# The Variance Information Manifold and the Functions on It

*A. T. JAMES*
UNIVERSITY OF ADELAIDE

## 1. THE VARIANCE INFORMATION MANIFOLD AND ITS BOUNDARY

The $m \times m$ variance matrices $\Sigma$ form a convex cone in $\frac{1}{2}m(m+1)$-dimensional Euclidean space $R^{1/2\, m(m+1)}$, whose interior consists of the positive definite matrices. Each nonsingular variance matrix $\Sigma$ is associated with an information matrix $J = \Sigma^{-1}$. We consider the variance information manifold (VIM) as a space in which each interior point has alternative coordinate matrices $\Sigma$ or $J = \Sigma^{-1}$.

The singular positive semidefinite matrices $\Sigma$, which form the surface of the cone of variance matrices in $R^{1/2\, m(m+1)}$, are allotted corresponding points in the variance information manifold to form part of its boundary. Likewise, the singular information matrices $J$ are allotted points in VIM to form a disjoint part of its boundary. The boundary of VIM is completed by a set of points of singular variance and information.

The variance information manifold may be realized by the mapping

$$(\Sigma, J) \to \Sigma(I + \Sigma)^{-1} = (I + J)^{-1} = \Lambda$$

of variance and information matrices into the space of $m \times m$ symmetric matrices whose latent roots $\lambda_i$ all lie between 0 and 1;

$$0 \le \lambda_i \le 1.$$

VIM points whose latent roots $\lambda_i$ are all less than one correspond to a variance matrix $\Sigma = \Lambda(I - \Lambda)^{-1}$; those with latent roots all greater than zero have an information matrix $J = \Lambda^{-1}(I - \Lambda)$.

If both zero and one occur among the latent roots of a VIM point, then it has neither a variance matrix $\Sigma$ nor an information matrix $J$, but it is singular in both variance and information. Nevertheless, one can consider the eigenspace of the zero root, $\lambda = 0$, as a subspace of sample space $R^n$ in which there is zero variance and infinite information, and the eigenspace of the unit root, $\lambda = 1$, as a subspace of infinite variance and zero information.

The Euclidean topology of the space of symmetric matrices with latent roots between 0 and 1 supplies a topology for the variance information manifold. With the inclusion of its boundary, the VIM is compact.

## 2. THE BIVARIATE CASE

In the case $m = 2$, the VIM is a double cone in $R^3$, base to base. A slight transformation will put it up on end. We introduce cylindrical polar coordinates $(r, \theta, z)$ as the following functions of the elements of

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$
$$= \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} = J^{-1},$$
$$\lambda_1 \geq \lambda_2, \quad 0 \leq \varphi < \pi,$$
$$t = \operatorname{trace}\Sigma = \sigma_{11} + \sigma_{22}, \qquad \Delta = \det\Sigma,$$

namely,

$$r = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2 + \lambda_1\lambda_2} = \frac{(t^2 - 4\Delta)^{1/2}}{1 + t + \Delta},$$
$$\theta = 2\varphi = \arctan\frac{2\sigma_{12}}{\sigma_{11} - \sigma_{22}} = \arctan\frac{2j_{12}}{j_{11} - j_{22}},$$
$$z = \frac{\lambda_1\lambda_2 - 1}{1 + \lambda_1 + \lambda_2 + \lambda_1\lambda_2} = \frac{\Delta - 1}{1 + t + \Delta}.$$

VIM is represented as a double cone with singular variance as the lower surface and singular information as the upper. Points on the circle where the upper and lower surfaces intersect have both singular information and variance. In terms of the distribution of a variate

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

in sample space $R^2$, singular information implies that probability statements only apply to some linear function

$$l_1 x_1 + l_2 x_2.$$

If this were to have nonzero variance $\sigma^2$, the information matrix would be

$$J = \frac{1}{\sigma^2}\begin{bmatrix} l_1{}^2 & l_1 l_2 \\ l_2 l_1 & l_2{}^2 \end{bmatrix}.$$

However, if there is singular variance as well as singular information, then the linear function is constant

$$l_1 x_1 + l_2 x_2 = c$$

and hence a distribution in $R^2$ of singular variance and information is equivalent to a linear constraint (Fig. 1). We have

$$\tan\theta = \tan 2\varphi = \frac{2l_1 l_2}{l_1^2 - l_2^2}$$

hence

$$\tan\varphi = -\frac{l_1}{l_2}.$$

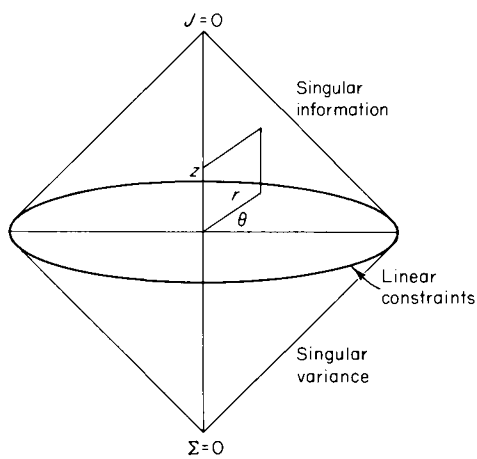


**Fig. 1**

The multinormal distribution with singular variance matrix is well known; in the following section, we introduce the multinormal distribution with singular information matrix.

## 3. THE MULTINORMAL DISTRIBUTION WITH SINGULAR INFORMATION MATRIX

### Heuristic Introduction

The multinormal distribution with a singular variance matrix $\Sigma$ is a distribution of a vector variate $\mathbf{x} \in R^n$ such that with probability one, its deviation $\mathbf{x} - \boldsymbol{\mu}$ from expectation $\boldsymbol{\mu}$ lies in a subspace. Now the dual of a subspace

$\mathscr{L}$, i.e., the set of linear functionals on it, is a quotient space (not a subspace), namely, the quotient space $R^{n*}/\mathscr{L}^{\perp}$ of the dual space $R^{n*}$ of $R^n$ over the annihilator $\mathscr{L}^{\perp} = \mathscr{S}$ of $\mathscr{L}$, $\mathscr{S} \subset R^{n*}$ being the set of all linear functionals which vanish on $\mathscr{L}$. A point $x^* + \mathscr{S}$ of the quotient space $R^{n*}/\mathscr{S}$ is a coset

$$\mathbf{x}^* + \mathscr{S} \overset{\text{def}}{=} \{\mathbf{x}^* + y^* | y^* \in \mathscr{S}\}.$$

The coset $x^* + \mathscr{S}$ clearly consists of all functionals $x^* \in R^{n*}$ which have the same values when restricted to $\mathscr{L}$. Any element of the coset determines it. Geometrically, the cosets may be represented by hyperplanes (Fig. 2).

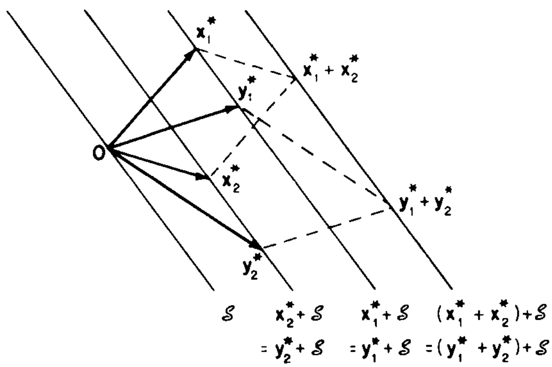


**Fig. 2**

The quotient space $R^{n*}/\mathscr{S}$ forms a vector space under addition and scalar multiplication (Fig. 3):

$$(\mathbf{x}_1^* + \mathscr{S}) + (\mathbf{x}_2^* + \mathscr{S}) = ((\mathbf{x}_1^* + \mathbf{x}_2^*) + \mathscr{S}),$$
$$\lambda(\mathbf{x}^* + \mathscr{S}) = (\lambda\mathbf{x}^* + \mathscr{S}).$$

The dual of a multinormal distribution with a singular variance matrix will then be a distribution with singular information matrix which will be a distribution on the quotient space $R^n/\mathscr{S}$. Notice that we now find it convenient to switch to the space, $R^n$, rather than its dual, $R^{n*}$.

If $\mathscr{S}$ is a subspace of $R^n$, then we have, from the definition of conditional probability, $P(\mathbf{x}|\mathbf{x} + \mathscr{S})$,

$$P(\mathbf{x}) = P(\mathbf{x} + \mathscr{S})P(\mathbf{x}|\mathbf{x} + \mathscr{S}).$$

The second factor is a distribution with singular variance matrix; the first factor will be an example of a distribution with singular information matrix that we are about to define.

The situation in normal regression theory suggests a method of definition. Suppose a vector variate $\mathbf{y} \in R^n$ is distributed as $N(X\boldsymbol{\beta}, \sigma^2 I_n)$ where $X$ is an $N \times p$ matrix of rank $r < p$, and $\boldsymbol{\beta} \in R^p$.

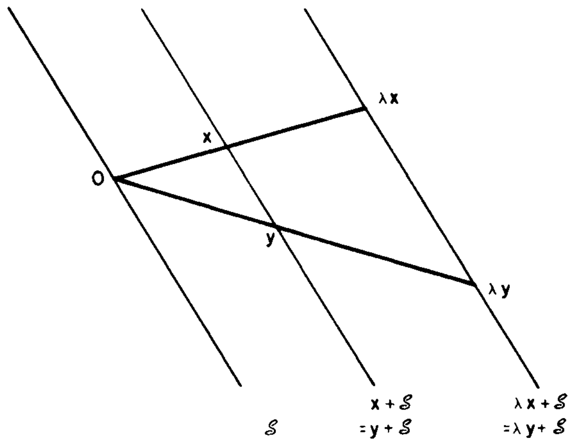


**Fig. 3**

Since all vectors in the coset $\boldsymbol{\beta} + \mathscr{K}(X)$ determine the same expectation of $\mathbf{y}$, where $\mathscr{K}(X)$ is the kernel or null space of $X$ consisting of all vectors annihilated by $X$, our estimator can be considered as a coset $\mathbf{b} + \mathscr{R}(X)$ consisting of all solutions of the normal equations

$$X'X\mathbf{b} = X'\mathbf{y}.$$

Note that $\mathscr{K}(X'X) = \mathscr{K}(X)$.

When $r = p$, the matrix $J = (1/\sigma^2)X'X$ is the inverse of the variance matrix of $\mathbf{b}$. Hence when $r < p$ and $J$ is singular, we can define that $\mathbf{b}$ or $\mathbf{b} + \mathscr{K}(X)$ has a singular information matrix $J = (1/\sigma^2)X'X$.

By dividing the normal equations by $\sigma^2$,

$$J\mathbf{b} = \frac{1}{\sigma^2} X'\mathbf{y},$$

we can see how to proceed with our definition in the general case, because $J\mathbf{b}$ is distributed as

$$N\left(\frac{1}{\sigma^2} X'X\boldsymbol{\beta}, \frac{1}{\sigma^2} X'X\right) = N(J\boldsymbol{\beta}, J).$$

**Definition.** *A vector variate* $\mathbf{x} \in R^n$ *is normally distributed with information matrix* $J$, *which may be singular, if* $J\mathbf{x}$ *is distributed as* $N(\boldsymbol{\mu}^*, J)$ *for some* $\boldsymbol{\mu}^* \in \mathscr{R}(J)$.

Since the distribution is only defined by the values of $J\mathbf{x}$, we define an equivalence relation, $\sim$, on $R^n$, namely, $\mathbf{x}_1 \sim \mathbf{x}_2$ if $J\mathbf{x}_1 = J\mathbf{x}_2$, i.e., if

$J(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, i.e., if $\mathbf{x}_1 - \mathbf{x}_2 \in \mathscr{K}(J)$. The equivalence classes are then the cosets $\mathbf{x} + \mathscr{K}(J) \in R^n/\mathscr{K}(J)$.

Just as a distribution with singular variance matrix may formally be regarded as either a distribution within $R^n$ which has all its probability on a hyperplane or alternatively as a distribution defined on a hyperplane; similarly, we may regard the sample space of a distribution with singular information matrix as either $R^n$ or $R^n/\mathscr{K}(J)$, according to convenience. If the sample space is taken to be $R^n$, then the only measurable sets are unions of cosets.

As a distribution on cosets, the expectation coset $\boldsymbol{\mu} + \mathscr{K}(J)$ is defined to be the set of all solutions $\boldsymbol{\mu}$ of $J\boldsymbol{\mu} = \boldsymbol{\mu}^*$.

Another way of considering the distribution in $R^n$ is as an improper distribution, because in some directions it is uniform with infinite variance.

## 4. DERIVATION VIA THE DISTRIBUTION OF LINEAR FUNCTIONS

Suppose that $L'\mathbf{x}$ has variance matrix $\Sigma_1$, where $L$ is an $n \times p$ constant matrix and $\Sigma_1$ is $p \times p$. If $M$ were square and nonsingular, then $\mathbf{x}$ would have variance matrix $L'^{-1}\Sigma_1 L^{-1}$. If $\Sigma_1$ is nonsingular, then $\mathbf{x}$ would have information matrix $L\Sigma_1^{-1}L'$. This suggests the

**Lemma.** *If the distribution of* $\mathbf{x} \in R^n$ *is given by the fact that* $L'\mathbf{x}$ *has a nonsingular variance* $\Sigma_1$, *where* $L$ *is an* $n \times p$ *constant matrix, then* $\mathbf{x}$ *has information matrix* $L\Sigma_1^{-1}L'$.

*Proof.* To prove that $\mathbf{x}$ has information matrix $J$, we must prove that $J\mathbf{x}$ has variance matrix $J$. Now

$$\begin{aligned} V(L\Sigma_1^{-1}L'\mathbf{x}) &= V((L\Sigma_1^{-1})(L'\mathbf{x})) \\ &= L\Sigma_1^{-1}\Sigma_1\Sigma_1^{-1}L' \\ &= L\Sigma_1^{-1}L', \end{aligned}$$

and the lemma follows.

## 5. APPLICATION TO THE ANALYSIS OF EXPERIMENTAL DESIGNS

The multinormal distributions with singular information matrices are useful in the analysis of experimental designs. The experiment often splits into $r$ easily analyzable parts or groups, each of which supplies information about the vector $\boldsymbol{\beta}$ of treatment effects. The normal equations can be written separately for each of the $i$th parts of the experiment, and after eliminating the estimates of block effects one obtains *reduced* normal equations

$$J_i \mathbf{b}_i = \mathbf{h}_i, \qquad i = 1, \ldots, r,$$

of which the solutions $\mathbf{b}_i + \mathscr{K}(J_i)$ are estimates of $\boldsymbol{\beta}$ with singular information matrix $J_i$, quasi-sufficient for each $i$th part of the experiment. The combined quasi-sufficient estimator is then given by the solution of the equation

$$\left(\sum_{i=1}^{r} J_i\right)\mathbf{b} = \sum_{i=1}^{r} \mathbf{h}_i .$$

## 6. REPRESENTATION AS THE MARGINAL DISTRIBUTION OF A NONSINGULAR DISTRIBUTION

The distribution with singular information matrix $J$ may be represented as the marginal distribution of a nonsingular distribution. Namely, if $\Sigma$ is a nonsingular generalized inverse of $J$ and $\boldsymbol{\mu}$ is any vector in the expectation coset of the singular distribution, then if $\mathbf{x}$ is a variate distributed as $N(\boldsymbol{\mu}, \Sigma)$, then $J\mathbf{x}$ is distributed as $N(J\boldsymbol{\mu}, J)$ because $J\Sigma J = J$. Hence the marginal distribution of $\mathbf{x} + \mathscr{K}(J)$ has information matrix $J$.

The singular information matrices appear in the general decomposition of a multinormal distribution.

## 7. DECOMPOSITION OF A MULTINORMAL DISTRIBUTION

Let the vectors $\mathbf{x}$ in the sample space $R^n$ be called covariant. Then the (column) vectors $\mathbf{x}^*$ of coefficients of linear functions or contrasts belong to the dual space $R^{n*}$ and are thus contravariant. The variance matrix $\Sigma$ gives the natural inner product $\mathbf{x}^{*\prime}\Sigma\mathbf{y}^*$ in the space $R^{n*}$ of contrasts. As its indices must both contract with indices of contravariant vectors $\mathbf{x}^*$, $\mathbf{y}^*$, it must be doubly covariant.

The information matrix $\Sigma^{-1}$ supplies the natural inner product $\mathbf{x}'\Sigma^{-1}\mathbf{y}$ in the sample space $R^n$ and is hence doubly contravariant.

The matrix of a linear transformation of sample space is covariant $\times$ contravariant and for a linear transformation of contrast space the matrix is contravariant $\times$ covariant.

Let $\Sigma = TT'$. Then the equation

$$I = E_1 + E_2$$

for the decomposition of the identity matrix $I$ into symmetric idempotent matrices $E_1$, $E_2$ of ranks $r$ and $n - r$ has four clearly distinguishable meanings which are shown up by the four different ways in which it can be transformed; namely,

| | |
|---|---|
| $E_i \to TE_iT'$ | cogrediently × cogrediently, |
| $E_i \to T^{-1\prime}E_iT^{-1}$ | contragrediently × contragrediently, |
| $E_i \to TE_iT^{-1}$ | cogrediently × contragrediently, |
| $E_i \to T^{-1\prime}E_iT'$ | contragrediently × cogrediently. |

We obtain a set of useful expressions for the four transforms of the decomposition equation if we let $Q$ be an $n \times r$ matrix whose columns are covariant vectors spanning $\mathscr{S} \subset R^n$ and $L$ be an $n \times (n-r)$ matrix of contravariant column vectors spanning $\mathscr{L} = \mathscr{S}^\perp \subset R^{n*}$; $L'Q = \mathbf{0}$:

| | |
|---|---|
| cov. × cov. | $\Sigma = Q(Q'\Sigma^{-1}Q)^{-1}Q' + \Sigma L(L'\Sigma L)^{-1}L'\Sigma,$ |
| contr. × contr. | $\Sigma^{-1} = \Sigma^{-1}Q(Q'\Sigma^{-1}Q)Q'\Sigma^{-1} + L(L'\Sigma L)^{-1}L',$ |
| linear transformations of $R^n$, cov. × contr. | $I_n = Q(Q'\Sigma^{-1}Q)^{-1}Q'\Sigma^{-1} + \Sigma L(L'\Sigma L)^{-1}L',$ |
| linear transformations of $R^{n*}$, contr. × cov. | $I_n = \Sigma^{-1}Q(Q'\Sigma^{-1}Q)^{-1}Q' + L(L'\Sigma L)^{-1}L'\Sigma.$ |

The first two equations are the decompositions of the variance and information matrices into singular variance and information components associated with the subspace $\mathscr{S}$ and its conjugate space

$$\Sigma\mathscr{L} = \{\mathbf{x} = \Sigma\mathbf{x}^* \in R^n \mid \mathbf{x}^* \in \mathscr{L}\} = \{\mathbf{x} \in R^n \mid \mathbf{y}'\Sigma^{-1}\mathbf{x} = 0 \quad \text{for all } \mathbf{y} \in \mathscr{S}\}.$$

Notice that a component, e.g., $Q(Q'\Sigma^{-1}Q)Q'$, of the variance matrix is uniquely determined by two of its properties:

1. the information matrix $\Sigma^{-1}$ is a generalized inverse of it;
2. its range is $\mathscr{R}(Q) = \mathscr{S}$ or its kernel is $\mathscr{L}$.

The same is true for the components of the information matrix, the variance matrix $\Sigma$ being their generalized inverse.

The third equation gives the idempotents projecting on $\mathscr{S}$ and $\Sigma\mathscr{L}$ and the fourth equation, the idempotents projecting on $\Sigma^{-1}\mathscr{S}$ and $\mathscr{L}$.

The structure of the matrices follows from some simple rules:

1. Matrix multiplication of indices of order $n$ must always involve summation (or contraction) of a cogredient index with a contragredient one.
2. As the matrices depend only upon the ranges of $Q$ and $L$, they must be invariant under transformations

$$Q \to QP', \qquad L \to LM'$$

where $P$ and $M$ are nonsingular matrices of respective orders $r$ and $(n-r)$.

3. The covariance or contravariance of the first and last indices depends upon which of the four equations that one wants.

## 8. INVARIANT METRIC

From the Wishart density

$$\text{const.}\det \Sigma^{-1/2n}\,\text{etr}(-\tfrac{1}{2}n\Sigma^{-1}S)$$

we have the log likelihood function

$$\mathscr{L} = \text{const.} - \tfrac{1}{2}n(\log\det\Sigma + \text{tr}\,\Sigma^{-1}S)$$

and its two differentials

$$d\mathscr{L} = -\tfrac{1}{2}n\,\text{tr}(\Sigma^{-1}d\Sigma - \Sigma^{-1}d\Sigma\Sigma^{-1}S),$$
$$d^2\mathscr{L} = -\tfrac{1}{2}n\,\text{tr}(-\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma + 2\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma\Sigma^{-1}S),$$

and the differential form in the information matrix as the expectation of $(-d^2\mathscr{L})$

$$E[-d^2\mathscr{L}] = \tfrac{1}{2}n\,\text{tr}(\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma).$$

The differential form

$$(ds)^2 = \text{tr}(\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma) = \text{tr}(J^{-1}dJJ^{-1}dJ),$$

where $J = \Sigma^{-1}$, is a useful metric on VIM.

Maass [1] has shown that such a metric differential form $(ds)^2 = \text{tr}(X^{-1}dXX^{-1}dX)$ on the space of $m \times m$ positive definite symmetric matrices $X$ is invariant under congruence transformation

$$X \to LXL'$$

where $L$ is an $m \times M$ nonsingular matrix.

If we write

$$X = HYH'$$

with $H$ orthogonal and $Y = \text{diag}(y_i)$, then the metric differential form becomes

$$(ds)^2 = \sum_{i=1}^{m} \frac{(dy_i)^2}{y_i^2} + \sum_{i<j}^{m} \frac{(y_i - y_j)^2}{y_i y_j}(d\theta_{ij})^2$$

where the $d\theta_{ij}$ denote the differential forms of the skew symmetric matrix $H'dH$. Upon putting $z_i = \log y_i$ for $i = 1, \ldots, m$, we have

$$(ds)^2 = \sum_{i=1}^{m}(dz_i)^2 + \sum_{i<j}^{m}(\sinh(\tfrac{1}{2}(z_i - z_j)))^2\,d\theta_{ij}.$$

From the metric differential form, we obtain the Laplace–Beltrami operator as

$$\sum_{i=1}^{m}\left\{y_i^2\left(\frac{\partial^2}{\partial y_i^2}\right)+\left(\sum_{j=1,\,j\neq i}^{m} y_i^2(y_i-y_j)-\tfrac{1}{2}(m-3)y_i\right)\left(\frac{\partial}{\partial y_i}\right)\right\}$$
$$+\frac{1}{4}\sum_{i<j} y_i y_j (y_i-y_j)^{-1}\left(\frac{\partial^2}{\partial\theta_{ij}^2}\right).$$

## 9. GEODESIC DISTANCE BETWEEN TWO MATRICES

To find the distance from $X_1$ to $X_2$, we let $L$ be a nonsingular matrix which by congruence transformation reduces $X_1$ to the identity matrix and $X_2$ to a diagonal matrix $Y = \operatorname{diag}(y_i)$ whose elements $y_i$ are the latent roots of the determinantal equation

$$\det(X_2 - yX_1) = 0; \tag{9.1}$$

$LX_1L' = I_m$, $LX_2L' = Y$.

Then since the distance $ds$ is invariant under congruence transformation, the distance between $X_1$ and $X_2$ is given by the following integral taken along a path which will minimize it. Hence we have

$$d(X_1, X_2) = \int_{X_1}^{X_2} ds = \int_{I_m}^{Y} ds = \int_{I_m}^{Y}\left(\sum_{i=1}^{m}\left(\frac{dy_i}{y_i}\right)^2\right)^{1/2}$$
$$= \int_{I_m}^{Z}\left(\sum_{i=1}^{m}(dz_i)^2\right)^{1/2} = \left(\sum_{i=1}^{m} z_i^2\right)^{1/2}.$$

It is fairly clear that this is the minimum and it can be checked by the calculus of variations. Hence we have the

**Theorem.** *If $X_1$, $X_2$ are two positive definite symmetric matrices, the geodesic distance between them is*

$$\left(\sum_{i=1}^{m}(\log y_i)^2\right)^{1/2}$$

*where the $y_i$ are the roots of the determinantal equation* (9.1).

Since $\frac{1}{2}n$ times our metric differential form is a quadratic form in the information matrix for $S$, it follows that:

**Theorem.** *If $y_1, \ldots, y_m$ are the latent roots of the equation*

$$\det(S - y_i\Sigma) = 0,$$

*then the statistic*

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

*is asymptotically distributed as* $\chi^2$ *on* $\tfrac{1}{2}m(m+1)$ *degrees of freedom.*

The statistic $d^2$ could be used as a test of a hypothesis which prescribes $\Sigma$.

## 10. ZONAL POLYNOMIALS

**Definition.** *A zonal polynomial is a symmetric homogeneous polynomial of degree k in the latent roots* $y_1, \ldots, y_k$ *of a matrix X which is an eigenfunction of the operator*

$$\Delta = \sum_{i=1}^{m} y_i^2\left(\frac{\partial^2}{\partial y_i^2}\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{m} y_i^2(y_i - y_j)^{-1}\left(\frac{\partial}{\partial y_i}\right) \tag{10.1}$$

*derived from the Laplace–Beltrami operator in the latent roots* $y_i$ *omitting the multiple of the Euler operator,* $\sum y_i(\partial/\partial y_i)$.

**Theorem.** *For each monomial* $y_1^{k_1} \cdots y_m^{k_m}$ *with* $k_1 \geq k_2 \geq \cdots \geq k_m$, *there is exactly one zonal polynomial with this as the term of highest weight. The eigenvalue is* $\rho_\kappa + k(m-1)$ *where* $\rho_\kappa = \sum_{i=1}^{m} k_i(k_i - i)$, $\kappa = (k_1, \ldots, k_m)$, $k_1 + \cdots + k_m = k$.

*Proof.* One can verify that

$$\Delta y_1^{k_1} \cdots y_m^{k_m} = (\rho_\kappa + k(m-1)) y_1^{k_1} \cdots y_m^{k_m} + \text{terms of lower weight.}$$

The zonal polynomial

$$C_\kappa(Y) = c_\kappa y_1^{k_1} \cdots y_1^{k_m} + \text{terms of lower weight}$$

satisfies the differential equation

$$\Delta C_k(Y) = (\rho_\kappa + k(m-1)) C_\kappa(Y), \tag{10.2}$$

which yields a recurrence relation from which the coefficients of terms of lower weight can be determined from the first coefficient $c_\kappa$.

Let $C_\kappa^*(Y)$ be a renormalized zonal polynomial, so that it is unity at the identity matrix

$$C_\kappa^*(I_m) = 1.$$

**Theorem.** *If X and Y are arbitrary symmetric matrices, then*

$$\int_{O(m)} C_\kappa^*(XHYH')(dH) = C_\kappa^*(X) C_\kappa^*(Y). \tag{10.3}$$

*Proof.* Since

$$C_\kappa^*(XHYH') = C_\kappa^*(X^{1/2}HY(X^{1/2}H)') = C_\kappa^*(LYL')$$

where $L = X^{1/2}H$, and the operator $\Delta$ is invariant under congruence transformation

$$\Delta_{LYL'} = \Delta_Y$$

and can be taken under the integral sign, we see that the left-hand side of (10.3), when considered as a function of $Y$ for fixed $X$, is an eigenfunction of $\Delta$, namely,

$$\begin{aligned}\Delta_Y \int_{O(m)} C_\kappa^*(XHYH')(dH) &= \int_{O(m)} \Delta_{LYL'} C_\kappa^*(LYL')(dH) \\ &= (\rho_\kappa + k(m-1)) \int_{O(m)} C_\kappa^*(LYL')(dH) \\ &= (\rho_\kappa + k(m-1)) \int C_\kappa^*(XHYH')(dH).\end{aligned}$$

Hence the left-hand side of (10.3) must be a multiple of the zonal polynomial

$$\int_{O(m)} C_\kappa^*(XHYH')(dH) = \lambda C_\kappa^*(Y).$$

Putting $Y = I_m$, we have $C_\kappa^*(X) = \lambda$ and the theorem follows.

Now suppose both $X = \operatorname{diag}(x_i)$ and $Y = \operatorname{diag}(y_i)$ are diagonal matrices. By equating coefficients of $x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$ on both sides of Eq. (10.3), one obtains the following integral representation of the zonal polynomial

$$C_\kappa^*(Y) = \int_{O(m)} \left(\sum_i h_{i1}^2 y_i\right)^{k_1-k_2} \left(\sum_{i_1<i_2} \det\begin{bmatrix} h_{i_11} & h_{i_12} \\ h_{i_21} & h_{i_22}\end{bmatrix}^2 y_i y_i\right)^{k_2-k_3} \cdots \det Y^{k_m}(dH). \tag{10.4}$$

This integral over the orthogonal group $O(m)$ can be transformed to an integral over $\frac{1}{2}m(m-1)$-dimensional Euclidean space by representing the orthogonal matrix $H$ as a Gramm–Schmidt orthonormalization of a triangular matrix with unities on the diagonal:

$$Z = (z_{ij}) = \begin{bmatrix} 1 & 0 & 0 \\ z_{21} & 1 & \\ \vdots & & \ddots & \\ z_{m1} & \cdots & & 1\end{bmatrix} = [\mathbf{z}_1 \mathbf{z}_2 \cdots \mathbf{z}_m].$$

The volume element $(dH)$ becomes

$$(dH) = \frac{\Gamma_m(\frac{1}{2}m)}{\pi^{1/2\,m^2}} \frac{1}{D_1 D_2 \cdots D_{m-1}} (dZ)$$

where $D_k$ is the first $k$th-order principal minor of $Z'Z$:

$$D_k = \det \begin{bmatrix} \mathbf{z}_1'\mathbf{z}_1 & \cdots & \mathbf{z}_1'\mathbf{z}_k \\ \vdots & & \vdots \\ \mathbf{z}_k'\mathbf{z}_1 & \cdots & \mathbf{z}_k'\mathbf{z}_k \end{bmatrix}$$

and the integral transforms to a form similar to one given by Bhanu Murti [2]

$$C_\kappa^*(Y) = \frac{\Gamma_m(\frac{1}{2}m)}{\pi^{1/2\, m^2}}$$

$$\cdot \int_{R^{1/2\, m(m-1)}} \frac{\left(\sum_i z_{i1}^2 y_i\right)^{k_1-k_2} \left(\sum_{i_1<i_2} \det \begin{bmatrix} z_{i_1 1} & z_{i_1 2} \\ z_{i_2 1} & z_{i_2 2} \end{bmatrix}^2 y_{i_1} y_{i_2}\right)^{k_2-k_3} \cdots \det Y^{k_n}}{D_1^{k_1-k_2+1} D_2^{k_2-k_3+1} \cdots D_{n-1}^{k_{n-1}-k_n+1}} (dZ). \tag{10.5}$$

## REFERENCES


1. Maass, H. (1955). Die Bestimmung der dirichletreihen mit Grossencharakteren zu den Modulformen $n$-ten Grades. *J. Indian Math. Soc.* **19** 1–23.
2. 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.