MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION

BY

T. W. ANDERSON and I. OLKIN

TECHNICAL REPORT NO. 38
JULY 1979

PREPARED UNDER CONTRACT
NO0014-75-C-0442 (NR-042-034)
FOR THE OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION.

by

T. W. ANDERSON and I. OLKIN

TECHNICAL REPORT NO. 38
JULY 1979

PREPARED UNDER CONTRACT N00014-75-C-0442 (NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.

Also issued as Technical Report No. 144 under National Science Foundation
Grant MCS 78-07736, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Accession For
NIIS GRAI
DCS TAB
Unannounced
Justification

By
Distribution
Availability Codes

Dist. Available and/or
special
MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS
OF A MULTIVARIATE NORMAL DISTRIBUTION

by

T. W. Anderson and I. Olkin
Stanford University

ABSTRACT

This paper provides an exposition of several alternative techniques
used to obtain maximum likelihood estimators for the parameters of a
multivariate normal distribution. In particular, matrix differentiation,
matrix transformations and induction are treated. These techniques are used
to derive the maximum likelihood estimators of the covariances of a Wishart
distribution, of the covariances when there are missing observations,
and of the means under a rank constraint. Although the paper is mainly
expository, some of the proofs are new.

Key words: multivariate normal distribution; Wishart distribution;
maximum likelihood estimation; maximization subject to
a rank constraint; matrix transformations; matrix
differentiation; multivariate inequalities.
Maximum Likelihood Estimation of the Parameters of a Multivariate Normal Distribution

1. Introduction.

Consider the maximum likelihood estimation of the mean $\mu$ and covariance matrix $\Sigma$ of a normal $p$-variate distribution based on observations $x_1, \ldots, x_N$. The maximum likelihood estimator of the mean is the sample mean $\bar{x}$, and the logarithm of the concentrated likelihood is

$$-\frac{1}{2} \log(2\pi) + \frac{1}{2} \text{tr} G^{-1}V$$

or

$$-\log|H| + \text{tr} HV,$$

where

$$V = \frac{1}{N} \sum_{\alpha=1}^{N} (x_{\alpha} - \bar{x})(x_{\alpha} - \bar{x})'$$

is positive definite, $G$ represents the covariance matrix $\Sigma$ and $H$ the inverse $\Sigma^{-1}$. The problem is to maximize (1) with respect to positive definite $G$. Because $H$ is a one-to-one transformation of $G$, it is equivalent to maximize (2) with respect to positive definite $H$. (This is Lemma 3.2.2 of Anderson (1958).)

We use the notation $A > 0$ to mean that the symmetric matrix $A$ is positive definite.
We shall describe three alternative approaches: differentiation, matrix transformations, and induction, each of which can be applied to $f(G)$ or $g(H)$. It is expected that these techniques will be useful in other problems of maximization. Three examples are given to illustrate the ideas.

Remark. It is somewhat surprising that an early reference to the fact that the sample covariance matrix is the maximum likelihood estimator of $\Omega$ is elusive. The general result is implicit in the work of Wilks (1932, p. 476) dealing with likelihood ratio tests.

2. The Method of Differentiation.

The functions $f(G)$ and $g(H)$ go to $\infty$ as $G$ or $H$ approaches the boundary of positive definite matrices or as one or more elements increases without bound. From the facts that $\log|H|$ is concave (see Bellman (1970), p. 128) and $\text{tr} \ H$ is linear, it follows that $g(H)$ is concave in $H > 0$, so that a maximum exists, and it is unique. The function $f(G)$ is neither concave nor convex. However, since there is only one solution to the derivative equations, these equations yield the maximum.

To obtain the derivative equations we use the differential form

\[
(3) \quad d[f(G)]_{ij} = (2-\delta_{ij}) \left\{ \frac{G_{ij}}{G} \right\} d_{ij} + (\text{tr} \ G^{-1} E_{ij} G^{-1} V) d_{ij} = 0, \quad i \leq j,
\]

where $\delta_{ij}$ is the Kronecker delta, $E_{ij}$ is a matrix with 1 in the $(i,j)$th position and 0 elsewhere, and $G_{ij}$ is the cofactor of $g_{ij}$. This yields the matrix equation

\[
(4) \quad -G^{-1} + G^{-1} V G^{-1} = 0,
\]

with the unique solution $\hat{G} = V$. 
Similarly, using $g(H)$,

$$(5) \quad d[g(H)]_{ij} = (2 - \delta_{ij}) \left\{ \frac{H_{ij}}{|H|} \frac{dh}{|H|} - (\text{tr} \ E_{ij} V) \frac{dh}{|H|} \right\} = 0, \quad i \leq j,$$

which yields the equation

$$(6) \quad H^{-1} - V = 0,$$

with the unique solution $\tilde{H} = V^{-1}$.

In essence, this approach applied to $f(G)$ has been discussed by Smith (1978); the approach applied to $g(H)$ was used by Anderson (1958). For a more detailed discussion of differentials see Deemer and Olkin (1951), or Anderson (1958) p. 310.

3. The Method of Matrix Transformations.

The functions $f(G)$ and $g(H)$ can be written in canonical forms. For any matrix $C$ such that $CC' = V$, let

$$\tilde{G} = C^{-1}GC^{-1}, \quad \tilde{H} = C'H'C.$$

Then (1) and (2) yield

$$f(G) - \log |V| = - \log |\tilde{G}| - \text{tr} \tilde{G}^{-1},$$

$$g(H) - \log |V| = \log |\tilde{H}| - \text{tr} \tilde{H}.$$ 

Hence we shall take $V = I$ and drop the tilde.
We use three well-known representations for a positive definite matrix, namely,

(i) \( H = TT' \),
where \( T \) is upper (or lower) triangular;

(ii) \( H = DRD \),
where \( D = \text{diag}(\sqrt{h_{11}}, \ldots, \sqrt{h_{pp}}) \), and \( R = (r_{ij}) \) is a correlation matrix, (i.e., \( r_{ii} = 1 \));

(iii) \( H = \Gamma D_d \Gamma' \),
where \( \Gamma \) is orthogonal, \( D_d = \text{diag}(d_1, \ldots, d_p) \) and \( d_1, \ldots, d_p \) are the characteristic roots of \( H \).

In each case the problem reduces to

\[
\text{Max}_{z>0} \ (\log z - z).
\]

The function \( \log z - z \) is concave and has the unique maximum of \(-1\) at \( z = 1 \).

3.1. Transformation to Rectangular Coordinates: \( H = TT' \). Now the maximization of \( g(H) \) becomes

\[
\text{Max} \ \{ \sum_{i} \ (\log t_{ii}^2 - t_{ii}^2 - \sum_{i<j} t_{ij}^2) \}.
\]

Clearly, the maximum over \( t_{ij}, \ i < j \) occurs at \( t_{ij} = 0 \), so that

(7) reduces to a sum of terms like (6).
3.2. **Transformation to a Correlation Matrix**: \( H = DRD \). Now the maximization of \( g(H) \) becomes

\[
\begin{align*}
\text{(8)} \quad \max_{R>0} & \left\{ \sum_{i} \left( \log h_{ii} - h_{ii} \right) + \log |R| \right\} \\
& \quad \text{subject to } h_{ii} > 0 \\
\end{align*}
\]

Since \( R \) is a correlation matrix, by Hadamard's determinant inequality
\[
|R| \leq \prod_{i} r_{ii} = 1, \quad \text{with equality for } R = I.
\]
Alternatively, we can write
\[
R = UU', \quad \text{where } U \text{ is upper triangular with } \sum_{j=1}^{i} u_{ij}^2 = 1; \quad \text{consequently,}
\]
\[
|R| = \prod_{i} u_{ii}^2 < 1. \quad \text{Now (8) reduces to a sum of terms like (6).}
\]

3.3. **Transformation to Characteristic Roots**: \( H = \Gamma D \Gamma' \). Now the maximization of \( g(H) \) becomes

\[
\begin{align*}
\max_{d_i > 0} & \left\{ \sum_{i} \left( \log d_i - d_i \right) \right\}, \\
& \quad \text{which is of the form (6).}
\end{align*}
\]

The transformation 3.3 has been suggested by Anderson (1958) Problem 4, Chapter 3, and has been used by Watson (1964); it is the essence of the method of Khatri (1979) and Tamhane (1979).

4. **The Method of Induction**.

Write

\[
H = \begin{pmatrix}
H_{11} & H_{12} \\
H_{21} & H_{22}
\end{pmatrix}, \quad H_{11}: \ p-1 \times p-1;
\]

then the maximization of \( g(H) \) becomes
\[
\max_{H_{11}, h_{21}, h_{22}, H > 0} \left\{ \left( \log |H_{11}| - \text{tr } H_{11} \right) + \log \left( h_{22} - H_{21} H_{11}^{-1} H_{12} \right) - h_{22} \right\}.
\]

The maximum with respect to \( H_{21} \) is achieved at \( H_{21} = 0 \) for any \( H_{11} \) and \( h_{22} \), which yields

\[
\max \left\{ \left( \log |H_{11}| - \text{tr } H_{11} \right) + \left( \log h_{22} - h_{22} \right) \right\}.
\]

The result now follows from the inductive hypothesis.

5. Other Methods.

If \( f(G) \) and \( g(H) \) are not put into canonical form several other methods may be employed. There is a nonsingular matrix \( Q \) such that

\[
V = QQ', \quad G = Q\Lambda Q',
\]

where \( \Lambda \) is a diagonal matrix with diagonal elements the roots of \(|G - \lambda V| = 0\). Then

\[
f(G) = -2 \log |Q| - \sum_{i=1}^{p} \left( \log \lambda_i + \frac{1}{\lambda_i} \right),
\]

which is maximized for \( \lambda_i = 1 \); that is, \( \hat{\Lambda} = I, \hat{G} = V \).

Another method invokes a theorem of von Neumann (1937):

\[
\text{tr } H V \geq \sum_{i=1}^{p} n_{i} v_{p-i+1}
\]
where \( \eta_1 > \ldots > \eta_p \) are the ordered characteristic roots of \( H \) and \( \nu_1 > \ldots > \nu_p \) are the ordered characteristic roots of \( V \); equality is attained when the characteristic vectors of \( H \) are identical to those of \( V \). We have

\[
g(H) \leq \sum_{i=1}^{p} (\log \eta_i - \eta_i \nu_{p-i+1})
\]

with equality when the characteristic vectors of \( H \) are identical to those of \( V \). Then \((\log \eta_i - \eta_i \nu_{p-i+1})\) is maximized at \( \eta_i = 1/\nu_{p-i+1} \), \( i = 1, \ldots, p \). Thus \( g(H) \) is maximized at \( \hat{H} = \hat{V}^{-1} \). This method is used by Theobald (1975).

6. A Modified Model with Missing or Additional Observations.

Suppose a sample of size \( N \) is observed from a \( p \)-variate normal distribution with covariance matrix \( \Sigma \), and an additional sample of size \( M \) is observed on the first \( k \) (out of \( p \)) variates. Alternatively, this model can be viewed as a sample of size \( N+M \) from a \( p \)-variate normal distribution, where \( M \) observations on the last \( p-k \) (out of \( p \)) variates are missing. This problem was considered by Anderson (1957) and in another context by Olkin and Sylvan (1977).

The problem now is to

\[
\text{Max}_{G>0} \left\{ -N \log |G| - \text{tr} G^{-1} V - M \log |G_{11}| - \text{tr} G_{11}^{-1} W \right\},
\]

where \( G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \), \( G_{11}: k \times k \), \( V: p \times p \) and \( W: k \times k \) are positive definite.
6.1. **Reduction to a Canonical Form.** In terms of \( H = G^{-1} \), the problem becomes

\[
\begin{align*}
\text{(10)} \quad & \max_{H>0} \left( N \log|H| - \text{tr} \ H V + M \log|H_{11} - H_{12}^{-1} H_{22}^{-1} H_{21}| - \text{tr}(H_{11} - H_{12}^{-1} H_{22}^{-1} H_{21}) W \right) \\
& = \max_{H>0} \left( N \log|H_{22}| + (N + M) \log|H_{11} - H_{12}^{-1} H_{22}^{-1} H_{21}| - \text{tr} P_{11} V_{11} - \text{tr} H_{12} V_{21} \right. \\
& \quad \left. - \text{tr} H_{21} V_{12} - \text{tr} H_{22} V_{22} - \text{tr}(H_{11} - H_{12}^{-1} H_{22}^{-1} H_{21}) W \right).
\end{align*}
\]

6.2. **Transformation.**

One can think of this likelihood as composed of the marginal likelihood of the \( N + M \) observations on the \( k \) variables and the conditional likelihood function of the \( N \) observations on the \( p - k \) variables conditional on the \( N \) observations on the \( k \) variables. The appropriate parameters are

\[
H_{11} - H_{12}^{-1} H_{22}^{-1} H_{21} = J, \quad H_{22}^{-1} H_{21} = K, \quad H_{22} = H_{22}.
\]

(If \( H = G^{-1} \), then \( J = G_{11}^{-1} \), \( K = -G_{21} G_{11}^{-1} \), \( H_{22} = (G_{22} - G_{21} G_{11}^{-1} G_{12})^{-1} \).)

The corresponding transformation of the covariance matrix of \( N \) observations is

\[
V_{22} - V_{21} V_{11}^{-1} V_{12} = F, \quad V_{21} V_{11}^{-1} = E, \quad V_{11} = V_{11}.
\]
Then (10) is transformed to

\[
\text{Max}_{H_{22}>0, J>0, K} \{ N \log H_{22} - \text{tr} H_{22} F + (N+M) \log |J| - \text{tr} J(V_{11} + W) - \text{tr} H_{22}(K+E) V_{11}(K+E) \}.
\]

The last term is nonpositive because $H_{22}$ and $V_{11}$ are positive definite; its maximum of 0 occurs at $\hat{K} = -E = -V_{21} V_{11}^{-1}$. Then (11) becomes the sum of two terms like (2). The maximum over $H_{22} > 0$ and $J > 0$ occurs at $\hat{H}_{22} = NF^{-1}$ and $\hat{J} = (N+M)(V_{11}+W)^{-1}$.

It might be noted that this last analysis is applicable to the following problem. Given a matrix $X$ with regression $KZ$ and covariance matrix $H_{22}^{-1}$, the log likelihood function is (11) with $J = 0$, $E = X'X (Z'Z)^{-1}$ and $F = X'X - EZ'ZE'$.

6.3. Differentiation.

We now solve (9) by differentiation; this alternative may have some intrinsic interest. Using differential forms, we obtain

\[
(2-\delta_{ij}) \left\{ \frac{G^*}{G} \partial g_{ij} + (\text{tr} G^{-1} E_{ij} G^{-1} V) \partial g_{ij} - M \left( \begin{array}{cc} A_{ij}^* & \partial a_{ij} \\ G_{11} & 0 \end{array} \right) + \partial a_{ij} \text{tr} \left( \begin{array}{cc} G_{11}^{-1} E_{ij} G_{11}^{-1} W & 0 \\ 0 & 0 \end{array} \right) \right\} = 0,
\]

where $A \equiv G_{11}$ and $A_{ij}^*$ is the cofactor of $a_{ij}$ in $A$. This yields the matrix equation

\[
-NG^{-1} + G^{-1} W G^{-1} - M \left( \begin{array}{cc} G_{11}^{-1} W & 0 \\ 0 & 0 \end{array} \right) + \left( \begin{array}{cc} G_{11}^{-1} W & 0 \\ 0 & 0 \end{array} \right) = 0.
\]
Pre- and post-multiplication by \( G \) in (12) yields

\[
\begin{pmatrix}
G^{-1} & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
\hat{G}^{-1} & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{pmatrix}
\begin{pmatrix}
-\Lambda & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
0 & 0 \\
0 & \Lambda
\end{pmatrix}
\begin{pmatrix}
G_1 & G_2 \\
G_2 & G_3
\end{pmatrix}
= 0,
\]

which simplifies to the set of equations

\[
\begin{align*}
(14a) \quad & -\Lambda_{11} + \Lambda_{12} - \Lambda_{21} + \Lambda_{22} = 0, \\
(14b) \quad & -\Lambda_{12} + \Lambda_{12} - \Lambda_{21} + \Lambda_{22} = 0, \\
(14c) \quad & -\Lambda_{12} + \Lambda_{12} - \Lambda_{21} + \Lambda_{22} = 0.
\end{align*}
\]

The equations (14) can be solved in sequence to yield

\[
\begin{align*}
(15a) \quad & \hat{G}_{11} = \frac{1}{N+H} (\Lambda_{11} + \Lambda), \\
(15b) \quad & \hat{G}_{12} = \frac{1}{N+H} (\Lambda+\Lambda)_{12} , \\
(15c) \quad & \hat{G}_{22} = \frac{1}{N} (\Lambda_{22} + \frac{1}{N+H} (\Lambda+\Lambda)_{12} - \frac{M}{N(N+H)} (\Lambda+\Lambda)_{12}) - 12.
\end{align*}
\]

7. A Problem of Rank.

Let the \( N \) columns of \( X \) and the \( M \) columns of \( Y \) be independently distributed according to a \( p \)-variate normal distribution with covariance matrix \( \Sigma \) and \( EX = 0, \ EY = \Phi \), where \( \Phi \) is of rank \( r < p < M \). This model is considered by Anderson (1951).
To obtain the maximum likelihood estimators of $\Sigma$ and $\phi$ we start with the likelihood function

\begin{equation}
(16) \quad c |\Sigma|^{-\frac{1}{2}(N+M)} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1}[XX' + (Y-\phi)(Y-\phi)']\right),
\end{equation}

where $c$ is a normalizing constant. From Sections 1, 2, and 3, the maximum of (16) with respect to $\Sigma$ (for fixed $\phi$) is

\[ \hat{\Sigma} = [XX' + (Y-\phi)(Y-\phi)]/(N+M), \]

so that we need to determine

\begin{equation}
(17) \quad \min_{\phi} |\hat{\Sigma}| = \min_{\phi} |XX' + (Y-\phi)(Y-\phi)|.
\end{equation}

To simplify notation, write

\begin{equation}
(18) \quad \tilde{Y} = (XX')^{-\frac{1}{2}}Y, \quad \tilde{\phi} = (XX')^{-\frac{1}{2}}\phi.
\end{equation}

then (17) becomes (except for the term $|XX'|$)

\begin{equation}
(19) \quad \min_{\phi} |I + (\tilde{Y} - \tilde{\phi})(\tilde{Y} - \tilde{\phi})|.
\end{equation}

Since $\tilde{\phi}: p \times M$ is of rank $r$, we can write

\begin{equation}
(20) \quad \tilde{\phi} = (T_1 T_2) \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} = T_1 \Lambda_1.
\end{equation}
where \((T_1, T_2): p \times p, T_1: p \times r; \Delta = \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix}: M \times M, \Delta_1: r \times M\) and \(\Delta\) is orthogonal. In the representation (20), \(\phi \phi' = T_1 T_1^T\). If \(T_1 = \begin{pmatrix} T_{11} \\ T_{21} \end{pmatrix}\), where \(T_{11}: r \times r\) is lower triangular with positive diagonal elements, then the representation \(\bar{\phi} = T_1 \Delta_1\) is unique.

In (20) note that

\[
\begin{align*}
|I_p + (\bar{Y} - \phi) (\bar{Y} - \phi)^T| \\
= |I_p + \bar{Y} \bar{Y}^T + T_1 T_1^T - T_1 \Lambda_1 \bar{Y}^T - \bar{Y} \Lambda_1^T T_1^T| \\
= |(I_p + \bar{Y} \bar{Y}^T - \bar{Y} \Lambda_1^T \bar{Y}^T) + (T_1 - \bar{Y} \Lambda_1^T) (T_1 - \bar{Y} \Lambda_1^T)^T| \\
= |I_p + \bar{Y} \Lambda_2^T \bar{Y}^T + (T_1 - \bar{Y} \Lambda_1^T) (T_1 - \bar{Y} \Lambda_1^T)^T| .
\end{align*}
\]

Since \(I_p + \bar{Y} \Lambda_2^T \bar{Y}^T\) is positive definite, the minimum of (21) over \(T_1\) is achieved at \(\bar{\phi}_1 = \bar{Y} \Lambda_1\), in which case we need to determine

\[
\begin{align*}
\text{Min}_{\Lambda_2} |I_p + \bar{Y} \Lambda_2^T \bar{Y}^T| &= \text{Min}_{\Lambda_2} |I_{M-\Gamma} + \Delta_2 \bar{Y}^T \bar{Y} \Lambda_2^T| \\
&= \text{Min}_{\Lambda_2} |\Lambda_2 (I_{M-\Gamma} + \bar{Y} \bar{Y}^T) \Lambda_2^T| .
\end{align*}
\]

The minimum of (22) is \(p_{\lambda_1}^p\), where \(\lambda_1 > \ldots > \lambda_p\) are the ordered characteristic roots of \(I + \bar{Y} \bar{Y}^T = I + \bar{Y} \bar{X}^T \bar{X}^T\), that is, the roots of \(|(X'X + Y'Y) - \lambda X'X| = 0\). (See, e.g., Bellman (1970),
Theorem 10, p. 132.) The minimum is attained at

$$\Delta_2 = (c_{r+1}, \ldots, c_p),$$

where $c_1$ is the vector satisfying $(I + Y'QX')^{-1}Yc_1 = \lambda_1c_1$ and $c_1'c_1 = 1$.

Combining our results we obtain as the maximized likelihood

$$c e^{-\frac{1}{2}p} \frac{(N+H)^{\frac{1}{2}p(N+H)}}{|XX'|^{\frac{1}{2}(N+H)} \left( P \prod_{r+1} \lambda_i \right)^{\frac{1}{2}(N+H)}}.$$

The model (16) is considered by Healy (1979). His procedure is to first make a series of transformations motivated by the rank condition on $\phi$, which transforms the model to a canonical form, and then to carry out the maximizations. However, by first maximizing with respect to the covariances (as in the above derivation), the required transformations (and proof) become considerably simpler.
REFERENCES


17. "General Exponential Models for Discrete Observations,"

18. "On the Interrelationships among Sufficiency, Total Sufficiency and


20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models

21. "Asymptotic Properties of Some Estimators in Moving Average Models,
    Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,
    Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear
    Discriminant Function," Chien-Pai Han, April 1976.


    September 1976.


28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
    Javad Behboodian, October 1976.

29. "Identification of Parameters by the Distribution of a Maximum

30. "Discrimination Between Stationary Gaussian Processes, Large Sample

31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
    Christine M. Waternaux, October 1977.

32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a


<table>
<thead>
<tr>
<th>1. REPORT NUMBER</th>
<th>2. GOVT ACCESSION NO.</th>
<th>3. RECIPIENT'S CATALOG NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE (and Subtitle)</th>
<th>5. TYPE OF REPORT &amp; PERIOD COVERED</th>
<th>6. PERFORMING ORG. REPORT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF A MULTIVARIATE NORMAL DISTRIBUTION</td>
<td>Technical Report</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. AUTHOR(s)</th>
<th>8. CONTRACT OR GRANT NUMBER(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. W. ANDERSON and I. OLKIN</td>
<td>N00014-75-C-0442</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. PERFORMING ORGANIZATION NAME AND ADDRESS</th>
<th>10. PROGRAM ELEMENT PROJECT, TASK AREA &amp; WORK UNIT NUMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Statistics</td>
<td>(NR-042-034)</td>
</tr>
<tr>
<td>Stanford University</td>
<td></td>
</tr>
<tr>
<td>Stanford, California</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. CONTROLLING OFFICE NAME AND ADDRESS</th>
<th>12. REPORT DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office of Naval Research</td>
<td>JULY 1979</td>
</tr>
<tr>
<td>Statistics &amp; Probability Program Code 436</td>
<td></td>
</tr>
<tr>
<td>Arlington, Virginia 22217</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. NUMBER OF PAGES</th>
<th>14. SECURITY CLASS. (of this report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>UNCLASSIFIED</td>
</tr>
</tbody>
</table>

**DISTRIBUTION STATEMENT (of this Report)**

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)**

**SUPPLEMENTARY NOTES**

Issued also as Technical Report No. 144 under National Science Foundation Grant MCS 78-07736, Department of Statistics, Stanford University.

**KEY WORDS** (Continue on reverse side if necessary and identify by block number)

Multivariate normal distribution; Wishart distribution; maximum likelihood estimation; maximisation subject to a rank constraint; matrix transformations; matrix differentiation; multivariate inequalities.

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

SEE REVERSE SIDE.
ABSTRACT.

This paper provides an exposition of several alternative techniques used to obtain maximum likelihood estimators for the parameters of a multivariate normal distribution. In particular, matrix differentiation, matrix transformations and induction are treated. These techniques are used to derive the maximum likelihood estimators of the covariances of a Wishart distribution, of the covariances when there are missing observations, and of the means under a rank constraint. Although the paper is mainly expository, some of the proofs are new.