papers, and the Constantine (1963) paper. A comprehensive survey of the statistical uses of zonal polynomials can be found in Subrahmaniam

(1974), which also contains an excellent statistical bibliography.

The mathematical literature is more of a problem. Unfortunately the early landmark papers and texts on semisimple groups and representation theory are either untranslated (e.g. Cartan (1914), (1929), and Weyl (1925), (1926 a and b)) or unreadable (see Weyl (1939)). This theory is quite adequately covered in the recent text by Varadarajan (1974), from basic Lie theory up to the Weyl character formulae for representations of compact groups, although so much time is spent on elementary results of Lie groups that he does not get to the representation theory till the last chapter. For the actual root space decompositions of semisimple algebras over C, and their representation theory, I recommend highly the remarkably clear and concise paperback by Samelson (1969). If one has, by any chance, access to the Princeton mimeographed notes of Weyl (1935), these are much easier going than the famous book, the Classical Groups, Weyl (1939), where he attempts to do without the analytic methods, and becomes as a result extremely dry and difficult.

Miller (1972) suffers to a certain extent from this too, although it does give a nice description of the representation relationships between the symmetric groups and $G\ell(m,R)$, and goes through the representation theory of all the classical groups, showing corrections for $O(m,R)$ from $SO(m,R)$, and tricky things of this sort. One interesting book for the dabbler is by Hermann (1966), entitled 'Lie Groups for Physicists' which gives a very pragmatic and low level interpretation of Lie theory in general.

The modern classic for the theory of ordinary special functions, from the group representation point of view, is the encyclopaedic work of Vilenkin (1968). The first chapter of this book also contains a surprisingly clear and short account of much of the structure and general mechanics of this subject, which beautifully motivates the study of spherical functions. Two other texts, Miller (1968) and Talman (1968) I find too tediously orientated towards applied physics.

The deeper results of representation theory, dealing with the existence and form of spherical functions and generalized fourier analysis, rests squarely with Harish-Chandra. Unfortunately his original papers are quite impossible (see e.g. Harish-Chandra (1953) (1954 II and III), (1958)). Luckily several authors have made valiant attempts to understand and explain his work to the lower classes - notably Helgason (1962) chapter 10, (1973); Gangolli (1972); and Varadarajan (1973). The Helgason (1973) paper is particularly valuable for it having clear statements of many results, and for its excellent guide to the literature.

4.4 Spherical Functions on the Grassmann Manifold

Since it had been found possible to construct the zonal spherical functions (of finite dimension) for $M(m,R)$, it is natural to wonder what the situation is for the other symmetric space of interest statistically, the Grassmann manifold, $G_{p,q-p}$. This question was attacked and solved in James and Constantine (1974). Not only were the spherical functions found, as generalizations of the Gegenbaur polynomials, but a new class of generalized Jacobi polynomials were also found, the 'intertwining functions'.

These derivations are well set out in James and Constantine (1974). We have $G_{p,q-p} = O(q,R)/O(p,R) \times O(q-p,R)$ the unorientated Grassmann manifold, which can be obtained from the orientated space of planes, $SO(q,R)/SO(p,R) \times SO(q-p,R)$ by identifying elements of opposite orientations. The spherical functions on $G_{p,q-p}$, since they must be constant on the double cosets,

$$O(p,R) \times O(q-p,R)\cdot H\cdot O(p,R) \times O(q-p,R) \quad , \; H \in O(q,R)$$

thus turn out to be functions of the critical angles. $G_{p,q-p}$ is a compact space, so in contrast to the non-compact space, $M(m,R)$, the spherical functions on $G_{p,q-p}$ possess orthogonal properties with respect to the group invariant measure on the space. James and Constantine (1974) derived the generalized Gegenbaur polynomials as a complete set of polynomials in the squares of the cosines of the critical angles, $\theta_1,\ldots,\theta_p$, $p \leqslant q/2$, orthogonal with respect to

$$\left(\prod_i y_i\right)^{-\frac{1}{2}} \left(\prod_i (1-y_i)\right)^{\frac{1}{2}(q-2p-1)} \prod_{i<j} (y_i-y_j) \quad , \quad y_i = \cos^2 \theta_i ,$$

the measure derived from the measure on $G_{p,q-p}$ (see James (1954)). These collapse to the ordinary Gegenbaur polynomials for $p = 1$ (Erdelyi *et al* (1953)).

It was actually just as simple to find the extensions of these functions, the generalized Jacobi polynomials, which are constants on the double cosets

$$O(m,R) \times O(q-m,R)\cdot H\cdot O(p,R) \times O(q-p,R),$$

so functions of the critical angles $\theta_1, \theta_2, \ldots, \theta_m$, $(m \leqslant p)$, which are orthogonal with respect to

$$\Pi\, y_i^{\frac{1}{2}(p-m-1)}\ (1-y_i)^{\frac{1}{2}(q-p-m-1)}\ \prod_{i<j} (y_i-y_j)\ .$$

This is of course related to the ease of finding invariants between two hyperplanes of different dimension, m and p, in q-space, as mentioned earlier. For $m = 1$ these become the ordinary Jacobi polynomials.

The generalized Laguerre polynomials ${}_1F_1\ (-k,a,X)$ which had been discovered earlier (see Constantine (1966)) possessed expansions in zonal polynomials

$${}_1F_1\ (-k,a,X) = \sum_{\sigma \leqslant k} \frac{(-1)^s \binom{k}{\sigma}}{(a)_\sigma}\ C^*_\sigma(X)$$

where $C^*_\sigma(X) = C_\sigma(X)/C_\sigma(I)$, $\binom{k}{\sigma}$ is defined via $C^*_k(I_m+X) = \sum_{\sigma \leqslant k} \binom{k}{\sigma}\ C^*_\sigma(X)$, and $(a)_\sigma = \prod_{i=1}^{m} \left(a-\frac{1}{2}(i-1)\right)_{s_i}$ for $\sigma = (S_1 S_2 \ldots S_m)$, and the usual ordering of partitions. These functions were orthogonal with respect to the measure $\operatorname{etr}(-X)\det X^{\frac{1}{2}(p-m-1)}$, which is a limiting case of the Jacobi polynomial measure $(q \to \infty)$. This suggested a similar form for the Jacobi polynomials, $P_k(X)$ (for roots of X the $\cos^2 \theta_i$), and James and Constantine proved

$$P_k(X) = \sum_{\sigma \leqslant k} \frac{(-1)^s \binom{k}{\sigma}\ (C)_{k,\sigma}}{(a)_\sigma}\ C^*_\sigma(X)\ ,$$

with the help of the Laplace-Beltrami operator on $G_{m,q-m}$ to find recurrence formulae for the $(C)_{k,\sigma}$. The confluence of $P_k(X) \to {}_1F_1\ (-k,a;X)$ follows a completely analogous result for the ordinary polynomials.

The unexpected result in this work, as I see it, is the very fact that these functions do turn out to be expressible in terms of the zonal polynomials of $M(m,R)$, even though the two spaces have little in common besides their symmetry. This suggests to me that perhaps the $C_k(Y)$'s may prove of great value in general symmetric space theory as the fundamental building blocks of all the spherical functions.

4.5 Symmetric Space Root decompositions

The root space decomposition theory of the semi-simple Lie groups can be carried over to the symmetric spaces to a quite remarkable degree - even down to the Dynkin diagrams. I have not been able to find any explicit calculations in the literature of the type which follow, although the results can certainly be checked in Loos (1969), Araki (1962), or Satake (1960). These all derive the graphs by considering the involutive automorphisms of a simple algebra, G - those transformations, σ, which leave a maximally compact sub-algebra $\mathcal{H}$ fixed, with $\sigma^2 = I$, a theoretically more elegant and general method. But I find this approach confusing in its generality, and much harder to comprehend, especially when a simple explicit example can give a startlingly clear picture of what is really going on.

These decomposition results follow from the way in which each space decomposes into fundamental blocks. The only place that I know of where the decomposition theory of symmetric spaces can be found is Helgason (1962), chapter 4-8, where it is done with complete rigour, as far as he goes. A brief survey of the contents of this landmark of a book will provide us with the results we need. Chapters 1 and 2 cover the basic facts of differential geometry and elementary Lie theory, but with such rigour and detail, in such a short space, as to be quite impossible if one is not already familiar with the concepts involved. Chapter 3 proves the compact real form imbedding theorems, and shows the important Cartan decomposition of any semi-simple algebra G into $G = \mathcal{H} \oplus M$ for $\mathcal{H}$ the subalgebra corresponding to a maximally compact subgroup of G, M the orthogonal complement of $\mathcal{H}$ with respect to the Killing form. Chapter 4 and 5 then prove that every symmetric space, M, decomposes as

$$M = M_+ \times M_0 \times M_-$$

where M_0 is a flat euclidean space, $M_+ = G_+/H_+$ for G_+ a compact semi-simple Lie group over R, H_+ maximally compact in G_+, and $M_- = G_-/H_-$ for G_- a non-compact semi-simple Lie group over R, H_- maximally compact in G_-. These results follow from the Cartan decomposition theorems. The important duality between the compact and non-compact spaces is also investigated. Chapters 6 and 7 discuss both types, and chapter 8 covers the Hermitian spaces, which are important in complex analysis. But hidden in chapter 8, §5, are the crucial theorems on the decomposition of each (non-)compact space into irreducible components, $M_i = G_i/H_i$, for G_i a (non-)compact simple Lie group over R, H_i maximally compact in G_i. So all the irreducible components are calculable from the various real forms of the 9 simple algebras over C, and their maximally compact subalgebras. Chapter 9 lists these spaces, but omits any mention of Dynkin diagrams or any detailed calculations, to make room for chapter 10, a discussion of modern trends in symmetric space theory, basically the work on spherical functions and fourier analysis by Harish-Chandra.

Each complex simple algebra has only one real compact form, and several non-compact ones. But each non-compact real form has only one maximally compact subalgebra (Helgason (1962), chapter 6), whereas each compact form may have several different maximally compact subalgebras (Helgason (1962), chapter 7). Luckily the two sets match up under the duality. Schematically:

$$\begin{array}{llcccl}
 & G\ (\text{compact}) & : \mathcal{H}_1 & \mathcal{H}_2 \ldots & \mathcal{H}_m & \left\{\begin{array}{l}\text{maximally compact}\\ \text{subalgebras}\end{array}\right. \\
\nearrow & & \downarrow & & \downarrow & \\
G_C & & M_1 = G/H_1 & \ldots & M_m = G/H_m & \\
(\text{simple} & & \updownarrow & & \updownarrow & \\
\text{over } C) & & M^* = G_1/H_1^* & \ldots & M_m^* = G_m/H_m^* & \\
\searrow & & \uparrow & & \uparrow & \\
 & & G_1 & G_2 \ldots & G_m & \left\{\begin{array}{l}\text{non-compact real}\\ \text{forms}\end{array}\right.
\end{array}$$

where M^* is the dual of M.

So one need only deal with the compact real forms, and find all maximally compact subalgebras (subgroup) for each one. The root space decomposition for a symmetric space G/H takes place over a Cartan subalgebra, Γ, of $G = \mathcal{H} \oplus M$, where Γ is as much as possible inside M, and only hangs over into $\mathcal{H}$ minimally. One finds the roots of G with respect to Γ as for the simple algebras, then restricts these to M to get the root structure of M (eventually let the parameters in $\mathcal{H}$ go to zero). The number of parameters one can squeeze into M is the rank of M.

We shall now explicitly derive the Dynkin diagrams for the orientated Grassmann family of compact symmetric spaces, the

$$SO(q,R)/SO(p,R) \times SO(q-p,R).$$

Several different cases arrive. For example, q can be even or odd, corresponding to two different complex simple algebras, and every partition of q into two integers gives a symmetric space, so there are several symmetric spaces associated with each $SO(q,R)$.

Choosing $SO(5,R)$ for example, the compact real form of $SO(5,C)$, then there are two symmetric spaces associated with it,

$$\frac{SO(5,R)}{SO(2,R) \times SO(3,R} \quad \text{and} \quad \frac{SO(5,R)}{SO(1,R) \times SO(4,R)} = \frac{SO(5,R)}{SO(4,R)}.$$

Here is a small snag. In the previous decomposition of $\mathcal{L}SO(5,C)$, we used the Cartan subalgebra

$$\Gamma = \left(\begin{bmatrix} 0 & -\theta_1 & 0 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\theta_2 & 0 \\ 0 & 0 & \theta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \ \theta_i \in C \right)$$

which is completely inside the algebra of $SO(2,C) \times SO(3,C)$ namely:

$$\mathcal{H} = \left(\begin{bmatrix} \begin{matrix} 0 & -\theta_1 \\ \theta_1 & 0 \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & -\theta_2 & -\theta_3 \\ \theta_2 & 0 & -\theta_4 \\ \theta_3 & \theta_4 & 0 \end{matrix} \end{bmatrix}, \ \theta_i \in C \right).$$

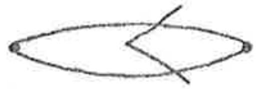
So I hunted around for another Cartan subalgebra (maximally abelian in G) which was as much as possible outside $\mathcal{H}$, and turned up

$$\Gamma = \left(\gamma = \begin{bmatrix} 0 & 0 & 0 & -\theta_1 & 0 \\ 0 & 0 & 0 & 0 & -\theta_2 \\ 0 & 0 & 0 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & 0 \end{bmatrix}, \ \theta_i \in C \right).$$

which one can easily verify is abelian and maximal, since it is the same rank as an old Cartan subalgebra. This is completely outside $\mathcal{H}$ - in these cases M is said to be maximal, the same rank as the associated simple algebra. Furthermore, the adjoint representation $[\gamma,g]$ turned out to be exactly the same form as for the regular decomposition $\mathcal{L}SO(5,C)$, except for permuted rows and columns, so the roots of $[\gamma,g] = \lambda(\gamma)g$ turn out the same, namely:

$$0,\ 0,\ i(\theta_1+\theta_2),\ i(\theta_1-\theta_2),\ i\theta_1,\ i\theta_2$$
$$-i(\theta_1+\theta_2),\ -i(\theta_1-\theta_2),\ -i\theta_1,\ -i\theta_2\ .$$

Choosing the fundamental system of positive roots as before gives the Dynkin diagram for $SO(5,R)/SO(2,R) \times SO(3,R)$ as



$i(\theta_1 - \theta_2)$ $\qquad$ $i\theta_2$

the same as for $SO(5,C)$, with no multiple roots. This is exactly what one would expect for a maximal rank space. [Note, it is a matter of individual preference whether one chooses to point the arrow from the lower root (the one that is added on) to the higher, or vice-versa. Samelson (1972) uses the above notation, Araki (1962) and Loos (1969) prefer the other way.]

The situation changes for $SO(5,R)/SO(4,R)$. Here the complexified Cartan subalgebra is again

$$\Gamma = \left(\gamma = \begin{bmatrix} 0 & 0 & 0 & -\theta_1 & 0 \\ 0 & 0 & 0 & 0 & -\theta_2 \\ 0 & 0 & 0 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & 0 \end{bmatrix}, \ \theta_i \in C \right)$$

but

$$\underset{\text{complexified}}{\mathcal{H}} \left(h = \left[\begin{array}{c:cccc} 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & & & & \\ 0 & & \mathcal{L}SO(4,C) & & \\ 0 & & & & \\ 0 & & & & \end{array} \right] \right)$$

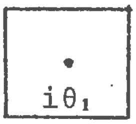
and θ_2 overlaps into $\mathcal{H}$. The unrestricted roots are of course just as they are above, but when restricted to

$$M = \left(m = \left[\begin{array}{c:cccc} 0 & * & * & * & * \\ \hdashline * & & & & \\ * & & 0 & & \\ * & & & & \\ * & & & & \end{array} \right] \right)$$

the non-zero roots $(\theta_2 = 0)$ are just:

$$\begin{array}{ccc} i\theta_1, & i\theta_1, & i\theta_1, \\ -i\theta_1, & -i\theta_1, & -i\theta_1, \end{array}$$

or just the two roots $\pm i\theta_1$ repeated 3 times, with Dynkin diagram,



.

A less trivial example is $SO(7,R)$. $\mathcal{L}SO(7,C)$ has non-zero roots $\pm i(\theta_1 \pm \theta_2)$, $\pm i(\theta_2 \pm \theta_3)$, $\pm i(\theta_1 \pm \theta_3)$, $\pm i\theta_1$, $\pm i\theta_2$, $\pm i\theta_3$ for

$$\Gamma = \left(\gamma = \begin{bmatrix} & & & 0 & -\theta_1 & & \\ & & & 0 & & -\theta_2 & \\ & & & \vdots & & & \theta_3 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ \theta_1 & & & \vdots & & & \\ & \theta_2 & & \vdots & & & \\ & & \theta_3 & 0 & & & \end{bmatrix}, \quad \theta_i \in C \right) .$$

Then

$\mathcal{H}_1 = \begin{bmatrix} 0 & \vdots \\ \cdots & \cdots \\ \vdots & \mathcal{L}SO(6,C) \end{bmatrix}$ gives roots $\pm i\theta$, repeated 5 times,

$\mathcal{H}_2 = \begin{bmatrix} \mathcal{L}SO(2,C) & \vdots \\ \cdots & \cdots \\ \vdots & \mathcal{L}SO(5,C) \end{bmatrix}$ gives roots $(\theta_3 = 0)$, $\pm i(\theta_1 \pm \theta_2)$, $\pm i\theta_1$, $\pm i\theta_2$

with $\pm i(\theta_1 \pm \theta_2)$ of multiplicity 1, and $\pm i\theta_1$ and $\pm i\theta_2$ of multiplicity 3,

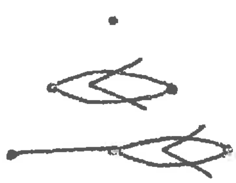
$\mathcal{H}_3 = \begin{bmatrix} \mathcal{L}SO(3,C) & \vdots \\ \cdots & \cdots \\ \vdots & \mathcal{L}SO(4,C) \end{bmatrix}$ gives the same roots as $\mathcal{L}SO(7,C)$ since it is maximal, multiplicities 1.

By inspecting their respective fundamental root systems, we get Dynkin diagrams:

$SO(7,R)/SO(6,R)$

$SO(7,R)/SO(2,R) \times SO(5,R)$

$SO(7,R)/SO(3,R) \times SO(4,R)$



.

For the general case, q odd, one gets the two kinds of roots $\pm i(\theta_i \pm \theta_j)$ with multiplcities 1, and the $\pm i\theta_i$ with multiplicities $2\left(\frac{q-1}{2} - p\right) + 1$ for $SO(q,R)/SO(p,R) \times SO(q-p,R)$, $p < q-p$. These results can be checked in Araki (1962), p.32, where his ℓ is the rank of $SO(q,C)$, $\ell = (q-1)/2$.

We can do the same things with the even dimensional spaces - those derived from $SO(2q,R)$. Take $SO(8,R)$ for example. Then $\mathcal{L}SO(8,C)$ is of rank 4, with Cartan subalgebra

$$\Gamma = \left(\gamma = \left[\begin{array}{cccc:cccc} & & & & -\theta_1 & & & \\ & & & & & -\theta_2 & & \\ & & & & & & -\theta_3 & \\ & & & & & & & -\theta_4 \\ \hdashline \theta_1 & & & & & & & \\ & \theta_2 & & & & & & \\ & & \theta_3 & & & & & \\ & & & \theta_4 & & & & \end{array} \right] , \quad \theta_i \in C \right)$$

and unrestricted non-zero roots, $\pm i(\theta_i \pm \theta_j)$, $i < j$. So the various symmetric spaces have restricted roots corresponding to their respective $\mathcal{H}$'s as follows:

$$\mathcal{H}_1 = \left[\begin{array}{c:c} 0 & \\ \hdashline & \mathcal{L}SO(7,C) \end{array} \right], \text{ roots } \pm i\theta_1, \text{ repeated } 2(4-1) \text{ times}$$

$$\mathcal{H}_2 = \left[\begin{array}{c:c} \mathcal{L}SO(2,C) & \\ \hdashline & \mathcal{L}SO(6,C) \end{array} \right],$$ roots $\pm i(\theta_1 \pm \theta_2)$, multiplicities 1 and $\pm i\theta_1, \pm i\theta_2$, multiplicities 4

$$\mathcal{H}_3 = \left[\begin{array}{c:c} \mathcal{L}SO(3,C) & \\ \hdashline & \mathcal{L}SO(5,C) \end{array} \right],$$ roots $\pm i(\theta_1 \pm \theta_2), i(\theta_1 \pm \theta_3)$, $i(\theta_2 \pm \theta_3)$, multiplicities 1 an $\pm i\theta_1, \pm i\theta_2, \pm i\theta_3$, multiplicities 2

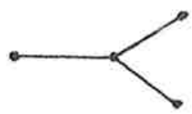
$$\mathcal{H}_4 = \left[\begin{array}{c:c} \mathcal{L}SO(4,C) & \\ \hdashline & \mathcal{L}SO(4,C) \end{array} \right],$$ roots the same as for $\mathcal{L}SO(8,C)$, multiplicities 1, maximal rank

and Dynkin diagrams

$SO(8,R)/SO(7,R)$

$SO(8,R)/SO(2,R) \times SO(6,R)$ 

$SO(8,R)/SO(3,R) \times SO(5,R)$ 

$SO(8,R)/SO(4,R) \times SO(4,R)$ 

In general, we see that $SO(q,R)/SO(p) \times SO(q-p,R)$, $p \leqslant q/2$, has (restricted) roots $\pm i(\theta_i \pm \theta_j)$ with multiplicities 1, and $i\theta_i$'s with multiplicities $2(\frac{1}{2}q - p)$, for q even.

The rules for arbitrary simple G then are to find all the usual roots of G, and restrict them to the space M, where $G = M \oplus \mathcal{H}$, and the Cartan subalgebra is maximal in M. Then take these restricted root forms and look up their diagram which must correspond to one of the 9 simple diagrams. Strangely, only the 4 diagrams of the 4 classical families ever turn up as Dynkin diagrams for the symmetric spaces derived from the classical simple groups. The odd exceptional space has the same diagram as a classical space however (see Araki (1962), p.33, type E III and E IV).

We may as well look at $S\ell(m,R)/SO(m,R)$, the semi-simple part of $G\ell(m,R)/O(m,R)$. $S\ell(m,R)$ is non-compact, so there is only the one maximally compact subgroup, $SO(m,R)$, and the Cartan subalgebra of diagonal matrices is completely outside $\mathcal{L}SO(m,R)$ anyway, so $S\ell(m,R)/SO(m,R)$ is of maximal rank, and has Dynkin diagram

with roots $(\theta_i - \theta_j)$ $i < j$ of multiplicities 1, and $\sum_{i=1}^{m} \theta_i = 0$.

These two families do not exhibit the double root forms, $2\lambda_i$, for λ_i a root, which can occur with some of the other spaces due to linear restrictions imposed on the root forms under the restriction to M. These can be kept track of when they occur since each irreducible space only has at most two types of roots (i.e. the $\pm i(\theta_i \pm \theta_j)$ and $i\theta_i$'s for $SO(q,R)$) and only one type ever exhibits the double forms (Araki (1962)).

4.6 Koornwinder Polynomials

The main reason (besides curiosity) for going so deeply into the structure of symmetric spaces, was to try and understand the work of Tom H. Koornwinder (1974, I, II, III and IV), and (1975). He has undertaken a comprehensive study of orthogonal polynomials in two variables. Many of his polynomials are quite simple extensions of univariate orthogonal polynomials, but two types have interpretations as spherical functions on compact symmetric spaces of rank two, and to understand this one needs to be able to follow the decomposition theory for symmetric spaces.

Koornwinder (1975) §3.3 defined the polynomials of type 6 as polynomials in u,v, orthogonal with respect to the weight function

$$(1-u+v)^{\alpha} \quad (1+u+v)^{\beta} \quad (u^2-4v)^{\gamma}.$$

A change of parameters, $u = \cos 2s + \cos 2t$, $v = \cos 2s \cos 2t$, gives functions in s,t, orthogonal with respect to

(9) $\quad \delta = (\sin s \sin t)^{2\alpha+1} \quad (\mathrm{Cos}\, s \,\mathrm{Cos}\, t)^{2\beta+1} \quad (\sin(s+t)\sin(s-t))^{2\gamma+1}$

He then claims that $D' = (\delta^{-\frac{1}{2}}(\partial_{tt}+\partial_{ss}))\delta^{\frac{1}{2}} - \delta^{-\frac{1}{2}}((\partial_{tt}+\partial_{ss})\delta^{\frac{1}{2}})$ is the radial part of the Laplace-Beltrami operator for symmetric spaces of rank 2 with Dynkin diagram

and root system $\lambda_1, 2\lambda_1, \lambda_2$ with multiplicities $m(\lambda_1) = 2\alpha-2\beta$, $m(2\lambda_1) = 2\beta+1$, and $m(\lambda_2) = 2\gamma+1$. The orthogonal polynomials, being eigenfunctions of D', are then the spherical functions on such a space. This appeared, to say the least, rather miraculous, but by following up the references, Harish-Chandra (1958), Helgason (1972) and Araki (1962), and by working out the ideas of symmetric space decompositions, I finally managed to elucidate what powers were being called on.

According to the lists, in Araki (1962), as we demonstrated earlier, the symmetric space $SO(q,R)/SO(2,R) \times SO(q-2,R)$ has Dynkin diagram

and multiplicities:

q odd : $m(\pm i(\theta_1 \pm \theta_2)) = 1$, $m(\pm i\theta_1) = m(\pm i\theta_2) = 2\left(\frac{q-1}{2} - 2\right)+1 = q-4$,

$m(2\lambda_i) = 0.$

q even : $m(\pm i(\theta_1 \pm \theta_2)) = 1$, $m(\pm i\theta_1) = m(\pm i\theta_2) = 2\left(\frac{q}{2} - 2\right) = q-4$,

$m(2\lambda_i) = 0.$

So the orthogonal polynomials of Koornwinder, type 6, for $\alpha = (q-5)/2$, $\beta = -\frac{1}{2}$, $\gamma = 0$ should be the same as our Gegenbaur polynomials, which from James and Constantine (1974) are polynomials in y_1, y_2 orthogonal with respect to

$$(y_1 y_2)^{-\frac{1}{2}}\left((1-y_1)(1-y_2)\right)^{\frac{1}{2}(q-5)} (y_1 - y_2) .$$

Actually the difference in looks is just a parameter change, which can be found from a deeply hidden result of Harish-Chandra's. For, in his 1958 paper, he derives (§7), a formula for the radial part of the Laplace-Beltrami operator

$$\Delta' = \delta^{-1} \sum_{i,j} g^{ij} H_i \cdot \delta H_j$$

for H_i a basis for the restriction of a Cartan subalgebra to M, and the Riemann metric $g_{ij} = K(H_i,H_j)$ - the Killing form again. Most importantly, δ is just $\prod_{\lambda \in p^+} \left(e^{\lambda(H)} - e^{-\lambda(H)}\right)$, for p^+ the set of positive restrictions, λ, of the roots to M including multiplicities (Harish-Chandra (1958) §5). But Δ' is just the D' defined by Koornwinder (see Helgason (1972) p. 415), and we can easily calculate δ for $SO(q,R)/SO(2,R) \times SO(q-2,R)$. That is, the positive restricted roots are

$$i(\theta_1+\theta_2),\quad i(\theta_1-\theta_2),\quad i\theta_1,\quad i\theta_2$$

with multiplicities 1,1,q-4,q-4 respectively. So

$$\delta = \left(e^{i(\theta_1-\theta_2)}-e^{-i(\theta_1-\theta_2)}\right)\left(e^{i(\theta_1+\theta_2)}-e^{-i(\theta_1+\theta_2)}\right).$$

$$\cdot\left(e^{i\theta_1}-e^{-i\theta_1}\right)^{q-4}\ \left(e^{i\theta_2}-e^{-i\theta_2}\right)^{q-4}$$

$$= k\ \sin(\theta_1-\theta_2)\ \sin(\theta_1+\theta_2)(\sin\theta_1\ \sin\theta_2)^{q-4}$$

which, ignoring constant terms which do not matter, is the same as (9), for $\alpha = (q-5)/2$, $\beta = -\frac{1}{2}$, $\gamma = 0$, and $s = \theta_1$, $t = \theta_2$. But these θ_i are just the critical angles of James and Constantine (1974), whose Gegenbaur polynomials are polynomials in y_1, y_2, which are the squares of the cosines of the critical angles. So, as a check, substituting $y_1 = \cos^2 s$, $y_2 = \cos^2 t$ in

$$(\sin s\ \sin t)^{q-4}\ \sin(s+t)\ \sin(s-t)\ dsdt$$

gives $\left((1-y_1)(1-y_2)\right)^{\frac{1}{2}(q-4)}(y_1-y_2)\left(y_1y_2(1-y_1)(1-y_2)\right)^{-\frac{1}{2}}\ dy_1dy_2$

$$= (y_1y_2)^{-\frac{1}{2}}\left((1-y_1)(1-y_2)\right)^{\frac{1}{2}(q-5)}\ (y_1-y_2)\ dy_1dy_2$$

as required, again ignoring constant terms.

Having recognized in this case that the Koornwinder variables, s and t, are just the critical angles, we can now interpret all the type 6 polynomials, orthogonal with respect to

$$\delta = (\sin s\ \sin t)^{2\alpha+1}(\mathrm{Cos}\ s\ \mathrm{Cos}\ t)^{2\beta+1}\left(\sin(s+t)\sin(s-t)\right)^{2\gamma+1}$$

for $\alpha = \frac{1}{2}(q-p-3)$, $\beta = \frac{1}{2}(p-3)$, $\gamma = 0$

as James and Constantine type Jacobi polynomials, which are orthogonal with respect to

$$(y_1y_2)^{\frac{1}{2}(p-3)}\ \left((1-y_1)(1-y_2)\right)^{\frac{1}{2}(q-p-3)}\ (y_1-y_2)$$

under the correspondence $y_1 = \mathrm{Cos}^2\ s$, $y_2 = \mathrm{Cos}^2\ t$. These are polynomials in y_1, y_2 constant on the cosets $(O(2,R)\times O(q-2,R)$ and $O(p,R)\times O(q-p,R)$, a subset of the total number of Jacobi polynomials. The variable change shifts us from functions on $SO(q,R)$ to functions on $O(q,R)$, where we now identify as one the opposite orientations of a hyper-plane.

Other values of α, β, γ also have symmetric space interpretations. Using the identity $\sin 2x = 2 \operatorname{Cos} x \operatorname{Sin} x$, we can re-write δ as

$$\delta = (\sin s \sin t)^{2\alpha-2\beta}(\sin 2s \sin 2t)^{2\beta+1}\{\sin(s+t)\sin)s-t)\}^{2\alpha}$$

which corresponds more closely with symmetric space roots as the measure for a symmetric space whose positive restricted roots are is, it, $2is$, $2it$, $i(s+t)$, $i(s-t)$ with multiplicities $m(is) = m(it) = 2\alpha-2\beta$, $m(2is) = m(2it) = 2\beta+1$, $m(i(s+t))=m(i(s-t))$ $= 2\gamma+1$. But the Dynkin diagram for fundamental roots it, $i(s-t)$ is

as we saw earlier. So, to find all symmetric space interpretations one need only work out all possible multiplicities for this diagram, or, even easier, look them up in a table such as Araki (1962), pp.32, 33. Koornwinder (1975) does just this, and lists the spaces including the null space $\alpha = \beta = \gamma = -\frac{1}{2}$, and the simple group $SO(5,R)$ (see §5). Each simple group, G, can be considered as a symmetric space in its own right, acting on $G \times G$, with the same roots as G considered as a semi-simple group, but with multiplicities 2 for each root (see Helgason (1962), chapter 4, and Loos (1969) p. 83). For some reason he leaves out the  spaces, which by relabelling the roots can be considered as a  space (this only happens for rank two spaces;  $\neq$ ).

These are:

$SU(4)/SU(2) \times SU(2)$	$m(\lambda_1) = 2$, $m(\lambda_2) = 1$
$SP(2)/U(2)$	$m(\lambda_1) = m(\lambda_2) = 1$
$SP(4)/SP(2) \times SP(2)$	$m(\lambda_1) = 4$, $m(\lambda_2) = 3$
$SO(8)/U(4)$	$m(\lambda_1) = 4$, $m(\lambda_2) = 1$

of which none have double roots.

Koornwinder's type 7 polynomials (Koornwinder (1972) §4.5) also have symmetric space interpretation. These are defined as polynomials in $Z = x + iy$ and $\bar{Z} = x - iy$, orthogonal with respect to the weight function.

$$\hat{\delta} = [-(x^2+y^2+9)^2 + 8(x^3-3xy^2) + 108]^{\alpha}$$

which under the transformation $Z = e^{i(s+t/\sqrt{3})} + e^{i(-s+t/\sqrt{3})} + e^{-2it/\sqrt{3}}$ become functions in s,t orthogonal with respect to

$$(\sin s)^{2\alpha+1} \quad \sin(\tfrac{1}{2}\, s+\tfrac{1}{2}\sqrt{3}t)^{2\alpha+1} \quad (\sin \tfrac{1}{2}\, s-\tfrac{1}{2}\sqrt{3}t)^{2\alpha+1} .$$

These functions are claimed to be spherical functions on compact symmetric spaces of rank 2, with Dynkin diagram , and multiplicities $2\alpha+1$ for each root. But we know that $S\ell(3,R)/SO(3,R)$ has Dynkin diagram , with positive roots $\theta_1-\theta_2$, $\theta_2-\theta_3$, $\theta_1-\theta_3$, multiplicities 1, where the Cartan subalgebra is the set of diagonal matrices

$$\left(\begin{pmatrix} \theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 \\ 0 & 0 & \theta_3 \end{pmatrix} \in \mathcal{L}S\ell(3,R), \quad \theta_i \in R \right)$$

which, since trace $X = 0 \ \forall\ X \in \mathcal{L}S\ell(3,R)$ means $\theta_1+\theta_2+\theta_3 = 0$. But this is a non-compact space. To find the dual of a space G/H, one simply takes the Cartan decomposition of $\mathcal{G} = \mathcal{K} + \mathcal{M}$ and transforms to $\mathcal{K} + i\mathcal{M}$, which then corresponds to the Cartan decomposition of the dual of G/H (see Helgason (1962) p. 199), and for G non-compact, the dual is compact, and vice-versa. For $S\ell(3,R)$, the decomposition is just

$\mathcal{L}S\ell(3,R) = \mathcal{L}SO(3,R)$ + symmetric matrices, trace zero

which transforms to

$\mathcal{L}SO(3,R)$ + i(symmetric matrices, trace zero).

If X is such an element, then $X + \bar{X}' = 0$. But this is just the rule for the Lie algebra of $SU(3)$, giving the compact dual of $S\ell(3,R)/SO(3,R)$ as $SU(3)/SO(3,R)$.

The Cartan subalgebra transforms to diagonal matrices, elements $i\theta_1$, $i\theta_2$, $i\theta_3$, with roots $i\theta_1 - i\theta_2$, $i\theta_2 - i\theta_3$, $i\theta_1 - i\theta_3$, multiplicities 1, and the same Dynkin diagram. So, for $\alpha = 0$, the type 7 polynomials should be spherical on $SU(3,R)/SO(3,R)$, which is the semi-simple part of the symmetric space $U(3,R)/O(3,R)$, the compact dual of $G\ell(3,R)/O(3,R) = M(3,R)$. Complete lists of the irreducible spaces and their duals can be found in Helgason (1962) pp. 346, 354.

Using the Harish-Chandra formula again, the measure on $SU(3,R)/SO(3,R)$ is just

$$\begin{aligned}\delta &= \left(e^{i(\theta_1-\theta_2)} - e^{-i(\theta_1-\theta_2)}\right)\left(e^{i(\theta_2-\theta_3)} - e^{-i(\theta_2-\theta_3)}\right)\left(e^{i(\theta_1-\theta_3)} - e^{-i(\theta_1-\theta_3)}\right) \\ &= \text{const.} \times \sin(\theta_1-\theta_2)\ \sin(\theta_2-\theta_3)\ \sin(\theta_1-\theta_3) \\ &= \text{const.} \times \sin(\theta_1-\theta_2)\ \sin(\theta_1+2\theta_2)\ \sin(2\theta_1+\theta_2).\end{aligned}$$

Ignoring constant terms, this corresponds with $\hat{\delta}$ for $\theta_1 = \frac{1}{2}(s-t/\sqrt{3})$, $\theta_2 = t/\sqrt{3}$, and $\alpha = 0$. Obviously any symmetric space with these roots, multiplicities $2\alpha+1$ will correspond with $\hat{\delta}$ for these θ_i, which justifies Koornwinder's claim. Z becomes simply $Z = e^{-2i\theta_1} + e^{-2i\theta_2} + e^{-2i\theta_3}$. (I have not figured out why such unwieldy parameters were used instead of the more natural θ_i's.)

He further states that polynomials in $Z = \zeta^{-\frac{1}{3}}\xi$, $\bar{Z} = \zeta^{-\frac{2}{3}}\xi$, multiplied by $\zeta^{k_1+k_2+k_3/3}$ can be identified with the zonal polynomials $C_k(X)$, roots x_1, x_2, x_3, for $\xi = x_1+x_2+x_3$, $\eta = x_1x_2+x_2x_3+x_3x_1$, $\zeta = x_1x_2x_3$. To see that this is in fact correct, let y_1, y_2, y_3 be the diagonal entries in in a member of the Cartan subgroup (of diagonal matrices) of $G\ell(3,R)$. Then the $y_j/(y_1y_2y_3)^{\frac{1}{3}}$'s can be considered as the entries in the members of the Cartan subgroup of $S\ell(3,R)$, (Det = 1) corresponding to the e^{θ_j}'s. If we identify the $\lambda_j = y_j/(y_1y_2y_3)^{\frac{1}{3}}$ with the $e^{i\theta_j}$'s in the Cartan subgroup of $SU(3)$ then the polynomials in the $e^{2i\theta_j}$'s become polynomials in the $\lambda_j{}^2$'s. These y_j's are the latent roots of elements in $G\ell(3,R)$, which correspond to $y_j y_j{}' = y_j{}^2 = x_j$, the latent roots of $gg' = m\varepsilon M(3,R)$. In $SU(3)$, the elements $\zeta^{-\frac{1}{3}}\xi$ and $\zeta^{-\frac{2}{3}}\eta$ become

$$\lambda^{-\frac{1}{3}}\xi = \frac{x_1+x_2+x_3}{(x_1 x_2 x_3)^{\frac{1}{3}}} = \lambda_1^2+\lambda_2^2+\lambda_3^2 = \lambda_1^2+\lambda_2^2+1/\lambda_1^2\ \lambda_2^2$$

$$\zeta^{-\frac{2}{3}}\eta = \frac{x_1x_2+x_2x_3+x_1x_3}{(x_1 x_2 x_3)^{\frac{2}{3}}} = \lambda_1^2\ \lambda_2^2 + \lambda_2^2\ \lambda_3^2 + \lambda_1^2\ \lambda_3^2$$

$$= \lambda_1^2\ \lambda_2^2 + 1/\lambda_1^2 + 1/\lambda_2^2$$

which for $\lambda_j = e^{i\theta j}$ means $\zeta^{-\frac{1}{3}}\xi = \overline{\zeta^{-\frac{2}{3}}\eta}$ as required. But considering λ_j as $y_j/(y_1y_2y_3)^{\frac{1}{3}}$ gives spherical functions on $S\ell(3,R)/SO(3,R)$, and multiplying by the determinent factor gives the $C_k(X)$'s, ignoring problems of normalization. Without making this identification, one would get the spherical functions on $U(3)/O(3,R)$, as studied by Hua (1963).

As before, one can interpret functions corresponding to other values of α as spherical functions on symmetric spaces, by looking up all •——• symmetric spaces (with multiplicities $2\alpha+1$).

It is perhaps worthwhile noting that explicit forms for the type 6 and 7 Koornwinder polynomials only exist for certain parameter values, in some cases as simple products of univariate polynomials, or in others because of their interpretation in terms of James' style polynomials.

These Koornwinder polynomials are a lovely example of the power of deeper symmetric space results. Similar techniques could easily be used to define higher dimension orthogonal polynomials for all the symmetric spaces from their root structures. Perhaps some combination of James/Koornwinder results will provide new general explicit functions.

4.7 Generalized Hermite Polynomials

Once the generalized special functions have been fitted into the overall scheme of special function theory, it becomes much easier to spot any gaps. For example, as mentioned earlier, the generalized Laguerre plynomials are a limiting case of the generalized Jacobi polynomials, which follows a well-known analogous result for the univariate polynomials. That is, from James and Constantine (1974), the $P_k(X)$, for $X \varepsilon M(m,R)$ are polynomials orthogonal with respect to the measure

$$w(X) = \det X^{a-\frac{1}{2}(m+1)} \det (I-X)^{c-a-\frac{1}{2}(m+1)}$$
$$(a = \tfrac{1}{2}p,\ c = \tfrac{1}{2}q).$$

(Transforming to the latent roots, x_i, of X adds the $\prod_{i<j} (x_i - x_j)$ factors.) Letting $X = c^{-1} X$, then

$$\lim_{c\to\infty} c^{am} w(c^{-1}Z) = \operatorname{etr}(-Z) \det Z^{a-\frac{1}{2}(m+1)}$$

which provides the measure for the ${}_1F_1(-k,a,Z)$. This is an obvious generalization of the univariate case, $m = 1$, for $w(x) = x^{a-1}(1-x)^{c-a-1}$ and

$$\lim_{c\to\infty} (c)^a w(c^{-1}z) = e^{-z} z^{a-1},$$

the measure for the Laguerre polynomials (see Erdelyi *et al* (1953) Vol. II, p. 164).

But it is also well-known that in the univariate case the Hermite polynomials, which are orthogonal with respect to $e^{-\frac{1}{2}y^2}$ are the limiting case of the Laguerre polynomials, taking the limit of $e^{-z} z^{a-1}$ for $z = a + \sqrt{a}\, y$ as $a \to \infty$. So what will be the generalized Hermite polynomials? Taking the analogous limit for $Z = aI + \sqrt{a}\, Y$ gives the measure $e^{-\frac{1}{2}Y^2}$, with respect to which the generalized Hermite polynomials $He_k(Y)$ will be orthogonal. Taking progressive limits of the Laplace-Beltrami operator on $G_{m,q-m}$ gives the differential equations obeyed by the generalized Laguerre, and then the Hermite polynomials, and provides a method of constructing the $He_k(Y)$

polynomials in terms of the zonal polynomials.

From James and Constantine (1974), the $P_k(X)$ are eigenfunctions of the Laplace-Beltrami operator, which gives the differential equation

$$\{D^* + (c-m+1)\,E - \delta^* - (a-\tfrac{1}{2}m+\tfrac{1}{2})\,\varepsilon - (\rho_k+kc)\}\,P_k(X) = 0$$

where

$$D^* = \sum_i x_i^2 \frac{\partial}{\partial x_i^2} + \sum_{\substack{i,j\\ i\neq j}} \frac{x_i^2}{(x_i-x_j)} \frac{\partial}{\partial x_i}$$

$$\delta^* = \sum_i x_i \frac{\partial^2}{\partial x_i^2} + \sum_{i,j} \frac{x_i}{(x_i-x_j)} \frac{\partial}{\partial x_i}$$

$$E = \sum_i x_i \frac{\partial}{\partial x_i^2}\ , \qquad \varepsilon = \sum_i \frac{\partial}{\partial x_i}$$

and the roots of X are $x_1, x_2, \ldots, x_m$, and $\rho_k = \sum_i \{k_i(k_i-i)\}$ with $k = (k_1, k_2, \ldots, k_m)$, a non-increasing partition of k into m parts. (These x_i's are just the canonical correlations, the $\cos^2\theta_i$.) Substituting $X = c^{-1} Z$ in the operator gives:

$$\Big\{ \sum_i c^2\left(\frac{z_i}{c}\right)^2 \frac{\partial^2}{\partial z_i^2} + \sum_{i,j} \left(\frac{z_i}{c}\right)^2 \left(\frac{c}{z_i-z_j}\right) c\,\frac{\partial}{\partial z_i}$$

$$+ (c-m+1) \sum_i \frac{z_i}{c}\cdot c\,\frac{\partial}{\partial z_i}$$

$$- \sum_i \frac{z_i}{c}\cdot c^2 \frac{\partial^2}{\partial z_i^2} - \sum_{i,j} \frac{c z_i}{c(z_i-z_j)}\cdot c\,\frac{\partial}{\partial z_i}$$

$$- (a-\tfrac{1}{2}m+\tfrac{1}{2}) \sum_i c\cdot\frac{\partial}{\partial z_i} - (\rho_k+ck)\Big\}$$

$$= \Big\{ \sum_i z_i^2 \frac{\partial^2}{\partial z_i^2} + \sum_{i,j} \frac{z_i^2}{(z_i-z_j)} \frac{\partial}{\partial z_i} + (c-m+1) \sum_i z_i \frac{\partial}{\partial z_i}$$

$$- c\sum_i z_i \frac{\partial}{\partial z_i^2} - c \sum_{i,j} \frac{z_i}{(z_i-z_j)} \frac{\partial}{\partial z_i}$$

$$- c(a-\tfrac{1}{2}m+\tfrac{1}{2}) \sum_i \frac{\partial}{\partial z_i} - (\rho_k+ck)\Big\} = \Delta$$

Taking

$$\lim_{c\to\infty} -\frac{\Delta}{c} = \sum_i z_i \frac{\partial^2}{\partial z_i^2} + \sum_{i,j} \frac{z_i}{(z_i-z_j)} \frac{\partial}{\partial z_i}$$

$$+ (a-\tfrac{1}{2}m+\tfrac{1}{2}) \sum_i \frac{\partial}{\partial z_i} - \sum z_i \frac{\partial}{\partial z_i} + k$$

$$= \delta^* + (a-\tfrac{1}{2}m+\tfrac{1}{2})\,\varepsilon - E + k = \Delta^*$$

But from the confluence $P_k(c^{-1}Z) \to {}_1F_1(-k,a,Z)$ we must have

$$\left(\delta^* + (a-\tfrac{1}{2}m+\tfrac{1}{2})\,\varepsilon - E + k\right)\, {}_1F_1\,(-k,a,Z) = 0.$$

Now substituting $Z = aI + \sqrt{a}\,Y$ in Δ^* gives

$$\sum_i \frac{(a+\sqrt{a}\,y_i)}{(\sqrt{a})^2}\frac{\partial^2}{\partial y_i^{\,2}} + \sum_{i,j}\frac{(a+\sqrt{a}\,y_i)}{(\sqrt{a}\,y_i-\sqrt{a}\,y_j)} \cdot \frac{1}{\sqrt{a}}\frac{\partial}{\partial y_i}$$

$$+ (a-\tfrac{1}{2}m+\tfrac{1}{2})\sum_i\left(\frac{1}{\sqrt{a}}\right)\frac{\partial}{\partial y_i} - \sum_i \frac{(a+\sqrt{a}\,y_i)}{\sqrt{a}}\frac{\partial}{\partial y_i} = k$$

$$= \sum_i\left(1+\frac{y_i}{\sqrt{a}}\right)\frac{\partial^2}{\partial y_i^{\,2}} + \sum_{i,j}\frac{\left(1+\frac{y_i}{\sqrt{a}}\right)}{(y_i-y_j)}\frac{\partial}{\partial y_i}$$

$$+ \left(\sqrt{a} - \frac{(\tfrac{1}{2}m-\tfrac{1}{2})}{\sqrt{a}}\right)\sum_i\frac{\partial}{\partial y_i} - \sum_i(\sqrt{a}+y_i)\frac{\partial}{\partial y_i} + k$$

$$= \sum_i\left(1+\frac{y_i}{\sqrt{a}}\right)\frac{\partial^2}{\partial y_i^{\,2}} + \sum_{i,j}\frac{\left(1+\frac{y_i}{\sqrt{a}}\right)}{(y_i-y_j)}\frac{\partial}{\partial y_i}$$

$$- \frac{\tfrac{1}{2}(m-1)}{\sqrt{a}}\sum_i\frac{\partial}{\partial y_i} - \sum_i y_i\frac{\partial}{\partial y_i} + k \quad .$$

Taking $\lim\limits_{a\to\infty}\Delta^* = \sum\limits_i\frac{\partial^2}{\partial y_i^{\,2}} + \sum\limits_{i,j}\frac{1}{(y_i-y_j)}\frac{\partial}{\partial y_i} - \sum\limits_i y_i\frac{\partial}{\partial y_i} + k$

$$= \Delta^{**} .$$

But $(\varepsilon\delta^*-\delta^*\varepsilon) = \sum\limits_i\frac{\partial^2}{\partial y_i^{\,2}} + \sum\limits_{i,j}\frac{1}{(y_i-y_j)}\frac{\partial}{\partial y_i}$.

Let $\delta^{**} = (\varepsilon\delta^*-\delta^*\varepsilon)$

Then

$$\Delta^{**} = \delta^{**} - E + k$$

and from the confluence ${}_1F_1(-k,a,aI+\sqrt{a}\,Y) \to H_{e_k}(Y)$ we must have

$$(\delta^{**} - E + k)\,H_{e_k}(Y) = 0 \quad .$$

For $m = 1$, this becomes

$$\frac{\partial^2}{\partial y^2} - y\frac{\partial}{\partial y} + k \quad .$$

If, instead, one had put $Z = aI + \sqrt{2a}\,Y$ and then taken the limit as $a \to \infty$, the operator becomes $\tfrac{1}{2}\delta^{**}-E+k$, and we get another normalization of the $H_{e_k}(Y)$'s, now orthogonal with respect to e^{-Y^2}.

Consider the action of δ^{**} on $C^*_k(Y)$.

$$\delta^{**}C^*_k(Y) = (\varepsilon\delta^* - \delta^*\varepsilon)\ C^*_k(Y)$$

But $\varepsilon C^*_k(Y) = \sum_i \binom{k^i}{k} C^*_{k^i}(Y)$

and $\delta^* C^*_k(Y) = \sum_i \left(k_i + \tfrac{1}{2}(m-i) - 1\right) \binom{k^i}{k} C^*_{k^i}(Y)$.

(see James and Constantine (1974)).

Then

$$\Delta^{**} = \delta^{**} - E + k$$

where $k^i = (k_1, \ldots, k_i - 1, \ldots k_m)$, the i^{th} part reduced by 1 .

So,
$$\begin{aligned}
\delta^{**}C^*_k(Y) &= \sum_i \left(k_i - 1 + \tfrac{1}{2}(m-i)\right)\binom{k^i}{k}\left(\varepsilon C^*_{k^i}(Y)\right) \\
&\quad - \sum_i \binom{k^i}{k} \delta^* C^*_{k^i}(Y) \\
&= \sum_i \left(k_i - 1 + \tfrac{1}{2}(m-i)\right)\binom{k^i}{k} \sum_j \binom{k^i}{k^{ij}} C^*_{k^{ij}}(Y) \\
&\quad - \sum_i \binom{k^i}{k} \sum_j \left(k_j^{(i)} + =(m-j) - 1\right)\binom{k^i}{k^{ij}} C^*_{k^{ij}}(Y) \\
&= \sum_i\sum_j \left(k_i - k_j^{(i)} - \left(\frac{i-j}{2}\right)\right)\binom{k^i}{k}\binom{k^i}{k^{ij}} C^*_{k^{ij}}(Y) \\
&= \sum_i\sum_j a_{kij}\ C^*_{k^{ij}}(Y)
\end{aligned}$$

where $k_j^{(i)}$ means the j^{th} element of k^i.

For $m = 1$, this becomes $\delta^{**}C^*_k(y) = 2\binom{k}{k-2} C^*_{k-2}(y)$. If we now define $He_k(Y) = C^*_k(Y) + \sum_\lambda b_\lambda C^*_\lambda(Y)$ where $k - \ell = 2n$ for some n, and $\lambda \leqslant k$, then for $He_k(Y)$ to obey $(\delta^{**} - E + k)\ He_k(Y) = 0$, the b_λ must obey the recurrence relation:

$$b_\lambda = -\frac{1}{(k-\ell)} \sum_{\lambda_{ij}} b_{\lambda_{ij}}\ a_{\lambda_{ij}^{ij}}\ ,\quad b_k = 1\ . \tag{10}$$

For $m = 1$, this gives:

$$\begin{aligned}
b_\ell &= -\frac{1}{(k-\ell)}\ b_{\ell+2}\ a_\ell \\
&= -\frac{2\binom{\ell+2}{\ell}}{(k-\ell)}\ b_{\ell+2}
\end{aligned}$$

which is precisely the recurrence formula for the coefficients of the ordinary Hermite polynomials,

$$H_{e_k}(y) = k! \sum_{m=o}^{[k/2]} \frac{(-1)^m \, (x)^{k-2m}}{2^m m! \, (k-2m)!}$$

where $k - \ell = 2(m+1)$.

Also, $\left(\frac{\partial^2}{\partial y^2} - y \frac{\partial}{\partial y} = k\right) H_{e_k}(y) = 0$ which is $\Delta^{**} H_{e_k}(y) = 0$ for $m = 1$.

Thus our $H_{e_k}(Y)$ defined above, with the recurrence relation (10), can be considered as the multivariate analogue of the Hermite polynomials $H_{e_k}(y)$.

The following are the generalized Hermite polynomials, up to order 4.

$$H_{e_0}(X) = 1$$
$$H_{e_1}(X) = C_1^*(X)$$

$$H_{e_2}(X) = C_2^*(X) - 1$$
$$H_{e_{11}}(X) = C_{11}^*(X) + \tfrac{1}{2}$$

$$H_{e_3}(X) = C_3^*(X) - 3\, C_1^*(X)$$
$$H_{e_{21}}(X) = C_{21}^*(X) - \tfrac{1}{2}\, C_1^*(X)$$
$$H_{e_{111}}(X) = C_{111}^*(X) + \tfrac{3}{2}\, C_1^*(X)$$

$$H_{e_4}(X) = C_4^*(X) - 6\, C_2^*(X) + 3$$
$$H_{e_{31}}(X) = C_{31}^*(X) - \tfrac{1}{6}\, C_2^*(X) - \tfrac{7}{3}\, C_{11}^*(X) - \tfrac{1}{2}$$
$$H_{e_{22}}(X) = C_{22}^*(X) - \tfrac{8}{3}\, C_2^*(X) + \tfrac{5}{3}\, C_{11}^*(X) + \tfrac{7}{4}$$
$$H_{e_{211}}(X) = C_{211}^*(X) + \tfrac{5}{6}\, C_2^*(X) - \tfrac{1}{3}\, C_{11}^*(X) - \tfrac{1}{2}$$
$$H_{e_{1111}}(X) = C_{1111}^*(X) + 3\, C_{11}^*(X) + \tfrac{3}{4}$$

4.8 Some further possibilities

At this stage it appears likely that the deeper theory of Lie groups and symmetric spaces may prove useful in several different areas of multivariate theory. The works of Harish-Chandra (1957),(1958) on generalized fourier analysis of bi-invariant functions on symmetric spaces opens up the possibility of a theory of generalized characteristic functions. Instead of finding transforms using $e^{it\theta}$ over R, one could find transforms with respect to the spherical functions of arbitrary symmetric spaces. The Plancheral formulae relating the transforms to their inverse transforms are beginning to be calculated (see Bhanu-Murti (1960)).

Also, several of the characteristic differential equations obeyed by the ${}_aF_b\ (\ldots,\ldots,X)$ and ${}_aF_b\ (\ldots;\ldots;X,Y)$ functions have been found for small a,b values (Muirhead (1970), Constantine and Muirhead (1972), Fujikoshi (1974)). It is becoming clear that a basis for each equation, fixed a,b is in terms of commutators (Lie products) of the operators for the equations for lower a,b values. At this point there is not much use in carrying out the tedious and messy calculations to verify the equations for higher functions, but it may be possible to use the Lie algebra properties that are emerging to find a general formula for arbitrary a,b, such as exists for the univariate functions (Erdelyi *et al* (1953), Vol. I, p. 184).

APPENDIX

Coefficients of the Zonal Polynomials

The zonal polynomials, $C_k(Y)$ for $Y \in M(m,R)$, and $k = (k_1,\ldots,k_m)$, $\sum_i k_i = k$, being symmetric homogeneous polynomials in the y_i's, can be expressed as linear combinations of some basis set of symmetric functions in the y_i's. Subroutine POLLY calculates the coefficients C_λ required to write the renormalized zonal polynomials $Z_k(Y)$, where

$$C_k(Y) = [\chi_{[2k]}(I)] \; \frac{2^k \, k!}{(2k)!} \; Z_k(Y)$$

as

$$Z_k(Y) = \sum_{\lambda \leq k} c_\lambda M_\lambda$$

for $M_\lambda = y_1^{\ell_1} \, y_2^{\ell_2} \ldots y_m^{\ell_m}$ + symmetric terms, and the sum is taken over all partitions $\lambda = (\ell_1,\ldots,\ell_m) \leq k$, where $\lambda \leq k$ means that the first ℓ_i different from k_i has $\ell_i < k_i$ (the standard lexicographical ordering). This normalization is chosen because the c_λ's are integers, which facilitates the programming.

This routine follows entirely the method developed in James (1968), and wherever possible the same notation is used. Zonal polynomials of low order can be checked against the values given in the above paper. The final coefficient c_λ, $\lambda = (1,1,\ldots,1)$ is always $k!$ which also serves as a check. POLLY requires an auxiliary subroutine DOWN which, for a given k,m, and partition k, calculates the next lowest allowable partition. This subroutine could be used to generate partitions from the driver routine, so that complete tables of coefficients for a set k and m could be produced at one run. It is envisaged that the computed coefficients would be stored on magnetic tape, or perhaps punched on data cards, if m is small, to be used as required. This seems necessary because of the appreciable time involved in calculating high degree coefficients. The large numbers of such coefficients also preclude hand calculations from printed tables.

Language: ANSI Standard Fortran.

Structure

SUBROUTINE POLLY(K,M,KAPPA,N,PART,KOEF,IFAULT)

Formal Parameters

K	Integer	input:	degree of partition, k
M	Integer	input:	maximum number of non-zero parts to partition; the rank of matrix Y in $C_{\kappa}(Y)$.
KAPPA	Integer array (14)	input:	partition $\kappa = (k_1, k_2, \ldots, k_m)$ where KAPPA(i)=k_i and $\sum_{i=1}^{m} k_i = k$
N	Integer	output:	number of allowed partitions between input κ and lowest one.
PART	Integer array (101)	output:	each allowed partition stored as an integer without zeros, numbered 1 to N from highest partition to lowest, for use as a label (e.g. 3,1,1,0,0) stored as 311.)
KOEF	Integer array (101)	output:	the coefficient corresponding to the monomial symmetric function $M_{PART(i)}$ (e.g. $c_{PART(i)}$ = KOEF(i), i=1,...,N)
IFAULT	Integer	output:	see 'Failure Indications' below.

Failure Indications

IFAULT = 0	indicates	no errors found in input values
= 1	indicates	K exceeds 13
= 2	indicates	$M \leqslant 0$
= 3	indicates	$\sum_{i=1}^{M} \text{KAPPA}(i) \neq K$
= 4	indicates	KAPPA is not in non-increasing order. (e.g. KAPPA(i) < KAPPA(i+1) for some i.)

Auxiliary Algorithm

SUBROUTINE DOWN(K,M,KAPPA,L,IFLAG)

Formal Parameters

K	Integer	input:	same as above
M	Integer	input:	same as above
KAPPA	Integer array (14)	input:	same as above
		output:	next allowable partition below the input KAPPA
L	Integer	output:	non-zero length of the returned partition.
IFLAG	Integer	output:	takes value 1 from 0 if there is no lower acceptable partition.

Double Precision

The zonal polynomial coefficients, by construction, are integers, but they are also of the order K! Thus the maximum value of K is limited by the word length of the machine, which must be capable of storing K! and somewhat larger numbers as integers. This program is limited to, and dimensioned for, $K \leqslant 13$, the maximum size for a CDC 6400 machine. The labelling mechanism starts to give out around the same value, since the lowest partition (1,1,...,1) is stored as 111...1, K times.

This bound is of little practical importance, since the possibility of taking a power sum beyond $K = 12$ or 13, or of wanting to, is remote - there are simply too many partitions, (i.e. 627 for $K = 20$). However, the higher degree coefficients may be of use to the theoretical worker, and this routine has been successfully adapted to double precision, with coefficients up to order $K = 21$ evaluated. The necessary alterations are extensive though. The variables A,B,C,IDUM,KI,KJ,KOEF,PI,PN, PART,R,RI,RN,RHO,S,XI,XL must be changed to double precision, plus all the constant values used in any statements involving these variables. DOWN requires no changes.

Restrictions

The arrays PART, KOEF, RHO are dimensioned for full sets of coefficients up to $K = 13$. (There are 101 partitions of 13 into 13 or fewer parts.) These variables must be dimensioned to at least the number of partitions of the user's maximum K into M or fewer parts. If one is working with small matrices, and thus small values of M, the savings in time and space can be considerable. For example, in order to include the $K = 13$ terms in a power sum for a matrix of rank $M \geqslant 13$, $\frac{101\times102}{2}$ coefficients are required. But for a matrix of rank $M = 3$, only $\frac{21\times22}{2}$ of the coefficients are needed, since there are only 21 partitions of 13 into 3 or fewer parts. Whole $C_k(Y)$'s are unnecessary when k is of length greater than M, and also individual terms in required $C_k(Y)$'s, when their corresponding M_λ term would be zero for λ of greater length than M. For the case $M = 3$, PART, KOEF, and RHO would only need to be dimensioned to 21.

```
      SUBROUTINE POLLY(K,M,KAPPA,N,PART,KOEF,IFAULT)
C
C
C        CALCULATES COEFFICIENTS OF ZONAL POLYNOMIALS
C              USING THE METHOD OF A. T. JAMES   IN
C            - CALCULATION OF COEFFICIENTS OF ZONAL POLYNOMIALS
C              BY MEANS OF THE LAPLACE - BELTRAMI OPERATOR -
C                   ANNALS OF MATHEMATICAL STATISTICS  1968.
C
C
C
      DIMENSION KAPPA(14),KAPPB(14),KAPPC(14)
      INTEGER PART(101),KOEF(101),RHO(101),P,P1,PN,R1,RN
C
C       CHECK OF INITIAL VALUES AND KAPPA EXTENDED WITH ZEROS
C
      IFAULT = 0
      IF(K.GE.14)IFAULT = 1
      IF(M.LE.0)IFAULT = 2
      IF(IFAULT.NE.0)RETURN
      I = MINO(M,K)
      DO 1 J = 1, I
    1 KAPPB(J) = KAPPA(J)
      I = I + 1
      II = K + 1
      DO 2 J = I,II
    2 KAPPB(J) = 0
      N = 1
C
C        CALCULATES LABEL, PART(1), FOR THE FIRST COEFFICIENT
C
      P1 = 0
      P = 0
      NTOT = 0
      DO 4 I = 1, K
      II = I + 1
      IF(KAPPB(I).LT.KAPPB(II))GO TO 5
      IF(KAPPB(I).EQ.0)GO TO 6
      P1 = P1*10 +KAPPB(I)
      NTOT = NTOT + KAPPB(I)
      P = P + 1
    4 CONTINUE
      GO TO 6
    5 IFAULT = 4
      RETURN
    6 IF(NTOT.EQ.K)GO TO 7
      IFAULT = 3
      RETURN
    7 PART(1) = P1
C
C                  CALCULATES FIRST COEFFICIENT FOR POLYNOMIAL
C
      K1 = 1
      J = 0
      DO 10 L = 1, P
      DO 10 I = 1, L
      LP = L + 1
```

```
      XL = FLOAT(L)
      XI = FLOAT(I)
      A = FLOAT(KAPPB(I))
      B = FLOAT(KAPPB(L))
      C = FLOAT(KAPPB(LP))
      R = ( XL - XI + 1.0)*0.5 + A
      S = R - B
      R = R - C
    8 IF(R.EQ.S) GO TO 10
C
C
      R = R - 1.
      LL = INT(R*2.0)
      K1 = K1 * LL
      J = J + 1
      GO TO 8
   10 CONTINUE
C
      J = K - J
      KOEF(1) = K1*2**J
C
C                    CALCULATES VALUE OF RHO FOR KAPPA
C
      R1 = 0
      DO 15 I = 1, K
   15 R1 = R1 + KAPPB(I)*(KAPPB(I) - I)
      RHO(1) = R1
C
C         FINDS AND LABELS THE LOWER ORDER PARTITIONS
C
   20 CALL DOWN(K,M,KAPPB,P,IFLAG)
      IF(IFLAG.EQ.1)RETURN
      N = N + 1
      PN = 0
      DO 30 I = 1, P
   30 PN = PN*10 + KAPPB(I)
      PART(N) = PN
      RN = 0
      DO 50 I = 1, K
   50 RN = RN + KAPPB(I)*(KAPPB(I) - I)
      RHO(N) = RN
C
C         CALCULATES VALUES OF THE LOWER ORDER COEFFICIENTS
C
      KOEF(N) = 0
      DO 100 I = 2,P
      ITT = I - 1
      MA = KAPPB(I)
C
      DO 100 II = 1, MA
      DO 100 J = 1, ITT
C
      KS = K + 1
      DO 66 JJ = 1, KS
   66 KAPPC(JJ) = KAPPB(JJ)
      KAPPC(J) = KAPPC(J) + II
      KAPPC(I) = KAPPC(I) - II
```

```
      KI = KAPPC(I)
      KJ = KAPPC(J)
      ML = J
   68 IF(ML.EQ.1) GO TO 69
      MM = ML - 1
      IF(KAPPC(ML).LE.KAPPC(MM)) GO TO 69
      KK = KAPPC(ML)
      KAPPC(ML) = KAPPC(MM)
      KAPPC(MM) = KK
      ML = ML - 1
      GO TO 68
C
   69 ML = I
   72 MM = ML + 1
      IF(KAPPC(ML).GE.KAPPC(MM)) GO TO 80
      KK = KAPPC(ML)
      KAPPC(ML) = KAPPC(MM)
      KAPPC(MM) = KK
      ML = ML + 1
      GO TO 72
C
   80 IDUM = 0
      DO 81 JJ = 1, K
      IF(KAPPC(JJ).EQ.0) GO TO 85
      IDUM = IDUM * 10 + KAPPC(JJ)
   81 CONTINUE
   85 DO 90 JJ = 1, N
      IF(IDUM.EQ.PART(JJ)) GO TO 91
   90 CONTINUE
      GO TO 100
   91 KOEF(N) = KOEF(N) + (KJ - KI) * KOEF(JJ)
  100 CONTINUE
C
      IF(KOEF(N).EQ.0) GO TO 20
      KOEF(N) = KOEF(N) / (RHO(1) - RHO(N))
      GO TO 20
      END
```

```
      SUBROUTINE DOWN(K,M,KAPPA,L,IFLAG)
C
C           CALCULATES NEXT LOWEST PARTITION
C        IF NO ALLOWABLE LOWER PARTITION, IFLAG SET TO 1
C
      DIMENSION KAPPA(14)
      IFLAG=0
      I = M + 1
      DO 1 J = I, 14
    1 KAPPA(J) = 0
    2 IF(KAPPA(1).GT.1)GO TO 3
      IFLAG = 1
      RETURN
C
    3 DO 4 L = 2, 14
      IF(KAPPA(L).EQ.0)GO TO 5
      IF(KAPPA(L).EQ.1)GO TO 6
    4 CONTINUE
C
    5 I = L - 1
      KAPPA(I) = KAPPA(I) - 1
      KAPPA(L) = 1
      GO TO 11
C
    6 I = L - 1
      KK = KAPPA(I) - 1
      KAPPA(I) = KK
      J = K
      DO 7 JJ = 1,I
    7 J = J - KAPPA(JJ)
    8 IF(J.LE.KK) GO TO 9
      KAPPA(L) = KK
      J = J - KK
      L = L + 1
      GO TO 8
C
    9 KAPPA(L) = J
      I = L + 1
      DO 10 J = I, K
   10 KAPPA(J) = 0
   11 IF(L.GT.M)GO TO 2
      RETURN
      END
```

BIBLIOGRAPHY

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Anderson, T.W. (1963). Asymptotic theory for principal component analysis. Ann. Math. Statist. 34, 122-148.

Araki, S. (1962). On root systems and infinitesimal classification of irreducible symmetric spaces. J. Math. Osaka City Univ. 13, 1-34.

Bellman, R. (1960). Introduction to Matrix Analysis. McGraw-Hill, New York.

Bhanu Murti, T.S. (1960). Plancherel's measure for the factor space SL(n,R)/SO(n,R). Dokl. Akad. Nauk. S.S.S.R. 133, 503-506.

Blackith, R.E. and Reyment, R.A. (1971). Multivariate Morphometrics. Academic Press, London and New York.

Boerner, H. (1963). Representations of Group. North-Holland Publishing Co., Amsterdam.

Cartan, É. (1894). Sur la structure des groupes de transformations finis et continus. Thèse, Paris, Nony.

Cartan, É. (1914). Les groupes réels simples finis et continus. Ann. Sci. École Norm. Sup. 31, 263-355.

Cartan, É. (1927. Sur une classe remarquable d'espaces de Riemann. Bull. Soc. Math. France. 55, 114-134.

Cartan, É. (1929). Sur la determination d'un system orthogonal complet dans un espace de Riemann symmetriques clos. Rend. Circ. Mat. Palermo. 53, 217-252.

Chevalley, C. (1946). Theory of Lie Groups. Princeton Univ. Press, New Jersey.

Cohn, P.M. (1957). Lie Groups. Cambridge Univ. Press, Cambridge.

Constantine, A.G. (1963). Some non-central distribution problems in multivariate analysis. Ann. Math. Statist. 34, 1270-1285.

Constantine, A.G. (1966). The distribution of Hotelling's generalized T_0^2. Ann. Math. Statist. 37, 215-225.

Constantine, A.G. and Muirhead, R.J. (1972). Partial differential equations for hypergeometric functions of two argument matrices. J. Multivariate Anal. 2, 332-338.

Erdelyi, A. et al (1953). Higher Transcendental Functions, 1, 2, and 3. McGraw-Hill, New York.

Fujikoshi, Y. (1974). Partial differential equations for hypergeometric functions ${}_3F_2$ of matrix argument. To appear in J. Multivariate Statist.

Gangolli, R. (1972). Spherical functions on semi-simple Lie groups. In Symmetric Spaces (Boothby, W.M. and Weiss, G.L. editors). Marcel Dekker, Inc., New York.

Handscomb, D.C. (1958). On the random disorientation of two cubes. Canad. J. Math. 10, 85-88.

Harish-Chandra. (1953). Representations of semi-simple Lie groups, I. Trans. Amer. Math. Soc. 75, 185-243.

Harish-Chandra. (1954). Representations of semi-simple Lie groups, II and III. Trans. Amer. Math. Soc. 76, 26-65, 234-254.

Harish-Chandra. (1957). Fourier transforms on a semi-simple Lie algebra, I, II. Amer. J. Math. 79, 193-257, 653-686.

Harish-Chandra. (1958). Spherical functions on a semi-simple Lie Group, I, II. Amer. J. Math. 80, 241-310, 553-613.

Helgason, S. (1962). Differential Geometry and Symmetric Spaces. Academic Press, New York.

Helgason, S. (1972). A formula for the radial part of the Laplace-Beltrami operator. J. Diff. Geometry 6, 411-419.

Helgason, S. (1973). Functions on symmetric spaces. In Harmonic Analysis on Homogeneous Spaces. American Mathematical Society, Providence, Rhode Island.

Hermann, R. (1966). Lie Groups for Physicists. Benjamin, New York.

Herz, C.S. (1955). Bessel functions of matrix argument. Ann. Math. 61, 474-523.

Hotelling, H. (1936). Relations between two sets of variates. Biometrika. 28, 321-377.

Hua, L.K. (1963). Harmonic Analysis of Functions of Several Complex Variables in Classical Regions. Amer. Math. Soc.

James, A.T. (1954). Normal multivariate analysis and the orthogonal group. Ann. Math. Statist. 25, 40-75.

James, A.T. (1960). Distribution of the latent roots of the covariance matrix. Ann. Math. Statist. 31, 151-158.

James, A.T. (1961). Zonal polynomials of the real positive definite symmetric matrices. Ann. Math. 74, 3, 456-469.

James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. Ann. Math. Statist. 35, 475-501.

James, A.T. (1966). Inference on latent roots by calculation of hypergeometric functions of matrix argument. In Multivariate Analysis I (P.R. Krishnaiah ed.) Academic Press, New York.

James, A.T. (1968). Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator. Ann. Math. Statist. 39, 1711-1718.

James, A.T. (1969). Tests for equality of latent roots of the covariance matrix. In Multivariate Analysis II (P.R. Krishnaiah, ed.) Academic Press, New York.

James, A.T. (1973). The variance information manifold and the functions on it. In Multivariate Analysis III (P.R. Krishnaiah, ed.) Academic Press, New York.

James, A.T. and Constantine, A.G. (1974). Generalized Jacobi polynomials as spherical functions of the Grassmann manifold. Proc. London Math. Soc. (3). 29, 174-192.

Jolicoeur, P. and Mosimann, J.E. (1960). Size and shape variation in the painted turtle. A principal component analysis. Growth. 24, 339-354.

Kendall, M.G. and Stuart, A. (1958). The Advanced Theory of Statistics. Vol. 2. 3rd ed. C. Griffin, London.

Koornwinder, T.H. (1974). Orthogonal polynomials in two variables which are eigenfunctions of two algebraically independent partial differential operators, I, II, III, IV. Nederl. Akad. Wetensch. Proc. Ser. A77 - Jndag. Math. 36, 48-66, 357-381.

Koornwinder, T.H. (1975). Two variable analogues of the classical orthogonal polynomials. To appear in Proceedings Advanced Seminar on Special Functions, Madison, 1975.

Loos, O. (1969). Symmetric Spaces, I and II. W.A. Benjamin, Inc., Amsterdam.

Maass, H. (1955). Die bestimmung der dirichletreihen mit grössencharacteren zu den modulformen n-ten grades. J. Indian Math. Soc. 19, 1-23.

McLaren, M.L. (1976). Coefficients of the zonal polynomials. An algorithm to appear in Applied Statistics, JRSS (series C).

Miller, W. (1968). Lie Theory and Special Functions. Academic Press, New York.

Miller, W. (1972). Symmetry Groups and Their Applications. Academic Press, New York.

Montgomery, D. and Zippin, L. (1955). Topological Transformation Groups. Interscience Publishers. New York and London.

Muirhead, R.J. (1970). Systems of partial differential equations for hypergeometric functions of matrix argument. Ann. Math. Statist. 41, 991-1001.

Pontrjagin, L. (1958). Topological Groups. Princeton University Press, New Jersey.

Roy, S.N. (1947). A note on critical angles between two flats in hyperspace with certain statistical applications. Sankhyā. 8, 177-194.

Samelson, H. (1969). Notes on Lie Algebras. Van Nostrand Reinhold Co., New York.

Satake, I. (1960). On representations and compactifications of symmetric Riemannian spaces. Ann. Math. 71, 77-110.

Subrahmaniam, Kocherlakota, (1974). Recent trends in multivariate normal distribution theory: on the zonal polynomials and other functions of matrix argument. Report from University of Manitoba, Dept. of Statistics, Winnipeg.

Talman, J. (1968). Special Functions. A Group Theoretic Approach. Benjamin, New York.

Varadarajan (1973). The theory of characters and the discrete series for semi-simple Lie groups. In Harmonic Analysis on Homogeneous Spaces. American Mathematical Society, Providence, Rhode Island.

Varadarajan, V.S. (1974). Lie Groups, Lie Algebras, and their Representations. Prentice-Hall, Inc., New Jersey.

Vilenkin, N. (1968). Special Functions and the Theory of Group Representations. Amer. Math. Soc. Transl. Providence, Rhode Is.

Weyl, H. (1925). Theorie der darstellung kontinuerlicher halb einfacher gruppen durch lineare transformationen, I. Mathematische Zeitschrift 23, 327-309.

Weyl, H. (1926 a and b). Theorie der darstellung kontinuerlicher halbeinfacher gruppen durch lineare transformationen, II, III. Mathematische Zeitschrift 24, 328-376, 377-395.

Weyl, H. (1935). The Structure and Representation of Continuous Groups. (Lecture notes). Princeton Univ. Press, New Jersey.

Weyl, H. (1939). The Classical Groups. Princeton University Press, New Jersey.