The Annals of Statistics
1980, Vol. 8, No. 4, 802–810

NECESSARY AND SUFFICIENT CONDITIONS FOR EXPLICIT SOLUTIONS IN THE MULTIVARIATE NORMAL ESTIMATION PROBLEM FOR PATTERNED MEANS AND COVARIANCES[1]

BY TED H. SZATROWSKI

Rutgers University

The problem of finding maximum likelihood estimates for patterned means and covariance matrices in multivariate analysis is considered. Necessary and sufficient conditions are presented for the existence of explicit solutions and the obtaining of these explicit solutions in one iteration of the scoring equations from any positive definite starting point. Cases in which averaging yields the explicit maximum likelihood estimates are discussed. These results can be applied to the problems of finding maximum likelihood estimates for the parameters in the complete, compound and circular symmetry patterns; mixed models in the analysis of variance; and for finding asymptotic distributions of likelihood ratio statistics when the parameters under the null hypothesis have explicit maximum likelihood estimates.

1. Introduction. We consider the problem of estimation of the mean vector and covariance matrix of a multivariate normal distribution when the mean vector and covariance matrix have linear structure. Anderson (1969, 1970, 1973) studies this problem and in the (1973) paper he presents the likelihood equations and suggests an iterative algorithm for finding the solutions of the likelihood equations based on the method of scoring. In the present study, necessary and sufficient conditions are presented for (1) the existence of explicit maximum likelihood estimates and (2) the convergence of the iterative procedure proposed by Anderson (1973) in one iteration from any positive definite starting point and these results are then applied to some well-known problems.

In Section two, the problem under consideration is described in detail including the likelihood equations and the scoring algorithm. Also in this section is a discussion of a canonical form for the problem that is used in the remainder of this paper and a short discussion on averaging. The results for estimating the mean vector are given in Section three, those for the covariance matrix in Section four. Applications are described in Section five for patterned symmetry problems and asymptotic distributions of likelihood ratio statistics under alternative hypotheses.

Received June 1976; revised August 1978.

[1]Research partially supported by the National Science Foundation under contract MPS-7509450-000, grant MCS77-28184, and N.S.F. Competitive Fellowship and a Rutgers University Faculty Summer Fellowship. This paper is a revised and extended version of part of the author's doctoral dissertation at Stanford University.

AMS 1970 *subject classifications.* 62H05, 62H15.

Key words and phrases. Analysis of variance, averaging, circular symmetry, complete symmetry, compound symmetry, convergence, explicit solutions, maximum likelihood estimation, mixed model, patterned covariance matrices, patterned means.

Applications for mixed models of the analysis of variance appear in Szatrowski and Miller (1980).

2. Likelihood equations, canonical forms and averaging. Let $\mathbf{X}$ be a p-component column vector with multivariate normal distribution such that the mean vector $\boldsymbol{\mu} = \mathcal{E}\mathbf{X}$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = \mathcal{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$ have the linear structure considered by Anderson (1973). Specifically, $\boldsymbol{\mu} = \Sigma_{j=1}^{r} \beta_j \mathbf{z}_j = \mathbf{Z}\boldsymbol{\beta}$, $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_r]$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_r)'$, $\boldsymbol{\beta} \in R^r$, where the $\mathbf{z}$'s are known, linearly independent column vectors and the β's are unknown scalars. The covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \Sigma_{g=0}^{m} \sigma_g \mathbf{G}_g$, $\boldsymbol{\sigma} = (\sigma_0, \cdots, \sigma_m)'$, where the $\mathbf{G}$'s are known, linearly independent symmetric matrices and the σ's are unknown scalars such that $\boldsymbol{\sigma} \in \Theta$, $\Theta = \{\boldsymbol{\theta} \in R^{m+1} | \boldsymbol{\Sigma}(\boldsymbol{\theta}) > \mathbf{0}\}$ where $\boldsymbol{\Sigma} > \mathbf{0}$ denotes $\boldsymbol{\Sigma}$ positive definite. We assume that Θ is nonempty so that there exists at least one value of $\boldsymbol{\sigma}$ that results in $\boldsymbol{\Sigma}(\boldsymbol{\sigma})$ being positive definite. Maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in these forms are desired based on N independent p-dimensional observations $\mathbf{x}_1, \cdots, \mathbf{x}_N$.

2.1 Likelihood equations. Let $[d_{gh}]$ denote a matrix whose g, h element is d_{gh} and let (d_g) denote a column vector whose gth element is d_g. The likelihood equations for this problem as given by Anderson (1973) are

$$(\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{x}}, \tag{2.1}$$

$$\left[\mathrm{tr}\, \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h\right]\hat{\boldsymbol{\sigma}} = \left(\mathrm{tr}\, \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C}\right), \tag{2.2}$$

where $[\mathrm{tr}\, \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_h]$ is an $(m+1) \times (m+1)$ matrix, $(\mathrm{tr}\, \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{G}_g\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{C})$ is an $(m+1) \times 1$ column vector, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are $r \times 1$ and $(m+1) \times 1$ column vectors of unknowns,

$$\bar{\mathbf{x}} = (1/N)\Sigma_{\alpha=1}^{N}\mathbf{x}_\alpha \text{ and } \mathbf{C} = (1/N)\Sigma_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})' + (\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})', \quad \text{for } \hat{\boldsymbol{\mu}} = \mathbf{Z}\hat{\boldsymbol{\beta}}.$$

Anderson points out that the likelihood equations written in this form suggest an iterative scheme (noted by J. N. K. Rao (1973) to correspond to the method of scoring) wherein from an initial estimate of $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\Sigma}}$, one can solve the linear equations in $\boldsymbol{\beta}$, compute $\mathbf{C}$ and then solve the linear equations in $\hat{\boldsymbol{\sigma}}$ to yield the next estimate of $\hat{\boldsymbol{\Sigma}}$. Note that, in general, one should monitor the values of $\hat{\boldsymbol{\sigma}}$ to insure that $\hat{\boldsymbol{\sigma}} \in \Theta$, i.e., $\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}}) > \mathbf{0}$. Convergence of this iterative procedure is not guaranteed.

2.2 Canonical form. We assume, without loss of generality, that the problem is in the canonical form when there exists a value $\boldsymbol{\sigma}^* \in \Theta$ such that $\boldsymbol{\Sigma}(\boldsymbol{\sigma}^*) = \mathbf{I}$, the identity matrix. If the problem is not in canonical form, it can be trivially rotated into this form.

2.3 Averaging. The likelihood equations (2.1) and (2.2) evaluated at the initial estimate $\boldsymbol{\Sigma} = \mathbf{I}$ yield the values

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\bar{\mathbf{x}}, \quad \hat{\boldsymbol{\sigma}} = \left[\mathrm{tr}\, \mathbf{G}_g\mathbf{G}_h\right]^{-1}(\mathrm{tr}\, \mathbf{G}_g\mathbf{C}). \tag{2.3}$$

We note that the value $\hat{\boldsymbol{\beta}}$ minimizes $(\bar{\mathbf{x}} - \mathbf{Z}\boldsymbol{\beta})'(\bar{\mathbf{x}} - \mathbf{Z}\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in R^r$ and that $\hat{\boldsymbol{\sigma}}$ (using the value of $\hat{\boldsymbol{\beta}}$ for calculating $\mathbf{C}$) minimizes $\mathrm{tr}(\boldsymbol{\Sigma}(\boldsymbol{\sigma}) - \mathbf{C})^2$ for $\boldsymbol{\sigma} \in R^{m+1}$. In general, the estimates in (2.3) do not correspond to the maximum likelihood estimates and need not even yield a positive definite estimate of $\boldsymbol{\Sigma}$. However, in certain cases we can reparameterize $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ so that the $\mathbf{z}$ vectors and $\mathbf{G}$ matrices consist only of zeroes and ones. In these cases, the estimates in (2.3) may be obtained by first estimating each of the r distinct elements of the mean vector $\boldsymbol{\mu}$ by the average of their corresponding values in $\bar{\mathbf{x}}$, using these estimates to compute $\mathbf{C}$, and finally estimating each of the $m + 1$ distinct elements of $\boldsymbol{\Sigma}$ by the average of their corresponding values in $\mathbf{C}$. Thus, when (2.3) yields the explicit maximum likelihood estimates (see Theorem 4 and Corollary 1) and the above reparameterization holds, the explicit maximum likelihood estimates can be obtained directly by averaging.

3. **Mean vector results.** In this section results are derived for the maximum likelihood estimates of the unknown parameters $\boldsymbol{\beta}$ of the mean vector $\boldsymbol{\mu}$. The following result, with $\mathbf{A} = \boldsymbol{\Sigma}$ and $\mathbf{B} = \mathbf{I}$ is used for showing when the least squares and Markov estimates for the mean vector coincide. The more general form is used for showing when the "least squares" and "Markov estimates" for the covariance matrix coincide.

THEOREM 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $p \times p$, *symmetric, positive definite matrices and let* $\mathbf{X}$ *be* $p \times r$, $r \leqslant p$ *of full rank. A necessary and sufficient condition for* $(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{-1} = (\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}^{-1}$ *is that the columns of* $\mathbf{X}$ *are linear combinations of* r *characteristic vectors of* $\mathbf{AB}^{-1}$.

PROOF. The proof uses techniques that may be found in Anderson (1971) in the proofs of Theorem 2.4.1 and Theorem 10.2.1. This proof is deferred to the Appendix.

Before stating the theorem giving necessary and sufficient conditions for explicit solutions and one iteration convergence for the maximum likelihood estimates of the mean vector, we define what we mean by an "allowable starting point" and an explicit representation for the mean. An "allowable starting point" is any value $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})$ which is positive definite. It is used in equation (2.1) to start the iterative procedure. An explicit representation for the mean is one in which the maximum likelihood estimate for $\boldsymbol{\beta}$, $\tilde{\boldsymbol{\beta}}$, can be expressed as $\tilde{\boldsymbol{\beta}} = \mathbf{A}\bar{\mathbf{x}}$ where $\mathbf{A}$ is a function only of the $\mathbf{z}$'s and $\mathbf{G}$'s. Results similar to conditions 1, 2 and 5 of Theorem 2 have been obtained by several authors including Zyskind (1967), Thomas (1968) and Mitra and Moore (1973).

THEOREM 2. *The following five conditions are equivalent for the problem described in Section two in its canonical form*:

1. $(\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ *for all allowable starting points.*
2. *The* r *columns of* $\mathbf{Z}$ *are spanned by* r *eigenvectors of* $\boldsymbol{\Sigma}$.

3. *The likelihood estimate of* $\boldsymbol{\beta}$ *(and thus of* $\boldsymbol{\mu}$*) has an explicit representation.*
4. *Equation* (2.1) *in the scoring algorithm converges in one iteration from any allowable starting point to the maximum likelihood estimate of* $\boldsymbol{\beta}$.
5. *The least squares and Markov estimators of* $\boldsymbol{\beta}$ *coincide.*

Note it is assumed for conditions 3, 4 and 5 that they do not depend on the specific data values obtained. Specifically, we ignore the cases (except when $\mathbf{Z}$ is $p \times p$) in which $\bar{\mathbf{x}}$ can be expressed in the form $\mathbf{Z}\boldsymbol{\beta}$ for some $\boldsymbol{\beta}$. Such events, of probability zero, yield trivial counterexamples of Theorem 2.

PROOF. Conditions 1 and 2 are equivalent using Theorem 1 with $\mathbf{A} = \boldsymbol{\Sigma}$, $\mathbf{B} = \mathbf{I}$ and $\mathbf{X} = \mathbf{Z}$. Conditions 2 and 5 are equivalent by Theorem 1 of Zyskind (1967). From our definition of explicit solutions, we observe that an explicit maximum likelihood estimate of $\boldsymbol{\beta}$ must be independent of allowable values of $\boldsymbol{\sigma}$, thus 1 and 3 are equivalent. The equivalence of 1 and 4 follows by noting the form of likelihood equation (2.1). □

4. Results for the covariance matrix. In this section results are derived for the maximum likelihood estimates of the covariance matrix $\boldsymbol{\Sigma}$. The elements in the upper triangle of $\boldsymbol{\Sigma}$ are written as a vector. The likelihood equation (2.2) is rewritten in this vector form using an extension of a result of Anderson (1969) given in Theorem 3. The necessary and sufficient conditions for explicit maximum likelihood estimates of $\boldsymbol{\Sigma}$ and one iteration convergence of the likelihood equation (2.2) are given in Corollary 1 of Theorem 4. Theorem 5 gives necessary and sufficient conditions for explicit maximum likelihood estimates of Σ when it is diagonalizable. Both Corollary 1 and Theorem 5 assume explicit maximum likelihood estimates for $\boldsymbol{\beta}$ exist.

DEFINITION 1. Let $\mathbf{A}$ be a symmetric $p \times p$ matrix. $\langle \mathbf{A} \rangle$ is defined to be a column vector consisting of the upper triangle of elements of $\mathbf{A}$, i.e.,

$$\langle \mathbf{A} \rangle = (a_{11}, a_{22}, \cdots, a_{pp}, a_{12}, a_{13}, \cdots, a_{1p}, a_{23}, \cdots, a_{p-1,p})^t.$$

DEFINITION 2. The maximum likelihood estimate of $\boldsymbol{\sigma}$, $\tilde{\boldsymbol{\sigma}}$, has an explicit representation if and only if $\tilde{\boldsymbol{\sigma}} = \mathbf{B}(\langle \mathbf{C} \rangle$ for some matrix B which is a function only of the G's. (Note $\mathbf{C}$ is the sample covariance matrix defined in Section 2.1.)

DEFINITION 3 (Anderson, 1969). Define $\boldsymbol{\Phi}$ as the $\{p(p+1)/2\} \times \{p(p \times 1)/2\}$ symmetric matrix with elements $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}(\boldsymbol{\Sigma}) = (\phi_{ij,kl}) = (\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$, $i \leqslant j$, $k \leqslant l$. The notation $\phi_{ij,kl}$ represents the element of Φ with row in the same position as the element a_{ij} in $\langle \mathbf{A} \rangle$ where $\mathbf{A}$ is a $p \times p$ symmetric matrix and column in the same position as a_{kl} in $\langle \mathbf{A} \rangle^t$.

We observe that if the $p \times p$ matrix $\mathbf{R} > 0$ has a Wishart distribution with parameters $\boldsymbol{\Sigma} > 0$ and $n(\mathcal{L}(\mathbf{R}) = \mathcal{W}(\boldsymbol{\Sigma}, n))$, then $n\boldsymbol{\Phi}(\boldsymbol{\Sigma}) = \mathrm{Cov}(\langle \mathbf{R} \rangle)$, (e.g., Szatrowski (1979, Lemma 1)).

THEOREM 3. *If* **E** *and* **F** *are* $p \times p$ *symmetric matrices, then*

$$\langle \mathbf{E} \rangle' \mathbf{\Phi}^{-1}(\mathbf{\Sigma}) \langle \mathbf{F} \rangle = \tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1} \mathbf{E} \mathbf{\Sigma}^{-1} \mathbf{F}. \tag{4.1}$$

PROOF. The proof is a straightforward extension of Anderson's (1969, page 61) proof with $\mathbf{G}_h$ replaced by **E** and **C** replaced by **F**. □

Five examples of possible **E** and/or **F** matrices that can be used in Theorem 3 are (a) $\mathbf{\Sigma}$, (b) $\mathbf{G}_g$, $g = 0, 1, \cdots, m$, (c) **C**, the sample covariance defined in Section 2.1, (d) $K\Sigma_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})'$, K a positive constant, (e) $(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{x}} - \hat{\boldsymbol{\mu}})'$, useful with mixed models in the analysis of variance when $N = 1$.

Using Theorem 3, we note the following identities useful in rewriting the likelihood equation (2.2), with $\hat{\mathbf{\Phi}} \equiv \mathbf{\Phi}(\hat{\mathbf{\Sigma}})$,

$$\langle \mathbf{G}_g \rangle' \hat{\mathbf{\Phi}}^{-1} \langle \mathbf{C} \rangle = \tfrac{1}{2} \operatorname{tr} \hat{\mathbf{\Sigma}}^{-1} \mathbf{G}_g \hat{\mathbf{\Sigma}}^{-1} \mathbf{C}, \quad g = 0, 1, \cdots, m, \tag{4.2}$$

$$\langle \mathbf{G}_g \rangle' \hat{\mathbf{\Phi}}^{-1} \langle \mathbf{G}_h \rangle = \tfrac{1}{2} \operatorname{tr} \hat{\mathbf{\Sigma}}^{-1} \mathbf{G}_g \hat{\mathbf{\Sigma}}^{-1} \mathbf{G}_h, \quad g, h = 0, 1, \cdots, m. \tag{4.3}$$

Using (4.2) and (4.3) and letting $\mathbf{W} = [\langle \mathbf{G}_0 \rangle, \cdots, \langle \mathbf{G}_m \rangle]$, a matrix with the vector forms of the **G**'s as columns, we can rewrite likelihood equation (2.2) as

$$\hat{\boldsymbol{\sigma}} = (\mathbf{W}' \hat{\mathbf{\Phi}}^{-1} \mathbf{W})^{-1} \mathbf{W}' \hat{\mathbf{\Phi}}^{-1} \langle \mathbf{C} \rangle. \tag{4.4}$$

Let $\mathbf{\Phi}_\mathbf{I}$ be defined by $\mathbf{\Phi}_\mathbf{I} \equiv \mathbf{\Phi}(\mathbf{I}) = \operatorname{diag}(2\mathbf{I}_p, \mathbf{I}_{p(p-1)/2})$.

In addition, following Seely (1971), we define a quadratic subspace (more appropriately termed a Jordan algebra) to be a subspace $\mathfrak{B}$ of the vector space of $p \times p$ real symmetric matrices $\mathcal{Q}$ with the property that $B \in \mathfrak{B}$ implies $B^2 \in \mathfrak{B}$.

THEOREM 4. *When* $\boldsymbol{\beta}$ *is known, the following seven conditions are equivalent for the problem described in Section two in its canonical form*:

1. $(\mathbf{W}'\mathbf{\Phi}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{\Phi}^{-1} = (\mathbf{W}'\mathbf{\Phi}_I^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{\Phi}_\mathbf{I}^{-1}$ *for* $\mathbf{\Phi} = \mathbf{\Phi}(\mathbf{\Sigma})$ *generated by all allowable starting points.*

2. *The* $m + 1$ *columns of* **W** *are spanned by* $m + 1$ *eigenvectors of* $\mathbf{\Phi}(\mathbf{\Sigma})\mathbf{\Phi}^{-1}(\mathbf{I})$.

3. *The maximum likelihood estimates of* $\boldsymbol{\sigma}$ (*and thus of* $\mathbf{\Sigma}$) *have explicit representations.*

4. *Equation* (2.2) *in the scoring algorithm converges in one iteration from any allowable starting point to the maximum likelihood estimate of* $\boldsymbol{\sigma}$.

5. *The set of* $\mathbf{\Sigma}$ *given by* $\{\mathbf{\Sigma} : \langle \mathbf{\Sigma} \rangle = \mathbf{W}\boldsymbol{\sigma}, \boldsymbol{\sigma} \in R^{m+1}\}$ *forms a quadratic subspace.*

6. *For any* $\boldsymbol{\sigma} \in R^{m+1}$ *such that* $(\mathbf{\Sigma}(\boldsymbol{\sigma}))^{-1}$ *exists, there exists a* $\boldsymbol{\gamma} \in R^{m+1}$ *such that* $\langle (\mathbf{\Sigma}(\boldsymbol{\sigma}))^{-1} \rangle = \mathbf{W}\boldsymbol{\gamma}$.

7. *For any* $\mathbf{a} \in R^{m+1}$, $\mathbf{a}'\tilde{\boldsymbol{\sigma}}$ *are uniformly minimum variance unbiased estimates of* $\mathbf{a}'\sigma$ *where* $\tilde{\boldsymbol{\sigma}}$ *is the* MLE *for* $\boldsymbol{\sigma}$.

Note that it is assumed that conditions 3 and 4 do not depend upon the specific data values obtained. We ignore the trivial counterexamples which occur when $\langle \mathbf{C} \rangle$ can be expressed as $\mathbf{W}\boldsymbol{\sigma}$ for some $\boldsymbol{\sigma}$ since (except in the case when **W** is square) this event occurs with probability zero.

PROOF. Conditions 1 and 2 are equivalent by Theorem 1 with $\mathbf{A} = \mathbf{\Phi}(\mathbf{\Sigma})$, $\mathbf{B} = \mathbf{\Phi}_{\mathbf{I}}$, $\mathbf{X} = \mathbf{W}$. The equivalence of 1-4 parallels the proof of Theorem 2. Condition 1 implies Condition 7 since the Cramér-Rao lower bound matrix is $(\mathbf{W}'\mathbf{\Phi}^{-1}(\mathbf{\Sigma}^*)\mathbf{W})^{-1}/N$, where $\mathbf{\Sigma}^*$ is the true value, which agrees with Cov $\tilde{\boldsymbol{\sigma}}$ since

$$\text{Cov}\,\tilde{\boldsymbol{\sigma}} = (\mathbf{W}'\mathbf{\Phi}^{-1}(\mathbf{\Sigma})\mathbf{W})^{-1}\mathbf{W}'\mathbf{\Phi}^{-1}(\mathbf{\Sigma})\text{Cov}\langle\mathbf{C}\rangle\mathbf{\Phi}^{-1}(\mathbf{\Sigma})\mathbf{W}(\mathbf{W}'\mathbf{\Phi}^{-1}(\mathbf{\Sigma})\mathbf{W})^{-1},$$

for any allowable $\mathbf{\Sigma}$. Thus the result follows by choosing $\mathbf{\Sigma} = \mathbf{\Sigma}^*$ and using $\text{Cov}\langle\mathbf{C}\rangle = \Phi(\mathbf{\Sigma}^*)/\mathbf{N}$ by the comment after Definition 3. Conditions 7 and 5 are equivalent by Seely (1971), Theorem 1. Condition 5 implies 6 by the discussion after Theorem 1 in Seely (1971). Condition 6 allows the likelihood function to be reparameterized with $\mathbf{\Sigma}^{-1} = \Sigma_{i=0}^{m}\gamma_i\mathbf{G}_i$ which yields explicit MLE thus implying condition 3. □

In the case where $\boldsymbol{\beta}$ is not known, we have

COROLLARY 1. *If the maximum likelihood estimate of $\boldsymbol{\beta}$ has an explicit representation, then statements $1 - 6$ of Theorem 4 are equivalent for the problem described in Section two in its canonical form.*

We note that in the case where $\boldsymbol{\beta}$ is not known, statement 7 of Theorem 4 does not hold because the MLE of $\boldsymbol{\sigma}$ is not unbiased.

A simplified version of Corollary 1 is given in Theorem 5 for the case where the **G**'s are simultaneously diagonalizable. Well-known necessary and sufficient conditions for the **G**'s being simultaneously diagonalizable are given in Lemma 1.

LEMMA. *The following statements are equivalent.*

a. *There exists an orthogonal matrix,* **P**, *independent of the σ's such that* $\mathbf{P}\mathbf{\Sigma}\mathbf{P}'$ *is diagonal.*

b. *The* **G** *matrices commute.*

c. $\mathbf{G}_i\mathbf{G}_j$ *is symmetric,* $i, j = 0, 1, \cdots, m$.

THEOREM 5. *Assume that the maximum likelihood estimate of $\boldsymbol{\beta}$ has an explicit representation and that the* **G**'s *in* $\mathbf{\Sigma} = \Sigma_{g=0}^{m}\sigma_g\mathbf{G}_g$ *are all diagonal in the canonical form. Then the maximum likelihood estimate of $\boldsymbol{\sigma}$ has an explicit representation if and only if the diagonal elements of $\mathbf{\Sigma}$ consist of exactly $m + 1$ linearly independent combinations of the σ's.*

PROOF. First note that there must be at least $m + 1$ linear combinations since the **G**'s are linearly independent. If there are exactly $m + 1$ combinations, then by representing these combinations by $\tau_0, \tau_1, \cdots, \tau_m$ where $\boldsymbol{\sigma} = \mathbf{A}\boldsymbol{\tau}$ for some nonsingular **A** and by grouping the same linear combinations together, we see that $\mathbf{\Sigma}$ can be put in the form $\mathbf{\Sigma} = \text{diag}(\tau_0\mathbf{I}_{p_0}, \tau_1\mathbf{I}_{p_1}, \cdots, \tau_m\mathbf{I}_{p_m})$. Here $\Sigma_{g=0}^{m}p_g = p$ and $\mathbf{I}_s$ is the $s \times s$ identity matrix. In this form, $\mathbf{\Phi}\mathbf{\Phi}_{\mathbf{I}}^{-1} = \text{diag}(\tau_0^2\mathbf{I}_{p_0}, \tau_1^2\mathbf{I}_{p_1}, \cdots, \tau_m^2\mathbf{I}_{p_m}, \mathbf{b})$ where **b** is a row vector of various positive elements representing variances of off-diagonal elements of the form $\sigma_{ii}\sigma_{jj}$. The $\langle\mathbf{G}\rangle$'s of $\mathbf{\Sigma}$ in this form are themselves eigenvectors of $\mathbf{\Phi\Phi}_{\mathbf{I}}^{-1}$, thus by Theorem 4, there are explicit maximum likelihood estimates of $\boldsymbol{\tau}$ and thus for $\boldsymbol{\sigma}$.

Suppose there are more than $m + 1$ combinations. Then one can represent $m + 1$ combinations by $\tau_0, \cdots, \tau_m$ where $\boldsymbol{\sigma} = \mathbf{A}\boldsymbol{\tau}$, $\mathbf{A}$ nonsingular, and the remaining different combinations by $\gamma_1, \cdots, \gamma_f$ where the γ's are linear combinations of two or more τ's. Again, rearranging the order of the diagonal elements yields $\mathbf{\Sigma}$ of the form $\mathbf{\Sigma} = \operatorname{diag}(\tau_0\mathbf{I}_{r_0}, \cdots, \tau_m\mathbf{I}_{r_m}, \gamma_1\mathbf{I}_{s_1}, \cdots, \gamma_f\mathbf{I}_{s_f})$ where $\Sigma_{g=0}^{m} r_g + \Sigma_{h=1}^{f} s_h = p$, $f \geqslant 1$. In this form, $\mathbf{\Phi\Phi}_{\mathbf{I}}^{-1} = \operatorname{diag}(\tau_0^2\mathbf{I}_{r_0}, \cdots, \tau_m^2\mathbf{I}_{r_m}, \gamma_1^2\mathbf{I}_{s_1}, \cdots, \gamma_f^2\mathbf{I}_{s_f}, \mathbf{b})$ where $\mathbf{b}$ is a row vector of the form described earlier in the proof. Clearly more than $m + 1$ eigenvectors of $\mathbf{\Phi\Phi}_{\mathbf{I}}^{-1}$ are needed in this case to span the $\langle \mathbf{G} \rangle$'s. Thus by Theorem 4, the $\boldsymbol{\tau}$'s do not have explicit maximum likelihood estimators and thus neither do the $\boldsymbol{\sigma}$'s.

5. Applications and examples.

5.1 COMPLETE, COMPOUND AND CIRCULAR SYMMETRY. The block and nonblock forms of complete (Wilks, (1946)), compound (Votaw (1948), Arnold (1973), (1976), Szatrowski (1976), (1977)) and circular symmetry (Olkin and Press (1969), Olkin (1972)) are all known to have explicit maximum likelihood estimates for the patterned means and covariances considered by these authors. Since all these forms include $\mathbf{\Sigma} = \mathbf{I}$, we know by Theorems 2 and 4 that the maximum likelihood estimates are given by the likelihood equations (2.1) and (2.2) with $\hat{\mathbf{\Sigma}} = \mathbf{I}$. In addition, since the $\mathbf{z}$'s and $\mathbf{G}$'s may be parameterized to consist of only zeroes and ones, we known that the maximum likelihood estimates for the elements of the patterned mean vector may be found by averaging corresponding elements of the sample mean vector $\bar{\mathbf{x}}$, and the maximum likelihood estimates for the patterned covariance matrix may be found by averaging corresponding elements in the sample covariance matrix $\mathbf{C}$. This last result has not been previously reported for block compound symmetry and for circular symmetry.

5.2. ASYMPTOTIC NONNULL DISTRIBUTIONS. Szatrowski (1979) obtains the asymptotic nonnull distribution of the likelihood ratio statistic for the problem of testing which of two nested models with patterned means and covariances matrices is appropriate. When we have explicit maximum likelihood estimates under the null hypothesis, one can use the likelihood equations (2.1) and (2.2) to give a simple expression for the maximum likelihood estimates as a function of the data since $\hat{\mathbf{\Sigma}}$ in these expressions can be any allowable starting point and thus is independent of the data. This greatly simplifies taking derivatives needed for the standard delta method as does the use of several versions of Theorem 3.

APPENDIX

PROOF OF THEOREM 1

Necessity. Taking the transpose and multiplying by $\mathbf{A}$ on the left yields $\mathbf{X}(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})^{-1} = \mathbf{A}\mathbf{B}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})^{-1}$. There exists a nonsingular matrix $\mathbf{P}$ with the properties $\mathbf{P}'(\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})\mathbf{P} = \mathbf{I}$ and $\mathbf{P}'(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})\mathbf{P} = \mathbf{D}^{-1}$ where $\mathbf{D}$ is a nonsingular diagonal matrix. Multiplying on the right by $(\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})\mathbf{P}$ yields, after freely inserting $\mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$ in several places,

$$\mathbf{XP}\left[\mathbf{P}^{-1}(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{P}^{-\prime}\right]\left[\mathbf{P}'(\mathbf{X}'\mathbf{B}^{-1}\mathbf{X})\mathbf{P}\right] = \mathbf{AB}^{-1}\mathbf{XP}.$$

By properties of $\mathbf{P}$, we note the first expression in [ ]'s is $\mathbf{D}$ and the second $\mathbf{I}$ yielding $(\mathbf{XP})\mathbf{D} = \mathbf{AB}^{-1}(\mathbf{XP})$. From this last expression we see that the columns of $\mathbf{XP}$ are eigenvectors of $\mathbf{AB}^{-1}$, thus the columns of X are linear combinations of eigenvectors of $\mathbf{AB}^{-1}$.

Sufficiency. Assuming the columns of $\mathbf{X}$ are linear combinations of the characteristic vectors of $\mathbf{AB}^{-1}$, then $\mathbf{X} = \mathbf{QF}$ where $\mathbf{F}$ is nonsingular and $\mathbf{AB}^{-1}\mathbf{Q} = \mathbf{Q}\Lambda$, Λ diagonal with positive elements. Substitution of $\mathbf{X} = \mathbf{QF}$ yields $(\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})\mathbf{X}'\mathbf{A}^{-1} = \mathbf{F}^{-1}(\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{A}^{-1}$ and a similar expression with $\mathbf{A}$ replaced by $\mathbf{B}$. Thus we need to show $(\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{A}^{-1} = (\mathbf{Q}'\mathbf{B}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{B}^{-1}$. This follows by noting $(\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q})^{-1}\mathbf{Q}' = (\Lambda\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q})^{-1}\Lambda\mathbf{Q}' = (\mathbf{Q}'\mathbf{B}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{Q})^{-1}\Lambda\mathbf{Q}' = (\mathbf{Q}'\mathbf{B}^{-1}\mathbf{Q})^{-1}\Lambda\mathbf{Q}' = (\mathbf{Q}'\mathbf{B}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{B}^{-1}\mathbf{A}$.

Acknowledgments. I would like to express my appreciation to my thesis advisor, Professor Ingram Olkin for suggesting this area of research. Thanks are also due to an anonymous referee who suggested that with careful formulation Zyskind's (1967) results might yield necessary and sufficient conditions for one-iteration solutions. Finally, special thanks go to Professor John Miller for numerous helpful comments and conversations on both this and an earlier version of this paper. His insight into this topic was most helpful.

REFERENCES

[1] Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. *Proc. Second Internat. Symp. Multivariate Anal.* (P. R. Krishnaiah, ed.), pages 55–66. Academic Press, New York.

[2] Anderson, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. *Essays in Probability and Statistics*, pages 1–24. Univ. North Carolina Press, Chapel Hill.

[3] Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[4] Anderson, T. W. (1973). Asymptotic efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

[5] Arnold, S. F. (1973). Application of the theory of products of problems to certain patterned covariance matrices. *Ann. Statist.* **1** 682–699.

[6] Arnold, S. F. (1976). Applications of products to the generalized compound symmetry problem. *Ann. Statist.* **4** 227–233.

[7] Miller, J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. *Technical Report No. 12*, Stanford Univ.

[8] Miller, J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

[9] Mitra, S. K. and Moore, B. J. (1973). Gauss-Markov estimation with an incorrect dispersion matrix. *Sankyā, Ser. A* **35** 139–152.

[10] Olkin, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks. *Proc. Symp. Multivariate Anal.* pages 183–195. Dalhousie, Nova Scotia.

[11] OLKIN, I. and PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Statist.* **40** 1358–1373.
[12] RAO, J. N. K. (1973). Personal communication to T. W. Anderson.
[13] SEELY, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist.* **42** 710–721.
[14] SZATROWSKI, T. H. (1976). Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure. *Technical Report No. 107*, Stanford Univ.
[15] SZATROWSKI, T. H. (1977). Testing and estimation in the block compound symmetry problem. *Rutgers Technical Report*, Rutgers Univ.
[16] SZATROWSKI, T. H. (1979). Asymptotic nonnull distributions for likelihood ratio statistics in the multivariate normal patterned mean and covariance matrix testing problem. *Ann. Statist.* **7** 823–837.
[17] SZATROWSKI, T. H. and MILLER, J. (1980). Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance. *Ann. Statist.* **8** 811–819.
[18] THOMAS, D. H. (1968). When do minimum variance estimators coincide? *Ann. Math. Statist.* **39** 1365.
[19] VOTAW, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *Ann. Math. Statist.* **19** 447–473.
[20] WILKS, S. S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist.* **17** 257–281.
[21] ZYSKIND, G. (1967). On canonical forms, non-negative covariances matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903