The Annals of Statistics
1991, Vol. 19, No. 2, 850–869

MAXIMUM LIKELIHOOD ESTIMATION OF A SET OF COVARIANCE MATRICES UNDER LÖWNER ORDER RESTRICTIONS WITH APPLICATIONS TO BALANCED MULTIVARIATE VARIANCE COMPONENTS MODELS

BY JAMES A. CALVIN[1] AND RICHARD L. DYKSTRA[2]

University of Iowa

The problem of maximum likelihood estimation of Löwner ordered covariance matrices is considered. It is shown that a dual formulation of this problem is tractable and important in its own right. The interplay between the primal and dual problems suggests a general algorithm for computing the solutions to these problems. This algorithm has application to some estimation problems in balanced multivariate variance components models. The speed of convergence is also discussed for the variance components models.

1. Introduction. The study of variance component estimation in univariate mixed models has long been a topic of interest to statisticians. Henderson (1953) proposed equating mean squares with their expected values to estimate individual variance components. Herbach (1959), Thompson (1962) and Patterson and Thompson (1971, 1975) have proposed procedures for maximum likelihood (MLE) and restricted maximum likelihood (REML) estimation of the components. Rao (1971a, 1971b) developed MINQU and MIVQU estimators which are related to REML estimation and are discussed in Rao and Kleffe (1988). Searle (1971) and Harville (1977) provide excellent summaries of several such techniques. Harville and Callanan (1990) discuss the computation problems associated with REML estimation.

One of the problems encountered in estimation can best be seen by considering a simple example. Consider the two-factor random effects model without interaction:

$$Y_{ijk} = \mu + a_i + b_j + e_{ijk}, \qquad i = 1,\dots,A;\ j = 1,\dots,B;\ k = 1,\dots,N, \tag{1.1}$$

where the a_i, b_j and e_{ijk} are independent with $a_i \sim N(0, \sigma_b^2)$, $b_j \sim N(0, \sigma_b^2)$ and $e_{ijk} \sim N(0, \sigma_e^2)$. A set of minimal sufficient statistics for this model is $\overline{Y}_{\dots}$,

Received August 1988; revised March 1990.

[1]Work partially supported by an Old Gold Fellowship from the University of Iowa. Partial support was also supplied by Texas A & M University while this author was visiting during the 1989/1990 academic year.

[2]Work partially supported by Office of Naval Research Contract N00014-83-K-0249. Partial support was also provided by the Army Research Office through the Mathematical Sciences Institute while this author was visiting Cornell University.

AMS 1980 *subject classifications*. Primary 62H12; secondary 62J10.

Key words and phrases. Isotonic regression, Wishart density, Fenchel duality, multivariate linear model, restricted maximum likelihood estimation, REML, variance components models, least squares estimation.

MS_a, MS_b and MS_e [Arnold (1981), Section 15.2], where MS_a, MS_b and MS_e are the standard mean squares from the analysis of variance table. The expected values of the mean squares (EMS) for a, b and e are $\sigma_e^2 + BN\sigma_a^2$, $\sigma_e^2 + AN\sigma_b^2$ and σ_e^2, respectively. The method-of-moments estimation procedure—equating the mean squares for a, b and e to their expected values and solving for σ_a^2, σ_b^2 and σ_e^2—may encounter the problem of negative estimates since either MS_a or MS_b may be smaller than MS_e. MINQUE and MIVQUE, which are unbiased, also have a positive probability of yielding negative variance estimates. In practice, one usually sets these estimates to zero [Arnold (1981), page 262]. Likelihood based estimation procedures will typically, in addition, adjust the other estimates for the fact that one or more of the estimates are zero. In the balanced one-factor model, the REML estimate of σ_e^2 is the weighted average of MS_a and MS_e when σ_a^2 is estimated to be zero.

The development of the equivalent estimation and computational techniques for multivariate variance components models has not proceeded at the same pace. Clearly, equating mean squares to their expected values is still feasible. However, there is little agreement about what procedures should be employed. Rao and Kleffe (1988) discuss multivariate extensions to MINQUE and MIVQUE procedures, but implementation is difficult and the iterated versions may be slow to converge.

The multivariate version of (1.1) is

$$\mathbf{Y}_{ijk} = \boldsymbol{\mu} + \mathbf{A}_i + \mathbf{B}_j + \mathbf{E}_{ijk}, \quad i = 1,\dots,A;\ j = 1,\dots,B;\ k = 1,\dots,N, \tag{1.2}$$

where $\mathbf{A}_i$, $\mathbf{B}_j$ and $\mathbf{E}_{ijk}$ are independent $p \times 1$ vectors, with $\mathbf{A}_i \sim N(\mathbf{0}, \Sigma_A)$, $\mathbf{B}_j \sim N(\mathbf{0}, \Sigma_B)$ and $\mathbf{E}_{ijk} \sim N(\mathbf{0}, \Sigma_E)$. For this model a set of minimal sufficient statistics is $\{\overline{\mathbf{Y}}_{\dots}, A_A, A_B, A_E\}$, where

$$\begin{aligned} A_A &= \sum_{i=1}^{A}\sum_{j=1}^{B}\sum_{k=1}^{N} (\overline{\mathbf{Y}}_{i\cdot\cdot} - \overline{\mathbf{Y}}_{\dots})(\overline{\mathbf{Y}}_{i\cdot\cdot} - \overline{\mathbf{Y}}_{\dots})', \\ A_B &= \sum_{i=1}^{A}\sum_{j=1}^{B}\sum_{k=1}^{N} (\overline{\mathbf{Y}}_{\cdot j\cdot} - \overline{\mathbf{Y}}_{\dots})(\overline{\mathbf{Y}}_{\cdot j\cdot} - \overline{\mathbf{Y}}_{\dots})' \end{aligned} \tag{1.3}$$

and

$$\begin{aligned} A_E = &\sum_{i=1}^{A}\sum_{j=1}^{B}\sum_{k=1}^{N} \left(\overline{\mathbf{Y}}_{ij\cdot} - \overline{\mathbf{Y}}_{i\cdot\cdot} - \overline{\mathbf{Y}}_{\cdot j\cdot} + \overline{\mathbf{Y}}_{\dots}\right)\left(\overline{\mathbf{Y}}_{ij\cdot} - \overline{\mathbf{Y}}_{i\cdot\cdot} - \overline{\mathbf{Y}}_{\cdot j\cdot} + \overline{\mathbf{Y}}_{\dots}\right)' \\ &+ \sum_{i=1}^{A}\sum_{j=1}^{B}\sum_{k=1}^{N} \left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij\cdot}\right)\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij\cdot}\right)'. \end{aligned}$$

The sums of squares matrices A_A, A_B and A_E are distributed $\mathrm{Wishart}_p(n_A = A - 1,\ \mathrm{EMS}_A = \Sigma_E + BN\Sigma_A)$, $\mathrm{Wishart}_p(n_B = B - 1,\ \mathrm{EMS}_B = \Sigma_E + AN\Sigma_B)$ and $\mathrm{Wishart}_p(n_E = ABN - A - B + 1,\ \mathrm{EMS}_E = \Sigma_E)$, respectively. If the mean square matrices are equated to their EMS matrices, it is easy to see that it is possible to obtain estimates for individual matrices that are not nonnegative definite (n.n.d.). Hill and Thompson (1978) and Bhargava and Disch (1982) show that for even relatively simple models, the probability of obtaining an estimate outside the parameter space can be quite high. In this case setting these matrix estimates to zero is clearly the wrong thing to do. One alternative is to project the estimates onto the space of n.n.d. matrices by taking a spectral decomposition of the matrices outside the parameter space, setting negative eigenvalues to zero and then recombining to get estimates in the parameter space. Klotz and Putter (1969), Amemiya (1985) and Anderson, Anderson and Olkin (1986) discuss maximum likelihood versions of this problem for models with only two variance component matrices. However, when more than two matrices are involved, it is not obvious what should be done.

In this paper we are concerned with maximum likelihood estimation of the parameter matrices of a set of Wishart matrices subject to the restriction that the differences between certain pairs of matrices are nonnegative definite. In practice, the Wishart matrices are typically the sums-of-squares and cross-products matrices from a balanced multivariate variance components model and the parameter matrices are the corresponding EMS's. (Note that model 1.2 satisfies these conditions when estimation is based on A_A, A_B and A_E.) The pairwise difference restrictions correspond to requiring that the estimates of the individual covariance matrices be in the parameter space. In the case of balanced variance components models, our procedure leads to REML estimates of the matrices. Our proposed procedure will work for any balanced model for which the covariance matrices can be described as the difference between two EMS matrices. All nested models and all two-factor models fall into this wide class of balanced multivariate variance component models.

As an example, we consider a subset of the data from Calvin and Sedransk (1991). They look at two response measures of the quality of care received by cancer patients. The data was collected from random samples of patients within each of a set of randomly sampled hospitals within each of seven strata classifications. In our subset we will use two randomly selected patients within each of two randomly selected hospitals within each of the seven strata. The bivariate response measure contains (1) the pretreatment score and (2) the therapy score, which were designed to measure the quality of care received during the two stages of patient care. [The data and the software necessary to perform the analysis described in this paper are available from the authors upon request. For a further description of the data, see Calvin and Sedransk, (1991).] Thus, a model for this data is the two-factor random effects nested model:

$$\mathbf{Y}_{ijk} = \boldsymbol{\mu} + \mathbf{A}_i + \mathbf{B}_{ij} + \mathbf{E}_{ijk}, \tag{1.4}$$

where the strata effect $\mathbf{A}_i$, the facility effect $\mathbf{B}_{ij}$ and the random error $\mathbf{E}_{ijk}$ are independent 2×1 vectors, with $\mathbf{A}_i \sim N(\mathbf{0}, \Sigma_A)$, $\mathbf{B}_{ij} \sim N(\mathbf{0}, \Sigma_B)$, and $\mathbf{E}_{ijk} \sim N(\mathbf{0}, \Sigma_E)$. Using our subset we find the EMS's for this model are $\mathrm{EMS}_A = \Sigma_E + 2\Sigma_B + 4\Sigma_A$, $\mathrm{EMS}_B = \Sigma_E + 2\Sigma_B$ and $\mathrm{EMS}_E = \Sigma_E$. Thus, $\Sigma_A = (\mathrm{EMS}_A - \mathrm{EMS}_B)/4$, $\Sigma_B = (\mathrm{EMS}_B - \mathrm{EMS}_e)/2$ and $\Sigma_e = \mathrm{EMS}_e$. The mean square matrices for the data and the estimates of Σ_A, Σ_B and Σ_E based on equating the mean squares to their expected values are listed next. Note that the solutions do not fall in the parameter space since the estimate of Σ_A, $(\mathrm{MS}_A - \mathrm{MS}_B)/4$, is not n.n.d.

$$\mathrm{MS}_A = \begin{bmatrix} 2420.68 & 1571.20 \\ 1571.20 & 1681.43 \end{bmatrix} \qquad \mathrm{MS}_B = \begin{bmatrix} 374.21 & 167.87 \\ 167.87 & 1136.51 \end{bmatrix}$$

$$\mathrm{MS}_E = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix}$$

(1.5)

$$\hat{\Sigma}_A = \begin{bmatrix} 511.62 & 350.83 \\ 350.83 & 136.23 \end{bmatrix} \qquad \hat{\Sigma}_B = \begin{bmatrix} 59.32 & 59.96 \\ 59.96 & 536.51 \end{bmatrix}$$

$$\hat{\Sigma}_E = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix}.$$

Projecting the individual estimates back into the parameter space yields

$$\hat{\Sigma}_A = \begin{bmatrix} 531.16 & 318.22 \\ 318.22 & 190.65 \end{bmatrix} \qquad \hat{\Sigma}_B = \begin{bmatrix} 59.32 & 59.96 \\ 59.96 & 536.51 \end{bmatrix}$$

$$\hat{\Sigma}_E = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix}.$$

However, it is not clear that $\hat{\Sigma}_A$ should be the only estimate which is modified. It is based on both MS_A and MS_B and, as we shall show, this simple adjustment does not yield restricted maximum likelihood estimates.

Section 2 provides background notation and defines the problem. Section 3 is devoted to developing the major building block in the estimation procedure and Section 4 describes the general estimation procedure. Section 5 discusses the speed of convergence of the algorithm and provides an application using the data in (1.5). The proof of the convergence of the algorithm is left to the Appendix.

2. Background and problem definition. To set notation, suppose $\Sigma_1, \dots, \Sigma_k$ are real symmetric $p \times p$ matrices and that $\lesssim$ is a partial-order on the index set $\{1, \dots, k\}$. We will say that the vector of matrices

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_1 \\ \vdots \\ \Sigma_k \end{bmatrix}$$

is isotonic with respect to $\lesssim$ if it is order preserving in the Löwner sense [Löwner, (1934)]. This means that if $i \lesssim j$, then $\Sigma_j - \Sigma_i$ is n.n.d, which we write as $\Sigma_i \leq \Sigma_j$. We say that $\mathbf{\Sigma}$ is antitonic if $i \lesssim j$ implies $\Sigma_j \leq \Sigma_i$. The set G, which we will refer to as the minimal set of pairwise restrictions, contains the smallest possible set of pairs (i, j) which describe the partial order. For (1.4), if we let $\Sigma_1 = \mathrm{EMS}_E$, $\Sigma_2 = \mathrm{EMS}_B$ and $\Sigma_3 = \mathrm{EMS}_A$, then $G = \{(1,2),(2,3)\}$. Thus, our goal is to construct the maximum likelihood estimate of $\mathbf{\Sigma}$ which is isotonic with respect to $\lesssim$, the partial order associated with G.

Let $A_1, \dots, A_k$ denote independent $p \times p$ Wishart random matrices with respective parameters $(n_1, \Sigma_1), \dots, (n_k, \Sigma_k)$, where the n_i are all known and greater than or equal to p and the Σ_i are all positive definite (p.d.). The log-likelihood of Σ can be expressed as

$$
\begin{aligned}
\ell_0(\Sigma, A) &= \ln \prod_{i=1}^{k} \frac{|A_i|^{(n_i-p-1)/2} \exp\{\operatorname{tr}(-\frac{1}{2}A_i\Sigma_i^{-1})\}}{2^{pn_i/2}|\Sigma_i|^{n_i/2}\pi^{p(p-1)/4}\prod_{j=1}^{p}\Gamma((n_i+1-j)/2)} \\
&= \sum_{i=1}^{k}\left\{-\frac{n_i}{2}\ln|\Sigma_i| - \frac{1}{2}\operatorname{tr}(A_i\Sigma_i^{-1})\right\} + c,
\end{aligned}
\tag{2.1}
$$

where c is a constant which does not depend upon Σ.

For our particular problem, it will be convenient to parameterize the likelihood in terms of $\Lambda_i = \Sigma_i^{-1}$, $i = 1, \dots, k$. Since $0 < \Sigma_i \le \Sigma_j$ iff $\Lambda_i \ge \Lambda_j > 0$ [Rao, (1973), page 70], our problem then becomes one of maximizing the likelihood over the class of antitonic rather than isotonic vectors of matrices. While this is a difficult optimization problem which cannot be solved in closed form, we are able to provide a tractable algorithm which is guaranteed to converge to the correct solution. Our approach hinges on the delicate interplay between the stated problem and a specific dual formulation based on the Fenchel duality theorem.

Fenchel-type duality theorems have been used in several statistical contexts to identify interesting dual problems, e.g., Barlow and Brunk (1972), Pukelsheim (1980, 1981), Müller-Funk, Pukelsheim and Witting (1985) and Dykstra and Lemke (1988). In this paper, however, it is the interplay between the primal and the dual problems which is of primary interest, since it suggests the algorithm and provides a vehicle for the proof of convergence. We now briefly review the needed Fenchel duality theorem, as stated in Rockafellar [(1970), page 335].

THEOREM 2.1 (Fenchel). *Suppose f is a closed proper concave function defined on $\mathbb{R}^n$ and K is a closed nonempty convex cone in $\mathbb{R}^n$. If we define the concave conjugate of f, f^*, and the dual cone of K, K^*, as*

$$
f^*(\mathbf{y}) = \inf_{\mathbf{x}\in\mathbb{R}^n}\left\{\sum_{i=1}^{n} x_i y_i - f(\mathbf{x})\right\}, \qquad K^* = \left\{\mathbf{y}\in\mathbb{R}^n : \sum_{i=1}^{n} x_i y_i \le 0\ \forall\, \mathbf{x}\in K\right\},
$$

then

$$
\sup_{\mathbf{x}\in K} f(\mathbf{x}) = -\sup_{\mathbf{y}\in K^*} f^*(\mathbf{y}) \tag{2.2}
$$

if either

(i) ri(dom f) $\cap$ ri(K) $\neq \varnothing$ *or*
(ii) ri(dom f^*) $\cap$ ri(K^*) $\neq \varnothing$,

where ri *means relative interior and* dom $f = \{\mathbf{x}\in\mathbb{R}^n : f(\mathbf{x}) > -\infty\}$.

If (i) holds, $\sup_{\mathbf{y} \in K^*} f^*(\mathbf{y})$ is attained and if (ii) holds, $\sup_{\mathbf{x} \in K} f(\mathbf{x})$ is attained. In general, $\mathbf{x}^*$ and $\mathbf{y}^*$ are respective solutions to (2.2) such that

$$f(\mathbf{x}^*) = \sup_{\mathbf{x} \in K} f(\mathbf{x}) = -\sup_{\mathbf{y} \in K^*} f^*(\mathbf{y}) = -f^*(\mathbf{y}^*)$$

if and only if

$$\mathbf{x}^* \in K, \tag{2.3a}$$

$$\mathbf{y}^* \in K^*, \tag{2.3b}$$

$$\sum_{i=1}^{n} x_i^* y_i^* = 0, \tag{2.3c}$$

$$-\mathbf{y}^* \text{ is a subgradient of } -f \text{ at } \mathbf{x}^*. \tag{2.3d}$$

Suppose now that K denotes the cone of antitonic vectors of symmetric $p \times p$ matrices. Our MLE problem is then of the same form as the primal problem, $\sup_{\mathbf{x} \in K} f(\mathbf{x})$, in Theorem 2.1; that is, find the supremum over $\boldsymbol{\Lambda} \in K$ of

$$\ell(\boldsymbol{\Lambda}, \mathbf{A}) = \begin{cases} \sum_{i=1}^{k} \frac{n_i}{2} \ln|\Lambda_i| - \frac{1}{2} \operatorname{tr}(A_i \Lambda_i), & \Lambda_i \text{ sym. p.d., } i = 1, \dots, k, \\ -\infty, & \text{elsewhere.} \end{cases} \tag{2.4}$$

Since $\ln|\Lambda_i|$ is a concave function of symmetric p.d. Λ_i [Farrell, (1985), page 326], ℓ will be a closed proper concave function. What makes the Fenchel dual approach appealing is the tractability of the concave conjugate of the log-likelihood. [Note: It is possible to formulate (2.4) either in terms of square $p \times p$ matrices (p^2 variables) or triangular $p \times p$ matrices ($p(p+1)/2$ variables) since the problem requires the Λ_i to be symmetric. We have chosen the former, since it makes the Fenchel duality theorem more transparent and is more consistent with standard matrix notation. We then add symmetry by restricting the domain of ℓ.]

LEMMA 2.1. *Let $\mathbf{A}$ and $\boldsymbol{\Lambda}$ be $k \times 1$ vectors of matrices, with each A_i symmetric. The concave conjugate of the function $\ell(\boldsymbol{\Lambda}, \mathbf{A})$ given in (2.4) is*

$$\ell^*(\boldsymbol{\Psi}, \mathbf{A}) = \begin{cases} \frac{1}{2} \sum_{i=1}^{k} n_i \ln|\psi_i + \psi_i' + A_i| + c^*, & \psi_i + \psi_i' + A_i \ p.d., i = 1, \dots, k, \\ -\infty, & \textit{otherwise}, \end{cases}$$

where

$$c^* = \frac{p}{2} \sum_{i=1}^{k} n_i [1 - \ln(n_i)].$$

PROOF. Observe that the matrix of partial derivatives with respect to the $p \times p$ matrix Z are given by

$$\frac{d}{dZ}\operatorname{tr}(A'Z) = A \quad \text{and} \quad \frac{d}{dZ}\ln|Z| = (Z')^{-1} \quad [\text{if } |Z| > 0, \text{Dwyer (1967)}].$$

Then

$$\begin{aligned}\ell^*(\mathbf{\Psi}, \mathbf{A}) &= \inf_{\mathbf{\Lambda}}\left[\sum_{i=1}^{k} \operatorname{tr}(\psi_i'\Lambda_i) - \ell(\mathbf{\Lambda}, \mathbf{A})\right] \\ &= \inf_{\substack{\mathbf{\Lambda}:\, \Lambda_i \text{ sym.} \\ \text{and p.d.}}} \frac{1}{2}\left[\sum_{i=1}^{k} \operatorname{tr}\big((\psi_i + \psi_i' + A_i)'\Lambda_i\big) - \sum_{i=1}^{k} n_i \ln|\Lambda_i|\right].\end{aligned}$$

The quantity in brackets is a convex function of symmetric p.d. $\mathbf{\Lambda}$ and is differentiable over the restricted region. Setting the partial derivatives of the quantity in brackets equal to zero results in the equations

$$(\psi_i + \psi_i' + A_i) = n_i \Lambda_i'^{-1} \quad \text{or} \quad \Lambda_i = n_i(\psi_i + \psi_i' + A_i)'^{-1}.$$

This easily yields the desired result since these Λ_i are symmetric. □

Note that all $\mathbf{\Psi}$ in $\{\mathbf{\Psi}: \psi_i + \psi_i' = M_i,\ i = 1, \dots, k\}$ yield the same value for $\ell^*(\mathbf{\Psi}, \mathbf{A})$. Thus, using components ψ_i, ψ_i' or $(\psi_i + \psi_i')/2$ will yield the same solution. Since our interest is in symmetric matrices, when considering dual problems we will always let ψ_i represent the symmetric matrix which yields the desired solution.

To apply Theorem 2.1 we will need to be able to find the form of the dual cone K^*. The following lemma [Rockafellar (1970), page 146] will be useful.

LEMMA 2.2. *If $K_i, \dots, K_m$ are convex cones such that $\bigcap_{i=1}^{m} \operatorname{ri}(K_i) \neq 0$, then*

$$\left(\bigcap_{i=1}^{m} K_i\right)^* = \underset{i=1}{\overset{m}{+}} K_i^*,$$

where $+$ indicates the sum $(A + B = \{a + b: a \in A,\ b \in B\})$.

If H denotes the cone of symmetric $p \times p$ n.n.d. matrices, then the dual cone of H (in $\mathbb{R}^{p^2}$) need not contain only symmetric matrices. However, if we restrict ourselves to matrices in the dual which are symmetric, then the resulting subset will be the set of symmetric nonpositive definite (n.p.d.) matrices.

LEMMA 2.3. *If H is the cone of symmetric $p \times p$ n.n.d. matrices and Z is the set of symmetric $p \times p$ matrices, then*

$$H^* \cap Z = \{\psi \in Z: \psi \text{ is n.p.d.}\}.$$

PROOF. Since $H^* = \{\psi\colon \operatorname{tr} \psi'\Lambda \le 0\ \forall\, \Lambda \in H\}$, it easily follows that $\psi \in H^*$ iff $\psi + \psi' \in H^* \cap Z$. Thus,

$$\begin{aligned}
H^* \cap Z &= \{\psi \in Z\colon \operatorname{tr} \psi\Lambda \le 0, \forall\, \Lambda \in H\}\\
&= \{\psi \in Z\colon \operatorname{tr} \psi O\Lambda O' \le 0, \forall\, \Lambda \in H, \forall\, O \text{ orthnormal}\}\\
&= \left\{\psi \in Z\colon \sup_{O \text{ orth}} \operatorname{tr} \psi O\Lambda O' \le 0, \forall\, \Lambda \in H\right\}\\
&= \left\{\psi \in Z\colon \sum_{i=1}^{p} \lambda_i(\psi)\lambda_i(\Lambda) \le 0, \forall\, \Lambda \in H\right\} \quad \text{[Farrell (1985), page 323]}\\
&= \{\psi \in Z\colon \lambda_1(\psi) \le 0\} = \{\psi \in Z\colon \psi \text{ is n.p.d.}\},
\end{aligned}$$

where $\lambda_i(\cdot)$ denotes the ith largest characteristic root of the symmetric argument. □

We can now identify the structure of the dual cone of K.

THEOREM 2.2. *Let K denote the set of vectors $\mathbf{\Lambda}' = [\Lambda_1, \dots, \Lambda_k]$ whose elements, symmetric $p \times p$ matrices, are antitonic with respect to $\lesssim$. Let $G = \{(i_l, j_l)\colon l = 1, \dots, m\}$ be a minimal set of pairwise orderings which generates the partial order $\lesssim$. For each element (i, j) of G, let H_{ij} be the set of vectors of matrices of the form*

$$\mathbf{\Psi}^{(i,j)} = \begin{bmatrix} 0 \cdots 0 & \psi_i & 0 \cdots 0 & \psi_j & 0 \cdots 0 \end{bmatrix}', \tag{2.5}$$

where $\psi_i = -\psi_j$, 0's are in all other positions and $\psi_i + \psi_i'$ is n.p.d. Then the dual cone of K is

$$K^* = \mathop{+}_{(i,j)\in G} H_{ij}.$$

PROOF. By Lemma 2.2, it will suffice to show that H_{ij} is the dual of the convex cone

$$K_{ij} = \{\mathbf{\Lambda}\colon \Lambda_l \text{ symmetric } \forall\, l,\ \Lambda_i \ge \Lambda_j\}.$$

To use Lemma 2.3, we need to show that every element in $K_{ij}^* \cap Z^k$ is of the form (2.5), where ψ_i is a symmetric $p \times p$ n.p.d. matrix and Z^k is the set of vectors of length k whose elements are symmetric $p \times p$ matrices. For the elements in K_{ij}, only the ith and the jth coordinates have restrictions. Thus, clearly all coordinates for the elements in K_{ij}^* must be 0 except in the ith and jth positions, and

$$\begin{aligned}
K_{ij}^* \cap Z^k = \{\mathbf{\Psi}\colon &\operatorname{tr} \psi_i\Lambda_i + \operatorname{tr} \psi_j\Lambda_j \le 0, \forall\, \mathbf{\Lambda} \in K_{ij},\\
&\psi_i \text{ and } \psi_j \text{ symmetric}, \psi_l = 0\ l \ne i, j\}\\
= \{\mathbf{\Psi}\colon &\operatorname{tr} \psi_i(\Lambda_i - \Lambda_j) + \operatorname{tr}(\psi_i + \psi_j)\Lambda_j \le 0,\\
&\forall\, \mathbf{\Lambda} \in K_{ij}, \psi_i \text{ and } \psi_j \text{ symmetric}, \psi_l = 0, l \ne i, j\}.
\end{aligned}$$

However, it can fairly easily be shown that for $\boldsymbol{\Psi}$ to be in $K_{ij}^* \cap Z^k$, ψ_i must equal $-\psi_j$. Thus,

$$K_{ij}^* \cap Z^k = \big\{\boldsymbol{\Psi}: \operatorname{tr} \psi_i(\Lambda_i - \Lambda_j) \le 0, \forall \boldsymbol{\Lambda} \in K_{ij}, \\ \psi_i = -\psi_j \text{ symmetric}, \psi_l = 0,\ l \ne i, j\big\}.$$

But then by Lemma 2.3, ψ_i must be n.p.d. which implies the desired result. □

We note that if $p = 1$, (2.5) is a multiple of an elementary contrast. Thus, the dual cone, K^* is the direct sum spanned by elementary contrasts between those elements whose subscripts appear in the minimal set G.

The original MLE problem now fits nicely into the dual formulation discussed in Theorem 2.1 with

$$\sup_{\substack{\boldsymbol{\Lambda} \in K \\ \Lambda_i \text{ p.d.}}} \sum_{i=1}^{k} \frac{n_i}{2} \ln|\Lambda_i| - \frac{1}{2} \operatorname{tr}(A_i \Lambda_i) \tag{2.6}$$

being the primal problem and

$$\sup_{\substack{\boldsymbol{\Psi} \in K^* \\ \psi_i + \psi_i' + A_i \text{ p.d.}}} \sum_{i=1}^{k} \frac{n_i}{2} \ln|\psi_i + \psi_i' + A_i| + c^* \tag{2.7}$$

being the dual problem, where c^* is the constant given in Lemma 2.1 and does not depend upon $\mathbf{A}$.

Both problems will have solutions since conditions (i) and (ii) of Theorem 2.1 are satisfied. Of course, there is a strong connection between the solutions to these two problems. The key in relating these solutions is condition (d) of Theorem 2.1. As mentioned after the definition of ℓ^* in Lemma 2.1, all $\boldsymbol{\Psi}$ such that $\boldsymbol{\Psi} + \boldsymbol{\Psi}'$ is a fixed vector yield the same value of ℓ^*. Thus, a $\boldsymbol{\Psi}$ which maximizes (2.7) is not unique. However, if $\boldsymbol{\Psi}$ is a solution, the vector with the ith element being $(\psi_i + \psi_i')/2$ will yield a unique solution within the set of vectors of symmetric matrices. It can be shown that the only symmetric subgradient of $-\ell$ at $\boldsymbol{\Lambda}$ is given by

$$\begin{bmatrix} \frac{1}{2} A_1 - \frac{n_1}{2} \Lambda_1^{-1} \\ \vdots \\ \frac{1}{2} A_k - \frac{n_k}{2} \Lambda_k^{-1} \end{bmatrix}. \tag{2.8}$$

Thus, by condition (d) of Theorem 2.1, if $\hat{\boldsymbol{\Lambda}}$ solves the primal problem, then the unique symmetric solution to the dual problem must be

$$\hat{\Psi} = \begin{bmatrix} \frac{n_1}{2}\hat{\Lambda}_1^{-1} - \frac{1}{2}A_1 \\ \vdots \\ \frac{n_k}{2}\hat{\Lambda}_k^{-1} - \frac{1}{2}A_k \end{bmatrix}. \tag{2.9}$$

Moreover, from Theorem 2.1 the conditions $\hat{\Psi} \in K^*$ and $\sum_{i=1}^{k} \operatorname{tr}(\hat{\Lambda}_i \hat{\psi}_i) = 0$ must also hold. Conversely, if $\hat{\Lambda}$ is a vector of symmetric p.d. matrices in K such that the $\hat{\Psi}$ defined in (2.9) satisfies $\hat{\Psi} \in K^*$ and $\sum_{i=1}^{k} \operatorname{tr}(\hat{\Lambda}_i \hat{\psi}_i) = 0$, then $\hat{\Lambda}$ must be a solution to the primal problem and $\hat{\Psi}$ must solve the dual problem.

3. The pairwise problem. The pairwise problem with the single constraint $\Lambda_1 \geq \Lambda_2$ can be nicely solved in closed form and also serves to illustrate the relationship between the primal and dual problems. More importantly, as we will illustrate, it is possible to implement an algorithm based only on the pairwise problem which is guaranteed to correctly converge to the solution corresponding to any partial ordering. Although the pairwise problem has been discussed in other contexts [see Klotz and Putter (1969), Amemiya (1985) and Anderson, Anderson and Olkin (1986)], we give an easy verification of the solution which also illustrates the duality aspect. The following procedure obtains the MLE of Λ subject to the restriction $\Lambda_1 \geq \Lambda_2$.

1. Choose a nonsingular matrix B which simultaneously diagonalizes $S_1 = A_1/n_1$ and $S_2 = A_2/n_2$ [see Rao (1973), page 41] so that

$$BS_1B' = C \quad \text{and} \quad BS_2B' = D,$$

where $C = \operatorname{diag}(c_1, \dots, c_p)$ and $D = \operatorname{diag}(d_1, \dots, d_p)$.

2. Let $\hat{C}$ and $\hat{D}$ be diagonal matrices with corresponding ith diagonal elements

$$(\hat{c}_i, \hat{d}_i) = \begin{cases} (c_i, d_i), & \text{if } c_i \leq d_i, \\ \left(\dfrac{n_1 c_i + n_2 d_i}{n_1 + n_2}, \dfrac{n_1 c_i + n_2 d_i}{n_1 + n_2} \right), & \text{otherwise.} \end{cases}$$

3. The MLE's are then given by

$$\hat{\Lambda}_1 = B'\hat{C}^{-1}B \quad \text{and} \quad \hat{\Lambda}_2 = B'\hat{D}^{-1}B.$$

The MLE's of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ subject to $\Sigma_1 \leq \Sigma_2$ are then $\hat{\Sigma}_1 = \hat{\Lambda}_1^{-1} = B^{-1}\hat{C}B'^{-1}$ and $\hat{\Sigma}_2 = \hat{\Lambda}_2^{-1} = B^{-1}\hat{D}B'^{-1}$. Note, that if no pooling takes place $\hat{\Sigma}_i = S_i$ as would be expected. From earlier discussion we know that if $(\hat{\Lambda}_1, \hat{\Lambda}_2)'$ solves the primal problem, then the unique symmetric solution to the dual problem is given by

$$\hat{\boldsymbol{\Psi}} = \begin{bmatrix} \frac{n_1}{2}\hat{\Lambda}_1^{-1} - \frac{1}{2}A_1 \\ \frac{n_2}{2}\hat{\Lambda}_2^{-1} - \frac{1}{2}A_2 \end{bmatrix} = \begin{bmatrix} \frac{n_1}{2}B^{-1}(\hat{C} - C)B'^{-1} \\ \frac{n_2}{2}B^{-1}(\hat{D} - D)B'^{-1} \end{bmatrix}.$$

Note that the ith diagonal elements of $n_1(\hat{C} - C)$ and $n_2(\hat{D} - D)$ are both zero if $c_i \le d_i$. If not, then the ith diagonal element of $n_1(\hat{C} - C)$ is

$$\begin{aligned} n_1(\hat{c}_i - c_i) &= n_1\left(\frac{n_1 c_i + n_2 d_i}{n_1 + n_2} - c_i\right) \\ &= -n_2\left(\frac{n_1 c_i + n_2 d_i}{n_1 + n_2} - d_i\right) = -n_2(\hat{d}_i - d_i), \end{aligned}$$

which is the ith diagonal element of $-n_2(\hat{D} - D)$. Moreover, since $\hat{D} - D$ is clearly n.n.d., it easily follows that $\hat{\boldsymbol{\Psi}} \in K^*$. Clearly $\hat{\boldsymbol{\Lambda}}$ lies in K and $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Lambda}}$ satisfy condition (d) of Theorem 2.1 by the way $\hat{\boldsymbol{\Psi}}$ is defined. Thus all that remains is to show that $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ satisfy condition (c) and, by Theorem 2.1, we will have proven that $\hat{\boldsymbol{\Lambda}}$ is the solution to the primal problem and $\hat{\boldsymbol{\Psi}}$ is a solution to the dual problem. To verify condition (c), note that

$$\begin{aligned} \sum_{i=1}^{k} \operatorname{tr}(\hat{\Lambda}_i \hat{\psi}_i) &= \frac{n_1}{2} \operatorname{tr}\{B'\hat{C}^{-1}BB^{-1}(\hat{C} - C)B'^{-1}\} \\ &\quad + \frac{n_2}{2} \operatorname{tr}\{B'\hat{D}^{-1}BB^{-1}(\hat{D} - D)B'^{-1}\} \\ &= \frac{1}{2} \operatorname{tr}\left\{B'\left[n_1 I + n_2 I - \left(n_1\hat{C}^{-1}C + n_2\hat{D}^{-1}D\right)\right]B'^{-1}\right\}. \end{aligned}$$

But inspection shows that $n_1 I + n_2 I - (n_1\hat{C}^{-1}C + n_2\hat{D}^{-1}D) = 0$, which ensures the desired result.

4. The general problem. The MLE bears a resemblance to a least squares-type estimator which is worth exploring. Consider the vector space V whose elements are $k \times 1$ vectors of real $p \times p$ matrices. An inner product is placed on V by defining

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{k} \operatorname{tr}\left(A_i' B_i\right) = \operatorname{tr}(\mathbf{A}'\mathbf{B}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary elements of V. We let $\|\mathbf{A}\| = \langle \mathbf{A}, \mathbf{A} \rangle^{1/2}$ denote the standard norm corresponding to $\langle \cdot, \cdot \rangle$. The set $-K = \{\mathbf{B} \in Z^k: \mathbf{B}$ is isotonic w.r.t. $\lesssim\}$ is a convex cone in the space V. For comparison with the MLE problem, we again assume that the $n_i S_i$ are distributed independently as Wishart(n_i, Σ_i) matrices. Now, the solution to

$$\min_{\boldsymbol{\Sigma} \in -K} \left\| \mathbf{n}^{1/2}(\mathbf{S} - \boldsymbol{\Sigma}) \right\| \tag{4.1}$$

is guaranteed to exist uniquely [Luenberger (1969), page 51], where $\mathbf{n}^{1/2}\mathbf{B} = [n_1^{1/2}B_1, \dots, n_k^{1/2}B_k]'$. It is well known that the solution to (4.1) is uniquely characterized as the only element Σ^* in $-K$ such that

$$(4.2)\quad \operatorname{tr}\{[\mathbf{n}(\mathbf{S}-\Sigma^*)]'\Sigma^*\} = 0 \quad \text{and} \quad \operatorname{tr}\{[\mathbf{n}(\mathbf{S}-\Sigma^*)]'\Gamma\} \geq 0 \qquad \forall\, \Gamma \in K$$

[see Robertson, Wright and Dykstra (1988), page 17]. The MLE of Σ can be characterized by analogous properties based upon the duality relationship.

THEOREM 4.1. *The MLE of Σ subject to the constraints that Σ be isotonic with respect to $\lesssim$ is characterized as the unique element $\hat{\Sigma}$ in $-K$ such that*

$$(4.2)'\quad \operatorname{tr}\{[\mathbf{n}(\mathbf{S}-\hat{\Sigma})]'\hat{\Sigma}^{-1}\} = 0 \quad \textit{and} \quad \operatorname{tr}\{[\mathbf{n}(\mathbf{S}-\hat{\Sigma})]'\Gamma\} \geq 0 \qquad \forall\, \Gamma \in K.$$

PROOF. If $\hat{\Sigma}$ denotes the MLE, $\hat{\Sigma}^{-1}$ must be the solution to the primal problem (2.6). Moreover, by condition (d) of Theorem 2.1, $\frac{1}{2}\mathbf{n}(\hat{\Sigma}-\mathbf{S})$ must then be a solution to the dual problem (2.7). But, (4.2)′ is then just conditions (c) and (b) of Theorem 2.1. □

In the event that $p = 1$ so that only variances are involved in the ordering, (4.2) is equivalent to (4.2)′ [Robertson, Wright and Dykstra (1988), page 21] and the MLE and least squares solutions to (4.1) are identical. While for any p, the least squares estimator Σ^* has the desirable property that

$$(4.3)\qquad \|\mathbf{n}^{1/2}(\mathbf{S}-\Sigma)\| \geq \|\mathbf{n}^{1/2}(\Sigma^*-\Sigma)\|$$

if $\Sigma \in -K$ [Robertson, Wright and Dykstra (1988), Section 1.6], this is not necessarily true for the MLE $\hat{\Sigma}$. However, indications are that (4.3) usually holds for the MLE as well (see Section 5). A condition that will guarantee that $\hat{\Sigma}$ is at least as close [in the sense of (4.3)] to Σ as is $\mathbf{S}$ for every value of Σ in $-K$ is given in the following theorem.

THEOREM 4.2. *If $\|\mathbf{n}^{1/2}\mathbf{S}\| \geq \|\mathbf{n}^{1/2}\hat{\Sigma}\|$, then $\|\mathbf{n}^{1/2}(\mathbf{S}-\Sigma)\| \geq \|\mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\|$ for every $\Sigma \in -K$.*

PROOF. Note that

$$\begin{aligned}
\|\mathbf{n}^{1/2}(\mathbf{S}-\Sigma)\|^2 &= \|\mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}+\hat{\Sigma}-\Sigma)\|^2 \\
&= \|\mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma})\|^2 + \|\mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\|^2 \\
&\quad + 2\langle \mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}), \mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\rangle \\
&\geq \langle \mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}), \mathbf{n}^{1/2}\mathbf{S}\rangle - \langle \mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}), \mathbf{n}^{1/2}\hat{\Sigma}\rangle \\
&\quad + 2\langle \mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}), \mathbf{n}^{1/2}\hat{\Sigma}\rangle + \|\mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\|^2 \\
&= \langle \mathbf{n}^{1/2}(\mathbf{S}-\hat{\Sigma}), \mathbf{n}^{1/2}(\mathbf{S}+\hat{\Sigma})\rangle + \|\mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\|^2 \\
&= \|\mathbf{n}^{1/2}\mathbf{S}\|^2 - \|\mathbf{n}^{1/2}\hat{\Sigma}\|^2 + \|\mathbf{n}^{1/2}(\hat{\Sigma}-\Sigma)\|^2
\end{aligned}$$

from which the result easily follows. □

The previous condition can be easily checked if $\hat{\mathbf{\Sigma}}$ can be computed. Thus, if the norm of $\mathbf{n}^{1/2}\mathbf{S}$ exceeds the norm of $\mathbf{n}^{1/2}\hat{\mathbf{\Sigma}}$, we are guaranteed that $\hat{\mathbf{\Sigma}}$ is closer to the true $\mathbf{\Sigma}$ than is $\mathbf{S}$, even though $\mathbf{\Sigma}$ is unknown.

THE ALGORITHM FOR THE GENERAL ISOTONIC PROBLEM. The form of the solution to the dual problem given in (2.7) suggests an algorithm for computing the solution to the general isotonic problem which depends only upon the pairwise procedure discussed in Section 3. The key is that the solutions to the dual pairwise problems can be conceptually combined with the data $A_1, \dots, A_k$ in constructing the MLE's. Moreover, the primal-dual machinery given in Theorem 2.1 provides an elegant proof that the proposed procedure must converge correctly.

Let $G = \{(i_s, j_s)\colon s = 1, \dots, m\}$ denote a minimal set of pairwise order restrictions which generate the partial order, where by the sth constraint we will mean the antitonic constraint $\Lambda_{i_s} \geq \Lambda_{j_s}$. The vectors of matrices $\mathbf{A}^{r,s}$, $\hat{\mathbf{\Lambda}}^{r,s}$ and $\hat{\mathbf{\Psi}}^{r,s}$ will represent the data adjusted for past solutions, the primal solution and dual solution associated with the rth iteration and the sth constraint, respectively. The algorithm can now be stated as follows:

0. Initialize by setting $\hat{\mathbf{\Psi}}^{0,s} = 0$ for $s = 1, \dots, m$, $\mathbf{A}^{0,m} = \mathbf{A}$, $r = 1$ and $s = 1$.
1. Let

$$\mathbf{A}^{r,s} = \begin{cases} \mathbf{A}^{r-1,m} + 2\hat{\mathbf{\Psi}}^{r-1,m} - 2\hat{\mathbf{\Psi}}^{r-1,s}, & s = 1, \\ \mathbf{A}^{r,s-1} + 2\hat{\mathbf{\Psi}}^{r,s-1} - 2\hat{\mathbf{\Psi}}^{r-1,s}, & s > 1. \end{cases}$$

Solve the sth pairwise constraint using the pseudodata $\mathbf{A}^{r,s}$.

2. Let $s = s + 1$. If $s < m + 1$, then go to 1.
3. Check for convergence. If convergence criterion is not met, set $s = 1$ and $r = r + 1$ and go to 1.

Thus, with each iteration, the data is being updated by the duality relationship suggested by the Fenchel duality theorem. The new $\mathbf{A}^{r,s}$ is updated from the previous pseudodata by adding twice the latest dual solution and subtracting twice the previous dual solution to the current constraint. The addition provides the new pairwise solution and the subtraction allows for the possibility that because of the other adjustments, the data for the current constraint may have been overadjusted. One may define convergence either in terms of the log-likelihood function or in terms of the inner product used for the dual space. In Section 5, we define convergence to have occurred when $\|\mathbf{n}^{1/2}(\hat{\mathbf{\Sigma}}^{l,m} - \hat{\mathbf{\Sigma}}^{l-1,m})\| \leq 0.1$. Clearly, this algorithm only has utility if it converges with reasonable speed to the MLE. Fortunately, the convergence rates appear to be quite fast and the primal-dual format leads to an elegant proof of the convergence.

THEOREM 4.3. *The previous algorithm for the solution to the general isotonic problem converges to the unique MLE and the convergence is monotonic in the log-likelihood function.*

PROOF. See the Appendix. □

We have now developed an algorithm for the maximum likelihood estimation of a set of covariance matrices. The concluding section is devoted to studying the speed of the algorithm and producing the REML estimates from the data in (1.5). A possible convergence criterion is also given.

5. Multivariate variance components models. As demonstrated in Section 1, multivariate variance components models for balanced data form a set of well-used models for analyzing multivariate data [see, additionally, Dahm, Melton and Fuller (1983) or Gregory, Swiger, Sumption, Koch, Ingalls, Rowden and Rothlisberger (1966)]. However, rather than discuss the application of the previous results to the general multivariate variance component model, we thought it would be more insightful to demonstrate the results for two simple models.

5.1. *The two-factor random effects model without interaction.* The two-factor model without interaction and its associated sufficient statistics are given in (1.2) and (1.3). The random matrices A_A, A_B, A_E are distributed $\text{Wishart}_p(n_A = A - 1, \text{EMS}_A = \Sigma_E + BN\Sigma_A)$, $\text{Wishart}_p(n_B = B - 1, \text{EMS}_B = \Sigma_E + AN\Sigma_B)$, and $\text{Wishart}_p(n_E = ABN - A - B + 1, \text{EMS}_E = \Sigma_E)$, respectively. Note that all three Wishart matrices are functions of contrasts of the data and are location invariant. Thus, estimating Σ_A, Σ_B, and Σ_E by maximizing the likelihood associated with the three independent Wishart matrices produces restricted maximum likelihood (REML) estimates for the multivariate variance components model [see Patterson and Thompson, (1974)].

By defining $\Sigma_1 = \text{EMS}_E$, $\Sigma_2 = \text{EMS}_B$ and $\Sigma_3 = \text{EMS}_A$ and letting n_i be the appropriate degrees of freedom, it is clear that the random matrices for (1.2) produce an estimation problem within the framework of the previous sections. The order restrictions correspond to a simple tree structure and the set of minimal pairwise orderings is $G = \{(1,2),(1,3)\}$. It is interesting to note that for this ordering, each cycle of the general algorithm in Section 4 will yield a set of estimates which is in the parameter space, but only the limiting value maximizes the likelihood.

With the use of an iterative estimation procedure, some information about the speed of convergence is needed before one can be comfortable using the procedure. A small simulation study was run to determine the number of cycles needed for convergence for (1.2). Six different sets of covariance matrices (see Table 1) and three sets of model sizes were studied for each model. With each of the 18 combinations, 1000 independent sets of independent Wishart matrices were generated and the general isotonic algorithm was used to estimate the EMS's under the defined order restriction. The algorithm was considered to have converged when $\|\mathbf{n}^{1/2}(\hat{\Sigma}^{l,m} - \hat{\Sigma}^{l-1,m})\| \leq 0.1$. The results of this study are summarized in Table 2. As can be seen from the table, in most cases the number of times that the Wishart matrices do not conform to

TABLE 1
Covariance patterns Tables 2 and 3

Covariance pattern	Σ_A	Σ_B	Σ_E
1	$0.2I$	I	$0.4I$
2	$0.3I + 0.1J$	0	I
3	$0.3I + 0.1J$	$I + 0.5J$	$0.5I + 0.3J$
4	$\sigma_{ij} = 0.5^{\lvert i-j\rvert}$	$\sigma_{ij} = 0.7^{\lvert i-j\rvert}$	I
5	$\sigma_{ij} = 0.7^{\lvert i-j\rvert}$	$\sigma_{ij} = -0.5^{\lvert i-j\rvert}$	$\sigma_{ij} = 0.9^{\lvert i-j\rvert}$
6	$\sigma_{ij} = 0.2^{\lvert i-j\rvert}$	$\sigma_{ij} = -0.1^{\lvert i-j\rvert}$	$\sigma_{ij} = 2\delta_{[i=j]} + 0.9^{\lvert i-j\rvert}$

Note: I represents the identity matrix, J is a matrix of all ones and δ is the indicator function.

the order restriction is quite large. The only time this is not true is when all three degrees of freedoms are large, so that estimation of the covariance matrices is quite accurate. The last two columns of the table show that when the estimators disagree, the REML estimate is nearly always closer to the true value of the parameters than is S. Note that when $\mathbf{A}$ is in the parameter space, $\mathbf{S} = \hat{\Sigma}$. Thus, $\hat{\Sigma}$ can only be closer when $\mathbf{A}$ does not conform to $\lesssim$. However, Theorem 4.2 is not always a reliable indicator of when that is the case. More importantly, the distribution of the number of cycles needed for

TABLE 2
Results from the simulation study of convergence rates for two-factor random effects model without interaction (1000 *independent trials for each model size, covariance pattern combination*)

Model sizes				Covariance	No. of times	No. of cycles to converge				$\hat{\Sigma}$ closer	Theorem 4.2
a	b	n	p	pattern	A does not conform to $\lesssim$	1	2	3	4 or more	than S	satisfied
5	5	3	3	1	558	39	519	0	0	558	310
				2	999	0	939	60	0	997	969
				3	574	25	547	2	0	570	69
				4	736	15	709	12	0	734	351
				5	577	11	532	34	0	464	47
				6	865	4	693	168	0	864	334
5	15	1	4	1	875	29	846	0	0	875	489
				2	1000	0	417	583	0	1000	962
				3	878	10	866	2	0	852	82
				4	909	3	776	130	0	908	362
				5	764	13	586	165	0	586	56
				6	967	3	575	389	0	956	329
20	20	1	5	1	2	0	2	0	0	2	0
				2	1000	0	1000	0	0	1000	1000
				3	1	0	1	0	0	1	0
				4	1	1	3	0	0	4	1
				5	43	2	41	0	0	42	5
				6	31	0	31	0	0	31	1

convergence is such that all sets of data allowed convergence in less than five cycles. Thus, the general isotonic algorithm is a viable estimation procedure for this model.

5.2. *The two-factor random effects nested model.* The second model we consider is the two-factor random effects nested model, which is the model proposed for the data from Calvin and Sedransk (1991). The set of sufficient statistics for the model [see (1.4)] is $\{\overline{\mathbf{Y}}_{\ldots}, A_A, A_B, A_E\}$ where $\overline{\mathbf{Y}}_{\ldots}$ is the grand mean and A_A, A_B and A_E are the standard sums-of-squares and cross-products matrices. The random matrices A_A, A_B, A_E are distributed $\mathrm{Wishart}_p(n_A = A - 1,\ \mathrm{EMS}_A = \Sigma_E + N\Sigma_B + BN\Sigma_A)$, $\mathrm{Wishart}_p(n_B = A(B-1),\ \mathrm{EMS}_B = \Sigma_E + N\Sigma_B)$ and $\mathrm{Wishart}_p(n_E = AB(N-1),\ \mathrm{EMS}_E = \Sigma_E)$, respectively. The order restriction corresponds to a simple linear ordering where $\Sigma_1 = \mathrm{EMS}_E$, $\Sigma_2 = \mathrm{EMS}_B$ and $\Sigma_3 = \mathrm{EMS}_A$ and $G = \{(1,2),(2,3)\}$. A study of the convergence rate for this model was also performed. The simulation procedure used for (1.2) was also used for (1.4) and the results are in Table 3. Again the results show that a large proportion of the Wishart matrices do not conform to the order restriction and the general isotonic algorithm generally converges in a small number of cycles. Thus, it appears that the isotonic procedure provides a computationally inexpensive procedure for accurately computing restricted maximum likelihood estimates of covariance matrices in the multivariate variance component model setting.

TABLE 3

Results from the simulation study of convergence rates for two-factor random effects nested model (1000 independent trials for each model size, covariance pattern combination)

Model sizes				**Covariance**	**No. of times**	**No. of cycles to converge**				**$\hat{\Sigma}$ closer**	**Theorem 4.2**
a	***b***	***n***	***p***	**pattern**	**A does not conform to $\leq$**	**1**	**2**	**3**	**4 or more**	**than S**	**satisfied**
5	5	3	3	1	947	2	945	0	0	943	822
				2	995	2	570	422	1	991	945
				3	939	2	937	0	0	902	577
				4	805	8	733	63	1	764	204
				5	897	4	825	60	8	851	441
				6	905	5	664	194	42	904	588
10	5	2	4	1	976	1	975	0	0	976	891
				2	1000	0	656	344	0	1000	993
				3	914	2	912	0	0	899	436
				4	757	24	658	75	0	742	181
				5	864	6	784	74	0	842	362
				6	933	2	615	270	46	933	574
20	20	2	5	1	126	4	122	0	0	126	96
				2	1000	0	1000	0	0	1000	1000
				3	39	0	39	0	0	39	1
				4	7	0	7	0	0	7	4
				5	128	8	120	0	0	128	23
				6	32	1	31	0	0	32	14

5.3. *Analysis of the sample data from Section* 1. Display (1.5) contains three mean square matrices from a two-factor random effects nested model which do not fall in the parameter space. These matrices were used to construct the **A** vector and along with the appropriate degrees of freedom ($\mathrm{df}_A = 6$, $\mathrm{df}_B = 7$, $\mathrm{df}_E = 14$) were used as input to the algorithm in Section 4. The resulting REML estimates of the covariance matrices are

$$\hat{\Sigma}_A = \begin{bmatrix} 512.24 & 343.25 \\ 343.25 & 230.01 \end{bmatrix} \qquad \hat{\Sigma}_B = \begin{bmatrix} 58.76 & 66.96 \\ 66.96 & 449.96 \end{bmatrix}$$

$$\hat{\Sigma}_E = \begin{bmatrix} 255.57 & 47.96 \\ 47.96 & 63.49 \end{bmatrix}.$$

The major adjustment is to the diagonal of $\hat{\Sigma}_A$. We also note that since the initial estimates were not in the parameter space, the REML estimates are on a boundary and the rank of $\hat{\Sigma}_A$ is one.

APPENDIX

PROOF OF THEOREM 4.3. Let us first show the monotonicity property

$$\ell(\hat{\mathbf{\Lambda}}^{l,j-1}, \mathbf{A}^{l,j-1}) \geq \ell(\hat{\mathbf{\Lambda}}^{l,j}, \mathbf{A}^{l,j}) \geq \cdots \geq \ell(\hat{\mathbf{\Lambda}}^{l,m}, \mathbf{A}^{l,m})$$
$$\geq \ell(\hat{\mathbf{\Lambda}}^{l+1,1}, \mathbf{A}^{l+1,1}) \geq \cdots .$$

Recall the relationship $2\hat{\mathbf{\Psi}}^{l,j} = \mathbf{n}\hat{\mathbf{\Sigma}}^{l,j} - \mathbf{A}^{l,j}$, or equivalently, $\mathbf{A}^{l,j} + 2\hat{\mathbf{\Psi}}^{l,j} = \mathbf{n}\hat{\mathbf{\Sigma}}^{l,j} = \mathbf{n}\hat{\mathbf{\Lambda}}^{l,j^{-1}}$. Now, note that if $j > 1$ and K_j is the cone in $V \cap Z^k$ where the jth constraint is satisfied,

$$\begin{aligned}
\ell(\hat{\mathbf{\Lambda}}^{l,j}, \mathbf{A}^{l,j}) &= \sup_{\mathbf{\Lambda} \in K_j} \ell(\mathbf{\Lambda}, \mathbf{A}^{l,j}) = -\sup_{\mathbf{\Psi} \in K_j^*} \ell^*(\mathbf{\Psi}, \mathbf{A}^{l,j}) \\
&= -\sup_{\mathbf{\Psi} \in K_j^*} \sum_{i=1}^{k} \frac{n_i}{2} \ln\left| A_i^{l,j} + 2\psi_i \right| + c^* \\
&= -\sup_{\mathbf{\Psi} \in K_j^*} \sum_{i=1}^{k} \frac{n_i}{2} \ln\left| A_i^{l,j-1} + 2\hat{\psi}_i^{l,j-1} - 2\hat{\psi}_i^{l-1,j} + 2\psi_i \right| + c^* \\
&\leq -\sum_{i=1}^{k} \frac{n_i}{2} \ln\left| A_i^{l,j-1} + 2\hat{\psi}_i^{l,j-1} \right| + c^* \quad \left(\text{since } \mathbf{\Psi}^{l-1,j} \in K_j^*\right) \\
&= -\ell^*(\hat{\mathbf{\Psi}}^{l,j-1}, \mathbf{A}^{l,j-1}) = -\sup_{\mathbf{\Psi} \in K_{j-1}^*} \ell^*(\mathbf{\Psi}, \mathbf{A}^{l,j-1}) \\
&= \sup_{\mathbf{\Lambda} \in K_{j-1}} \ell(\mathbf{\Lambda}, \mathbf{A}^{l,j-1}) = \ell(\hat{\mathbf{\Lambda}}^{l,j-1}, \mathbf{A}^{l,j-1}).
\end{aligned}$$

The case for $j = 1$ works similarly. Thus, the convergence is guaranteed to be monotone in the log-likelihood, which will hopefully lead to quick convergence.

Now, to prove convergence, consider a maximal index for the antitonic cone, say i_0. That is, $(i_0, j_0) \in G$ for some j_0, but $(j, i_0) \notin G$ for any j. Then one must have $\hat{\psi}_{i_0}^{l,j} \leq A_{i_0}$ (in the Löwner sense), otherwise $\hat{\mathbf{\Psi}}^{l,j}$ would not solve the dual problem. In a similar fashion, one can argue that all the $\hat{\psi}_i^{l,j}$ are uniformly bounded in the Löwner sense, which implies that all the elements of the $\hat{\psi}_i^{l,j}$ are uniformly bounded. From this, it easily follows that all the elements of the $A_i^{l,j}$ and $\hat{\Sigma}_i^{l,j}$ are also uniformly bounded. Finally, note that $\ell(\hat{\mathbf{\Lambda}}^{l,j}, \mathbf{A}^{l,j})$ is bounded below, or equivalently that $\ell^*(\hat{\mathbf{\Psi}}^{l,j}, \mathbf{A}^{l,j})$ is bounded above. This easily follows since $\ln|A_n|$ is bounded above, if all elements of A_n are bounded.

From this, one can argue that if $\{l_h\}$ is any subsequence of positive integers, then there exists a subsubsequence $\{l_{h'}\}$ such that

$$\hat{\mathbf{\Psi}}^{l_{h'},j} \to \mathbf{\Psi}^j \quad \text{and} \quad \hat{\mathbf{\Psi}}^{l_{h'}-1,j} - \hat{\mathbf{\Psi}}^{l_{h'},j} \to 0 \quad \text{as } l_{h'} \to \infty$$

for some vector of matrices, $\mathbf{\Psi}^j$, $j = 1, \dots, m$. Now, since

$$\mathbf{n}\hat{\mathbf{\Sigma}}^{l,j} = \mathbf{A} + 2\hat{\mathbf{\Psi}}^{l,1} + 2\hat{\mathbf{\Psi}}^{l,2} + \cdots + 2\hat{\mathbf{\Psi}}^{l,j} + 2\hat{\mathbf{\Psi}}^{l-1,j+1} + \cdots + 2\hat{\mathbf{\Psi}}^{l-1,m},$$

it follows that

$$\mathbf{n}\hat{\mathbf{\Sigma}}^{l_{h'},j} - \mathbf{n}\hat{\mathbf{\Sigma}}^{l_{h'},m} \to 0 \quad \text{as } l_{h'} \to \infty \; \forall j.$$

However, $\hat{\mathbf{\Sigma}}^{l_{h'},j}$ is in $-K_j$ and

$$\mathbf{n}\hat{\mathbf{\Sigma}}^{l_{h'},j} \to \mathbf{A} + 2\mathbf{\Psi}^1 + \cdots + 2\mathbf{\Psi}^m = \mathbf{n}\hat{\mathbf{\Sigma}}, \qquad \forall j,$$

so that $\hat{\mathbf{\Sigma}} \in \bigcap_{j=1}^m -K_j = -K$, since the K_j are closed. Thus, $\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Sigma}}^{-1} \in K$. Similarly,

$$\hat{\mathbf{\Psi}} = \sum_{j=1}^m \mathbf{\Psi}^j = \sum_{j=1}^m \lim \mathbf{\Psi}^{l_{h'},j} \in K^*,$$

by closure properties. Finally, note that

$$0 = \operatorname{tr}\big((\hat{\mathbf{\Psi}}^{l_{h'},j})'\hat{\mathbf{\Lambda}}^{l_{h'},j}\big) \to \operatorname{tr}\big((\mathbf{\Psi}^j)'\hat{\mathbf{\Lambda}}\big),$$

so that

$$0 = \sum_{j=1}^m \operatorname{tr}\big((\mathbf{\Psi}^j)'\hat{\mathbf{\Lambda}}\big) = \operatorname{tr}\left(\left(\sum_{j=1}^m \mathbf{\Psi}^j\right)'\hat{\mathbf{\Lambda}}\right) = \operatorname{tr}(\hat{\mathbf{\Psi}}'\hat{\mathbf{\Lambda}}).$$

Since $\mathbf{n}\hat{\mathbf{\Lambda}}^{-1} = n\hat{\mathbf{\Sigma}} = \mathbf{A} + 2\sum_{j=1}^m \mathbf{\Psi}^j = \mathbf{A} + 2\hat{\mathbf{\Psi}}$, $-\hat{\mathbf{\Psi}}$ is a subgradient of $-\ell(\mathbf{\Lambda}, \mathbf{A})$ at the point $\hat{\mathbf{\Lambda}}$ (see 2.8). Thus all four conditions of Theorem 2.1 hold for $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Psi}}$ and we know that $\hat{\mathbf{\Lambda}}$ must solve the primal problem and $\hat{\mathbf{\Psi}}$ must solve the dual problem. The solution $\hat{\mathbf{\Lambda}}$ is unique by the strict concavity of the objective function in the appropriate region. While the solution to the dual problem is not unique, $\hat{\mathbf{\Psi}}$ is the unique vector of symmetric matrices which solves the dual problem. Moreover, since every subsequence of $\{\mathbf{\Lambda}^{l,j}\}$ contains a sub-subsequence converging to $\hat{\mathbf{\Lambda}}$, the sequence $\{\mathbf{\Lambda}^{l,j}\}$ must converge to $\hat{\mathbf{\Lambda}}$. □

Acknowledgments. We would like to thank the Associate Editor and the referees for their careful reading of this paper and their many constructive comments which helped to clarify its presentation.

REFERENCES

AMEMIYA, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statist.* **39** 112–117.

ANDERSON, B. M., ANDERSON, T. W. and OLKIN, I. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *Ann. Statist.* **14** 405–417.

ARNOLD, S. (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147.

BHARGAVA, A. K. and DISCH, D. (1982). Exact probabilities of obtaining estimated non-positive definite between group covariance matrices. *J. Statist. Comput. Simulation* **15** 27–32.

CALVIN, J. A. and SEDRANSK, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *J. Amer. Statist. Assoc.* To appear.

DAHM, P. F., MELTON, B. E. and FULLER, W. A. (1983). Generalized least squares estimation of a genotypic covariance. *Biometrics* **39** 587–597.

DWYER, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. *J. Amer. Statist. Assoc.* **62** 607–625.

DYKSTRA, R. L. and LEMKE, J. H. (1988). Duality of I projections and maximum likelihood estimation for log-linear models under cone constraints. *J. Amer. Statist. Assoc.* **83** 546–554.

FARRELL, R. H. (1985). *Multivariate Calculation: Use of the Continuous Groups*. Springer, New York.

GREGORY, K. E., SWIGER, L. A., SUMPTION, L. J., KOCH, R. M., INGALLS, J. E., ROWDEN, W. W. and ROTHLISBERGER, J. A. (1966). Heterosis effects on carcass of beef steers. *J. Animal Sci.* **25** 311–322.

HARVILLE, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340.

HARVILLE, D. and CALLANAN, T. (1990). Computational aspects of likelihood-based inference for variance components. In *Advances in Statistical Methods for Genetic Improvement of Livestock* (D. Gianola and K. Hammonds, eds.) 136–176. Springer, New York.

HENDERSON, C. R. (1953). Estimation of variance and covariance components. *Biometrics* **9** 226–252.

HERBACH, L. H. (1959). Properties of model II-type analysis of variance tests, A: Optimum nature of the F-test for model II in the balanced case. *Ann. Math. Statist.* **30** 939–959.

HILL, W. G. and THOMPSON, R. (1978). Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* **34** 429–439.

KLOTZ, J. H. and PUTTER, J. (1969). Maximum likelihood estimation of multivariate covariance components for the balanced one-way layout. *Ann. Math. Statist.* **40** 1100–1105.

LÖWNER, K. (1934). Über monotone matrixfunktionen. *Math. Z.* **38** 177–216.

LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

MÜLLER-FUNK, V., PUKELSHEIM, F. and WITTING, A. (1985). On the duality between locally optimal tests and optimal experimental design. *Linear Algebra Appl.* **67** 19–34.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554.

PATTERSON, H. D. and THOMPSON, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings Eighth International Biometric Conference* (L. C. A. Corsten and T. Postelnica, eds) 197–207. Biometric Soc., Washington, D.C.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4** 339–364.

PUKELSHEIM, F. (1981). Linear models and convex geometry: Aspects of non-negative variance estimation. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 271–286.

RAO, C. R. (1971a). Estimation of variance covariance components—MINQUE theory. *J. Multivariate Anal.* **1** 257–276.

RAO, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Anal* **1** 445–456.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

RAO, C. R. and KLEFFE, J. (1988). *Estimation of Variance Components and Applications*. North-Holland, Amsterdam.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

SEARLE, S. R. (1971). *Linear Models*. Wiley, New York.

THOMPSON, W. A., JR. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242