

## INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.
2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.
3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.\*
4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.\*

\*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.

UMI University  
Microfilms  
International



8604452

**Callanan, Terrance Patrick**

RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION OF VARIANCE  
COMPONENTS: COMPUTATIONAL ASPECTS

*Iowa State University*

PH.D. 1985

**University  
Microfilms  
International** 300 N. Zeeb Road, Ann Arbor, MI 48106



Restricted maximum likelihood estimation  
of variance components: Computational aspects

by

Terrance Patrick Callanan

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of the  
Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University  
Ames, Iowa

1985

## TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
A. The General Problem	1
B. Models and Basic Notation	1
C. An Example	7
D. Estimation of $\sigma$ - Applications	8
E. Restricted Maximum Likelihood Estimation - Motivation	9
F. Restricted Maximum Likelihood Estimation - Discussion	11
G. A Preview of Subsequent Chapters	15
II. THE RESTRICTED LOG-LIKELIHOOD FUNCTION AND ITS DERIVATIVES	18
A. The Log-Likelihood Function	18
B. Derivatives of the Log-Likelihood Function	32
1. General case	32
2. Special case	35
C. Alternative Parameterization	37
1. General case	39
2. Special case	41
D. Some Computational Considerations	43
1. Diagonalization versus inversion	43
2. Choice of mixed-model equations	52
III. ITERATIVE ALGORITHMS FOR COMPUTING REML ESTIMATES	54
A. The EM Algorithm	54
1. General description and application to the mixed linear model	55

	Page
2. Computation of the iterates	60
3. Alternative implementation	68
4. Nonnegativity constraints	73
B. The Newton-Raphson Method	75
1. Description of the method	76
2. Linearization	84
3. The concentrated log-likelihood function	98
a. $\underline{\gamma}$ -parameterization	98
b. Alternative parameterization	118
4. Linearization as applied to the concentrated log-likelihood function	122
a. $\underline{\psi}$ -parameterization	123
b. The $\underline{\gamma}$ -parameterization	
C. Method of Scoring	137
D. Method of Successive Approximations	141
1. Applications	142
2. Nonnegativity constraints	149
E. Summary Table	149
F. Special Case: $c=1$	151
IV. MODIFICATIONS TO ACCOMMODATE PARAMETRIC CONSTRAINTS	160
A. Linear Inequality Constraints	165
1. Active set methods	165
a. Definitions and notation	165
b. Motivation and general strategy	166
c. Search direction	168
d. Steplength	169

	Page
e. Updating the working set	170
f. Summary	170
g. Variations	171
h. A special case - simple bounds	176
2. Bertsekas' method	177
a. Simple nonnegativity constraints	177
b. Simple bounds	181
c. General linear constraints	182
3. Transformations	188
a. Squared-variable transformation	189
b. Drawbacks	190
B. Nonlinear Constraints	191
1. Barrier function methods	192
2. Sequential quadratic programming (SQP) methods	195
V. EXTENSIONS TO CORRELATED RANDOM EFFECTS	198
A. Case 1: $A_1, \dots, A_c$ Positive Definite	198
1. Derivatives of $L_1(\underline{\gamma}; \underline{y})$	198
2. Derivatives of $L_v(\underline{g}; \underline{y})$	212
3. Derivatives of $L_c(\underline{\gamma}^+; \underline{y})$	215
4. Modification of algorithms	221
a. The Newton-Raphson and Method of Scoring algorithms	221
b. EM algorithm	221
c. Method of Successive Approximations	224

	Page
B. Case 2: $A_1, \dots, A_c$ Nonnegative Definite	225
1. Derivatives of $L_1(\underline{y}; \underline{y})$	227
2. Derivatives of $L_v(\underline{\sigma}; \underline{y})$	245
3. Derivatives of $L_c(\underline{y}^+; \underline{y})$	264
4. Algorithm modifications	275
a. The Newton-Raphson and Method of Scoring algorithms	275
b. EM algorithm	276
c. Method of Successive Approximations	284
VI. NUMERICAL RESULTS	288
A. Diagonalization Versus Inversion	288
B. Algorithm Implementation	290
C. The Data Sets	294
D. Numerical Results	301
E. Discussion	331
VII. BIBLIOGRAPHY	337
VIII. ACKNOWLEDGEMENTS	342
IX. APPENDIX	343
A. Derivation of the Partial Derivatives of the Log-Likelihood Function With Respect to $\sigma_1^2, \dots, \sigma_{c+1}^2$	343
1. General case	343
2. Special case	354
B. Derivation of Simplified Expressions for the Special Case $c=1$	363

## I. INTRODUCTION

### A. The General Problem

As a consequence of various theoretical developments, and of improvements in computing strategies, restricted maximum likelihood (REML) estimation has become a viable procedure for estimating the variance components in mixed linear models. Many of these developments and innovations were reviewed by Harville (1977). The increased interest in REML estimation was noted by Henderson (1980), who observed that it has become one of the most commonly used procedures for estimating variance and covariance components. However, he also indicated that the computation of REML estimates is still quite difficult.

The purpose of this dissertation is to develop improved procedures for computing REML estimates of variance components.

In the remainder of this chapter, we introduce various linear models, define and discuss REML estimation, and provide a preview of subsequent chapters.

### B. Models and Basic Notation

Suppose that  $\underline{y}$  is an observable  $n \times 1$  random vector and that

$$\underline{y} \stackrel{d}{\sim} N_n(X\underline{\alpha}, \sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i Z_i Z_i']). \quad (1.1)$$

That is, the distribution of  $\underline{y}$  is  $n$ -dimensional multivariate normal with mean vector  $X\underline{\alpha}$  and variance-covariance matrix  $\sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i Z_i Z_i']$ .

Here,

$I_n$  is an  $n \times n$  identity matrix,

$X$  is an  $n \times p$  known matrix of rank  $p^*$ ,

$\underline{\alpha}$  is a  $p \times 1$  unknown vector of unobservable parameters (sometimes called fixed effects),

$Z_i$  is an  $n \times q_i$  known matrix ( $i = 1, \dots, c$ ), and

$\gamma_1, \dots, \gamma_c$  and  $\sigma_{c+1}^2$  are unknown, unobservable parameters.

Let  $q = \sum_{i=1}^c q_i$ , and define

$$D \equiv \sigma_{c+1}^2 \text{diag}(\gamma_1 I_{q_1}, \dots, \gamma_c I_{q_c}),$$

$$Z \equiv (Z_1, \dots, Z_c),$$

$$R \equiv \sigma_{c+1}^2 I_n,$$

$$V_i \equiv \sigma_{c+1}^2 [I_n + \sum_{j=i}^c \gamma_j Z_j Z_j'] \quad (i = 1, \dots, c),$$

$$V \equiv V_1.$$

Note that  $\underline{y} \stackrel{d}{\sim} N_n(X\underline{\alpha}, V)$ , with  $V = R + ZDZ'$ . Let  $\gamma_{c+1} = \sigma_{c+1}^2$ , and define

$$\underline{\gamma} \equiv (\gamma_1, \dots, \gamma_c, \gamma_{c+1})',$$

$$\sigma_i^2 \equiv \sigma_{c+1}^2 \gamma_i \quad (i = 1, \dots, c),$$

$$\underline{\sigma} \equiv (\sigma_1^2, \dots, \sigma_{c+1}^2)'.$$

An alternative to parameterizing model (1.1) in terms of  $\underline{\alpha}$  and  $\underline{\gamma}$  is to

parameterize it in terms of  $\underline{\alpha}$  and  $\underline{\sigma}$ , in which case,  $V_i = \sigma_{c+1}^2 I_n + \sum_{j=i}^c \sigma_j^2 Z_j Z_j'$  ( $i = 1, \dots, c$ ).

In what follows, we assume that the vector of fixed effects  $\underline{\alpha}$  can assume any value in  $p$ -dimensional Euclidean space  $\mathbb{E}^p$ . We consider two possible parameter spaces for  $\underline{\gamma}$ :

$$\Omega_1 \equiv \{\underline{\gamma} : \sigma_{c+1}^2 > 0, \gamma_i \geq 0 \ (i = 1, \dots, c)\}, \text{ and}$$

$$\Omega_1^* \equiv \{\underline{\gamma} : \sigma_{c+1}^2 > 0, V_i \text{ is a positive definite matrix } (i = 1, \dots, c)\}.$$

Note that  $\Omega_1 \subset \Omega_1^*$ . The corresponding parameter spaces for  $\underline{\sigma}$  are

$$\Omega_2 \equiv \{\underline{\sigma} : \sigma_{c+1}^2 > 0, \sigma_i^2 \geq 0 \ (i = 1, \dots, c)\}, \text{ and}$$

$$\Omega_2^* \equiv \{\underline{\sigma} : \sigma_{c+1}^2 > 0, V_i \text{ is a positive definite matrix } (i = 1, \dots, c)\},$$

respectively.

If the parameter space for  $\underline{\sigma}$  is taken to be  $\Omega_2$ , then model (1.1) is equivalent to the model

$$\underline{y} = X\underline{\alpha} + \sum_{i=1}^c Z_i \underline{b}_i + \underline{e} = X\underline{\alpha} + Z\underline{b} + \underline{e}, \quad (1.2)$$

where

$\underline{b}_i$  is a  $q_i \times 1$  unobservable random vector such that  $\underline{b}_i \stackrel{d}{\sim} N_{q_i}(\underline{0}, \sigma_i^2 I_{q_i})$

with  $\sigma_i^2 \geq 0 \ (i = 1, \dots, c)$ , and

$\underline{e}$  is an  $n \times 1$  vector of unobservable random errors such that

$$\underline{e} \stackrel{d}{\sim} N_n(\underline{0}, \sigma_{c+1}^2 I_n),$$

and where  $\underline{b}_1, \dots, \underline{b}_c, \underline{e}$  are distributed independently. The equivalence is in the sense that both models generate the same family of distributions for  $\underline{y}$ . Under the assumptions of model (1.2), the distribution of  $\underline{y}$  is multivariate normal with  $E(\underline{y}) = X\underline{\alpha}$ , and  $\text{Var}(\underline{y}) = \sigma_{c+1}^2 \underline{I}_n + \sum_{i=1}^c \sigma_i^2 \underline{Z}_i \underline{Z}_i'$ , the same as under model (1.1).

Let  $\underline{b} = (\underline{b}_1', \dots, \underline{b}_c')'$ , and, for convenience, define  $D = \text{Var}(\underline{b}) = \text{diag}(\sigma_1^2 \underline{I}_{q_1}, \dots, \sigma_c^2 \underline{I}_{q_c})$ ,  $R = \text{Var}(\underline{e}) = \sigma_{c+1}^2 \underline{I}_n$ , and  $V = \text{Var}(\underline{y}) = R + \underline{Z} D \underline{Z}'$ . Under model (1.2), the parameters  $\sigma_1^2, \dots, \sigma_c^2$  are interpretable as variances and the quantities  $\gamma_1 = \sigma_1^2 / \sigma_{c+1}^2, \dots, \gamma_c = \sigma_c^2 / \sigma_{c+1}^2$  are interpretable as variance ratios. As is customary, we will refer to  $\sigma_1^2, \dots, \sigma_{c+1}^2$  as variance components and to model (1.2) as the general mixed linear model. The parameter spaces  $\Omega_1^*$  and  $\Omega_2^*$  include values of  $\underline{\gamma}$  and  $\underline{\sigma}$  that have negative elements. Harville and Fenech (1985) noted that for the case  $c = 1$ ,  $\Omega_1^* = \{\underline{\gamma} : \sigma_2^2 > 0, \gamma_1 > -1/\lambda^*\}$ , where  $\lambda^*$  is the largest characteristic root of  $\underline{Z}_1 \underline{Z}_1'$ . (Since  $\underline{Z}_1 \underline{Z}_1'$  is a nonnegative definite matrix of rank  $\geq 1$ ,  $-1/\lambda^* < 0$ .)

There is a multivariate generalization of model (1.2). Suppose that one or more of  $\bar{c}$  traits is measured on each of  $N$  individuals. Letting  $n^{(i)}$  represent the number of individuals on which the  $i$ -th trait is measured ( $i = 1, \dots, \bar{c}$ ), and letting  $n = \sum_{i=1}^{\bar{c}} n^{(i)}$ , take  $\underline{y}$  to be the  $n \times 1$  vector  $(\underline{y}^{(1)'} , \dots, \underline{y}^{(\bar{c})'})'$ , where  $\underline{y}^{(k)}$  is the  $n^{(k)} \times 1$  vector of observations on the  $k$ -th trait ( $k = 1, \dots, \bar{c}$ ). Suppose that there exists a separate mixed linear model

$$\underline{y}^{(k)} = \underline{X}^{(k)} \underline{\alpha}^{(k)} + \sum_{i=1}^{\bar{c}^{(k)}} \underline{Z}_i^{(k)} \underline{b}_i^{(k)} + \underline{e}^{(k)} = \underline{X}^{(k)} \underline{\alpha}^{(k)} + \underline{Z}^{(k)} \underline{b}^{(k)} + \underline{e}^{(k)}$$

for each trait ( $k = 1, \dots, \bar{c}$ ), where

$\underline{X}^{(k)}$  is an  $n^{(k)} \times p^{(k)}$  known matrix,

$\underline{\alpha}^{(k)}$  is a  $p^{(k)} \times 1$  unknown vector of unobservable parameters,

$\underline{Z}_i^{(k)}$  is an  $n^{(k)} \times q_i^{(k)}$  known matrix ( $i = 1, \dots, \bar{c}^{(k)}$ ),

$$\underline{Z}^{(k)} = (\underline{Z}_1^{(k)}, \dots, \underline{Z}_{\bar{c}^{(k)}}^{(k)}),$$

$\underline{b}_i^{(k)}$  is a  $q_i^{(k)} \times 1$  unobservable random vector such that

$$\underline{b}_i^{(k)} \stackrel{d}{\sim} N_{q_i^{(k)}}(0, \sigma_i^{(k)2} I_{q_i^{(k)}}) \text{ with } \sigma_i^{(k)2} \geq 0 \text{ } (i = 1, \dots, \bar{c}^{(k)}),$$

$$\underline{b}^{(k)} = (\underline{b}_1^{(k)'}, \dots, \underline{b}_{\bar{c}^{(k)}}^{(k)'})',$$

$\underline{e}^{(k)}$  is an  $n^{(k)} \times 1$  vector of unobservable random errors such that

$$\underline{e}^{(k)} \stackrel{d}{\sim} N_n(0, \sigma_{c^{(k)}+1}^2 I_n),$$

and where  $\underline{b}_1^{(k)}, \dots, \underline{b}_{\bar{c}^{(k)}}^{(k)}, \underline{e}^{(k)}$  are independently distributed. Letting

$$p = \sum_{k=1}^{\bar{c}} p^{(k)} \text{ and } q = \sum_{k=1}^{\bar{c}} \sum_{i=1}^{\bar{c}^{(k)}} q_i^{(k)}, \text{ take}$$

$$X = \text{diag}(X^{(1)}, \dots, X^{(\bar{c})}),$$

$$\underline{\alpha} = (\underline{\alpha}^{(1)'} , \dots, \underline{\alpha}^{(\bar{c})'})',$$

$$Z = \text{diag}(Z^{(1)}, \dots, Z^{(\bar{c})}),$$

$$\underline{b} = (\underline{b}^{(1)'} , \dots, \underline{b}^{(\bar{c})'})',$$

$$\underline{e} = (\underline{e}^{(1)'} , \dots, \underline{e}^{(\bar{c})'})',$$

in model (1.2). Put  $D_{kk} = \text{Var}(\underline{b}^{(k)}) = \text{diag}(\sigma_1^{(k)2} I_{q_1^{(k)}}, \dots, \sigma_c^{(k)2} I_{q_c^{(k)}})$

and  $R_{kk} = \text{Var}(\underline{e}^{(k)}) = \sigma_{c+1}^{(k)2} I_{n^{(k)}} \quad (k = 1, \dots, \bar{c})$ . For  $i, j \in \{1, \dots, \bar{c}\}$

with  $i \neq j$ , define  $D_{ij} = \text{Cov}(\underline{b}^{(i)}, \underline{b}^{(j)}) = \sigma_{ij} D_{ij}^*$ , where  $D_{ij}^*$  is a

$\left( \sum_{k=1}^c q_k^{(i)} \right) \times \left( \sum_{k=1}^c q_k^{(j)} \right)$  known matrix, and  $\sigma_{ij}$  is an unknown, unobservable

covariance component. Also, define  $R_{ij} = \text{Cov}(\underline{e}^{(i)}, \underline{e}^{(j)}) = \bar{\sigma}_{ij} R_{ij}^*$ ,

where  $R_{ij}^*$  is an  $n^{(i)} \times n^{(j)}$  known matrix, and  $\bar{\sigma}_{ij}$  is an unknown, unobservable covariance component. Then, in model (1.2), take

$$D = \text{Var}(\underline{b}) = \begin{pmatrix} D_{11} & \dots & D_{1\bar{c}} \\ \vdots & & \vdots \\ D_{\bar{c}1} & \dots & D_{\bar{c}\bar{c}} \end{pmatrix},$$

$$R = \text{Var}(\underline{e}) = \begin{bmatrix} R_{11} & \dots & R_{1c} \\ \vdots & & \vdots \\ R_{c1} & \dots & R_{cc} \end{bmatrix}.$$

### C. An Example

We now describe an example where, in applying model (1.1), it might be appropriate to adopt the extended parameter space  $\Omega_1^*$  or  $\Omega_2^*$ , instead of  $\Omega_1$  or  $\Omega_2$ . Snedecor and Cochran (1967, Section 10.20), discuss a situation where  $N$  pigs are housed in each of  $t$  pens, with the pigs within a pen competing for an insufficient food supply. Let  $y_{ij}$  represent the weight of the  $j$ -th pig in the  $i$ -th pen ( $i = 1, \dots, t$ ;  $j = 1, \dots, N$ ). A possible model for the pig weights is to assume that the vector of observed weights,  $\underline{y}$ , has a multivariate normal distribution  $N_{Nt}(\mu \underline{1}_{Nt}, V)$ , where  $\underline{1}_{Nt}$  denotes an  $Nt \times 1$  vector of ones, and the variance-covariance matrix  $V$  is such that (i) every observation has a common variance  $\sigma_T^2$ , say, (ii) observations from different pens are uncorrelated, and (iii) any two observations within a pen have a common correlation  $\rho_I \equiv \text{Cov}(y_{ij}, y_{ik})/\sigma_T^2$  ( $i = 1, \dots, t$ ;  $j, k \in \{1, \dots, N\}$ ,  $j \neq k$ ), known as the intraclass correlation [Fisher (1925)]. This model can be formulated as a special case of model (1.1) by putting  $n = Nt$ ,  $c = 1$ ,  $X = \underline{1}_n$ ,  $Z = \text{diag}(\underline{1}_N, \dots, \underline{1}_N)$  (an  $n \times t$  matrix),  $\sigma_2^2 \equiv \sigma_T^2(1 - \rho_I)$ , and  $\gamma \equiv \rho_I/(1 - \rho_I)$ , in which case the parameter  $\sigma_1^2 = \gamma \sigma_2^2 = \rho_I \sigma_T^2$  represents the covariance between any two observations within a pen.

Snedecor and Cochran present the following partial analysis-of variance table for this example.

<u>Source</u>	<u>d.f.</u>	<u>SS</u>	<u>MS</u>	<u>EMS</u>
Between pens	t-1	$SSB = N \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2$	$SSB/(t-1)$	$EMSB = \sigma_T^2 [1 + (N-1)\rho_I]$ $= \sigma_2^2 + N\sigma_1^2$
Within pens	t(N-1)	$SSW = \sum_{i=1}^t \sum_{j=1}^N (y_{ij} - \bar{y}_{i.})^2$	$SSW/t(N-1)$	$EMSW = \sigma_T^2(1 - \rho_I) = \sigma_2^2$

Here,  $\bar{y}_{i.} = \frac{1}{N} \sum_{j=1}^N y_{ij}$  and  $\bar{y}_{..} = \frac{1}{t} \sum_{i=1}^t \bar{y}_{i.}$

If  $\underline{\gamma}$  were restricted to lie in the parameter space  $\Omega_1$ , then  $EMSB \geq EMSW$ . As indicated by Snedecor and Cochran, this restriction may be inappropriate. If there is unfair competition within a pen for the limited food supply, then the within-pen variability may exceed the between-pen variability. The parameter space  $\Omega_1^* = \{\underline{\gamma} : \sigma_2^2 > 0, \gamma > -\frac{1}{N}\}$  allows for the possibility that  $EMSB < EMSW$  and, hence, may be more appropriate than the parameter space  $\Omega_1$ .

#### D. Estimation of $\underline{\sigma}$ - Applications

When model (1.1) or model (1.2) is adopted, estimates of  $\underline{\gamma}$  or  $\underline{\sigma}$ , or of functions of  $\underline{\gamma}$  or  $\underline{\sigma}$ , may be sought. For example, animal breeders may wish to estimate heritability or repeatability. Under certain assumptions, these quantities are expressible as functions of the

variance components  $\sigma_1^2, \dots, \sigma_{c+1}^2$  in a mixed linear model [e.g., Falconer (1976) and Kempthorne (1973)].

Corresponding to an observed value of  $\underline{y}$  in model (1.2) is a realized or sample value of the random vector  $\underline{b}$ . Denote this value by  $\underline{\beta}$ . Like  $\underline{\alpha}$ , it is a fixed, but unobservable, vector. A second major use for estimates of  $\underline{\gamma}$  or  $\underline{\sigma}$  in model (1.2) is in the estimation of a linear combination  $\underline{t} \equiv \underline{\lambda}_1' \underline{\alpha} + \underline{\lambda}_2' \underline{\beta}$ , where  $\underline{\lambda}_1' \underline{\alpha}$  is an estimable linear function [e.g., Searle (1974) and Harville (1975)]. Assuming that  $\underline{\gamma}$  is known, it may be appropriate to estimate  $\underline{t}$  by the best linear unbiased predictor (BLUP)  $\tilde{\underline{t}} \equiv \underline{\lambda}_1' \tilde{\underline{\alpha}} + \underline{\lambda}_2' DZ'V^{-1}(\underline{y} - X\tilde{\underline{\alpha}})$ , where  $\tilde{\underline{\alpha}}$  is any solution to the Aitken equations  $X'V^{-1}X\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$ . If  $\underline{\gamma}$  is unknown, then an estimate of it may be obtained from  $\tilde{\underline{t}}$  by replacing the true value of  $\underline{\gamma}$  with an estimated value.

#### E. Restricted Maximum Likelihood Estimation - Motivation

Henderson (1980) listed, as the most commonly used methods for estimating  $\underline{\sigma}$  : Henderson's Methods I, II, and III [Henderson (1953)], minimum norm quadratic unbiased estimation (MINQUE) [Rao (1971)], maximum likelihood (ML) estimation [Hartley and Rao (1967)], restricted ML (REML) estimation [Patterson and Thompson (1971)], and MINQUE(0) [Rao (1970)].

The variance-component estimators obtained from Henderson's methods, MINQUE, and MINQUE(0) are quadratic, unbiased, and translation invariant. However, these estimators can, in general, yield estimates

outside the parameter space. If the purpose of the estimation is to obtain a value of  $\underline{\gamma}$  for substitution into the BLUP of a linear combination  $\underline{\lambda}'_1 \underline{\alpha} + \underline{\lambda}'_2 \underline{\beta}$ , then the use of such estimates can produce inappropriate results. There are various ad hoc ways in which these estimators can be modified to eliminate estimates that are outside the parameter space, however, after modification, the estimators are not, in general, unbiased. An alternative approach is to estimate  $\underline{\gamma}$  or  $\underline{\sigma}$  by maximum likelihood.

Hartley and Rao (1967) discussed the maximum likelihood (ML) estimation of  $\underline{\alpha}$  and  $\underline{\sigma}$  in model (1.2). The ML estimation of  $\underline{\sigma}$  is also discussed by, for example, Harville (1977).

The ML estimators of  $\sigma_1^2, \dots, \sigma_{c+1}^2$  can be severely biased, even in the case of balanced data [e.g., Corbeil and Searle (1976b)]. This undesirable characteristic led W. A. Thompson (1962) to propose, for purposes of estimating  $\underline{\sigma}$ , a modification of ML, which has come to be known as restricted maximum likelihood (REML). The REML estimate of  $\underline{\sigma}$  is obtained by maximizing the likelihood function associated with the location-invariant sufficient statistics. In contrast, the ML estimate is obtained by maximizing the likelihood function associated with the entire set of sufficient statistics. Thompson (1962) obtained explicit representations for the REML estimators in the case of random-effects models. Patterson and R. Thompson (1971) indicated, in some detail, how to obtain REML estimates in the case of unbalanced mixed models. The REML approach is discussed in more detail in the following section.

F. Restricted Maximum Likelihood  
Estimation - Discussion

The following definitions are useful.

Definition 1.1: An error contrast is a linear function  $\underline{a}'\underline{y}$  of the observations such that  $E(\underline{a}'\underline{y}) \equiv 0$  for all  $\underline{\alpha}$ ,  $\underline{\gamma}$ .

Definition 1.2: A set of error contrasts  $\underline{a}_1'\underline{y}$ , ...,  $\underline{a}_k'\underline{y}$  are said to be linearly independent if the vectors  $\underline{a}_1$ , ...,  $\underline{a}_k$  are linearly independent, i.e., if  $\sum_{i=1}^k c_i \underline{a}_i = \underline{0}$  implies that  $c_i = 0$  ( $i = 1, \dots, k$ ).

Patterson and Thompson (1971) proposed that the REML estimate of  $\underline{\gamma}$  or  $\underline{\sigma}$  be obtained by maximizing the likelihood function associated with a particular set of  $n-p^*$  linearly independent error contrasts. Harville (1974, 1977) showed that the log-likelihood function associated with any set of  $n-p^*$  linearly independent error contrasts differs by no more than an additive constant (which is free of  $\underline{\alpha}$  and  $\underline{\gamma}$ ) from the function

$$L_1(\underline{\gamma}; \underline{y}) \equiv -\frac{1}{2} \log|\underline{V}| - \frac{1}{2} \log|\underline{X}^{*'} \underline{V}^{-1} \underline{X}^*|$$

$$- \frac{1}{2} (\underline{y} - \underline{X}\tilde{\underline{\alpha}})' \underline{V}^{-1} (\underline{y} - \underline{X}\tilde{\underline{\alpha}}), \quad \underline{\gamma} \in \Omega_1^*, \quad (1.3)$$

where  $\underline{X}^*$  represents any  $n \times p^*$  matrix whose columns are linearly independent columns of  $\underline{X}$ , and where  $\tilde{\underline{\alpha}}$  is any solution to the Aitken equations

$X'V^{-1}X\tilde{\alpha} = X'V^{-1}y$ . Accordingly, we formally define a REML estimate of  $\underline{\gamma}$  as follows.

Definition 1.3: A REML estimate of  $\underline{\gamma}$  is any value of  $\underline{\gamma}$  for which

$L_1(\underline{\gamma}; y)$  attains a maximum value over the parameter space.

For an arbitrary matrix  $M$ , let  $M^-$  denote an arbitrary generalized inverse of  $M$ , i.e., any solution to  $MM^-M = M$ . Also, let  $P_X = X(X'X)^-X'$ . Take  $A$  to be an  $n \times (n-p^*)$  matrix such that  $A'A = I_{n-p^*}$  and  $AA' = I - P_X$ . Then,  $A'y \stackrel{d}{\sim} N_{n-p^*}(0, A'VA)$ , implying, in particular, that the  $n-p^*$  elements of  $A'y$  form a set of  $n-p^*$  linearly independent error contrasts. A REML estimate of  $\underline{\gamma}$  can thus be obtained by maximizing the likelihood function associated with  $A'y$ . [In fact, this is the likelihood function considered by Patterson and Thompson (1971).]

In the special case  $c=1$ , we consider, in addition to  $\Omega_1$  and  $\Omega_1^*$ , a third possible parameter space for  $\underline{\gamma}$ . Recall, that, in this special case,  $\Omega_1^* = \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, \gamma_1 > -\frac{1}{\lambda^*}\}$ , where  $\lambda^*$  is the largest characteristic root of  $Z_1Z_1'$ . This set consists of all  $\underline{\gamma}$  values for which  $V$  is positive definite. Observe that the likelihood function associated with  $A'y$  is actually defined for a larger set of  $\underline{\gamma}$  values, namely, for the set

$$\Omega_3^* \equiv \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, A'VA \text{ is a positive definite matrix}\}.$$

Note that

$$\begin{aligned} A'VA &= \sigma_2^2 A'[I + \gamma_1 Z_1 Z_1']A \\ &= \sigma^2 [I + \gamma_1 (A'Z_1)(A'Z_1)'] . \end{aligned}$$

A necessary and sufficient condition for  $A'VA$  to be a positive definite matrix is that all its characteristic roots be strictly positive. Since  $\sigma_2^2 > 0$ , this condition is equivalent to all the characteristic roots of  $I + \gamma_1 (A'Z_1)(A'Z_1)'$  being strictly positive or to  $\gamma_1 > -\frac{1}{\Delta^*}$ , where  $\Delta^*$  represents the largest characteristic root of  $(A'Z_1)(A'Z_1)'$ . Since the nonzero characteristic roots of  $(A'Z_1)(A'Z_1)'$  are the same as those of  $(A'Z_1)'(A'Z_1)$ ,  $\Delta^*$  is the largest characteristic root of the matrix  $C_{11}$  defined by

$$C_{11} \equiv Z_1'(I - P_X)Z_1 = (A'Z_1)'(A'Z_1).$$

The set  $\Omega_3^*$  has the alternative representation

$$\Omega_3^* = \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, \gamma_1 > -\frac{1}{\Delta^*}\}.$$

Subsequently, in Lemma 2.9, we show that  $-\frac{1}{\Delta^*} \leq \frac{1}{\lambda^*}$ . It follows that  $\Omega_3^* \supset \Omega_1^*$ .

Note that the matrix  $V$  is singular for any value of  $\gamma_1 \in (-\frac{1}{\Delta^*}, -\frac{1}{\lambda^*}]$  which equals minus one times the reciprocal of some characteristic root of  $Z_1 Z_1'$ . For other values of  $(\gamma_1, \sigma_2^2)' \in \Omega_3^*$ ,  $V$  is nonsingular, though it is positive definite if, and only if,  $(\gamma_1, \sigma_2^2)' \in \Omega_1^*$ . At most,  $Z_1 Z_1'$  has  $q_1$  positive characteristic roots and consequently there are at most  $q_1$  points in  $\Omega_3^*$  for which  $V$  is singular.

If, in the case of balanced data, the ANOVA estimate of  $\underline{\sigma}$  is contained in the parameter space, then the ANOVA estimate is the same as the REML estimate of  $\underline{\sigma}$  [e.g., Harville (1977)].

If  $c=0$  in model (1.2), then the ML and REML estimators of  $\sigma_1^2$  are

$$\hat{\sigma}_{1,ML}^2 = (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'(\underline{y} - \underline{X}\tilde{\underline{\alpha}})/n ,$$

$$\hat{\sigma}_{1,REML}^2 = (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'(\underline{y} - \underline{X}\tilde{\underline{\alpha}})/(n-p^*) .$$

In contrast to the ML estimator  $\hat{\sigma}_{1,ML}^2$ ,  $\hat{\sigma}_{1,REML}^2$  accounts for the  $p^*$  degrees of freedom that are exhausted in estimating  $\underline{\alpha}$  and it is an unbiased estimator of  $\sigma_1^2$ . The ML estimator of the variance  $\sigma_1^2$  has uniformly smaller mean squared error than the REML estimator when  $p^* \leq 4$ . However, when  $p^* \geq 5$  and  $n-p^*$  is sufficiently large ( $n-p^* > 2$  suffices when  $p^* \geq 13$ ), then the REML estimator has the smaller mean squared error [Harville (1977)].

### G. A Preview of Subsequent Chapters

Henderson (1980) described three sources of difficulty in the computation of a REML estimate of  $\gamma$  : 1. The REML estimate of  $\gamma$  is the solution to a constrained optimization problem (that of maximizing the log-likelihood function  $L_1$  over the parameter space). Explicit expressions for REML estimators of  $\gamma$  exist only in very special cases. The estimates must, in general, be computed iteratively, and hence their use requires more sophistication on the part of the practitioner. 2. An iterative procedure may converge slowly, if at all. 3. There is no guarantee that an iterative procedure will converge to a point at which  $L_1(\gamma; \underline{y})$  attains its supremum over the parameter space.

In light of Henderson's comments, it would seem that the use of REML estimates of  $\gamma$  could be facilitated by the development of more effective iterative computational procedures. How quickly an iterative algorithm converges (and whether it converges at all) may depend on the choice of the initial guess for  $\gamma$ , on the value of  $\underline{y}$ , and how the model is parameterized. Thus, there is little hope of finding a single algorithm which is uniformly "best" in all applications. However, numerical comparisons may suggest algorithms which can be recommended for widespread use.

In Chapters III and IV, we present fourteen different algorithms for determining a REML estimate of  $\gamma$  in model (1.1). The algorithms consist of variations on four general methods that have been used by practitioners : the EM algorithm, the Newton-Raphson method, the

Method of Scoring, and the Method of Successive Approximations. Each algorithm is introduced in Chapter III. In this initial presentation, the constraints on  $\underline{\gamma}$  (imposed by its confinement to the parameter space) are ignored. In Chapter IV, we discuss results in the constrained optimization literature that indicate how the basic algorithms can be modified to accommodate parametric constraints. In Chapter VI, we apply the fourteen algorithms for the estimation of  $\underline{\gamma}$  to four different data sets, in each case taking the model to be a special case of (1.1) with  $c=1$ , and employing two different initial guesses for  $\underline{\gamma}$ .

The amount of computer time required for an iterative procedure to attain convergence depends on the time per iteration as well as the number of iterations. Contributing to the time per iteration is the computation of certain quantities, such as the first- and second-order partial derivatives of  $L_1$ , that must be re-evaluated on each iteration. Computationally efficient representations for these quantities are presented in Chapter II.

In general, there is no guarantee that a point to which an iterative algorithm converges is a global maximum. General methods for locating global maxima are described by Dixon, Gomulka, and Szegö (1975), but their usefulness is limited by highly restrictive conditions and generally by excessive computations. In maximizing  $L_1$ , it would seem to be good practice to let the iterative algorithm converge from several different starting values, and to then choose the point of convergence for which  $L_1$  has the largest value.

Model (1.1) can be generalized by supposing that  $\text{Var}(\underline{y}) =$

$\sigma_{c+1}^2 [\underline{I}_n + \sum_{i=1}^c \gamma_i \underline{Z}_i \underline{A}_i \underline{Z}_i']$ , where (i)  $\underline{A}_i$  is a  $q_i \times q_i$  known, positive

definite matrix or, more generally, where (ii)  $\underline{A}_i$  is a  $q_i \times q_i$  known,

nonnegative definite matrix ( $i = 1, \dots, c$ ), in which cases  $\text{Var}(\underline{y})$

is the matrix  $\underline{R} + \underline{Z} \underline{D}_A \underline{Z}'$ , where  $\underline{D}_A$  is defined by

$$\underline{D}_A = \begin{pmatrix} \sigma_1^2 \underline{A}_1 & & \emptyset \\ & \ddots & \\ \emptyset & & \sigma_c^2 \underline{A}_c \end{pmatrix}.$$

Model (1.2) can be generalized in an analogous manner by supposing that

$\text{Var}(\underline{b}) = \underline{D}_A$ . In Chapter V, we extend the results given in Chapter III

to these more general models.

## II. THE RESTRICTED LOG-LIKELIHOOD FUNCTION AND ITS DERIVATIVES

The objective in this chapter is to derive various functions (of  $\underline{\gamma}$ ) whose evaluation may be required in computing a REML estimate of  $\underline{\gamma}$ . These functions include the function  $L_1(\underline{\gamma}; \underline{y})$  (which is equal, up to an additive constant, to the log-likelihood function associated with any set of  $n-p^*$  linearly independent error contrasts), the first- and second-order partial derivatives of  $L_1$ , the expected values of the second-order partial derivatives, and various related functions. Results will be given for both the  $\underline{\gamma}$ - and the  $\underline{\sigma}$ -parameterizations.

Many of the results in this chapter were derived or summarized by, for example, Harville (1977) and Searle (1979). In doing so, they assumed that  $\underline{\gamma}$  belonged to the restricted parameter space  $\Omega_1$  (or, equivalently, that  $\underline{\sigma}$  belonged to  $\Omega_2$ ). In this chapter, we extend the results given by them to the larger parameter space  $\Omega_1^*$  (or  $\Omega_2^*$ ). In the special case  $c=1$ , these results can be further extended, in an obvious way, to the even larger parameter space  $\Omega_3^*$ .

### A. The Log-Likelihood Function

Let  $A$  denote a  $n \times (n-p^*)$  matrix which satisfies  $A'A = I_{n-p^*}$  and  $AA' = I - P_X$ . Observe that  $A(A'A) = A$ , implying (since  $X'AA' = \emptyset$ ) that

$$X'A = X'AA'A = \emptyset ,$$

and hence that

$$E(A'y) = A'X\alpha = 0 \text{ for all } \alpha \in \mathbb{R}^p \text{ and } \gamma \in \Omega_1^*.$$

By Definition 1.1,  $A'y$  is a vector of  $n-p^*$  error contrasts. Further, the  $n-p^*$  error contrasts in  $A'y$  are linearly independent since  $\text{rank}(A') = \text{rank}(AA') = n-p^*$ .

Under model (1.1),  $y \stackrel{d}{\sim} N_n(X\alpha, V)$ . Thus,  $A'y \stackrel{d}{\sim} N_{n-p^*}(0, A'VA)$ . By definition,  $V$  is positive definite for  $\gamma \in \Omega_1^*$ . Further, since  $A$  has full column rank,  $A'VA$  is also positive definite for  $\gamma \in \Omega_1^*$ . Thus,  $A'y$  has a well-defined likelihood function which is expressible as

$$f(\gamma; A'y) \equiv (2\pi)^{-\frac{1}{2}(n-p^*)} |A'VA|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(A'y)'(A'VA)^{-1}(A'y)\right\},$$

$$\gamma \in \Omega_1^* . \quad (2.1)$$

Let  $X^*$  denote an  $n \times p^*$  matrix whose columns are linearly independent columns of  $X$ , and for  $\gamma \in \Omega_1^*$ , define

$$P \equiv V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} . \quad (2.2)$$

The following lemma can be used to derive an alternative representation for  $f(\gamma; A'y)$ .

Lemma 2.1: For  $\gamma \in \Omega_1^*$ ,

- (i)  $A(A'VA)^{-1}A' = P$  ,
- (ii)  $X(X'V^{-1}X)^{-1}X' = X^*(X^{*'}V^{-1}X^*)^{-1}X^{*'} ,$

$$(iii) |A'VA| = |V| |X^{*'} V^{-1} X^*| |X^{*'} X^*|^{-1},$$

$$(iv) PVP' = P.$$

Detailed proofs of the identities in Lemma 2.1 are given by Searle (1979). He assumed that  $\underline{\gamma} \in \Omega_1$ , however he used this assumption only to show that  $V$  is nonsingular. Since  $V$  is also nonsingular for  $\underline{\gamma} \in \Omega_1^*$ , Lemma 2.1 follows from Searle's derivations.

Parts (i)-(iii) of Lemma 2.1 can be used to derive the following alternative expression for the log-likelihood function associated with  $f(\underline{\gamma}; A'\underline{y})$  :

$$\begin{aligned} f(\underline{\gamma}; A'\underline{y}) &= (2\pi)^{-\frac{1}{2}(n-p^*)} |X^{*'} V^{-1} X^*|^{-\frac{1}{2}} |V|^{-\frac{1}{2}} |X^{*'} X^*|^{\frac{1}{2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2} \underline{y}' [V^{-1} - V^{-1} X^* (X^{*'} V^{-1} X^*)^{-1} X^{*'} V^{-1}] \underline{y}\right\}, \\ &\quad \underline{\gamma} \in \Omega_1^*. \end{aligned}$$

Thus, to evaluate  $f(\underline{\gamma}; A'\underline{y})$ , it is not necessary to compute the matrix  $A$ .

Since  $(2\pi)^{-\frac{1}{2}(n-p^*)}$  and  $|X^{*'} X^*|$  do not depend on  $\underline{\alpha}$  and  $\underline{\gamma}$ , the log-likelihood function associated with  $A'\underline{y}$  is, apart from an additive constant,

$$\begin{aligned} L_1(\underline{\gamma}; \underline{y}) &\equiv L_1(\gamma_1, \dots, \gamma_c, \sigma_{c+1}^2; \underline{y}) \\ &= -\frac{1}{2} \log|V| - \frac{1}{2} \log|X^{*'} V^{-1} X^*| \\ &\quad - \frac{1}{2} \underline{y}' [V^{-1} - V^{-1} X^* (X^{*'} V^{-1} X^*)^{-1} X^{*'} V^{-1}] \underline{y} \end{aligned}$$

$$= -\frac{1}{2} \log|V| - \frac{1}{2} \log|X^{*'} V^{-1} X^*| - \frac{1}{2} \underline{y}' P \underline{y}, \quad \underline{y} \in \Omega_1^*. \quad (2.3)$$

Subsequently, we let  $\tilde{\alpha}$  represent an arbitrary solution to the Aitken equations  $X'V^{-1}X\tilde{\alpha} = X'V^{-1}\underline{y}$ . Note that, for  $\underline{y} \in \Omega_1^*$ ,

$$\begin{aligned} VP\underline{y} &= V[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}]\underline{y} \\ &= [I - X(X'V^{-1}X)^{-1}X'V^{-1}]\underline{y} \\ &= \underline{y} - X(X'V^{-1}X)^{-1}X'V^{-1}\underline{y} \\ &= \underline{y} - X\tilde{\alpha}. \end{aligned} \quad (2.4)$$

Thus,

$$\begin{aligned} \tilde{\alpha}'X'V^{-1}(\underline{y} - X\tilde{\alpha}) &= \tilde{\alpha}'X'V^{-1}\underline{y} - \tilde{\alpha}'X'V^{-1}X\tilde{\alpha} \\ &= \tilde{\alpha}'(X'V^{-1}\underline{y} - X'V^{-1}X\tilde{\alpha}) \\ &= \tilde{\alpha}'(X'V^{-1}\underline{y} - X'V^{-1}\underline{y}) \\ &= 0. \end{aligned}$$

It follows that, for  $\underline{y} \in \Omega_1^*$ ,

$$\underline{y}'\underline{P}\underline{y} = \underline{y}'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) = (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) . \quad (2.5)$$

Together, results (2.3) and (2.5) imply that

$$\begin{aligned} L_1(\underline{\gamma}; \underline{y}) = & -\frac{1}{2} \log|\underline{V}| - \frac{1}{2} \log|\underline{X}^*'\underline{V}^{-1}\underline{X}^*| \\ & - \frac{1}{2} (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}), \underline{\gamma} \in \Omega_1^* . \end{aligned} \quad (2.6)$$

This expression for  $L_1(\underline{\gamma}; \underline{y})$  is the same as that given by Harville (1974, 1977).

We now derive an expression for  $L_1(\underline{\gamma}; \underline{y})$  which contrasts with expression (2.6) in the important respect that it does not involve the inverse of the  $n \times n$  matrix  $\underline{V}$ .

Define

$$\underline{S} \equiv \underline{R}^{-1} - \underline{R}^{-1}\underline{X}(\underline{X}'\underline{R}^{-1}\underline{X})^{-1}\underline{X}'\underline{R}^{-1} , \quad (2.7)$$

$$\underline{D}_i \equiv \sigma_{c+1}^2 \text{diag}(\emptyset, \dots, \emptyset, \gamma_i \underline{I}_{q_i}, \dots, \gamma_c \underline{I}_{q_c}) (i = 1, \dots, c). \quad (2.8)$$

Since  $\underline{R}$  is nonsingular for  $\underline{\gamma} \in \Omega_1^*$  and  $\underline{A}$  has full column rank, the matrix  $\underline{A}'\underline{R}\underline{A}$  is necessarily nonsingular for  $\underline{\gamma} \in \Omega_1^*$ .

Lemma 2.2: For  $\underline{\gamma} \in \Omega_1^*$ ,

$$\underline{S} = \underline{A}(\underline{A}'\underline{R}\underline{A})^{-1}\underline{A}' . \quad (2.9)$$

Searle (1979) stated and proved identity (2.9) for  $\underline{\gamma} \in \Omega_1$ . That this identity holds for  $\underline{\gamma} \in \Omega_1^*$  can be proven in essentially the same way.

Lemma 2.3: For  $\underline{\gamma} \in \Omega_1^*$ , the matrix  $(I + Z'SZD_i)$  is nonsingular ( $i = 1, \dots, c$ ).

Proof: Suppose that  $\underline{\gamma} \in \Omega_1^*$ . By definition,  $V_i = \sigma_{c+1}^2 [I_n + \sum_{j=1}^c \gamma_j Z_j Z_j']$  is nonsingular, implying (since A has full column rank) that  $A'V_i A$  is nonsingular. Since  $V_i = R + ZD_i Z'$ , we have that

$$|A'RA + A'ZD_i Z'A| \neq 0 ,$$

implying (since  $|A'RA| \neq 0$ ) that

$$|A'RA| |I + (A'RA)^{-1} A'ZD_i Z'A| \neq 0$$

and, hence, that

$$|I + (A'RA)^{-1} A'ZD_i Z'A| \neq 0 .$$

Recalling that  $|I + SU| = |I + US|$  for "arbitrary" matrices S and U, it follows that

$$|I + Z'A(A'RA)^{-1} A'ZD_i| \neq 0$$

which, in light of Lemma 2.2, completes the proof. □

Since  $D_1 = D$ , Lemma 2.3 implies, in particular, that  $(I + Z'SZD)$  is nonsingular for  $\underline{\gamma} \in \Omega_1^*$ . Further, since  $S$  reduces to  $R^{-1}$  when  $X$  is the null matrix, it follows that  $(I + Z'R^{-1}ZD)$  is a nonsingular matrix for  $\underline{\gamma} \in \Omega_1^*$ . It is then easy to verify the identity

$$\underline{V}^{-1} = R^{-1} - R^{-1}ZD(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}, \quad \underline{\gamma} \in \Omega_1^*, \quad (2.10)$$

given by Harville (1977) for  $\underline{\gamma} \in \Omega_1$ . From identity (2.10), we obtain the further identity

$$Z'\underline{V}^{-1} = (I + Z'R^{-1}ZD)^{-1}Z'R^{-1}, \quad \underline{\gamma} \in \Omega_1^*. \quad (2.11)$$

Subsequently, we define

$$\underline{\tilde{b}} \equiv DZ'\underline{V}^{-1}(\underline{y} - X\underline{\tilde{\alpha}}). \quad (2.12)$$

Lemma 2.4: For  $\underline{\gamma} \in \Omega_1^*$ ,

- (i)  $P = S - SZD(I + Z'SZD)^{-1}Z'S,$
- (ii)  $Z'P = (I + Z'SZD)^{-1}Z'S,$
- (iii)  $Z'\underline{V}^{-1}(\underline{y} - X\underline{\tilde{\alpha}}) = Z'P\underline{y}$
- (iv)  $\underline{y}'P\underline{y} = \underline{y}'S(\underline{y} - Z\underline{\tilde{b}}),$
- (v)  $\underline{V}^{-1}(\underline{y} - X\underline{\tilde{\alpha}}) = S(\underline{y} - Z\underline{\tilde{b}}) = R^{-1}(\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}}).$

The identities in Lemma 2.4 were given by Harville (1977), and were proved in detail by Searle (1979), for  $\underline{\gamma} \in \Omega_1$ . That these identities hold for  $\underline{\gamma} \in \Omega_1^*$  can be proved in essentially the same way.

The use of formula (2.6) to evaluate the function  $L_1(\underline{\gamma}; \underline{y})$  at a point  $\underline{\gamma} \in \Omega_1^*$  would require the evaluation of  $(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'V^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}})$ . Using identity (2.5) and parts (iv) and (v) of Lemma 2.4, the quadratic form  $(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'V^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}})$  can be re-expressed as

$$(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'V^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) = \underline{y}'S(\underline{y} - \tilde{\underline{Z}}\tilde{\underline{b}}) \quad (2.13)$$

$$= \underline{y}'R^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \tilde{\underline{Z}}\tilde{\underline{b}})$$

$$= (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'S(\underline{y} - \tilde{\underline{X}}\tilde{\underline{b}}) \quad (\text{since } X'S = \emptyset) \quad (2.14)$$

$$= (\underline{y} - \underline{X}\tilde{\underline{\alpha}})'R^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \tilde{\underline{Z}}\tilde{\underline{b}}), \quad \underline{\gamma} \in \Omega_1^*.$$

Moreover, it follows from identity (2.11) and parts (ii) and (iii) of Lemma 2.4, that, for  $\underline{\gamma} \in \Omega_1^*$ ,

$$\begin{aligned} \tilde{\underline{b}} &= DZ'V^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) \\ &= D(I + Z'SZD)^{-1}Z'S\underline{y} \end{aligned} \quad (2.15)$$

$$= D(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}). \quad (2.16)$$

Combining results (2.13) and (2.15) we find that

$$(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'V^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) = \underline{y}'[S - SZD(I + Z'SZD)^{-1}Z'S]\underline{y} . \quad (2.17)$$

The matrix inverses (or generalized inverses) in the right hand side of (2.17) are those of  $R$ ,  $X'R^{-1}X$ , and  $I + Z'SZD$ . We see that  $R = \sigma_{c+1}^2 I_n$  so that  $R^{-1}$  is immediately obtainable. The matrices  $X'X$  and  $I + Z'SZD$  have dimensions  $p \times p$  and  $q \times q$ , respectively. Since the left hand side of (2.17) involves the inverse of the  $n \times n$  matrix  $V$  and the generalized inverse of the  $p \times p$  matrix  $X'V^{-1}X$ , the right hand side of (2.17) is more useful, from a computational standpoint, than the left hand side.

Consider now the following set of linear equations

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}ZD \\ Z'R^{-1}X & I + Z'R^{-1}ZD \end{pmatrix} \begin{pmatrix} \underline{\alpha}^* \\ \tilde{\underline{v}} \end{pmatrix} = \begin{pmatrix} X'R^{-1}\underline{y} \\ Z'R^{-1}\underline{y} \end{pmatrix} , \quad (2.18)$$

which we refer to as mixed-model equations (MME). For  $\underline{y} \in \Omega_1$ , certain properties of these equations can be deduced, for example, from Harville's (1976) Theorem 2. The MME also have these properties for other  $\underline{y} \in \Omega_1^*$ , as indicated by the following theorem:

Theorem 2.1: Suppose that  $\underline{y} \in \Omega_1^*$ . Then, equations (2.18) are consistent. Further, if  $\underline{\alpha}^*$  and  $\tilde{\underline{v}}$  are the "first" and "second" components of any solution to (2.18), then  $X'V^{-1}X\underline{\alpha}^* = X'V^{-1}\underline{y}$  (i.e.,  $\underline{\alpha}^*$  satisfies

the Aitken equations) and  $\tilde{\underline{v}} = \underline{Z}'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}^*)$ . Conversely, for any solution  $\tilde{\underline{\alpha}}$  to  $\underline{X}'\underline{V}^{-1}\underline{X}\tilde{\underline{\alpha}} = \underline{X}'\underline{y}$ , the system (2.18) has a solution whose first component is  $\tilde{\underline{\alpha}}$ .

Proof: Suppose that  $\underline{y} \in \Omega_1^*$  and that  $\underline{\alpha}^*$  and  $\tilde{\underline{v}}$  are the components of any solution to (2.18). According to Lemma 2.3, the matrix  $(\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D})$  is nonsingular. Thus, the second set of equations in (2.18) implies that

$$\underline{Z}'\underline{R}^{-1}\underline{X}\underline{\alpha}^* + (\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D}) \tilde{\underline{v}} = \underline{Z}'\underline{R}^{-1}\underline{y} ,$$

and hence that

$$\tilde{\underline{v}} = (\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D})^{-1}\underline{Z}'\underline{R}^{-1}(\underline{y} - \underline{X}\underline{\alpha}^*) , \quad (2.19)$$

or equivalently [in light of result (2.11)]

$$\tilde{\underline{v}} = \underline{Z}'\underline{V}^{-1}(\underline{y} - \underline{X}\underline{\alpha}^*) .$$

Moreover, substituting expression (2.19) into the first set of equations in (2.18), we find that

$$\begin{aligned} \underline{X}'\underline{R}^{-1}\underline{X}\underline{\alpha}^* + \underline{X}'\underline{R}^{-1}\underline{Z}\underline{D}(\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D})^{-1}\underline{Z}'\underline{R}^{-1}\underline{y} \\ - \underline{X}'\underline{R}^{-1}\underline{Z}\underline{D}(\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D})^{-1}\underline{Z}'\underline{R}^{-1}\underline{X}\underline{\alpha}^* = \underline{X}'\underline{R}^{-1}\underline{y} , \end{aligned}$$

and hence that

$$\begin{aligned} & [X'R^{-1}X - X'R^{-1}ZD(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}X]\underline{\alpha}^* \\ & = [X'R^{-1} - X'R^{-1}ZD(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}] \underline{y} , \end{aligned}$$

or equivalently [in light of result (2.10)] that

$$X'V^{-1}\underline{x\alpha}^* = X'V^{-1}\underline{y} .$$

Now, consider any solution  $\tilde{\underline{\alpha}}$  to  $X'V^{-1}X\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$ . To show that there exists a solution to (2.18) whose first component is  $\tilde{\underline{\alpha}}$ , we demonstrate

that the vector  $\begin{pmatrix} \tilde{\underline{\alpha}} \\ Z'V^{-1}(\underline{y} - X\tilde{\underline{\alpha}}) \end{pmatrix}$  is a solution. Let  $\underline{\delta} \equiv Z'V^{-1}(\underline{y} - X\tilde{\underline{\alpha}})$ .

Then,

$$\underline{\delta} = (I + Z'R^{-1}ZD)^{-1}Z'R^{-1}(\underline{y} - X\tilde{\underline{\alpha}}) ,$$

and hence that

$$Z'R^{-1}X\tilde{\underline{\alpha}} + (I + Z'R^{-1}ZD) \underline{\delta} = Z'R^{-1}\underline{y} , \quad (2.20)$$

i.e.,  $(\tilde{\underline{\alpha}}', \underline{\delta}')'$  satisfies the second set of equations in (2.18). By assumption,  $X'V^{-1}X\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$ , implying [in light of result (2.10)] that

$$\begin{aligned}
& X'R^{-1}\tilde{\underline{\alpha}} - X'R^{-1}ZD(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}\tilde{\underline{\alpha}} \\
& = X'R^{-1}\underline{y} - X'R^{-1}ZD(I + Z'R^{-1}ZD)^{-1}Z'R^{-1}\underline{y} ,
\end{aligned}$$

and hence that

$$X'R^{-1}\tilde{\underline{\alpha}} + X'R^{-1}ZD \underline{\delta} = X'R^{-1}\underline{y} , \quad (2.21)$$

i.e.,  $(\tilde{\underline{\alpha}}', \underline{\delta}')'$  satisfies the first set of equations in (2.18), as well as the second set.

The consistency of the Aitken equations  $X'V^{-1}\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$ , together with results (2.20) and (2.21), imply the consistency of equations (2.18). □

Subsequently, in addition to taking  $\tilde{\underline{\alpha}}$  to represent an arbitrary solution to the Aitken equations, we equivalently [in light of Theorem 2.1] take  $\tilde{\underline{\alpha}}$  to be the first component of an arbitrary solution to the MME (2.18). Also, we define

$$\tilde{\underline{v}} \equiv Z'V^{-1}(\underline{y} - Z\tilde{\underline{\alpha}}) , \quad (2.22)$$

or equivalently [in light of Theorem 2.1] take  $\tilde{\underline{v}}$  to be the second component of an arbitrary solution to the MME (2.18).

Observe that as a consequence of result (2.11) and Lemma 2.4,

$$\tilde{\underline{y}} = (\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D})^{-1}\underline{Z}'\underline{R}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) \quad (2.23)$$

$$= (\underline{I} + \underline{Z}'\underline{S}\underline{Z}\underline{D})^{-1}\underline{Z}'\underline{S}\underline{y} . \quad (2.24)$$

Observe also [in light of definition (2.12)] that  $\tilde{\underline{b}} = \underline{D}\tilde{\underline{y}}$ .

As noted by, for example, Harville (1977), the MME (2.18) allows us to exploit, for computational purposes, the simple (diagonal) form of the matrices  $\underline{R}$  and  $\underline{D}$ . If  $\underline{X}$  and  $\underline{Z}$  are incidence matrices (as is the case in many applications), then the matrices  $\underline{X}'\underline{R}^{-1}\underline{X}$  and  $\underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D}$  are diagonal, in which case the computations required to solve the MME can be substantially reduced through absorption.

To evaluate the function  $L_1(\underline{y}; \underline{y})$  for a particular value of  $\underline{y}$ , we must evaluate the determinant  $|\underline{X}^{*'}\underline{V}^{-1}\underline{X}^*|$ , as well as the quadratic form  $(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}})$ . In doing so, we can, as in evaluating  $(\underline{y} - \underline{X}\tilde{\underline{\alpha}})'\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}})$ , circumvent the direct inversion of the  $n \times n$  matrix  $\underline{V}$ , as we now demonstrate.

Let  $\underline{C}^*$  denote the  $(p^* + q) \times (p^* + q)$  matrix obtained from the coefficient matrix of the MME (2.18) by replacing  $\underline{X}$  with  $\underline{X}^*$ . That is, take

$$\underline{C}^* = \begin{bmatrix} \underline{X}^{*'}\underline{R}^{-1}\underline{X}^* & \underline{X}^{*'}\underline{R}^{-1}\underline{Z}\underline{D} \\ \underline{Z}'\underline{R}^{-1}\underline{X}^* & \underline{I} + \underline{Z}'\underline{R}^{-1}\underline{Z}\underline{D} \end{bmatrix} .$$

Harville (1975, 1977), observed that for  $\underline{y}$  such that  $\underline{V}$ ,  $\underline{R}$ , and  $\underline{I} + \underline{Z}'\underline{S}\underline{Z}\underline{D}$  are nonsingular matrices,

$$\begin{aligned}
|V| |X^{*'} V^{-1} X^*| &= |R| |C^*| \\
&= |R| |X^{*'} R^{-1} X^*| |I + Z'SZD| ,
\end{aligned}$$

and hence that

$$\begin{aligned}
\log|V| + \log|X^{*'} V^{-1} X^*| &= \log|R| + \log|C^*| \\
&= \log|R| + \log|X^{*'} R^{-1} X^*| + \log|I + Z'SZD| .
\end{aligned}$$

Together with result (2.14), this result implies that, for  $\underline{\gamma} \in \Omega_1^*$ ,

$$\begin{aligned}
L_1(\underline{\gamma}; \underline{y}) &= -\frac{1}{2} \log|R| - \frac{1}{2} \log|C^*| - \frac{1}{2} \underline{y}' R^{-1} (\underline{y} - X\tilde{\underline{\alpha}} - Z\tilde{\underline{b}}) \\
&\quad \text{[using (2.6)]} \tag{2.25}
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2} \log|R| - \frac{1}{2} \log|X^{*'} R^{-1} X^*| - \frac{1}{2} \log|I + Z'SZD| \\
&\quad - \underline{y}' S (\underline{y} - Z\tilde{\underline{b}}) , \tag{2.26}
\end{aligned}$$

as indicated, for example, by Harville (1977, p. 326). Representations (2.25) and (2.26) are computationally advantageous because (as a consequence of the simple form of  $R$ ) they allow us to avoid the numerical inversion of  $n \times n$  matrices.

### B. Derivatives of the Log-Likelihood Function

Let  $\hat{\underline{\gamma}}$  represent a REML estimate of  $\underline{\gamma}$ , that is, a value of  $\underline{\gamma}$  at which  $L_1(\underline{\gamma}; \underline{y})$  attains its supremum over the parameter space. The problem of computing  $\hat{\underline{\gamma}}$  must, in general, be computed numerically. We consider some possible algorithms for this purpose in Chapters III and IV. The best-known algorithms for optimization of general nonlinear functions are gradient methods, i.e., methods which use derivative information about the function. Accordingly, in this section, we derive computationally efficient forms for the first- and second-order partial derivatives of  $L_1(\underline{\gamma}; \underline{y})$  and also for the expected values of the second-order derivatives.

Expressions for the expected values of the second-order partial derivatives of  $L_1(\underline{\gamma}; \underline{y})$  are required in some of the algorithms for computing  $\hat{\underline{\gamma}}$ . Moreover, the Fisher information matrix  $I(\underline{\gamma})$  associated with the function  $L_1$  is the matrix whose  $(i,j)$ -th element is

$$- E \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] \quad (i, j = 1, \dots, c+1).$$

The "large-sample" variance-covariance matrix of the REML estimator  $\hat{\underline{\gamma}}$  of  $\underline{\gamma}$  is  $[I(\underline{\gamma})]^{-1}$ . The meaning of "large-sample" in a mixed model context is discussed by Miller (1973).

#### 1. General case

Define

$$T \equiv (I + Z'SZD)^{-1}, \quad \underline{\gamma} \in \Omega_1^* \quad (2.27)$$

(It follows from Lemma 2.3 that the matrix  $I + Z'SZD$  is nonsingular for  $\underline{\gamma} \in \Omega_1^*$ .) Further, let  $T_{ij}$  represent the  $(i,j)$ -th submatrix of dimension  $q_i \times q_j$  of  $T$  ( $i, j = 1, \dots, c$ ), let  $\tilde{\underline{v}}_i$  represent the  $i$ -th  $q_i \times 1$  subvector of  $\tilde{\underline{v}}$ , and take

$$G_{ij} \equiv \sigma_{c+1}^2 \sum_{k=1}^c T_{ik} Z_k' S Z_j \quad (i, j = 1, \dots, c). \quad (2.28)$$

Note that

$$G_{ii} = \sigma_{c+1}^2 \sum_{k=1}^c T_{ik} Z_k' S Z_i \quad (i = 1, \dots, c).$$

Lemma 2.5: For  $\underline{\gamma} \in \Omega_1^*$ , and  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{\gamma}' S (\underline{\gamma} - Z \tilde{\underline{b}})],$$

$$(ii) \quad \frac{\partial L_1}{\partial \gamma_i} = - \left(\frac{1}{2}\right) [\text{tr}(G_{ii}) - \sigma_{c+1}^2 \tilde{\underline{v}}_i' \tilde{\underline{v}}_i],$$

$$(iii) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{\gamma}' S (\underline{\gamma} - Z \tilde{\underline{b}})],$$

$$(iv) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} = \tilde{\underline{v}}_i' \tilde{\underline{v}}_{-i},$$

$$(v) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = - \text{tr}(G_{ij} G_{ji}) + 2 \sigma_{c+1}^2 \tilde{\underline{v}}_i' G_{ij} \tilde{\underline{v}}_j,$$

$$(vi) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} = - \operatorname{tr}(G_{ii}^2) + 2\sigma_{c+1}^2 \tilde{\gamma}_i' G_{ii} \tilde{\gamma}_i ,$$

$$(vii) \quad (-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right) = \operatorname{tr}(G_{ii}^2) ,$$

$$(viii) \quad (-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right) = \frac{1}{\sigma_{c+1}^2} \operatorname{tr}(G_{ii}) ,$$

$$(ix) \quad (-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right) = \operatorname{tr}(G_{ij} G_{ji}) ,$$

$$(x) \quad (-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) = \frac{1}{\sigma_{c+1}^4} (n-p^*) .$$

Searle (1979) essentially proved each of the identities in Lemma 2.5. The identities given by parts (i), (iii), (iv), and (x) of Lemma 2.5 are the same as those given by Harville (1977) for  $\underline{\gamma} \in \Omega_1$ . That these identities hold for  $\underline{\gamma} \in \Omega_1^*$  can be proved in essentially the same way.

Define

$$C \equiv Z'(I - P_X)Z.$$

Since  $R = \sigma_{c+1}^2 I_n$ , we have that  $S = \frac{1}{\sigma_{c+1}^2} (I - P_X)$ .

Thus,  $T = [I + C \text{diag}(\gamma_1 I_{q_1}, \dots, \gamma_c I_{q_c})]^{-1}$ , and the matrix  $G_{ij}$  is the  $(i,j)$ -th submatrix of the matrix  $TC$ . In the special case  $c=1$ ,  $T = (I + \gamma_1 C)^{-1}$  and  $G_{11} = (I + \gamma_1 C)^{-1}C$ .

## 2. Special case

For nonzero values of  $\gamma_1, \dots, \gamma_{c+1}$ , the expressions given in Lemma 2.5 for the partial derivatives of  $L_1$  can be further simplified, as we now show.

As a straightforward extension of results given by Harville (1977) and Searle (1979) for  $\underline{\gamma} \in \Omega_1$ , we obtain the following lemma.

Lemma 2.6: For  $\underline{\gamma} \in \Omega_1^*$ , and  $i, j \in \{1, \dots, c\}$ ,

$$T_{ii} + \gamma_i G_{ii} = I_{q_i},$$

$$T_{ij} + \gamma_j G_{ij} = 0, \text{ if } i \neq j.$$

If  $\underline{\gamma} \in \Omega_1^*$  and  $\gamma_i \neq 0$ , then

$$G_{ii} = \frac{1}{\gamma_i} (I - T_{ii}), \quad (2.29)$$

$$G_{ij} = -\frac{1}{\gamma_j} T_{ij}, \quad (2.30)$$

for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ .

The following lemma extends results given by Harville (1977).

Lemma 2.7: For  $\underline{\gamma} \in \Omega_1^*$ , and  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{\gamma}' S(\underline{y} - Z\tilde{\underline{b}})] ,$$

$$(ii) \quad \frac{\partial L_1}{\partial \gamma_i} = - \frac{1}{2} \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})] - \sigma_{c+1}^2 \tilde{\underline{v}}_i' \tilde{\underline{v}}_i \right\}, \text{ if } \gamma_i \neq 0 ,$$

$$(iii) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{\gamma}' S(\underline{y} - Z\tilde{\underline{b}})] ,$$

$$(iv) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} = \tilde{\underline{v}}_i' \tilde{\underline{v}}_i ,$$

$$(v) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij} T_{ji}) \\ - 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_j}\right) \tilde{\underline{v}}_i' T_{ij} \tilde{\underline{v}}_j ,$$

if  $\gamma_i \neq 0, \gamma_j \neq 0$  ,

$$(vi) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} = - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{ii})^2] \\ + 2\sigma_{c+1}^2 \left(\frac{1}{\gamma_i}\right) \tilde{\underline{v}}_i' (I - T_{ii}) \tilde{\underline{v}}_i, \text{ if } \gamma_i \neq 0 ,$$

$$(vii) \quad (-2)\mathbb{E}\left[\frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i}\right] = \frac{1}{\gamma_i^2} \text{tr}[(I - T_{ii})^2] , \text{ if } \gamma_i \neq 0 ,$$

$$(viii) \quad (-2)\mathbb{E}\left[\frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i}\right] = \frac{1}{\sigma_{c+1}^2} \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})] , \text{ if } \gamma_i \neq 0 ,$$

$$(ix) \quad (-2)\mathbb{E}\left[\frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j}\right] = \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij} T_{ji}) , \text{ if } \gamma_i \neq 0, \gamma_j \neq 0 ,$$

$$(x) \quad (-2)\mathbb{E}\left[\frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2}\right] = \frac{1}{\sigma_{c+1}^4} (n-p^*) .$$

Proof: Substitution of expressions (2.29) and (2.30) for  $G_{ii}$  and  $G_{ij}$ , respectively, in the general partial derivative expressions of Lemma 2.5, yields the above identities.  $\square$

### C. Alternative Parameterization

When model (1.1) is parameterized in terms of  $\underline{\sigma}$  (instead of  $\underline{\gamma}$ ), the log-likelihood function associated with a vector of  $n-p^*$  linearly independent error contrasts is (up to an additive constant)

$$\begin{aligned} L_v &\equiv L_v(\sigma_1^2, \dots, \sigma_c^2, \sigma_{c+1}^2; \underline{y}) \\ &= L_1\left(\frac{\sigma_1^2}{\sigma_{c+1}^2}, \dots, \frac{\sigma_c^2}{\sigma_{c+1}^2}, \sigma_{c+1}^2; \underline{y}\right) . \end{aligned}$$

Expressions for the partial derivatives of  $L_v$  can be obtained from those of  $L_1$  by applying the chain rule of calculus. In particular, it follows from the chain rule that, for  $i = 1, \dots, c+1$ ,

$$\frac{\partial L_v}{\partial \sigma_i^2} = \sum_{j=1}^{c+1} \frac{\partial L_1}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \sigma_i^2}. \quad (2.31)$$

Since

$$\gamma_j = \begin{cases} \frac{\sigma_j^2}{\sigma_{c+1}^2}, & j = 1, \dots, c \\ \sigma_{c+1}^2, & j = c+1, \end{cases}$$

we have that, for  $i = 1, \dots, c$ ,

$$\left. \begin{aligned} \frac{\partial \gamma_j}{\partial \sigma_i^2} &= \begin{cases} \frac{1}{\sigma_{c+1}^2}, & j = i \\ 0, & \text{otherwise} \end{cases} \\ \frac{\partial \gamma_j}{\partial \sigma_{c+1}^2} &= \begin{cases} -\frac{\sigma_j^2}{\sigma_{c+1}^4}, & j = 1, \dots, c \\ 1, & j = c+1 \end{cases} \end{aligned} \right\} \quad (2.32)$$

1. General case

For  $\underline{g} \in \Omega_2^*$ , and  $i, j \in \{1, \dots, c\}$ ,

$$\begin{aligned} \frac{\partial L_v}{\partial \sigma_{c+1}^2} = & - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \left[ n - p^* - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \operatorname{tr}(G_{jj}) \right] \\ & + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - X\tilde{\alpha} - Z\tilde{b})' (\underline{y} - X\tilde{\alpha} - Z\tilde{b}), \end{aligned} \quad (2.33)$$

$$\frac{\partial L_v}{\partial \sigma_i^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \left[ \operatorname{tr}(G_{ii}) - \sigma_{c+1}^2 \tilde{y}_i' \tilde{y}_i \right], \quad (2.34)$$

$$\begin{aligned} (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = & - \frac{1}{\sigma_{c+1}^4} \left[ n - p^* - 2\underline{y}' S (\underline{y} - Z\tilde{b}) \right] \\ & - 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}_j' \tilde{y}_j \\ & + 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \operatorname{tr}(G_{jj}) \\ & + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left[ - \operatorname{tr}(G_{jk} G_{kj}) \right. \\ & \left. + 2 \sigma_{c+1}^2 \tilde{y}_j' G_{jk} \tilde{y}_k \right], \end{aligned} \quad (2.35)$$

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} &= - \frac{1}{\sigma_{c+1}^4} [\text{tr}(G_{ii}) - 2\sigma_{c+1}^2 \tilde{v}_i' \tilde{v}_i] \\
&\quad - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \left( \frac{1}{\sigma_{c+1}^2} \right) [- \text{tr}(G_{ij} G_{ji}) \\
&\quad + 2\sigma_{c+1}^2 \tilde{v}_i' G_{ij} \tilde{v}_j] , \quad (2.36)
\end{aligned}$$

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{1}{\sigma_{c+1}^4} [- \text{tr}(G_{ij} G_{ji}) + 2\sigma_{c+1}^2 \tilde{v}_i' G_{ij} \tilde{v}_j] , \\
&\quad \text{for } i \neq j , \quad (2.37)
\end{aligned}$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} = \frac{1}{\sigma_{c+1}^4} [- \text{tr}(G_{ii}^2) + 2\sigma_{c+1}^2 \tilde{v}_i' G_{ii} \tilde{v}_i] , \quad (2.38)$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^4} \text{tr}(G_{ii}^2) , \quad (2.39)$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^4} \text{tr}(G_{ii}) - \frac{1}{\sigma_{c+1}^6} \sum_{j=1}^c \sigma_j^2 \text{tr}(G_{ij} G_{ji}) , \quad (2.40)$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{\sigma_{c+1}^4} \text{tr}(G_{ij} G_{ji}) , \text{ for } i \neq j , \quad (2.41)$$

$$\begin{aligned}
(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] &= \frac{1}{\sigma_{c+1}^4} (n-p^*) - 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \operatorname{tr}(G_{jj}) \\
&\quad + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \operatorname{tr}(G_{jk} G_{kj}) , \tag{2.42}
\end{aligned}$$

as shown in Section A.1 of the Appendix.

## 2. Special case

$$\left. \begin{aligned}
&\text{If } \underline{\sigma} \in \Omega_2^* \text{ and } \sigma_j^2 \neq 0, \text{ then} \\
&G_{jj} = \sigma_{c+1}^2 \left( \frac{1}{\sigma_j^2} \right) (I - T_{jj}) , \\
&G_{ij} = - \sigma_{c+1}^2 \left( \frac{1}{\sigma_j^2} \right) T_{ij} ,
\end{aligned} \right\} \tag{2.43}$$

for  $i, j \in \{1, \dots, c\}$  with  $j \neq i$ . As shown in Section A.2 of the Appendix, the following expressions for the partial derivatives of  $L_v$  can be obtained by substituting expressions (2.43) into the formulas given in Section II.C.1. For  $\underline{\sigma} \in \Omega_2^*$ , and  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$\begin{aligned}
\frac{\partial L_v}{\partial \sigma_{c+1}^2} &= - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \operatorname{tr}(T)] \\
&\quad + \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{x}\tilde{\alpha} - \underline{z}\tilde{b})' (\underline{y} - \underline{x}\tilde{\alpha} - \underline{z}\tilde{b}) ,
\end{aligned}$$

$$\frac{L_v}{\partial \sigma_i^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{ii})] + \left(\frac{1}{2}\right) \tilde{v}_i' \tilde{v}_i, \text{ if } \sigma_i^2 \neq 0 ,$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \tilde{Z}\tilde{b})]$$

$$+ 2 \left(\frac{1}{\sigma_{c+1}^4}\right) \sum_{j=1}^c \left[\left(\frac{1}{2}\right) q_j - \sigma_j^2 \tilde{v}_j' \tilde{v}_j\right]$$

$$- \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{jk} T_{kj}) + 2\sigma_j^2 \tilde{v}_j' T_{jk} \tilde{v}_k],$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} = - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{ii})$$

$$+ 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \frac{1}{\sigma_i^2} \sum_{j=1}^c \left[\left(\frac{1}{2}\right) \text{tr}(T_{ij} T_{ji})\right]$$

$$+ \sigma_i^2 \tilde{v}_i' T_{ij} \tilde{v}_j] , \text{ if } \sigma_i^2 \neq 0 ,$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} = - \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{ij} T_{ji})$$

$$- 2 \left(\frac{1}{\sigma_j^2}\right) \tilde{v}_i' T_{ij} \tilde{v}_j , \text{ if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0 ,$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} = -\frac{1}{\sigma_i^4} \text{tr}[(I - T_{ii})^2]$$

$$+ 2\left(\frac{1}{\sigma_i^2}\right) \tilde{y}_i' (I - T_{ii}) \tilde{y}_i, \text{ if } \sigma_i^2 \neq 0,$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_i^4} \text{tr}[(I - T_{ii})^2], \text{ if } \sigma_i^2 \neq 0,$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [\text{tr}(T_{ii}) - \sum_{j=1}^c \text{tr}(T_{ij} T_{jk})],$$

$$\text{if } \sigma_i^2 \neq 0,$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{ij} T_{ji}), \text{ if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0,$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} [n - p^* - q + \text{tr}(T^2)].$$

#### D. Some Computational Considerations

##### 1. Diagonalization versus inversion

Many of the expressions given in Lemmas 2.5 and 2.7, for the partial derivatives of  $L_1$ , involve the inverse  $T$  of the  $q \times q$  matrix  $(I + Z'SZD) =$

$(I + \frac{1}{\sigma_{c+1}^2} CD)$ . In this section, we consider the problem of computing

the value of  $T$  corresponding to one or more values of  $\underline{\gamma}$ .

For  $i, j \in \{1, \dots, c\}$ , define

$$C_{ij} \equiv Z_i'(I - P_X)Z_j$$

to be the  $(i, j)$ -th submatrix of the matrix  $C = Z'(I - P_X)Z$ . Observe that

$$T = (I + \frac{1}{\sigma_{c+1}^2} CD)^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1},$$

where

$$B_{11} = I + \begin{pmatrix} \gamma_1 C_{11} & \cdots & \gamma_{c-1} C_{1,c-1} \\ \vdots & & \\ \gamma_1 C_{c-1,1} & \cdots & \gamma_{c-1} C_{c-1,c-1} \end{pmatrix},$$

$$B_{12} = (\gamma_c C'_{1c}, \dots, \gamma_c C'_{c-1,c})',$$

$$B_{21} = (\gamma_1 C_{c1}, \dots, \gamma_{c-1} C_{c,c-1}),$$

$$B_{22} = I + \gamma_c C_{cc}.$$

Lemma 2.8: The matrix

$$F_i \equiv \begin{pmatrix} I + \gamma_i C_{ii} & \gamma_{i+1} C_{i,i+1} & \cdots & \gamma_c C_{ic} \\ \gamma_i C_{i+1,i} & I + \gamma_{i+1} C_{i+1,i+1} & \cdots & \gamma_c C_{i+1,c} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_i C_{ci} & \gamma_{i+1} C_{c,i+1} & \cdots & I + \gamma_c C_{cc} \end{pmatrix}$$

is nonsingular for  $\underline{\gamma} \in \Omega_1^*$  ( $i = 1, \dots, c$ ).

Proof: For  $\underline{\gamma} \in \Omega_1^*$ ,

$$\begin{aligned} |F_i| &= \left| I + \begin{pmatrix} \gamma_i C_{ii} & \cdots & \gamma_c C_{ic} \\ \vdots & & \vdots \\ \gamma_i C_{ci} & \cdots & \gamma_c C_{cc} \end{pmatrix} \right| \\ &= \left| I + \begin{pmatrix} C_{ii} & \cdots & C_{ic} \\ \vdots & & \vdots \\ C_{ci} & \cdots & C_{cc} \end{pmatrix} \text{diag}\{\gamma_i I_{q_i}, \dots, \gamma_c I_{q_c}\} \right| \\ &= \left| I + \frac{1}{\sigma_{c+1}^2} (\emptyset, I) \text{CD} \begin{pmatrix} \emptyset \\ I \end{pmatrix} \right| \\ &= \left| I + \frac{1}{\sigma_{c+1}^2} \text{CD} \begin{pmatrix} \emptyset \\ I \end{pmatrix} (\emptyset, I) \right| \end{aligned}$$

$$= \left| I + \frac{1}{\sigma_{c+1}^2} CD_i \right|$$

$$= |I + Z'SZD_i|$$

$$\neq 0 \text{ (using Lemma 2.3).}$$

□

Let  $r = \text{rank}(X, Z_c) - \text{rank}(X)$ . Then,

$$\text{rank}(C_{cc}) = \text{rank}[Z_c'(I - P_X)Z_c]$$

$$= \text{rank}(X, Z_c) - \text{rank}(X)$$

$$= r.$$

Since  $C_{cc}$  is a  $q_c \times q_c$  nonnegative definite matrix of rank  $r$ ,  $C_{cc}$  has exactly  $r$  positive characteristic roots and  $q_c - r$  zero characteristic roots. Let  $\Delta_1, \dots, \Delta_r$  denote the  $r$  nonzero characteristic roots of  $C_{cc}$ , and take  $\underline{\Delta} \equiv \text{diag}(\Delta_1, \dots, \Delta_r)$ . Further, let  $P^* = (R^*, U^*)$  be a  $q \times q$  orthogonal matrix, where  $R^*$  is a  $q_c \times r$  matrix whose columns are orthonormal characteristic vectors of  $C_{cc}$  corresponding to the roots  $\Delta_1, \dots, \Delta_r$ , respectively, and  $U^*$  is a  $q_c \times (q_c - r)$  matrix whose columns are orthonormal characteristic vectors of  $C_{cc}$  corresponding to the zero roots. Using the relationships  $C_{cc} R^* = R^* \underline{\Delta}$  and  $R^{*'} R^* = I$ , we have that

$$\left. \begin{aligned}
 R^{*'} C_{cc} R^* &= R^{*'} R^* \underline{\Delta} = \underline{\Delta} , \\
 C_{cc}^2 R^* &= C_{cc} R^* \underline{\Delta} = R^* \underline{\Delta}^2 \\
 R^{*'} C_{cc}^2 R^* &= \underline{\Delta}^2 .
 \end{aligned} \right\} \quad (2.44)$$

According to Lemma 2.8,  $I + \gamma_c C_{cc}$  is nonsingular for  $\gamma \in \Omega_1^*$ .

This result can be strengthened as follows:

Lemma 2.9: The matrix  $I + \gamma_c C_{cc}$  is positive definite for  $\gamma \in \Omega_1^*$ .

Proof: For  $\gamma \in \Omega_1^*$ ,  $I + \gamma_c Z_c Z_c'$  is, by definition, positive definite.

Let  $\lambda_1, \dots, \lambda_s$  denote the nonzero characteristic roots of  $Z_c Z_c'$ .

Since  $Z_c Z_c'$  is nonnegative definite,  $\lambda_i > 0$  ( $i = 1, \dots, s$ ). Let

$\lambda^* = \max\{\lambda_1, \dots, \lambda_s\}$ . The matrix  $I + \gamma_c Z_c Z_c'$  has characteristic roots  $1 + \gamma_c \lambda_1, \dots, 1 + \gamma_c \lambda_s$  and  $n-s$  roots of unity. Thus,  $I + \gamma_c Z_c Z_c'$  is a positive definite matrix if and only if  $1 + \gamma_c \lambda_i > 0$  for  $i = 1, \dots, s$ , i.e., if and only if  $\gamma_c > -\frac{1}{\lambda^*}$ . This implies that  $\gamma_c > -\frac{1}{\lambda^*}$  for  $\gamma \in \Omega_1^*$ .

Let  $\Delta^* = \max\{\Delta_1, \dots, \Delta_r\}$ . Then,  $I + \gamma_c C_{cc}$  is a positive definite matrix if and only if  $\gamma_c > -\frac{1}{\Delta^*}$ . Clearly, to show that  $I + \gamma_c C_{cc}$  is positive definite for  $\gamma \in \Omega_1^*$ , it suffices to show that  $-\frac{1}{\Delta^*} \leq -\frac{1}{\lambda^*}$ .

Following Harville and Fenech (1985), consider the matrix

$$A^* \equiv \underline{\Delta}^{-\frac{1}{2}} R^{*'} Z'_c (I - P_X) (I + \gamma_c Z_c Z'_c) (I - P_X) Z_c R^* \underline{\Delta}^{-\frac{1}{2}}.$$

If  $I + \gamma_c Z_c Z'_c$  is nonnegative definite, then  $A^*$  has a representation of the form  $\Gamma' \Gamma$  and, hence, is also nonnegative definite. Moreover, using (2.44), we find that

$$\begin{aligned} A^* &= \underline{\Delta}^{-\frac{1}{2}} R^{*'} C_{cc}^* \underline{\Delta}^{-\frac{1}{2}} + \gamma_c \underline{\Delta}^{-\frac{1}{2}} R^{*'} C_{cc}^2 R^* \underline{\Delta}^{-\frac{1}{2}} \\ &= I + \gamma_c \underline{\Delta}, \end{aligned}$$

so that  $A^*$  is nonnegative definite if and only if  $\gamma_c > -\frac{1}{\lambda^*}$ . Since  $I - (\frac{1}{\lambda^*}) Z_c Z'_c$  is nonnegative definite, it follows that  $-\frac{1}{\Delta^*} \leq -\frac{1}{\lambda^*}$ .  $\square$

Suppose that  $\underline{\gamma} \in \Omega_1^*$ . Since  $T$  and  $B_{22} = I + \gamma_c C_{cc}$  are nonsingular matrices,

$$\left| B_{11} - B_{12} B_{22}^{-1} B_{21} \right| = \frac{|T|}{|B_{22}|} \neq 0,$$

i.e.,  $B_{11} - B_{12} B_{22}^{-1} B_{21}$  is also a nonsingular matrix. Thus,

$$(I + \frac{1}{\sigma_{c+1}^2} CD)^{-1} = \begin{pmatrix} E^{-1} & -E^{-1} B_{12} B_{22}^{-1} \\ -B_{22}^{-1} B_{21} E^{-1} & B_{22}^{-1} + B_{22}^{-1} B_{21} E^{-1} B_{12} B_{22}^{-1} \end{pmatrix}, \quad (2.45)$$

where  $E \equiv B_{11} - B_{12} B_{22}^{-1} B_{21}$ .

Consider the problem of computing the inverse of  $B_{22}$ . Clearly, this problem is equivalent to that of computing the solution to the linear system  $M_1 N = M_2$ , where  $M_1 = B_{22}$  and  $M_2 = I$ .

Numerical techniques for solving a linear system  $M_1 N = M_2$  (in  $N$ ) can be classified as direct or iterative. In the absence of roundoff error, direct methods produce an exact solution in a finite number of arithmetic operations. Iterative methods begin with an initial approximation to the solution and repeatedly update the approximation until the process is terminated by the user. In choosing a method for solving  $M_1 N = M_2$ , the characteristics of  $M_1$  should be considered. If  $M_1$  has fewer than 100 rows and columns, and has few zero elements, then Conte and de Boor (1972) recommend a direct method. If  $M_1$  is very large, then an iterative method may be the only recourse.

The matrix  $B_{22}$  has dimensions  $q_c \times q_c$  and generally has few zero elements. Further, it is symmetric and positive definite for  $\underline{\gamma} \in \Omega_1^*$ . A computationally stable and direct method for solving a linear system whose coefficient matrix is positive definite is the Cholesky decomposition [e.g., Kennedy and Gentle (1980, p. 294)].

The matrix  $B_{22}$  depends on  $\underline{\gamma}$  (through  $\gamma_c$ ), which we sometimes emphasize by writing  $B_{22}(\underline{\gamma})$  for  $B_{22}$ . The Cholesky decomposition would seem to be a suitable procedure for computing  $B_{22}(\underline{\gamma})$  for a single value of  $\gamma_c$ . However, in computing  $\hat{\underline{\gamma}}$  via an iterative algorithm,  $B_{22}(\underline{\gamma})$  generally has to be computed for more than one such value. In this case, the use of the Cholesky decomposition may be inefficient. We

now describe an alternative approach which is due to Dempster et al. (1984) and is described by Harville and Fenech (1985).

Observe that

$$P^{*'}(I + \gamma_c C_{cc})P^* = \text{diag}\{I + \gamma_c \underline{\Delta}, I\}.$$

Thus

$$B_{22}^{-1}(\underline{\gamma}) = R^*(I + \gamma_c \underline{\Delta})^{-1}R^{*'} + U^*U^{*'} \quad (2.46)$$

If  $B_{22}^{-1}(\underline{\gamma})$  were to be computed for just one value of  $\gamma_c$ , it would be considerably more efficient to use the Cholesky decomposition than to base the computations on expression (2.46). However, if  $B_{22}^{-1}(\underline{\gamma})$  is to be computed for a sufficiently large number of different values of  $\gamma_c$ , basing the computations on expression (2.46) would be more efficient than using the Cholesky decomposition for reasons which we now discuss.

The usefulness of (2.46) for computing  $B_{22}^{-1}(\underline{\gamma})$  for each of a large number of  $\gamma_c$  values lies in the fact that  $\underline{\Delta}$ ,  $R^*$ , and  $U^*$  do not depend on  $\underline{\gamma}$ . Once  $\underline{\Delta}$ ,  $R^*$ , and  $U^*$  are computed,  $B_{22}^{-1}(\underline{\gamma})$  can, by using expression (2.46), be easily computed for a large number of  $\gamma_c$  values.

Once  $B_{22}^{-1}(\underline{\gamma})$  is computed, the  $q \times q$  matrix inverse  $(I + \frac{1}{\sigma_c^2 + 1} CD)^{-1}$  can be computed by computing the  $(q - q_c) \times (q - q_c)$  matrix inverse  $E^{-1}$  and applying (2.45). Thus, if  $\underline{\Delta}$ ,  $R^*$ , and  $U^*$  are computed, then the subsequent computations required to compute  $(I + \frac{1}{\sigma_c^2 + 1} CD)^{-1}$  for any

number of different  $\underline{\gamma}$  values are, essentially, those required to compute  $E^{-1}$  for the same  $\underline{\gamma}$  values. If the number of different  $\underline{\gamma}$  values for which  $(I + \frac{1}{\sigma_{c+1}^2} CD)^{-1}$  is to be computed is large, then it may be computationally advantageous to base the computation of  $(I + \frac{1}{\sigma_{c+1}^2} CD)^{-1}$  on formulas (2.45) and (2.46) rather to invert  $(I + \frac{1}{\sigma_{c+1}^2} CD)$  directly using a method like the Cholesky decomposition.

The question arises as to whether the computations required to form the matrix inverse  $(I + \frac{1}{\sigma_{c+1}^2} CD)^{-1}$  can be further reduced. Partition  $E$  as

$$E = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix},$$

where  $E_{22} = (I + \gamma_{c-1} C_{c-1, c-1}) - (\gamma_c C_{c-1, c})(I + \gamma_c C_{cc})^{-1}(\gamma_{c-1} C_{c, c-1})$

$$= I + \gamma_{c-1} [C_{c-1, c-1} - \gamma_c C_{c-1, c} (I + \gamma_c C_{cc})^{-1} C_{c, c-1}] .$$

It follows from Lemma 2.8 that, for  $\underline{\gamma} \in \Omega_1^*$ ,

$$|E_{22}| = |I + \gamma_c C_{cc}|^{-1} \left| \begin{pmatrix} I + \gamma_{c-1} C_{c-1, c-1} & \gamma_c C_{c-1, c} \\ \gamma_{c-1} C_{c, c-1} & I + \gamma_c C_{cc} \end{pmatrix} \right|$$

$$\neq 0 ,$$

that is,  $E_{22}$  is nonsingular. We find that

$$E^{-1} = \begin{pmatrix} F_0^{-1} & -F_0^{-1}E_{12}E_{22}^{-1} \\ -E_{22}^{-1}E_{21}F_0^{-1} & E_{22}^{-1} + E_{22}^{-1}E_{21}F_0^{-1}E_{12}E_{22}^{-1} \end{pmatrix}, \quad (2.47)$$

where  $F_0 \equiv E_{11} - E_{12}E_{22}^{-1}E_{21}$ .

In contrast to the matrix  $B_{22}$  which can be diagonalized by a matrix (namely, the matrix  $P^*$ ) that does not depend on  $\underline{\gamma}$ , any orthogonal matrix  $Q$  such that  $Q'E_{22}Q$  is diagonal will, in general, be functionally dependent of  $\gamma_c$ . Thus, in computing  $E^{-1}$  and, then,  $(I + \frac{1}{\sigma_{c+1}^2} CD)^{-1}$  for a number of different  $\underline{\gamma}$  values, it would be computationally advantageous to use formula (2.47) and diagonalize  $E_{22}$  if  $\gamma_c$  were the same for all the  $\underline{\gamma}$  values but, in general, it would not be advantageous.

## 2. Choice of mixed-model equations

Earlier, we discussed the relationship between the MME (2.18) and  $\tilde{\underline{y}}$  and the solution to  $X'V^{-1}X\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$ . The MME (2.18) are one in a class of MME discussed by Harville (1976, p. 393). This class consists of linear systems of the general form

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}ZST_0 \\ T_0UZ'R^{-1}X & T_0 + T_0UZ'R^{-1}ZST_0 \end{pmatrix} \begin{pmatrix} \underline{\alpha}^* \\ \underline{\psi}^* \end{pmatrix} = \begin{pmatrix} X'R^{-1}\underline{y} \\ T_0UZ'R^{-1}\underline{y} \end{pmatrix}, \quad (2.48)$$

where  $D = ST_0U$ . Putting  $S = D$  and  $T_0 = U = I_q$ , we obtain equations (2.18) as a special case. As discussed by Harville,  $S$ ,  $T_0$  and  $U$  should be chosen so that the equations (2.48) are "well-conditioned and, at the same time, easy to form and solve." For our purposes, they should also be chosen in such a way that it is possible to recover  $\tilde{V}$  from their solution (since  $\tilde{V}$  is an integral part of our expressions for  $L_1$  and  $L_V$ , and for their derivatives).

For  $\underline{\gamma} \in \Omega_1^*$  such that  $D$  is positive definite, Henderson (1963) proposed the set of MME corresponding to  $S = U = D$  and  $T_0 = D^{-1}$ . Henderson showed that if  $\underline{\alpha}^*$  and  $\underline{\psi}^*$  are the first and second components of any solution to these equations, then  $\underline{\alpha}^*$  satisfies  $X'V^{-1}X\underline{\alpha}^* = X'V^{-1}\underline{y}$  and  $\underline{\psi}^* = D\tilde{V} = \tilde{b}$ . When  $S = U = D$  and  $T_0 = D^{-1}$ , the coefficient matrix of the MME is symmetric and nonnegative definite, which can be computationally advantageous. [See, for example, Westlake (1968).]

The MME (2.18) are applicable for all  $\underline{\gamma} \in \Omega_1^*$ . Moreover, as discussed in Section II.A,  $\tilde{V}$  is a component of any solution. However, the coefficient matrix for this system is not symmetric.

### III. ITERATIVE ALGORITHMS FOR COMPUTING REML ESTIMATES

To obtain a REML estimate of the vector  $\underline{\gamma}$  in model (1.1), we must find a value of  $\underline{\gamma}$  at which  $L_1(\underline{\gamma}; \underline{y})$  attains its supremum over the parameter space. In general, explicit expressions for a REML estimator do not exist, though W. A. Thompson (1962) obtained such expressions in the special case where the model is balanced and the parameter space for  $\underline{\gamma}$  is  $\Omega_1$ .

In this chapter, we present fourteen algorithms that can be used to determine a REML estimate of  $\underline{\gamma}$ . To simplify the presentation, we initially (in this chapter) ignore complications occasioned by the algorithm giving rise to an iterate that lies outside the parameter space. (These complications are considered in Chapter IV.) Each algorithm is a variation on one of four general iterative methods: the EM algorithm (Section III.A), the Newton-Raphson method (Section III.B), the Method of Scoring (Section III.C), and the Method of Successive Approximations (Section III.D).

In Section III.E, we collect the fourteen algorithms in a single table.

In Section III.F, we give simplified expressions for each of the fourteen algorithms, applicable to the special case  $c = 1$ .

#### A. The EM Algorithm

Dempster et al. (1977), presented the EM algorithm as a general iterative method for computing maximum likelihood estimates from "incomplete" data. Wu (1983) presented it in a more general context,

viewing it as a "special optimization algorithm." In Sections III.A.1-III.A.3, we apply the EM algorithm to the REML estimation of  $\underline{\sigma}$  in the mixed linear model (1.2).

In Section III.A.1, we discuss the concepts of incomplete and complete data as applied to this mixed linear model. We also discuss the intuitive appeal of the EM algorithm and review some of its convergence properties. In Sections III.A.2-III.A.3 we present the iterates for two different implementations of the EM algorithm. In Section III.A.4, we consider whether the iterates generated by the EM algorithm belong to the parameter space  $\Omega_2$ .

#### 1. General description and application to the mixed linear model

We discuss the EM algorithm as applied to the computation of a REML estimate of  $\underline{\gamma}$  under model (1.2). Recalling the invariance property of maximum likelihood estimates, note that if  $(\hat{\sigma}_1^2, \dots, \hat{\sigma}_{c+1}^2)'$  is a REML estimate of  $\underline{\gamma}$  under model (1.2), then  $(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_{c+1}^2}, \dots, \frac{\hat{\sigma}_c^2}{\hat{\sigma}_{c+1}^2}, \hat{\sigma}_{c+1}^2)$  is a REML estimate of  $\underline{\gamma}$  under that model.

As discussed in Section II.A, a REML estimate of  $\underline{\sigma}$  can be obtained by maximizing the likelihood function associated with the observable random vector  $A'\underline{y}$ , where  $A$  is a  $n \times (n-p^*)$  matrix which satisfies  $A'A = I$  and  $AA' = I - P_X$ . Recall that  $A'X = \emptyset$ , and observe that a REML estimate of  $\underline{\sigma}$  under model (1.2) is the same as a maximum likelihood estimate of

$\underline{\sigma}$  under the completely random linear model

$$A'\underline{y} = A'Z\underline{b} + A'\underline{e} . \quad (3.1)$$

Note that the model equation (3.1) can be re-written as

$$A'\underline{y} = (A'Z, A') \begin{pmatrix} \underline{b} \\ \underline{e} \end{pmatrix} .$$

To compute a maximum likelihood estimate of  $\underline{\sigma}$  based on model (3.1), we can apply the EM algorithm of Dempster et al. (1977). This algorithm is a general iterative algorithm for computing a maximum likelihood estimate from "incomplete" data.

We now discuss the concepts of complete data and incomplete data, as applied to model (3.1). Define

$$\underline{x}_i = \begin{cases} \underline{b}_i & , \quad i = 1, \dots, c , \\ \underline{e} & , \quad i = c+1 \end{cases}$$

and take  $\underline{x} \equiv (\underline{x}'_1, \dots, \underline{x}'_{c+1})' = (\underline{b}', \underline{e}')'$  . We have that

$$\underline{x} \stackrel{d}{\sim} N_{q+n} \left[ \begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix} , \begin{pmatrix} D & \emptyset \\ \emptyset & R \end{pmatrix} \right]$$

so that the probability density function for  $\underline{x}$  is

$$\begin{aligned}
 g(\underline{x}; \underline{\sigma}) &\equiv (2\pi)^{-\frac{1}{2}(q+n)} |D|^{-\frac{1}{2}} |R|^{-\frac{1}{2}} \\
 &\cdot \exp\left\{-\left(\frac{1}{2}\right) \underline{x}' \begin{pmatrix} D & \emptyset \\ \emptyset & R \end{pmatrix}^{-1} \underline{x}\right\} \\
 &= (2\pi)^{-\frac{1}{2}(q+n)} \left[ \prod_{i=1}^c (\sigma_i^2)^{q_i} \right]^{-\frac{1}{2}} \left[ (\sigma_{c+1}^2)^n \right]^{-\frac{1}{2}} \\
 &\cdot \exp\left\{-\left(\frac{1}{2}\right) \sum_{i=1}^{c+1} \left(\frac{1}{\sigma_i^2}\right) \underline{x}_i' \underline{x}_i\right\}, \quad \underline{\sigma} \in \Omega_2. \quad (3.2)
 \end{aligned}$$

Let  $\phi_i = -\left(\frac{1}{2}\right) \frac{1}{\sigma_i^2}$  and  $t_i(\underline{x}) = \underline{x}_i' \underline{x}_i$  ( $i = 1, \dots, c+1$ ), and take

$$\underline{\phi} \equiv (\phi_1, \dots, \phi_{c+1})', \text{ and}$$

$$\underline{t}(\underline{x}) \equiv (\underline{x}_1' \underline{x}_1, \dots, \underline{x}_{c+1}' \underline{x}_{c+1})'.$$

The probability density function for  $\underline{x}$  can be rewritten in the form

$$g(\underline{x}; \underline{\phi}) = k(\underline{\phi}) b(\underline{x}) \exp\{\underline{\phi}' \underline{t}(\underline{x})\},$$

where

$$k(\underline{\phi}) = (2\pi)^{-\frac{1}{2}(q+n)} \left[ \prod_{i=1}^c (\sigma_i^2)^{q_i} \right]^{-\frac{1}{2}} \left[ (\sigma_{c+1}^2)^n \right]^{-\frac{1}{2}}, \text{ and}$$

$$\underline{b}(\underline{x}) = 1.$$

In terms of  $\underline{\phi}$ , the parameter space is

$$\Omega_{\underline{\phi}} \equiv \{ \underline{\phi} : -\infty < \phi_i \leq 0 \quad (i = 1, \dots, c+1) \}.$$

Thus, when re-expressed in terms of  $\underline{\phi}$ , the distribution of  $\underline{x}$  is of the regular exponential family form and, if  $\underline{x}$  were observable,  $\underline{t}(\underline{x})$  would be a vector of complete sufficient statistics for  $\underline{\phi}$  [e.g., Ferguson (1967)]. Further, if  $\underline{x}$  were observable, the likelihood equations for estimating  $\underline{\phi}$  from  $\underline{x}$  would be

$$E[\underline{t}(\underline{x})] = \underline{t}(\underline{x})$$

[e.g., Sundberg (1974)].

In reality, the vector  $\underline{x}$  is unobservable, however, the vector  $A'\underline{y}$ , whose elements are linear functions of the elements of  $\underline{x}$ , is observable. In the context of the EM algorithm, it is customary to regard the elements of  $\underline{x}$  as the "complete data" and those of  $A'\underline{y}$  as the "incomplete" data.

Each iteration of the EM algorithm involves two steps - the E-step (Expectation step) and the M-step (Maximization step). The E-step consists of estimating  $\underline{t}(\underline{x})$ , which would be a complete sufficient statistic if the vector  $\underline{x}$  of complete data were available, by its conditional mean given the vector  $A'\underline{y}$  of incomplete data, i.e., by the vector  $\underline{\tau} \equiv \mathbb{E}[\underline{t}(\underline{x})|A'\underline{y}]$ . In carrying out the E-step, i.e., in computing  $\mathbb{E}[\underline{t}(\underline{x})|A'\underline{y}]$ , the current iterate for  $\underline{\sigma}$  is regarded as the value of  $\underline{\sigma}$ . The M-step consists of solving the equations  $\mathbb{E}[\underline{t}(\underline{x})] = \underline{\tau}$  for  $\underline{\sigma}$ . The equations that are solved in the M-step can be regarded as an approximation to what would be the likelihood equations if the vector  $\underline{x}$  of complete data were observable.

The EM algorithm begins with the specification of an initial guess for  $\underline{\sigma}$ , say  $\underline{\sigma}^{(0)}$ . Denoting by  $\underline{\sigma}^{(p)}$  the guess for  $\underline{\sigma}$  computed on the  $p$ -th iteration, the  $(p+1)$ -st iteration of the EM algorithm consists of:

E-step: Find  $\mathbb{E}[\underline{t}(\underline{x})|A'\underline{y}]$ , acting as though  
 $\underline{\sigma} = \underline{\sigma}^{(p)}$ , and call this expected value  $\underline{\tau}^{(p)}$  ;

M-step: Solve  $\mathbb{E}[\underline{t}(\underline{x})] = \underline{\tau}^{(p)}$  for  $\underline{\sigma}$  and set  $\underline{\sigma}^{(p+1)}$   
 equal to the solution.

The algorithm proceeds until some user-specified termination criteria are satisfied. For example, the algorithm might be terminated when

successive iterates differ from each other by no more than some number  $\varepsilon (\varepsilon > 0)$ .

Several authors have studied the convergence properties of the EM algorithm [e.g., Sundberg (1976), Dempster et al. (1977), Broyles (1983), and Wu (1983)]. It follows from the results of Dempster et al. (1977) that, for any sequence of iterates  $\{\underline{\sigma}^{(p)}\}$  generated by (3.3), the corresponding sequence of function values  $\{L_V(\underline{\sigma}^{(p)}; \underline{y})\}$  is nondecreasing. Moreover, it follows from the results of Wu (1983) that, if there exists at most a countable number of local maxima of  $L_V(\underline{\sigma}; \underline{y})$  having the same  $L_V$  value, then the sequence  $\{\underline{\sigma}^{(p)}\}$  converges to a local maximum  $\hat{\underline{\sigma}}$  of  $L_V(\underline{\sigma}; \underline{y})$  and the sequence  $\{L_V(\underline{\sigma}^{(p)}; \underline{y})\}$  converges monotonically to  $L_V(\hat{\underline{\sigma}}; \underline{y})$ .

In general, there is no guarantee that a limit point  $\underline{\sigma}^*$  of  $\{\underline{\sigma}^{(p)}\}$  is a global maximum of  $L_V(\underline{\sigma}; \underline{y})$ , i.e., a REML estimate of  $\underline{\sigma}$ . Since the limit point of the sequence  $\{\underline{\sigma}^{(p)}\}$  may depend on the choice of the starting value  $\underline{\sigma}^{(0)}$ , it may be desirable, in applying the EM algorithm, to try several different starting values and to then compare the values of  $L_V$  at the resulting limit points.

We next construct the iterates generated by algorithm (3.3), for two different interpretations of the vector  $\underline{x}$  of complete data.

## 2. Computation of the iterates

In carrying out the E-step of EM algorithm (3.3) for computing a REML estimate of  $\underline{\sigma}$ , we require expressions for

$$\begin{aligned}
\mathbb{E}(\underline{x}_i' \underline{x}_i | A' \underline{y}) &= \mathbb{E}(\underline{b}_i' \underline{b}_i | A' \underline{y}) \\
&= [\mathbb{E}(\underline{b}_i | A' \underline{y})]' [\mathbb{E}(\underline{b}_i | A' \underline{y})] \\
&\quad + \text{tr}[\text{Var}(\underline{b}_i | A' \underline{y})] \quad (i = 1, \dots, c).
\end{aligned} \tag{3.4}$$

Under model (1.2),

$$\begin{pmatrix} \underline{b} \\ A' \underline{y} \end{pmatrix} \stackrel{d}{\sim} N_{n-p+q} \left[ \begin{pmatrix} \underline{0} \\ A' Z D \end{pmatrix}, \begin{pmatrix} D & DZ'A \\ A'ZD & A'VA \end{pmatrix} \right], \quad \underline{\sigma} \in \Omega_2.$$

Using standard properties of the multivariate normal distribution, we find that

$$\begin{aligned}
\mathbb{E}(\underline{b} | A' \underline{y}) &= DZ'A(A'VA)^{-1}A'\underline{y} \\
&= DZ'P_{\underline{y}} \quad [\text{using part (i) of Lemma 2.1}] \\
&= DZ'V^{-1}(\underline{y} - X\tilde{\alpha}) \quad [\text{using part (iii) of Lemma 2.4}] \\
&= \tilde{\underline{b}},
\end{aligned} \tag{3.5}$$

and since

$$\begin{aligned}
\text{Cov}(\underline{b}, \tilde{\underline{b}}) &= \text{Cov}(\underline{b}, \underline{DZ}'\underline{Py}) \\
&= \text{Cov}(\underline{b}, \underline{DZ}'\underline{PX}\underline{\alpha} + \underline{DZ}'\underline{PZ}\underline{b} + \underline{DZ}'\underline{Pe}) \\
&= \text{Var}(\underline{b})\underline{Z}'\underline{PZD} \\
&= \underline{DZ}'\underline{PZD} \\
&= \underline{DZ}'\underline{PVP}'\underline{ZD} \text{ [using part (iv) of Lemma 2.1]} \\
&= \text{Var}(\underline{DZ}'\underline{Py}) \\
&= \text{Var}(\tilde{\underline{b}}),
\end{aligned}$$

that

$$\begin{aligned}
\text{Var}(\underline{b}|\underline{A}'\underline{y}) &= \underline{D} - \underline{DZ}'\underline{A}(\underline{A}'\underline{VA})^{-1}\underline{A}'\underline{ZD} \\
&= \underline{D} - \underline{DZ}'\underline{PZD} \text{ [using part (i) of Lemma 2.1]} \tag{3.6} \\
&= \underline{D} - \underline{DZ}'\underline{PVP}'\underline{ZD} \\
&= \text{Var}(\underline{b}) - \text{Var}(\tilde{\underline{b}}) \\
&= \text{Var}(\underline{b} - \tilde{\underline{b}}).
\end{aligned}$$

Substituting from expressions (3.5) and (3.6) into expression (3.4), we find that

$$E(\underline{x}'_i \underline{x}_i | A' \underline{y}) = \tilde{\underline{b}}'_i \tilde{\underline{b}}_i + \text{tr}(H_{ii}) \quad (i = 1, \dots, c) ,$$

where

$$\tilde{\underline{b}} = (\tilde{\underline{b}}'_1, \dots, \tilde{\underline{b}}'_c)' ,$$

$$D - DZ'PZD = \begin{pmatrix} H_{11} & \dots & H_{1c} \\ \vdots & & \vdots \\ H_{c1} & \dots & H_{cc} \end{pmatrix}$$

with  $\tilde{\underline{b}}_i$  having dimensions  $q_i \times 1$  and  $H_{ij}$  having dimensions  $q_i \times q_j$ . Moreover, since [according to part (ii) of Lemma 2.4]  $D - DZ'PZD = D - DTZ'SZD$ ,

$$\begin{aligned} \text{tr}(H_{ii}) &= \text{tr}(\sigma_i^2 \mathbf{I}_{q_i}) \\ &= \text{tr}(\sigma_i^2 \sum_{j=1}^c \mathbf{T}_{ij} \mathbf{Z}'_j \mathbf{S} \mathbf{Z}_i \sigma_i^2) \\ &= q_i \sigma_i^2 - \sigma_i^2 \text{tr}[\sigma_i^2 (\frac{1}{\sigma_{c+1}^2}) \mathbf{G}_{ii}] \end{aligned}$$

$$= q_i \sigma_i^2 - \sigma_i^2 \operatorname{tr}(I_{q_i} - T_{ii}) \text{ [using Lemma 2.6]}$$

$$= \sigma_i^2 \operatorname{tr}(T_{ii}) \quad (i = 1, \dots, c).$$

Thus

$$\mathbb{E}(\underline{x}'_{-i} \underline{x}_i | A' \underline{y}) = \tilde{\underline{b}}_{-i}' \tilde{\underline{b}}_i + \sigma_i^2 \operatorname{tr}(T_{ii}) \quad (i = 1, \dots, c). \quad (3.7)$$

Since  $t_i(\underline{x}) = \underline{b}_{-i}' \underline{b}_i$ ,

$$\begin{aligned} \mathbb{E}[t_i(\underline{x})] &= [\mathbb{E}(\underline{b}_{-i})]' [\mathbb{E}(\underline{b}_i)] + \operatorname{tr}[\operatorname{Var}(\underline{b}_i)] \\ &= 0 + \operatorname{tr}(\sigma_i^2 I_{q_i}) \\ &= q_i \sigma_i^2. \end{aligned} \quad (3.8)$$

Now, let  $\sigma_i^{2(p)}$  represent the  $i$ -th component of  $\underline{\sigma}^{(p)}$ . Taking  $\tilde{\underline{b}}_i^{(p)}$  and  $T_{ii}^{(p)}$  to be the values of  $\underline{b}_i$  and  $T_{ii}$  ( $i = 1, \dots, c$ ), respectively, at  $\underline{\sigma} = \underline{\sigma}^{(p)}$ , we conclude, on the basis of expressions (3.7) and (3.8), that

$$\sigma_i^{2(p+1)} = \frac{\tilde{\underline{b}}_{-i}^{(p)'} \tilde{\underline{b}}_i^{(p)} + \sigma_i^{2(p)} \operatorname{tr}(T_{ii}^{(p)})}{q_i} \quad (p = 0, 1, \dots). \quad (3.9)$$

Consider now the computation of  $\sigma_{c+1}^{2(p+1)}$ , the  $(c+1)$ -st component of  $\underline{\sigma}^{(p+1)}$ . Observe that

$$\begin{aligned} \mathbb{E}(\underline{e}'\underline{e}|\underline{A}'\underline{y}) &= [\mathbb{E}(\underline{e}|\underline{A}'\underline{y})]' [\mathbb{E}(\underline{e}|\underline{A}'\underline{y})] \\ &\quad + \text{tr}[\text{Var}(\underline{e}|\underline{A}'\underline{y})], \end{aligned}$$

and that, under model (1.2),

$$\begin{pmatrix} \underline{e} \\ \underline{A}'\underline{y} \end{pmatrix} \stackrel{d}{\sim} N_{2n-p*} \left[ \begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} \underline{R} & \underline{R}\underline{A} \\ \underline{A}'\underline{R} & \underline{A}'\underline{V}\underline{A} \end{pmatrix} \right], \quad \underline{\sigma} \in \Omega_2.$$

Thus,

$$\begin{aligned} \mathbb{E}(\underline{e}|\underline{A}'\underline{y}) &= \underline{R}\underline{A}(\underline{A}'\underline{V}\underline{A})^{-1}\underline{A}'\underline{y} \\ &= \underline{R}\underline{P}\underline{y} \quad [\text{using part (i) of Lemma 2.1}] \\ &= \underline{R}\underline{V}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}}) \quad [\text{using (2.4)}] \\ &= \underline{R}\underline{R}^{-1}(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) \quad [\text{using part (v) of Lemma 2.4}] \\ &= \underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}, \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} \text{Var}(\underline{e}|\underline{A}'\underline{y}) &= \underline{R} - \underline{R}\underline{A}(\underline{A}'\underline{V}\underline{A})^{-1}\underline{A}'\underline{R} \\ &= \underline{R} - \underline{R}\underline{P}\underline{R} \quad [\text{using part (i) of Lemma 2.1}]. \end{aligned} \tag{3.11}$$

It follows that

$$\begin{aligned} E(\underline{e}'\underline{e}|A'\underline{y}) &= (\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}})'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) \\ &\quad + \text{tr}(\underline{R} - \underline{RPR}). \end{aligned}$$

Moreover,

$$\text{tr}(\underline{R} - \underline{RPR}) = \text{tr}\{\underline{R} - \underline{R}[\underline{S} - \underline{SZD}(\underline{I} + \underline{Z}'\underline{SZD})^{-1}\underline{Z}'\underline{S}]\underline{R}\}$$

[using part (i) of Lemma 2.1]

$$= \text{tr}(\underline{R} - \underline{RSR} + \underline{RSZDTZ}'\underline{SR})$$

$$= \sigma_{c+1}^2 \text{tr}(\underline{I} - \underline{SR} + \underline{SZDTZ}'\underline{SR})$$

$$= \sigma_{c+1}^2 [\text{tr}(\underline{I} - \underline{SR}) + \text{tr}(\underline{SZDTZ}'\underline{SR})]$$

$$= \sigma_{c+1}^2 [\text{rank}(\underline{I} - \underline{SR}) + \text{tr}(\underline{SZDTZ}'\underline{SR})]$$

[since  $\underline{I} - \underline{SR}$  is idempotent]

$$= \sigma_{c+1}^2 \{\text{rank}[\underline{I} - \underline{A}(\underline{A}'\underline{RA})^{-1}\underline{A}'\underline{R}] + \text{tr}(\underline{TZ}'\underline{SRSZD})\}$$

[since Lemma 2.2]

$$= \sigma_{c+1}^2 \{ \text{rank}[I - A(A'A)^{-1}A'] + \text{tr}(TZ'SZD) \}$$

$$[\text{since } SRS = S]$$

$$= \sigma_{c+1}^2 [\text{rank}(I - AA') + \text{tr}(I_q - T)]$$

$$[\text{since } TZ'SZD = T(T^{-1} - I) = I - T]$$

$$= \sigma_{c+1}^2 [\text{rank}(P_X) + q - \text{tr}(T)]$$

$$= \sigma_{c+1}^2 [p^* + q - \text{tr}(T)] . \quad (3.12)$$

Thus,

$$\mathbb{E}(\underline{e}'\underline{e} | A'\underline{y}) = (\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}})'(\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}})$$

$$+ \sigma_{c+1}^2 [p^* + q - \text{tr}(T)] .$$

Since

$$\mathbb{E}[t_{c+1}(\underline{x})] = \mathbb{E}(\underline{e}'\underline{e})$$

$$= [\mathbb{E}(\underline{e})]'[\mathbb{E}(\underline{e})] + \text{tr}[\text{Var}(\underline{e})]$$

$$= 0 + \text{tr}(\sigma_{c+1}^2 \mathbf{I}_n)$$

$$= n\sigma_{c+1}^2 ,$$

we conclude that

$$\sigma_{c+1}^{2(p+1)} = \frac{(\underline{y} - \underline{X}\tilde{\underline{a}}^{(p)} - \underline{Z}\tilde{\underline{b}}^{(p)})'(\underline{y} - \underline{X}\tilde{\underline{a}}^{(p)} - \underline{Z}\tilde{\underline{b}}^{(p)}) + \sigma_{c+1}^{2(p)} [p^* + q - \text{tr}(\mathbf{T}^{(p)})]}{n} \quad (3.13)$$

Formulas (3.9) and (3.13) were given previously by, for example, Laird [1982, equations (3.12), (3.13)].

### 3. Alternative implementation

In applying the EM algorithm to the REML estimation of  $\underline{\sigma}$ , we have regarded the elements of the vector  $\underline{x} = (\underline{b}', \underline{e}')'$  as the complete data. Consider now the unobservable random vector  $\underline{x}^* = [\underline{b}', (\mathbf{A}'\underline{e})']'$ . Since the elements of the observable random vector  $\mathbf{A}'\underline{y}$  are linear functions of  $\underline{x}^*$ , we could, in implementing the EM algorithm, regard the elements of  $\underline{x}^*$ , rather than those of  $\underline{x}$ , as the complete data. We now consider this alternative implementation.

When parameterized in terms of the vector  $\underline{\phi}$ , the distribution of  $\underline{x}^*$ , like that of  $\underline{x}$ , is of the regular exponential form. If  $\underline{x}^*$  were observable, the vector  $\underline{t}^*(\underline{x}^*) = [\underline{b}'\underline{b}_1, \dots, \underline{b}'\underline{b}_c, (\mathbf{A}'\underline{e})'(\mathbf{A}'\underline{e})']'$  would

be a vector of complete sufficient statistics for  $\underline{\phi}$ . Note that  $\underline{t}(\underline{x})$  and  $\underline{t}^*(\underline{x}^*)$  differ only in their last element.

Observe that

$$\begin{aligned} \mathbb{E}[(\underline{A}'\underline{e})'(\underline{A}'\underline{e})|\underline{A}'\underline{y}] &= [\mathbb{E}(\underline{A}'\underline{e}|\underline{A}'\underline{y})]'[\mathbb{E}(\underline{A}'\underline{e}|\underline{A}'\underline{y})] \\ &\quad + \text{tr}[\text{Var}(\underline{A}'\underline{e}|\underline{A}'\underline{y})], \end{aligned} \quad (3.14)$$

and that

$$\begin{aligned} \mathbb{E}(\underline{A}'\underline{e}|\underline{A}'\underline{y}) &= \underline{A}'\mathbb{E}(\underline{e}|\underline{A}'\underline{y}) \\ &= \underline{A}'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) \quad [\text{using (3.10)}] \end{aligned} \quad (3.15)$$

and

$$\begin{aligned} \text{Var}(\underline{A}'\underline{e}|\underline{A}'\underline{y}) &= \underline{A}'\text{Var}(\underline{e}|\underline{A}'\underline{y})\underline{A} \\ &= \underline{A}'(\underline{R} - \underline{RPR})\underline{A} \quad [\text{using (3.11)}]. \end{aligned}$$

We find that

$$\begin{aligned} \text{tr}[\text{Var}(\underline{A}'\underline{e}|\underline{A}'\underline{y})] &= \text{tr}[\underline{A}'\underline{R}\underline{A} - \underline{A}'\underline{RPR}\underline{A}] \\ &= \text{tr}[\sigma_{c+1}^2 \underline{I}_{n-p*} - \underline{A}'\underline{R}\underline{S}\underline{R}\underline{A} \\ &\quad + \underline{A}'\underline{R}\underline{S}\underline{Z}\underline{D}\underline{T}\underline{Z}'\underline{S}\underline{R}\underline{A}] \quad [\text{using part (i) of Lemma 2.1}] \end{aligned}$$

$$= \sigma_{c+1}^2 [\text{tr}(\mathbf{I} - \mathbf{A}'\mathbf{SRA}) + \text{tr}(\mathbf{TZ}'\mathbf{SRAA}'\mathbf{SZD})].$$

Moreover,

$$\text{tr}(\mathbf{I} - \mathbf{A}'\mathbf{SRA}) = n - p^* - \text{tr}(\mathbf{A}'\mathbf{SRA})$$

$$= n - p^* - \sigma_{c+1}^2 \text{tr}(\mathbf{SAA}')$$

$$= n - p^* - \sigma_{c+1}^2 \text{tr}\left[\frac{1}{\sigma_{c+1}^2} (\mathbf{I} - \mathbf{P}_X)(\mathbf{I} - \mathbf{P}_X)\right]$$

$$= n - p^* - \text{tr}(\mathbf{I} - \mathbf{P}_X)$$

$$= 0 \quad [\text{since } \text{tr}(\mathbf{I} - \mathbf{P}_X) = n - \text{tr}(\mathbf{P}_X)]$$

$$= n - \text{rank}(X) = n - p^*]$$

and

$$\text{tr}(\mathbf{TZ}'\mathbf{SRAA}'\mathbf{SZD}) = \text{tr}(\mathbf{TZ}'\mathbf{SRSZD})$$

$$= \text{tr}(\mathbf{TZ}'\mathbf{SZD})$$

$$= q - \text{tr}(\mathbf{T}) \quad [\text{as shown in the derivation of (3.12)}].$$

Thus,

$$\text{tr}[\text{Var}(\underline{A}'\underline{e}|\underline{A}'\underline{y})] = \sigma_{c+1}^2[q - \text{tr}(T)]. \quad (3.16)$$

Substituting expressions (3.15) and (3.16) into expression (3.14), we obtain

$$\begin{aligned} \mathbb{E}[(\underline{A}'\underline{e})'(\underline{A}'\underline{e})|\underline{A}'\underline{y}] &= [\underline{A}'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}})]'[(\underline{A}'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}})] \\ &\quad + \sigma_{c+1}^2[q - \text{tr}(T)]. \end{aligned}$$

Now, recalling results on the MME (2.18) discussed in Section II.A, we find that

$$\underline{X}'\underline{R}^{-1}\underline{X}\tilde{\underline{\alpha}} + \underline{X}'\underline{R}^{-1}\underline{Z}\tilde{\underline{b}} = \underline{X}'\underline{R}^{-1}\underline{X}\tilde{\underline{\alpha}} + \underline{X}'\underline{R}^{-1}\underline{Z}\underline{D}\tilde{\underline{v}} = \underline{X}'\underline{R}^{-1}\underline{y}$$

implying that

$$\left(\frac{1}{\sigma_{c+1}^2}\right) \underline{X}'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) = 0$$

and, hence, that

$$\underline{P}_{\underline{X}}(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) = 0$$

so that

$$AA'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) = (I - P_X)(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) = \underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}.$$

It follows that

$$\begin{aligned} \mathbb{E}[(A'e)'(A'e) | A'y] &= (\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{X}\tilde{\underline{b}})'(\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) \\ &\quad + \sigma_{c+1}^2 [q - \text{tr}(T)]. \end{aligned} \quad (3.17)$$

Also,

$$\begin{aligned} \mathbb{E}[(A'e)'(A'e)] &= \mathbb{E}[\underline{e}'(AA')\underline{e}] \\ &= [\mathbb{E}(\underline{e})]'(AA')[\mathbb{E}(\underline{e})] \\ &\quad + \text{tr}[(AA') \text{Var}(\underline{e})] \\ &= 0 + \text{tr}(AA'R) \\ &= \sigma_{c+1}^2 \text{tr}(I - P_X) \\ &= \sigma_{c+1}^2 (n-p^*). \end{aligned} \quad (3.18)$$

Let  $\underline{\sigma}^{(p)} = (\sigma_1^{(p)}, \dots, \sigma_c^{(p)}, \sigma_{c+1}^{(p)})'$  represent the  $p$ -th iterate of the EM algorithm, and take  $\tilde{\alpha}^{(p)}$ ,  $\tilde{b}_i^{(p)}$ , and  $T_{ii}^{(p)}$  to be the values of  $\tilde{\alpha}$ ,  $\tilde{b}_i$ , and  $T_{ii}$  at  $\underline{\sigma} = \underline{\sigma}^{(p)}$ . When  $\underline{x}^*$ , rather than  $\underline{x}$ , is regarded as the vector of complete data, then it follows from results (3.17) and (3.18), together with the results of Section III.A.2, that the  $(p+1)$ -st iterate of the EM algorithm is given by

$$\sigma_i^{(p+1)} = \frac{\tilde{b}_i^{(p)'} \tilde{b}_i^{(p)} + \sigma_i^{(p)} \text{tr}(T_{ii}^{(p)})}{q_i} \quad (i = 1, \dots, c), \quad (3.19)$$

$$\sigma_{c+1}^{(p+1)} = \frac{(\underline{y} - \underline{X}\tilde{\alpha}^{(p)} - \underline{Z}\tilde{b}^{(p)})' (\underline{y} - \underline{X}\tilde{\alpha}^{(p)} - \underline{Z}\tilde{b}^{(p)}) + \sigma_{c+1}^{(p)} [q - \text{tr}(T^{(p)})]}{n - p^*}. \quad (3.20)$$

Note that the formula for the last component of the  $(p+1)$ -st iterate of the EM algorithm differs from that [formula (3.13)] in the previous implementation of this algorithm. The two formulas differ in the relative weight assigned to the previous iterate  $\sigma_{c+1}^{(p)}$ . More weight is assigned to  $\sigma_{c+1}^{(p)}$  in formula (3.13) than in formula (3.20).

#### 4. Nonnegativity constraints

We now consider whether the iterates generated by the EM algorithm belong to the parameter space  $\Omega_2 = \{\underline{\sigma} : \sigma_{c+1}^2 > 0, \sigma_i^2 \geq 0 \ (i = 1, \dots, c)\}$ . It follows from Harville's (1975) Lemma 1 that, for  $\underline{\sigma} \in \Omega_2$ ,

$$(i) \quad \text{tr}(T_{ii}) > 0 \quad (i = 1, \dots, c), \text{ and} \quad (3.21)$$

$$(ii) \quad q_i \geq \text{tr}(T_{ii}), \text{ with strict inequality holding if}$$

$$\sigma_i^2 > 0 \text{ and } \text{rank}(X, Z_i) > p^* \quad (i = 1, \dots, c).$$

[Note that if  $\sigma_i^2 = 0$ , then  $T_{ii} = I_{q_i}$ , implying that

$$q_i = \text{tr}(T_{ii}).] \quad (3.22)$$

Result (3.21) implies that, for  $\sigma_i^2 > 0$ ,

$$\tilde{b}_{i-i}' \tilde{b}_{i-i} + \sigma_i^2 \text{tr}(T_{ii}) > 0 \quad (i = 1, \dots, c).$$

Since  $\tilde{b}_{i-i}' \tilde{b}_{i-i} = \sigma_i^4 \tilde{v}_{i-i}' \tilde{v}_{i-i}$ , both terms of the sum  $\tilde{b}_{i-i}' \tilde{b}_{i-i} + \sigma_i^2 \text{tr}(T_{ii})$  equal 0 if  $\sigma_i^2 = 0$  ( $i = 1, \dots, c$ ). Thus, the  $i$ -th component  $\sigma_i^{2(p+1)}$  of the  $(p+1)$ -st iterate generated by either implementation of the EM algorithm is greater than (equal to) zero if  $\sigma_i^{2(p)}$  is greater than (equal to) zero. Hence, if all of the components of  $\underline{\sigma}^{(0)}$  are strictly positive, then  $\sigma_i^{2(p)}$  is strictly positive for all  $p$  ( $i = 1, \dots, c$ ). [If  $\underline{\sigma}^{(0)}$  is such that  $\sigma_i^{2(0)} = 0$ , then  $\sigma_i^{2(p)} = 0$  for all  $p$  ( $i = 1, \dots, c$ ).]

Equation (3.22) implies that, for  $\underline{\sigma} \in \Omega_2$ ,

$$q = \sum_{i=1}^c q_i \geq \sum_{i=1}^c \text{tr}(T_{ii}) = \text{tr}(T)$$

that is,

$$q - \text{tr}(T) \geq 0.$$

Note also that  $\underline{y} - X\underline{\tilde{\alpha}} = 0$  if, and only if,  $\underline{y} \in \mathcal{C}(X)$  and, hence, in light of part (v) of Lemma 2.4, that  $\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}} = \underline{0}$  if, and only if,  $\underline{y} \in \mathcal{C}(X)$ . Thus, if  $\underline{y} \notin \mathcal{C}(X)$ , then  $\sigma_{c+1}^{2(p)} > 0$  for all  $p$ .

We conclude that if the initial value  $\underline{\sigma}^{(0)}$  belongs to  $\Omega_2$  and if  $\underline{y} \notin \mathcal{C}(X)$ , then each of the iterates  $\underline{\sigma}^{(p)}$  ( $p = 1, 2, \dots$ ) belongs to  $\Omega_2$ .

### B. The Newton-Raphson Method

The Newton-Raphson method is a general iterative method for finding stationary points of an arbitrary nonlinear function [e.g., Bard (1974), and Gill, Murray and Wright (henceforth GMW) (1981)]. It can be used to find stationary points of the nonlinear function  $L_1(\underline{\gamma}; \underline{y})$ . If a REML estimate  $\hat{\underline{\gamma}}$  of  $\underline{\gamma}$  is located in the interior of the parameter space, then  $\hat{\underline{\gamma}}$  is a stationary point of  $L_1(\underline{\gamma}; \underline{y})$ .

In Section III.C.1, we will present the Newton-Raphson method as one method in a large class of numerical optimization techniques known as line-search methods. We then discuss the method as applied to the maximization of  $L_1(\underline{\gamma}; \underline{y})$ , or  $L_v(\underline{\sigma}; \underline{y})$ . In Section III.C.2-III.C.4, we introduce several modifications of the Newton-Raphson method [as applied to the maximization of  $L_1(\underline{\gamma}; \underline{y})$  or  $L_v(\underline{\sigma}; \underline{y})$ ] with the intent of improving its rate of convergence.

### 1. Description of the method

We describe a general class of iterative methods, known as line-search methods (or step-length methods), for locating the maximum of a nonlinear function. We describe them in the context of maximizing  $L_1(\underline{\gamma}; \underline{y})$ .

Let  $\underline{\gamma}^{(0)}$  represent the initial guess and  $\underline{\gamma}^{(p)}$  the  $p$ -th iterate of an iterative procedure for maximizing  $L_1(\underline{\gamma}; \underline{y})$ . In a line-search method,

$$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} + \alpha_p \underline{\ell}_p,$$

where  $\underline{\ell}_p$  is a vector in  $\mathbb{R}^{c+1}$ , and  $\alpha_p$  is a nonnegative real number ( $p = 0, 1, 2, \dots$ ). The vector  $\underline{\ell}_p$  defines a direction of movement away from  $\underline{\gamma}^{(p)}$  toward  $\underline{\gamma}^{(p+1)}$  (called a search direction), and the scalar  $\alpha_p$  defines the stepsize in that direction. Line-search methods for maximizing  $L_1(\underline{\gamma}; \underline{y})$  differ in their choice of  $\alpha_p$  and  $\underline{\ell}_p$ . [See, e.g., GMW (1981) for a review of several such methods.]

Subsequently, we use the notation  $[[m_i]]$  ( $i = 1, \dots, I$ ) to represent an  $I \times 1$  vector whose  $i$ -th element is  $m_i$ , and the notation  $[[m_{ij}]]$  ( $i = 1, \dots, I; j = 1, \dots, J$ ) to represent an  $I \times J$  matrix whose  $(i,j)$ -th element is  $m_{ij}$ .

Let  $\frac{\partial L_1}{\partial \underline{\gamma}}$  denote the  $(c+1) \times 1$  vector whose  $i$ -th element is the partial derivative of  $L_1(\underline{\gamma}; \underline{y})$  with respect to  $\gamma_i$  ( $i = 1, \dots, c+1$ ).

Define

$$\underline{b}(\underline{\gamma}) = \frac{\partial L_1}{\partial \underline{\gamma}},$$

$$B(\underline{\gamma}) = \left[ \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] \right] (i, j = 1, \dots, c+1).$$

Line search methods in which  $\underline{\ell}_p = R_p \underline{b}(\underline{\gamma}^{(p)})$ , for some  $(c+1) \times (c+1)$  real matrix  $R_p$ , are called gradient methods. Thus, the  $(p+1)$ -st iterate of a gradient method is

$$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} + \alpha_p R_p \underline{b}(\underline{\gamma}^{(p)}) . \quad (3.23)$$

If  $R_p$  is a positive definite matrix, and  $\underline{\gamma}^{(p)}$  is not a stationary point of  $L_1(\underline{\gamma}; \underline{y})$ , then there exists a sufficiently small step size  $\alpha_p$  ( $\alpha_p > 0$ ) such that  $L_1(\underline{\gamma}^{(p+1)}; \underline{y}) > L_1(\underline{\gamma}^{(p)}; \underline{y})$  ( $p = 0, 1, 2, \dots$ ) [e.g., GMW (1981)].

Once a search direction  $\underline{\ell}_p$  for the  $p$ -th iteration of a line-search algorithm has been decided upon, the stepsize  $\alpha_p$  is determined. Bertsekas (1982b) identified the following five strategies for choosing  $\alpha_p$  [assuming that  $\underline{\gamma}^{(p)}$  is not a stationary point of  $L_1(\underline{\gamma}; \underline{y})$ ]:

1. Maximization rule: Choose  $\alpha_p$  so that

$$L_1(\underline{\gamma}^{(p)} + \alpha_p \underline{\ell}_p; \underline{y}) = \max_{\alpha > 0} L_1(\underline{\gamma}^{(p)} + \alpha \underline{\ell}_p; \underline{y}).$$

2. Limited maximization rule: For some fixed number  $s$  ( $s > 0$ ), choose  $\alpha_p$  so that  $L_1(\underline{Y}^{(p)} + \alpha_p \underline{\ell}_p; \underline{y}) = \max_{\alpha \in [0, s]} L_1(\underline{Y}^{(p)} + \alpha \underline{\ell}_p; \underline{y})$ .
3. Armijo rule: For fixed real numbers  $s$ ,  $\beta$ , and  $\sigma$  [ $s > 0$ ,  $\beta \in (0, 1)$ , and  $\sigma \in (0, \frac{1}{2}]$ ], put  $\alpha_p = \beta^{m_p} s$ , where  $m_p$  is the first nonnegative integer  $m$  for which
 
$$L_1(\underline{Y}^{(p)} + \beta^m s \underline{\ell}_p; \underline{y}) - L_1(\underline{Y}^{(p)}; \underline{y}) \geq \alpha \beta^m s \underline{\ell}_p' \underline{b}(\underline{Y}^{(p)}).$$
4. Goldstein rule: For a fixed scalar  $\sigma$  [ $\sigma \in (0, \frac{1}{2}]$ ], choose  $\alpha_p$  so that
 
$$\sigma \leq \frac{L_1(\underline{Y}^{(p)} + \alpha_p \underline{\ell}_p; \underline{y}) - L_1(\underline{Y}^{(p)}; \underline{y})}{\alpha_p \underline{\ell}_p' \underline{b}(\underline{Y}^{(p)})} \leq 1 - \sigma.$$
5. Constant stepsize rule: For a fixed number  $s$  ( $s > 0$ ),  $\alpha_p = s$  ( $p = 1, 2, \dots$ ).

Rules 3 and 4 are designed for gradient methods for which the matrix  $R_p$  is positive definite. The use of Rule 1 or 2 requires the availability of a univariate maximization method. Kennedy and Gentle (1980) review some such methods. Line-search methods tend to converge in the fewest number of iterations when Rules 1 or 2 are used. However, they may

require that  $L_1(\underline{y}^{(p)} + \alpha \underline{g}_p; \underline{y})$  be evaluated at so many values of  $\alpha$  that they are not cost efficient. Bertsekas (1982b) and GMW (1981) discuss, in conjunction with Rules 1 and 2, the use of approximate univariate maximization methods. These approximate methods generally require fewer function evaluations.

Rules 3 and 4 were devised to insure sufficient progress from one iteration to another. Failure to maintain sufficient function value increases over successive iterations may result in the sequence of iterates  $\{\underline{y}^{(p)}\}$  converging to a stationary point which is not a local maximum [e.g., Fletcher (1980), GMW (1981)]. The constant stepsize rule is the easiest rule to implement, and it is also the least expensive (in terms of function and gradient evaluations). However, Bard (1970) suggested that, in terms of function and gradient evaluations, approximate versions of Rules 1 and 2 are better than Rules 1 and 2 themselves, or than Rule 5.

The Newton-Raphson method is a gradient method, that is, has iterates of the form (3.21). In the Newton-Raphson algorithm,  $R_p = -B^{-1}(\underline{y}^{(p)})$ , so that its  $(p+1)$ -st iterate is

$$\underline{y}^{(p+1)} = \underline{y}^{(p)} - \alpha_p B^{-1}(\underline{y}^{(p)}) \underline{b}(\underline{y}^{(p)}) \quad (p = 0, 1, 2, \dots)$$

[provided that  $B(\underline{y}^{(p)})$  is nonsingular] [e.g., GMW (1981)]. Often the constant stepsize  $\alpha_p = 1$  ( $p = 0, 1, 2, \dots$ ) is used, in which we refer to the Newton-Raphson algorithm as the "traditional" Newton-Raphson algorithm.

If  $L_1(\underline{\gamma}; \underline{y})$  were a quadratic function and  $-B(\underline{\gamma}^{(0)})$  were a positive definite matrix, then the traditional Newton-Raphson method would locate the maximizing value of  $L_1(\underline{\gamma}; \underline{y})$  in exactly one iteration. More generally,  $L_1(\underline{\gamma}; \underline{y})$  can be expected to behave like a quadratic function in a sufficiently small neighborhood of its maximizing value [e.g., Fletcher (1980)]. Thus, the traditional Newton-Raphson algorithm tends to exhibit a quadratic rate of convergence once its iterates are near a stationary point. Unfortunately, the traditional Newton-Raphson method may fail to converge or may converge to a stationary point which is not a local or global maximum of  $L_1(\underline{\gamma}; \underline{y})$ . This problem can be alleviated somewhat by using a version of the Newton-Raphson algorithm in which  $\alpha_p$  is chosen in a way that insures that some progress is made on the  $(p+1)$ -st iteration [though, if  $B(\underline{\gamma}^{(p)})$  is not positive definite, no such  $\alpha_p$  may exist].

Several authors have suggested that if the matrix  $-B(\underline{\gamma}^{(p)})$  is not positive definite, then a positive definite matrix  $-\bar{B}(\underline{\gamma}^{(p)})$ , which is closely related to  $-B(\underline{\gamma}^{(p)})$ , should be used in its place. Doing so insures the existence of a stepsize  $\alpha_p$  ( $\alpha_p > 0$ ) such that  $L_1(\underline{\gamma}^{(p+1)}; \underline{y}) > L_1(\underline{\gamma}^{(p)}; \underline{y})$ . Various choices for  $-\bar{B}(\underline{\gamma}^{(p)})$  have been proposed by Greenstadt (1967), Levenberg (1944), Marquardt (1963), and Murray (1972). Murray's approach is discussed by Gill and Murray (1974) and GMW (1981).

Modifications of the Newton-Raphson algorithm have also been developed for situations where  $-B(\underline{\gamma}^{(p)})$  may be too costly to compute. These methods approximate  $-B(\underline{\gamma}^{(p)})$ , or its inverse, at each iteration.

GMW (1981) describe the discrete Newton method; it uses  $\underline{b}(\underline{y}^{(p)})$  and forward-difference derivative formulas to approximate each column in  $-\underline{B}(\underline{y}^{(p)})$ . GMW (1981) also discuss variable-metric algorithms. The latter algorithms "carry-over" information about the approximate inverse Hessian matrix from one iteration to another. More specifically, if  $\underline{H}(\underline{y}^{(p)})$  is the approximation to  $-\underline{B}^{-1}(\underline{y}^{(p)})$ , then  $-\underline{B}^{-1}(\underline{y}^{(p+1)})$  is approximated by a matrix of the form  $\underline{H}(\underline{y}^{(p+1)}) = \underline{H}(\underline{y}^{(p)}) + \underline{U}(\underline{y}^{(p)})$ . Two of the most commonly used variable-metric methods are the Davidon-Fletcher-Powell [Fletcher and Powell (1963)], and the Broyden-Fletcher-Goldfarb-Shanno [e.g., Kennedy and Gentle (1980)]. In both methods,  $\underline{H}(\underline{y}^{(p)})$  is symmetric, and if  $\alpha_p$  is chosen by Rule 1,  $\underline{H}(\underline{y}^{(p)})$  is positive definite. The convergence properties of these two methods have been studied by, for example, Dixon (1972), Brodlie (1977), and Schnabel (1982).

The traditional Newton-Raphson method can be viewed as an application of Newton's iterative method for finding the roots of a system of one or more nonlinear equations.

For purpose of describing Newton's method, let  $\underline{z} = (z_1, \dots, z_n)'$  represent a real-valued vector, and let  $f_1(\underline{z}), \dots, f_n(\underline{z})$  represent nonlinear functions of  $\underline{z}$ . Define

$$\underline{f}(\underline{z}) = \begin{pmatrix} f_1(\underline{z}) \\ \vdots \\ f_n(\underline{z}) \end{pmatrix},$$

$$F(\underline{z}) = \begin{pmatrix} \frac{\partial f_1(\underline{z})}{\partial z_1} & \dots & \frac{\partial f_1(\underline{z})}{\partial z_n} \\ \vdots & & \vdots \\ \frac{\partial f_n(\underline{z})}{\partial z_1} & \dots & \frac{\partial f_n(\underline{z})}{\partial z_n} \end{pmatrix}.$$

The (p+1)-st iterate of Newton's method for solving the system of equations

$$\underline{f}(\underline{z}) = \underline{0} \quad (3.24)$$

is

$$\underline{z}^{(p+1)} = \underline{z}^{(p)} - F^{-1}(\underline{z}^{(p)}) \underline{f}(\underline{z}^{(p)}) \quad (p = 0, 1, 2, \dots). \quad (3.25)$$

In the special case where the functions  $f_1, \dots, f_n$  are linear, an explicit solution to system (3.24) exists and Newton's method will converge in a single iteration (provided, of course, that the system is consistent). As the degree of nonlinearity of  $f_1, \dots, f_n$  increases, we can expect the rate of convergence of the sequence of iterates (3.25) to decrease.

By definition, the problem of locating the stationary points of the function  $L_1(\underline{\gamma}; \underline{y})$  is that of solving the likelihood equations

$$\frac{\partial L_1}{\partial \underline{\gamma}} = \underline{0} . \quad (3.26)$$

If we apply Newton's method to the likelihood equations, we obtain the same sequence of iterates

$$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} - B^{-1}(\underline{\gamma}^{(p)}) \underline{b}(\underline{\gamma}^{(p)}) \quad (p = 0, 1, 2, \dots) \quad (3.27)$$

as in applying the Newton-Raphson method to the maximization of  $L_1(\underline{\gamma}; \underline{y})$ .

When the likelihood function is parameterized in terms of  $\underline{\sigma}$ , rather than  $\underline{\gamma}$ , the likelihood equations are

$$\frac{\partial L_v}{\partial \underline{\sigma}} = \underline{0} . \quad (3.28)$$

These equations are, of course, equivalent to equations (3.26).

Define

$$\underline{g}(\underline{\sigma}) = \left[ \left[ \frac{\partial L_v}{\partial \sigma_i^2} \right] \right] \quad (i = 1, \dots, c+1) ,$$

$$G(\underline{\sigma}) = \left[ \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] \right] \quad (i, j = 1, \dots, c+1) .$$

The (p+1)-st iterate of the traditional Newton-Raphson method, as applied to  $L_v(\underline{\sigma}; \underline{y})$ , or equivalently the (p+1)-st iterate of Newton's

method, as applied to equations (3.28), is

$$\underline{\sigma}^{(p+1)} = \underline{\sigma}^{(p)} - G^{-1}(\underline{\sigma}^{(p)}) \underline{g}(\underline{\sigma}^{(p)}) \quad (p = 0, 1, 2, \dots). \quad (3.29)$$

In general, the sequence of iterates (3.29) is not equivalent to the sequence (3.27).

Algorithms (3.27) and (3.29) can both produce iterates outside the parameter space. This problem can be eliminated by modifying these algorithms in accordance with various results in the constrained optimization literature. These modifications are discussed in Chapter IV.

## 2. Linearization

Following Fenech and Harville (1985), define

$$X_0^* = X,$$

$$X_i^* = (X, Z_1, \dots, Z_i) \quad (i = 1, \dots, c),$$

$$r_0 = \text{rank}(X),$$

$$r_i = \text{rank}(X_i^*) - \text{rank}(X_{i-1}^*) \quad (i = 1, \dots, c),$$

$$r_{c+1} = n - \text{rank}(X_c^*),$$

$$P_i = X_i^* (X_i^{*'} X_i^*)^{-1} X_i^{*'} \quad (i = 0, 1, \dots, c).$$

Assume that  $r_i > 0$  ( $i = 1, \dots, c+1$ ). Let

$$\lambda_{ji} = \begin{cases} 0, & \text{if } j < i \\ \frac{1}{r_i} \operatorname{tr}[Z_j'(P_i - P_{i-1})Z_j], & \text{if } j \geq i, (i, j = 1, \dots, c). \end{cases}$$

Observe that, for  $i = 1, \dots, c$ ,

$$\begin{aligned} \lambda_{ii} &= \frac{1}{r_i} \operatorname{tr}[Z_i'(P_i - P_{i-1})Z_i] \\ &= \frac{1}{r_i} \operatorname{tr}[Z_i'(I - P_{i-1})Z_i] \quad [\text{since } P_i Z_i = Z_i] \end{aligned}$$

and, in particular, that

$$\lambda_{11} = \frac{1}{r_1} \operatorname{tr}[Z_1'(I - P_0)Z_1] = \frac{1}{r_1} \operatorname{tr}(C_{11}).$$

Following Brown (1984), we adopt the following definition.

**Definition 3.1:** With respect to model (1.1), an ANOVA( $\sigma^2$ ) is a partitioning

$$\underline{y}'(I - P_X)\underline{y} = \underline{y}'A_1\underline{y} + \dots + \underline{y}'A_s\underline{y},$$

where  $A_1, \dots, A_s$  are  $n \times n$ , symmetric, known matrices such that

- (i)  $\underline{y}'A_1\underline{y}, \dots, \underline{y}'A_s\underline{y}$  are independently distributed,
- (ii)  $\frac{\underline{y}'A_i\underline{y}}{\sigma_{c+1}^2 c_i} \stackrel{d}{\sim} \chi^2(f_i)$ , for some positive integer  $f_i$  and some scalar  $c_i$  ( $i = 1, \dots, s$ ), and
- (iii) the scalars  $c_1, \dots, c_s$  are distinct.

An ANOVA( $\sigma^2$ ) exists if, and only if,  $A'Z_1Z_1'A, \dots, A'Z_cZ_c'A$  commute in pairs [Brown (1984)] or, equivalently, if, and only if,

$(I - P_X)Z_1Z_1'(I - P_X), \dots, (I - P_X)Z_cZ_c'(I - P_X)$  commute in pairs [Fenech and Harville (1985)]. Moreover, if an ANOVA( $\sigma^2$ ) exists, the sums of squares  $\underline{y}'A_1\underline{y}, \dots, \underline{y}'A_s\underline{y}$  are unique (up to order) [Brown (1984)].

Harville and Fenech (1985) showed that, if an ANOVA( $\sigma^2$ ) exists and if  $s = c+1$ , then

$$A_i = P_i - P_{i-1} \quad (i = 1, \dots, c),$$

$$A_{c+1} = I - P_c,$$

$$c_i = 1 + \sum_{j=1}^c \lambda_{ji} \gamma_j \quad (i = 1, \dots, c),$$

$$c_{c+1} = 1,$$

$$f_i = r_i \quad (i = 1, \dots, c+1)$$

(up to order).

Define  $S_i = \underline{y}'(P_i - P_{i-1})\underline{y}$  ( $i = 1, \dots, c$ ) and  $S_{c+1} = \underline{y}'(I - P_c)\underline{y}$ .

Note that

$$\begin{aligned} E(S_i) &= [\underline{y}'(P_i - P_{i-1})\underline{y}] \\ &= (\underline{X}\underline{\alpha})'(P_i - P_{i-1})(\underline{X}\underline{\alpha}) \\ &\quad + \text{tr}[(P_i - P_{i-1})(\sigma_{c+1}^2 I_n + \sum_{j=1}^c \sigma_{c+1}^2 \gamma_j Z_j Z_j')] \\ &= \sigma_{c+1}^2 r_i + \sum_{j=1}^c \sigma_{c+1}^2 \gamma_j \text{tr}[(P_i - P_{i-1})Z_j Z_j'] \\ &\quad [\text{since } (P_i - P_{i-1})X = \emptyset \text{ and} \\ &\quad \text{tr}(P_i - P_{i-1}) = \text{tr}(P_i) - \text{tr}(P_{i-1}) \\ &\quad = \text{rank}(P_i) - \text{rank}(P_{i-1}) \\ &\quad = \text{rank}(X_i^*) - \text{rank}(X_{i-1}^*) = r_i] \end{aligned}$$

$$\begin{aligned}
&= \sigma_{c+1}^2 \left\{ r_i + \sum_{j=1}^c \gamma_j \operatorname{tr}[Z_j'(P_i - P_{i-1})Z_j] \right\} \\
&= r_i \sigma_{c+1}^2 \left\{ 1 + \frac{1}{r_i} \sum_{j=1}^c \gamma_j \operatorname{tr}[Z_j'(P_i - P_{i-1})Z_j] \right\} \\
&= r_i \sigma_{c+1}^2 \left( 1 + \sum_{j=1}^c \lambda_{ji} \gamma_j \right)
\end{aligned}$$

and that

$$\begin{aligned}
E(S_{c+1}) &= E[\underline{y}'(I - P_c)\underline{y}] \\
&= \operatorname{tr}[(I - P_c)(\sigma_{c+1}^2 I_n + \sum_{j=1}^c \gamma_j \sigma_{c+1}^2 Z_j Z_j')] \\
&= \sigma_{c+1}^2 r_{c+1} + \sum_{j=1}^c \sigma_{c+1}^2 \gamma_j \operatorname{tr}[(I - P_c)Z_j Z_j'] \\
&= r_{c+1} \sigma_{c+1}^2 \left\{ 1 + \frac{1}{r_{c+1}} \sum_{j=1}^c \gamma_j \operatorname{tr}[Z_j'(I - P_c)Z_j] \right\} \\
&= r_{c+1} \sigma_{c+1}^2 \quad [\text{since } (I - P_c)Z_j = \emptyset].
\end{aligned}$$

Let  $m_i$  represent the expected mean square associated with the sum of squares  $S_i$  ( $i = 1, \dots, c+1$ ), i.e., let

$$\begin{aligned}
m_i &= E\left(\frac{1}{r_i} S_i\right) = \sigma_{c+1}^2 \left(1 + \sum_{j=i}^c \lambda_{ji} \gamma_j\right) \\
&= \sigma_{c+1}^2 + \sum_{j=i}^c \lambda_{ji} \sigma_j^2 \quad (i = 1, \dots, c),
\end{aligned}$$

$$m_{c+1} = E\left(\frac{1}{r_{c+1}} S_{c+1}\right) = \sigma_{c+1}^2.$$

Take  $\underline{m} = (m_1, \dots, m_{c+1})'$ . Let  $\Omega_m^*$  represent the set of  $\underline{m}$  values that correspond to the set  $\Omega_1^*$  of  $\underline{\gamma}$  values.

Let  $L_s(\underline{\gamma}; \underline{y})$  and  $L_s^*(\underline{m}; \underline{y})$  represent the log-likelihood functions associated with  $S_1, \dots, S_{c+1}$  when the model is parameterized in terms of  $\underline{\gamma}$  and  $\underline{m}$ , respectively. Also, let  $L_1^*(\underline{m}; \underline{y})$  represent the log-likelihood function associated with  $A'\underline{y}$  when the model is parameterized in terms of  $\underline{m}$ .

If an ANOVA( $\sigma^2$ ) exists and if  $s = c+1$ , then it follows from the results of Brown (1984) that  $S_1, \dots, S_{c+1}$  form a minimal sufficient set of statistics for the family of distributions  $\{N(\underline{0}, A'VA) : \underline{\gamma} \in \Omega_1^*\}$ , that is, for the family of possible distributions of the vector  $A'\underline{y}$  of error contrasts. Thus, if an ANOVA( $\sigma^2$ ) exists and if  $s = c+1$ , then the likelihood function associated with  $A'\underline{y}$  is proportional to that associated with the ANOVA( $\sigma^2$ ) sums of squares  $S_1, \dots, S_{c+1}$ . Or, equivalently, if an ANOVA( $\sigma^2$ ) exists and if  $s = c+1$ , then the function  $L_1^*$  is equal (up to an additive constant) to the log-likelihood function  $L_s^*$  associated with the sums of squares  $S_1, \dots, S_{c+1}$ .

Suppose now that an ANOVA( $\sigma^2$ ) exists and that  $s = c+1$ . Let  $g^{(i)}(s_i; \underline{\gamma})$  represent the probability density function of  $S_i$  ( $i = 1, \dots, c+1$ ), and let  $g^*(s_1, \dots, s_{c+1}; \underline{\gamma})$  represent the joint probability density function of  $S_1, \dots, S_{c+1}$ . Also, let  $\chi^2(u:r)$  denote the probability density function for a random variable  $u$  whose distribution is central chi-square with  $r$  degrees of freedom. Take

$$Q_i = \frac{1}{m_i} S_i \quad (i = 1, \dots, c+1).$$

Then, since  $\frac{\partial Q_i}{\partial S_i} = \frac{1}{m_i}$  ( $i = 1, \dots, c+1$ ),

$$g^{(i)}(s_i; \underline{\gamma}) = \chi^2\left(\frac{s_i}{m_i} : r_i\right) \frac{1}{m_i} \quad (i = 1, \dots, c+1),$$

for  $\underline{\gamma} \in \Omega_1^*$ . Further, since  $S_1, \dots, S_{c+1}$  are independently distributed,

$$g^*(s_1, \dots, s_{c+1}; \underline{\gamma}) = \prod_{i=1}^{c+1} g^{(i)}(s_i; \underline{\gamma}), \quad \underline{\gamma} \in \Omega_1^*.$$

The log-likelihood function associated with  $S_1, \dots, S_{c+1}$  is then

$$L_S(\underline{\gamma}; \underline{y}) = \sum_{i=1}^{c+1} \log[g_i^{(i)}(S_i; \underline{\gamma})], \underline{\gamma} \in \Omega_1^*.$$

When re-expressed in terms of  $m_1, \dots, m_{c+1}$ , the log-likelihood function associated with  $S_1, \dots, S_{c+1}$  is

$$L_S^*(\underline{m}; \underline{y}) = \sum_{i=1}^{c+1} \log \left[ \frac{\left(\frac{1}{m_i} S_i\right)^{\frac{1}{2}(r_i-2)} \exp\left[-\left(\frac{1}{2}\right)\frac{1}{m_i} S_i\right]}{\Gamma\left(\frac{r_i}{2}\right) \cdot (2)^{\frac{1}{2}(r_i)} (m_i)} \right]$$

$$= \sum_{i=1}^{c+1} \left\{ \left(\frac{1}{2}\right)(r_i-2) [\log(S_i) - \log(m_i)] \right.$$

$$\left. - \left(\frac{1}{2}\right) \frac{1}{m_i} S_i - \log\left[\Gamma\left(\frac{r_i}{2}\right) \cdot (2)^{\frac{1}{2}(r_i)}\right] \right.$$

$$\left. - \log(m_i) \right\}$$

$$= \sum_{i=1}^{c+1} \left(\frac{1}{2}\right) (r_i-2) \log(S_i)$$

$$- \sum_{i=1}^{c+1} \left(\frac{1}{2}\right) r_i \log(m_i)$$

$$- \sum_{i=1}^{c+1} \left(\frac{1}{2}\right) \frac{1}{m_i} s_i$$

$$- \sum_{i=1}^{c+1} \log \left[ \Gamma\left(\frac{r_i}{2}\right) \cdot (2)^{\frac{1}{2}(r_i)} \right], \underline{m} \in \Omega_m^*.$$

As noted earlier,

$$L_1^*(\underline{m}; \underline{y}) = L_s^*(\underline{m}; \underline{y}) + k_1,$$

for some scalar  $k_1$  that is free of  $\underline{m}$ . It follows that

$$\begin{aligned} \frac{\partial L_1^*}{\partial m_j} &= -\left(\frac{1}{2}\right) \frac{1}{m_j} r_j + \left(\frac{1}{2}\right) \frac{1}{m_j^2} s_j \\ &= \frac{1}{m_j^2} \left[ \left(\frac{1}{2}\right) s_j - \left(\frac{1}{2}\right) r_j m_j \right] \\ &= -\left(\frac{1}{2}\right) \frac{1}{m_j^2} (r_j m_j - s_j) \quad (j = 1, \dots, c+1), \underline{y} \in \Omega_1^*, \end{aligned} \quad (3.30)$$

in agreement with the results given by Thompson [(1962), equations (1.2) and (1.3)] for the special case  $c=1$ .

Equating the partial derivatives (3.30) to zero, we obtain the likelihood equations

$$\left[ \left[ \frac{\partial L_1^*}{\partial m_j} \right] \right] = \underline{0} \quad (j = 1, \dots, c+1). \quad (3.31)$$

Note that if we multiply both sides of the  $j$ -th of these equations by  $m_j^2$  ( $j = 1, \dots, c+1$ ), we obtain the equivalent set of equations

$$\left[ \left[ m_j^2 \frac{L_1^*}{m_j} \right] \right] = \underline{0} \quad (j = 1, \dots, c+1). \quad (3.32)$$

Clearly, the solution to equations (3.31) or (3.32) is  $\hat{m}_j = \frac{1}{r_j} s_j$  ( $j = 1, \dots, c+1$ ).

Note that the left-hand sides of equations (3.31) are nonlinear in  $m_1, \dots, m_{c+1}$ , respectively, while the left-hand sides of equations (3.32) are linear. Consequently, if Newton's method were applied to equations (3.32), it would converge to the solution in a single iteration. In contrast, if it were applied directly to the likelihood equations (3.31), additional iterations would be required (and, even then, only an approximate solution would be obtained).

Let us now consider the general problem of computing a REML estimate of  $\underline{m}$  or, equivalently, of  $\underline{\gamma}$  (dropping the supposition that an ANOVA( $\sigma^2$ ) exists). Suppose that the Newton-Raphson method were applied to the problem of maximizing the function  $L_1$  or, equivalently, that Newton's method were applied to the problem of solving the system of

equations  $\frac{\partial L_1^*}{\partial \underline{m}} = \underline{0}$ . In general, these equations are nonlinear in  $\underline{m}$ .

$$\text{Let } q_j^*(\underline{m}) = m_j^2 \frac{\partial L_1^*}{\partial m_j} \quad (j = 1, \dots, c+1).$$

Instead of applying Newton's method directly to the likelihood equations, we can apply it instead to the equivalent system of equations  $q_j^*(\underline{m}) = 0$  ( $j = 1, \dots, c+1$ ). We do so in the hope that the functions  $q_j^*(\underline{m})$  ( $j = 1, \dots, c+1$ ) will be "more nearly linear" than the functions  $\frac{\partial L_1^*}{\partial m_j}$  ( $j = 1, \dots, c+1$ ) and, hence, that Newton's method will converge more rapidly when applied to the equations  $q_j^*(\underline{m}) = 0$  ( $j = 1, \dots, c+1$ ) than when applied to the likelihood equations  $\frac{\partial L_1^*}{\partial m} = \underline{0}$ .

Define

$$\underline{q}^*(\underline{m}) = [ [ q_i^*(\underline{m}) ] ] \quad (i = 1, \dots, c+1),$$

$$Q^*(\underline{m}) = [ [ \frac{\partial q_i^*}{\partial m_j} ] ] \quad (i, j = 1, \dots, c+1).$$

Let  $\underline{m}^{(p)}$  represent the  $p$ -th iterate of Newton's method when this method is applied to the system of equations  $q_j^*(\underline{m}) = 0$  ( $j = 1, \dots, c+1$ ).

Then,

$$\underline{m}^{(p+1)} = \underline{m}^{(p)} - Q^{*-1}(\underline{m}^{(p)}) \underline{q}^*(\underline{m}^{(p)}), \quad (3.33)$$

for  $p = 0, 1, 2, \dots$ . We refer to the iterative method whose  $(p+1)$ -st

iterate is given by equation (3.33) as the linearized Newton-Raphson method.

Note that for  $i, j \in \{1, \dots, c+1\}$ ,

$$\begin{aligned} \frac{\partial q_i^*}{\partial m_j} &= \frac{\partial}{\partial m_j} \left[ m_i^2 \frac{\partial L_1^*}{\partial m_i} \right] \\ &= \frac{\partial m_i^2}{\partial m_j} \frac{\partial L_1^*}{\partial m_i} + m_i^2 \frac{\partial^2 L_1^*}{\partial m_i \partial m_j} \\ &= \begin{cases} 2m_i \frac{\partial L_1^*}{\partial m_i} + m_i^2 \frac{\partial^2 L_1^*}{\partial m_i \partial m_i}, & \text{if } i = j \\ m_i^2 \frac{\partial^2 L_1^*}{\partial m_i \partial m_j}, & \text{if } i \neq j. \end{cases} \end{aligned}$$

Thus, to implement algorithm (3.33), we require expressions for  $\frac{\partial L_1^*}{\partial m_i}$  ( $i = 1, \dots, c+1$ ) and  $\frac{\partial^2 L_1^*}{\partial m_i \partial m_j}$  ( $i, j = 1, \dots, c+1$ ).

According to the chain rule of calculus,

$$\frac{\partial L_1^*}{\partial m_i} = \sum_{j=1}^{c+1} \frac{\partial \sigma_j^2}{\partial m_i} \frac{\partial L_v}{\partial \sigma_j^2} \quad (i = 1, \dots, c+1). \quad (3.34)$$

Expressions for  $\frac{\partial L_v}{\partial \sigma_j^2}$  ( $j = 1, \dots, c+1$ ) are given in Section II.C.

To obtain convenient expressions for  $\frac{\partial \sigma^2}{\partial \underline{m}_i}$  ( $i, j = 1, \dots, c+1$ ), we require some additional notation. By definition,

$$\underline{m} = \Lambda' \underline{\sigma}$$

where

$$\Lambda = \begin{pmatrix} \lambda_{11} & & & 0 \\ \vdots & \ddots & \emptyset & \vdots \\ \lambda_{c1} & \dots & \lambda_{cc} & 0 \\ 1 & \dots & 1 & 1 \end{pmatrix}.$$

Then,

$$\underline{\sigma} = W' \underline{m},$$

where  $W = \Lambda^{-1}$ .

Let  $w_{ij}$  represent the  $(i,j)$ -th element of  $W$ . The matrix  $W$  is lower triangular, that is,  $w_{ij} = 0$  for  $i < j$ . Further,

$$w_{c+1,c+1} = 1, \text{ and}$$

$$w_{ii} = \frac{1}{\lambda_{ii}} \quad (i = 1, \dots, c).$$

For  $i, j \in \{1, \dots, c+1\}$ , we have that

$$\frac{\partial \sigma_j^2}{\partial m_i} = w_{ij}.$$

Substituting into expression (3.34), we find that

$$\frac{\partial L_1^*}{\partial m_i} = \sum_{j=1}^i w_{ij} \frac{\partial L_v}{\partial \sigma_j^2} \quad (i = 1, \dots, c+1). \quad (3.35)$$

Consider now the second-order partial derivatives of  $L_1^*$ . It follows from result (3.35) that, for  $i, k = 1, \dots, c+1$ ,

$$\frac{\partial^2 L_1^*}{\partial m_i \partial m_k} = \sum_{j=1}^i w_{ij} \sum_{\ell=1}^{c+1} \frac{\partial^2 L_v}{\partial \sigma_j^2 \partial \sigma_\ell^2} \frac{\partial \sigma_\ell^2}{\partial m_k}$$

$$= \sum_{j=1}^i w_{ij} \sum_{\ell=1}^{c+1} w_{k\ell} \frac{\partial^2 L_v}{\partial \sigma_j^2 \partial \sigma_\ell^2}$$

Expressions for  $\frac{\partial^2 L_v}{\partial \sigma_j^2 \partial \sigma_\ell^2}$  ( $j, \ell = 1, \dots, c+1$ ) are given in Section II.C.

Note that, if  $\hat{m}_1, \dots, \hat{m}_{c+1}$  are REML estimates of  $m_1, \dots, m_{c+1}$ , respectively, then

$$\hat{\sigma}_j^2 = \sum_{i=j}^{c+1} w_{ij} \hat{m}_i \quad (j = 1, \dots, c+1)$$

are REML estimates of  $\sigma_1^2, \dots, \sigma_{c+1}^2$ , respectively. Similarly,

$$\hat{\gamma}_j = \frac{\hat{\sigma}_j^2}{\hat{\sigma}_{c+1}^2} = \frac{1}{\hat{m}_{c+1}} \sum_{i=j}^{c+1} w_{ij} \hat{m}_i \quad (j = 1, \dots, c), \text{ and}$$

$$\hat{\gamma}_{c+1} = \hat{\sigma}_{c+1}^2 = \hat{m}_{c+1}$$

are REML estimates of  $\gamma_1, \dots, \gamma_{c+1}$ , respectively. Note also that, because of the linearity of the transformation  $\underline{\sigma} = W'\underline{m}$ , the Newton-Raphson method produces the same set of iterates when applied to the problem of maximizing the function  $L_V$ .

### 3. The concentrated log-likelihood function

a.  $\underline{\gamma}$ -parameterization      Instead of applying the Newton-Raphson method directly to the problem of maximizing the function  $L_1$  of the  $c+1$  variables  $\gamma_1, \dots, \gamma_{c+1}$ , we can apply it instead to the problem of maximizing a certain function of the  $c$  variables  $\gamma_1, \dots, \gamma_c$ , which, following Bard (1974), we call the concentrated log-likelihood function.

Let  $\underline{\gamma}^+ = (\gamma_1, \dots, \gamma_c)'$ , so that  $\underline{\gamma} = (\underline{\gamma}^+, \sigma_{c+1}^2)'$ , and recall (from Lemma 2.5) that

$$\frac{\partial L_1}{\partial \sigma_{c+1}^2} = -\left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - Z\underline{b})].$$

Letting  $H = I_n + \sum_{i=1}^c \gamma_i Z_i Z_i' = \frac{1}{\sigma_{c+1}^2} V$ , we find, using result (2.13), that

$$\begin{aligned} \frac{\partial L_1}{\partial \sigma_{c+1}^2} &= -\left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - (\underline{y} - X\underline{\tilde{\alpha}})' V^{-1} (\underline{y} - X\underline{\tilde{\alpha}})] \\ &= -\left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} (n - p^*) \left[ \sigma_{c+1}^2 - \frac{1}{n - p^*} (\underline{y} - X\underline{\tilde{\alpha}})' H^{-1} (\underline{y} - X\underline{\tilde{\alpha}}) \right] \\ &= -\left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} (n - p^*) [\sigma_{c+1}^2 - \hat{\sigma}_{c+1}^2(\underline{Y}^+)] \end{aligned} \quad (3.37)$$

where

$$\hat{\sigma}_{c+1}^2(\underline{Y}^+) = \frac{1}{n - p^*} (\underline{y} - X\underline{\tilde{\alpha}})' H^{-1} (\underline{y} - X\underline{\tilde{\alpha}}). \quad (3.38)$$

It follows from expression (3.37) that, for an arbitrary (fixed) value

of  $\underline{Y}^+$ , the function  $L_1$  attains its maximum, over the interval

$0 < \sigma_{c+1}^2 < \infty$ , uniquely at the value  $\hat{\sigma}_{c+1}^2(\underline{Y}^+)$ .

Recall that

$$L_1(\underline{Y}^+, \sigma_{c+1}^2; \underline{y}) = -\frac{1}{2} \log |V|$$

$$- \frac{1}{2} \log |X^{*'} V^{-1} X^*|$$

$$- \frac{1}{2} \underline{y}' P \underline{y}, \quad \underline{y} \in \Omega_1^*.$$

Letting

$$P^* \equiv H^{-1} - H^{-1} X (X' H^{-1} X)^{-1} X' H^{-1} = \sigma_{c+1}^2 P,$$

$L_1$  can be re-expressed in the form

$$L_1(\underline{y}^+, \sigma_{c+1}^2; \underline{y}) = - \frac{1}{2} \log |\sigma_{c+1}^2 H|$$

$$- \frac{1}{2} \log |X^{*'} \frac{1}{\sigma_{c+1}^2} H^{-1} X^*|$$

$$- \frac{1}{2} \underline{y}' \frac{1}{\sigma_{c+1}^2} P^* \underline{y}.$$

Define

$$L_c(\underline{y}^+; \underline{y}) \equiv L_1(\underline{y}^+, \hat{\sigma}_{c+1}^2(\underline{y}^+); \underline{y}).$$

Let  $\Omega_c^*$  represent the set of  $\underline{y}^+$  values that correspond to the set of  $\underline{y}^+$  values in  $\Omega_1^*$ . The function  $L_c$  can be regarded as a "concentrated" version of the function  $L_1$ . Note that

$$L_c(\underline{Y}^+; \underline{y}) = -\frac{1}{2} (n-p^*) \log \hat{\sigma}_{c+1}^2(\underline{Y}^+) - \frac{1}{2} \log |H|$$

$$- \frac{1}{2} \log |X^{*'} H^{-1} X^*|$$

$$- \left(\frac{1}{2}\right) \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \underline{y}' P^* \underline{y}, \quad \underline{Y}^+ \in \Omega_c^*.$$

Observing that

$$\underline{y}' P^* \underline{y} = \underline{y}' (\sigma_{c+1}^2 P) \underline{y}$$

$$= \sigma_{c+1}^2 (\underline{y} - X\tilde{\alpha})' V^{-1} (\underline{y} - X\tilde{\alpha}) \quad [\text{using (2.5)}]$$

$$= (\underline{y} - X\tilde{\alpha})' H^{-1} (\underline{y} - X\tilde{\alpha})$$

$$= (n-p^*) \hat{\sigma}_{c+1}^2(\underline{Y}^+), \quad (3.39)$$

the function  $L_c$  can be re-expressed in the form

$$L_c(\underline{Y}^+; \underline{y}) = -\frac{1}{2} (n-p^*) [1 + \log \hat{\sigma}_{c+1}^2(\underline{Y}^+)]$$

$$- \frac{1}{2} \log |H| - \frac{1}{2} |X^{*'} H^{-1} X^*|, \quad \underline{Y}^+ \in \Omega_c^*. \quad (3.40)$$

If  $L_c$  attains its maximum, for  $\underline{\gamma}^+ \in \Omega_c^*$ , at  $\hat{\underline{\gamma}}^+$ , then, clearly, a REML estimate of  $\underline{\gamma}$  is  $(\hat{\underline{\gamma}}^+, \hat{\sigma}_{c+1}^2(\hat{\underline{\gamma}}^+))'$ . Thus, by analytical means, the original  $(c+1)$ -dimensional maximization problem (the maximization of  $L_1$ ) has been reduced to a  $c$ -dimensional maximization problem (the maximization of  $L_c$ ).

The Newton-Raphson method can be applied to the problem of maximizing  $L_c$ . To do so, we require expressions for the first- and second-order partial derivatives of  $L_c$ . Based on expression (3.40), we have that

$$\begin{aligned} \frac{\partial L_c}{\partial \gamma_i} = & -\frac{1}{2} (n-p^*) \frac{\partial \log \hat{\sigma}_{c+1}^2(\underline{\gamma}^+)}{\partial \gamma_i} \\ & - \frac{1}{2} \left[ \frac{\partial \log |H|}{\partial \gamma_i} + \frac{\partial \log |X^{*'} H^{-1} X^*|}{\partial \gamma_i} \right] \quad (i = 1, \dots, c). \end{aligned} \quad (3.41)$$

Recalling that, for an "arbitrary" matrix  $M$  whose elements are functions of a variable  $t$ ,  $\frac{\partial M^{-1}}{\partial t} = -M^{-1} \left( \frac{\partial M}{\partial t} \right) M^{-1}$  and  $\frac{\partial \log |M|}{\partial t} = \text{tr} [M^{-1} \left( \frac{\partial M}{\partial t} \right)]$ , we find, for  $i = 1, \dots, c$ , that

$$\frac{\partial \log |H|}{\partial \gamma_i} = \text{tr} (H^{-1} \frac{\partial H}{\partial \gamma_i}) = \text{tr} (H^{-1} Z_i Z_i'),$$

$$\begin{aligned}
\frac{\partial \log |X^{*'} H^{-1} X^*|}{\partial \gamma_i} &= \text{tr}[(X^{*'} H^{-1} X^*)^{-1} \frac{\partial (X^{*'} H^{-1} X^*)}{\partial \gamma_i}] \\
&= \text{tr}[(X^{*'} H^{-1} X^*)^{-1} X^{*'} (\frac{\partial H^{-1}}{\partial \gamma_i}) X^*] \\
&= \text{tr}[(X^{*'} H^{-1} X^*)^{-1} X^{*'} (-H^{-1} \frac{\partial H}{\partial \gamma_i} H^{-1}) X^*] \\
&= - \text{tr}[(X^{*'} H^{-1} X^*)^{-1} X^{*'} H^{-1} Z_i Z_i' H^{-1} X^*] \\
&= - \text{tr}[H^{-1} X^* (X^{*'} H^{-1} X^*)^{-1} X^{*'} H^{-1} Z_i Z_i'],
\end{aligned}$$

and, hence, that

$$\begin{aligned}
\frac{\partial \log |H|}{\partial \gamma_i} + \frac{\partial \log |X^{*'} H^{-1} X^*|}{\partial \gamma_i} &= \text{tr}\{[H^{-1} - H^{-1} X^* (X^{*'} H^{-1} X^*)^{-1} X^{*'} H^{-1}] Z_i Z_i'\} \\
&= \text{tr}(P^* Z_i Z_i') \\
&= \text{tr}(Z_i' P^* Z_i). \tag{3.42}
\end{aligned}$$

Also, for  $i = 1, \dots, c$ ,

$$\begin{aligned}
\frac{\partial \log \hat{\sigma}_{c+1}^2(\underline{Y}^+)}{\partial \gamma_i} &= \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \frac{\partial \hat{\sigma}_{c+1}^2(\underline{Y}^+)}{\partial \gamma_i} \\
&= \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \left( \frac{1}{n-p^*} \right) \frac{\partial}{\partial \gamma_i} [(\underline{y} - \underline{X}\tilde{\alpha})' H^{-1} (\underline{y} - \underline{X}\tilde{\alpha})] \\
&\quad \text{[using (3.38)]}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \left( \frac{1}{n-p^*} \right) \frac{\partial}{\partial \gamma_i} (\underline{y}' P^* \underline{y}) \\
&\quad \text{[using (3.39)]}
\end{aligned}$$

$$= \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \left( \frac{1}{n-p^*} \right) \underline{y}' \frac{\partial P^*}{\partial \gamma_i} \underline{y} .$$

Since [according to part (i) of Lemma 2.1]

$$P^* = \sigma_{c+1}^2 P = \sigma_{c+1}^2 A' (A' V A)^{-1} A = A (A' H A)^{-1} A' ,$$

implying that

$$\begin{aligned}
\frac{\partial P^*}{\partial \gamma_i} &= A \frac{\partial (A' H A)^{-1}}{\partial \gamma_i} A' \\
&= A [-(A' H A)^{-1} \frac{\partial (A' H A)}{\partial \gamma_i} (A' H A)^{-1}] A'
\end{aligned}$$

$$= - A(A'HA)^{-1} A' Z_i Z_i' A(A'HA)^{-1} A'$$

$$= - P^* Z_i Z_i' P^* \quad (i = 1, \dots, c) ,$$

we have that

$$\frac{\partial \log \hat{\sigma}_{c+1}^2(Y^+)}{\partial \gamma_i} = - \frac{1}{\hat{\sigma}_{c+1}^2(Y^+)} \left( \frac{1}{n-p^*} \right) \underline{y}' P^* Z_i Z_i' P^* \underline{y}$$

$$(i = 1, \dots, c) . \quad (3.43)$$

Substituting from expressions (3.42) and (3.43) into expression (3.41), we obtain

$$\frac{\partial L_c}{\partial \gamma_i} = - \frac{1}{2} [\text{tr}(Z_i' P^* Z_i) - \frac{1}{\hat{\sigma}_{c+1}^2(Y^+)} (\underline{y}' P^* Z_i Z_i' P^* \underline{y})]$$

$$(i = 1, \dots, c) . \quad (3.44)$$

Differentiating expression (3.44) with respect to  $\gamma_j$ , we see that

$$\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = - \frac{1}{2} \left\{ \text{tr} \left( Z_i' \frac{\partial P^*}{\partial \gamma_j} Z_i \right) \right.$$

$$\begin{aligned}
& - \frac{1}{\hat{\sigma}_{c+1}^2 (\underline{Y}^+)} \frac{\partial (\underline{y}' P^* Z_i Z_i' P^* \underline{y})}{\partial \gamma_j} \\
& - \left[ \frac{\partial}{\partial \gamma_j} \left( \frac{1}{\hat{\sigma}_{c+1}^2 (\underline{Y}^+)} \right) \right] (\underline{y}' P^* Z_i Z_i' P^* \underline{y}) \} .
\end{aligned}$$

Observing that

$$\begin{aligned}
\frac{\partial}{\partial \gamma_j} (\underline{y}' P^* Z_i Z_i' P^* \underline{y}) &= \underline{y}' P^* Z_i \frac{\partial (Z_i' P^* \underline{y})}{\partial \gamma_j} \\
&+ \frac{\partial (\underline{y}' P^* Z_i)}{\partial \gamma_j} Z_i' P^* \underline{y} \\
&= \underline{y}' P^* Z_i Z_i' (-P^* Z_j Z_j' P^*) \underline{y} \\
&+ \underline{y}' (-P^* Z_j Z_j' P^*) Z_i Z_i' P^* \underline{y} \\
&= -2 \underline{y}' P^* Z_i Z_i' P^* Z_j Z_j' P^* \underline{y}
\end{aligned}$$

and that

$$\begin{aligned}
\frac{\partial}{\partial \gamma_j} \left( \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right) &= - \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right]^2 \frac{\partial \hat{\sigma}_{c+1}^2(\underline{\gamma}^+)}{\partial \gamma_j} \\
&= \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right]^2 \left( \frac{1}{n-p^*} \right) \underline{y}' P^* Z_j Z_j' P^* \underline{y} ,
\end{aligned}$$

we find that

$$\begin{aligned}
\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} &= - \frac{1}{2} \{ - \text{tr}(Z_i' P^* Z_j Z_j' P^* Z_i) \\
&\quad + 2 \left( \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right) \underline{y}' P^* Z_i Z_i' P^* Z_j Z_j' P^* \underline{y} \\
&\quad - \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right]^2 \left( \frac{1}{n-p^*} \right) (\underline{y}' P^* Z_j Z_j' P^* \underline{y}) (\underline{y}' P^* Z_i Z_i' P^* \underline{y}) \} \\
&\quad (i, j = 1, \dots, c) .
\end{aligned} \tag{3.45}$$

Expressions (3.44) and (3.45) involve the matrix  $P^*$  and, hence, the matrix inverse  $H^{-1}$ . The numerical inversion of  $H$  would be costly for large values of  $n$ . Alternative expressions for  $\frac{\partial L_c}{\partial \gamma_i}$  and  $\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j}$

( $i, j = 1, \dots, c$ ) can be obtained from the following lemma.

Lemma 3.1: For  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ , and  $\underline{\gamma}^+ \in \Omega_c^*$ ,

$$(i) \quad \text{tr}(Z_i' P^* Z_i) = \begin{cases} \text{tr}(G_{ii}) \\ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})], \text{ if } \gamma_i \neq 0, \end{cases}$$

$$(ii) \quad \underline{y}' P^* Z_i Z_i' P^* \underline{y} = \begin{cases} \sigma_{c+1}^4 \tilde{v}_{i-i}' \tilde{v}_{i-i} \\ \left(\frac{1}{\gamma_i}\right)^2 \tilde{b}_{i-i}' \tilde{b}_{i-i}, \text{ if } \gamma_i \neq 0, \end{cases}$$

$$(iii) \quad \text{tr}(Z_i' P^* Z_j Z_j' P^* Z_i) = \begin{cases} \text{tr}(G_{ij} G_{ji}) \\ \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij} T_{ji}), \text{ if } \gamma_i \neq 0, \gamma_j \neq 0, \end{cases}$$

$$(iv) \quad \text{tr}(Z_i' P^* Z_i Z_i' P^* Z_i) = \begin{cases} \text{tr}(G_{ii}^2) \\ \left(\frac{1}{\gamma_i}\right)^2 \text{tr}[(I - T_{ii})^2], \text{ if } \gamma_i \neq 0, \end{cases}$$

$$(v) \quad \underline{y}' P^* Z_i Z_i' P^* Z_j Z_j' P^* \underline{y} = \begin{cases} \sigma_{c+1}^4 \tilde{v}_{i-i}' G_{ij} \tilde{v}_{j-j} \\ - \left(\frac{1}{\gamma_j}\right)^2 \frac{1}{\gamma_i} \tilde{b}_{i-i}' T_{ij} \tilde{b}_{j-j}, \text{ if } \gamma_i \neq 0, \gamma_j \neq 0, \end{cases}$$

$$(vi) \quad \underline{y}' P^* Z_i Z_i' P^* Z_i Z_i' P^* \underline{y} = \begin{cases} \sigma_{c+1}^4 \tilde{\underline{y}}_i' G_{ii} \tilde{\underline{y}}_i \\ (\frac{1}{\gamma_i})^3 \tilde{\underline{b}}_i' (I - T_{ii}) \tilde{\underline{b}}_i, \text{ if } \gamma_i \neq 0. \end{cases}$$

Proof: Define

$$\underline{\Delta}_i \equiv \text{diag}\{\emptyset_{q_1 \times q_1}, \dots, \emptyset_{q_{i-1} \times q_{i-1}}, I_{q_i}, \emptyset_{q_{i+1} \times q_{i+1}}, \dots, \emptyset_{q_c \times q_c}\}$$

$$(i = 1, \dots, c).$$

Suppose that  $\underline{\gamma}^+ \in \Omega_c^*$ , and  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ .

$$\begin{aligned} (i) \quad \text{tr}(Z_i' P^* Z_i) &= \sigma_{c+1}^2 \text{tr}(Z_i' P Z_i) \\ &= \sigma_{c+1}^2 \text{tr}(\underline{\Delta}_i Z_i' P Z_i \underline{\Delta}_i) \\ &= \sigma_{c+1}^2 \text{tr}(\underline{\Delta}_i T Z_i' S Z_i \underline{\Delta}_i) \end{aligned}$$

[using part (ii) of Lemma 2.4]

$$\begin{aligned} &= \sigma_{c+1}^2 \text{tr}(i\text{-th } q_i \times q_i \text{ diagonal submatrix of } T Z_i' S Z_i) \\ &= \sigma_{c+1}^2 \text{tr}\left(\sum_{j=1}^c T_{ij} Z_j' S Z_i\right) \end{aligned}$$

$$= \text{tr}(G_{ii})$$

$$= \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})], \text{ if } \gamma_i \neq 0 \text{ [using (2.29)]}.$$

$$(ii) \quad \underline{y}' P^* Z_i Z_i' P^* \underline{y} = \sigma_{c+1}^4 (\underline{y}' P Z_i) (Z_i' P \underline{y})$$

$$= \sigma_{c+1}^4 \tilde{y}_{-i-i}' \tilde{y}_{-i-i} \text{ [using part (iii) of Lemma 2.4]}$$

$$= \left(\frac{1}{\gamma_i}\right)^2 \tilde{b}_{-i-i}' \tilde{b}_{-i-i}, \text{ if } \gamma_i \neq 0.$$

$$(iii) \quad \text{tr}(Z_i' P^* Z_j Z_j' P^* Z_i) = \sigma_{c+1}^4 \text{tr}(Z_i' P Z_j Z_j' P Z_i)$$

$$= \sigma_{c+1}^4 \text{tr}(\underline{\Delta}_{-i} Z_i' P Z_j Z_j' P \underline{\Delta}_{-i})$$

$$= \sigma_{c+1}^4 \text{tr}(Z_j' P Z \underline{\Delta}_{-i-i} \underline{\Delta}_{-i-i} Z_i' P Z_j)$$

$$= \sigma_{c+1}^4 \text{tr}(\underline{\Delta}_{-j} Z_j' P Z \underline{\Delta}_{-i-i} \underline{\Delta}_{-i-i} Z_i' P Z \underline{\Delta}_{-j})$$

$$= \sigma_{c+1}^4 \text{tr}[(\underline{\Delta}_{-j} T Z' S Z \underline{\Delta}_{-i}) (\underline{\Delta}_{-i} T Z' S Z \underline{\Delta}_{-j})]$$

[using part (ii) of Lemma 2.4]

$$= \sigma_{c+1}^4 \operatorname{tr}\{[(j,i)\text{-th } q_j \times q_i \text{ submatrix of}$$

$$TZ'SZ] \cdot [(i,j)\text{-th } q_i \times q_j \text{ submatrix of}$$

$$TZ'SZ]\}$$

$$= \sigma_{c+1}^4 \operatorname{tr}\left[\left(\sum_{k=1}^c T_{jk} Z_k' S Z_i\right) \left(\sum_{k=1}^c T_{ik} Z_k' S Z_j\right)\right] \quad (3.46)$$

$$= \operatorname{tr}(G_{ji} G_{ij})$$

$$= \frac{1}{\gamma_i} \frac{1}{\gamma_j} \operatorname{tr}(T_{ij} T_{ji}) \quad [\text{using (2.30)}].$$

(iv) Essentially the same derivation that led to result (3.46)

gives

$$\operatorname{tr}(Z_i' P_i^* Z_i Z_i' P_i^* Z_i) = \sigma_{c+1}^4 \operatorname{tr}\left[\left(\sum_{k=1}^c T_{ik} Z_k' S Z_i\right)^2\right]$$

$$= \operatorname{tr}(G_{ii}^2)$$

$$= \left(\frac{1}{\gamma_i}\right)^2 \operatorname{tr}[(I - T_{ii})^2], \text{ if } \gamma_i \neq 0$$

[using (2.29)].

$$(v) \quad \underline{y}' P^* \underline{z}_i \underline{z}_i' P^* \underline{z}_j \underline{z}_j' P^* \underline{y} = \sigma_{c+1}^6 \underline{y}' P \underline{z}_i \underline{z}_i' P \underline{z}_j \underline{z}_j' P \underline{y}$$

$$= \sigma_{c+1}^6 \text{tr}(\underline{z}_i' P \underline{z}_j \underline{z}_j' P \underline{y} \underline{y}' P \underline{z}_i)$$

$$= \sigma_{c+1}^6 \text{tr}(\underline{\Delta}_i \underline{z}_i' P \underline{z}_j \underline{z}_j' P \underline{y} \underline{y}' P \underline{\Delta}_i)$$

$$= \sigma_{c+1}^6 \text{tr}(\underline{\Delta}_j \underline{z}_j' P \underline{y} \underline{y}' P \underline{\Delta}_i \underline{\Delta}_i \underline{z}_i' P \underline{\Delta}_j)$$

$$= \sigma_{c+1}^6 \underline{y}' P \underline{\Delta}_i \underline{z}_i' P \underline{\Delta}_j \underline{z}_j' P \underline{y}$$

$$[\text{since } \underline{\Delta}_i \underline{\Delta}_i = \underline{\Delta}_i]$$

$$= \sigma_{c+1}^6 \tilde{\underline{y}}' \underline{\Delta}_i \underline{z}_i' P \underline{\Delta}_j \tilde{\underline{y}}$$

$$[\text{using part (iii) of Lemma 2.4}]$$

$$= \sigma_{c+1}^6 \tilde{\underline{y}}' [\underline{\Delta}_i T \underline{z}_i' S \underline{\Delta}_j] \tilde{\underline{y}}$$

$$[\text{using part (ii) of Lemma 2.4}]$$

$$= \sigma_{c+1}^6 \tilde{\underline{y}}_i' \left( \sum_{k=1}^c T_{ik} \underline{z}_k' S \underline{z}_j \right) \tilde{\underline{y}}_j \quad (3.47)$$

$$= \sigma_{c+1}^4 \tilde{v}'_{-i} G_{ij} \tilde{v}_{-j}$$

$$= - \sigma_{c+1}^4 \left(\frac{1}{\gamma_j}\right) \tilde{v}'_{-i} T_{ij} \tilde{v}_{-j}, \text{ if } \gamma_j \neq 0$$

[using (2.30)]

$$= - \left(\frac{1}{\gamma_j}\right)^2 \frac{1}{\gamma_i} \tilde{b}'_{-i} T_{ij} \tilde{b}_{-j}, \text{ if } \gamma_i \neq 0, \gamma_j \neq 0.$$

(vi) Essentially the same derivation that led to result (3.47)

gives

$$\underline{y}' P^* Z_i Z_i' P^* Z_i Z_i' P^* \underline{y} = \sigma_{c+1}^6 \tilde{v}'_{-i} \left( \sum_{k=1}^c T_{ik} Z_k' S Z_i \right) \tilde{v}_{-i}$$

$$= \sigma_{c+1}^4 \tilde{v}'_{-i} G_{ii} \tilde{v}_{-i}$$

$$= \sigma_{c+1}^4 \left(\frac{1}{\gamma_i}\right) \tilde{v}'_{-i} (I - T_{ii}) \tilde{v}_{-i}, \text{ if } \gamma_i \neq 0$$

[using (2.29)]

$$= \left(\frac{1}{\gamma_i}\right)^3 \tilde{b}'_{-i} (I - T_{ii}) \tilde{b}_{-i}, \text{ if } \gamma_i \neq 0. \quad \square$$

The expressions given in Lemma 3.1 are in terms of quantities that do not depend on  $\sigma_{c+1}^2$  or can be re-expressed in terms of quantities that do not depend on  $\sigma_{c+1}^2$ , as we now discuss. The matrix  $T$  can be re-expressed as

$$T[I + C \cdot \text{diag}\{\gamma_1 I_{q_1}, \dots, \gamma_c I_{q_c}\}]^{-1}. \quad (3.48)$$

Since  $C = Z'(I - P_X)Z$ , this expression for  $T$  is in terms of quantities that do not depend on  $\sigma_{c+1}^2$ . Also, for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$G_{ij} = \sum_{k=1}^c T_{ik} Z_k'(I - P_X)Z_j, \text{ and}$$

$$G_{ii} = \sum_{k=1}^c T_{ik} Z_k'(I - P_X)Z_i.$$

These expressions, like expression (3.48) for  $T$ , do not depend on  $\sigma_{c+1}^2$ . Observe that, for  $\underline{y} \in \Omega_1^*$ ,

$$\begin{aligned} \underline{\tilde{y}} &= Z'V^{-1}(\underline{y} - X\underline{\tilde{\alpha}}) \\ &= Z'(\sigma_{c+1}^2 H)^{-1}(\underline{y} - X\underline{\tilde{\alpha}}) \\ &= \frac{1}{\sigma_{c+1}^2} Z'H^{-1}(\underline{y} - X\underline{\tilde{\alpha}}) \end{aligned} \quad (3.49)$$

implying that

$$\sigma_{c+1}^4 \tilde{\underline{v}}_{i-i} \tilde{\underline{v}}_{i-i} = [Z_i' H^{-1} (\underline{y} - X \tilde{\underline{\alpha}})]' [Z_i' H^{-1} (\underline{y} - X \tilde{\underline{\alpha}})] \quad (3.50)$$

and

$$\sigma_{c+1}^4 \tilde{\underline{v}}_{i-i} G_{ij} \tilde{\underline{v}}_{j-j} = [Z_i' H^{-1} (\underline{y} - X \tilde{\underline{\alpha}})]' G_{ij} [Z_j' H^{-1} (\underline{y} - X \tilde{\underline{\alpha}})]. \quad (3.51)$$

Since  $H = I_n + \sum_{i=1}^c \gamma_i Z_i Z_i'$  and since  $\tilde{\underline{\alpha}}$  represents any solution to  $X'V^{-1}X\tilde{\underline{\alpha}} = X'V^{-1}\underline{y}$  or, equivalently, to  $X'H^{-1}X\tilde{\underline{\alpha}} = X'H^{-1}\underline{y}$ , the right-hand sides of equations (3.50) and (3.51) are in terms of quantities that do not depend on  $\sigma_{c+1}^2$ . Further, since  $\tilde{\underline{b}}_i = \sigma_{c+1}^2 \gamma_i \tilde{\underline{v}}_{i-i}$  ( $i = 1, \dots, c$ ), (3.49) implies that

$$\tilde{\underline{b}}_i = \gamma_i Z_i' H^{-1} (\underline{y} - X \tilde{\underline{\alpha}}) \quad (i = 1, \dots, c).$$

Note that this expression for  $\tilde{\underline{b}}_i$  does not depend on  $\sigma_{c+1}^2$ .

If  $D$  is a nonsingular matrix, then, as discussed in Section II.D.2,  $\tilde{\underline{\alpha}}$  and  $\tilde{\underline{b}}$  can be obtained as the solution to Henderson's (1963) mixed-model

$$\text{equations } \begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & D^{-1} + Z'R^{-1}X \end{pmatrix} \begin{pmatrix} \tilde{\underline{\alpha}} \\ \tilde{\underline{b}} \end{pmatrix} = \begin{pmatrix} X'R^{-1}\underline{y} \\ Z'R^{-1}\underline{y} \end{pmatrix}. \quad \text{Since } R = \sigma_{c+1}^2 I_n,$$

this system of equations is equivalent to the system

$$\begin{pmatrix} X'X & X'Z \\ Z'X & [\text{diag}\{\gamma_1 I_{q_1}, \dots, \gamma_c I_{q_c}\}]^{-1} + Z'X \end{pmatrix} \begin{pmatrix} \tilde{\underline{\alpha}} \\ \tilde{\underline{b}} \end{pmatrix} = \begin{pmatrix} X'\underline{y} \\ Z'\underline{y} \end{pmatrix}.$$

Note that the coefficient matrix and the right-hand side of the latter system does not depend on  $\sigma_{c+1}^2$ .

Substituting from Lemma 3.1 into expressions (3.44) and (3.45), we find, in particular, that, for  $\underline{\gamma} \in \Omega_1^*$  and  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$\frac{\partial L_c}{\partial \gamma_i} = -\frac{1}{2} \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})] \right.$$

$$\left. - \frac{1}{\hat{\sigma}_{c+1}^2 (\underline{\gamma}^+)} \left( \frac{1}{\gamma_i} \right)^2 \tilde{b}_{i-i} \tilde{b}_{i-i} \right\}, \text{ if } \gamma_i \neq 0,$$

$$\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = -\frac{1}{2} \left\{ -\frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij} T_{ji}) \right.$$

$$\left. - 2 \left[ \frac{1}{\hat{\sigma}_{c+1}^2 (\underline{\gamma}^+)} \right] \left( \frac{1}{\gamma_j} \right)^2 \frac{1}{\gamma_i} \tilde{b}_{i-i} T_{ij} \tilde{b}_{j-j} \right.$$

$$\left. - \left[ \frac{1}{\hat{\sigma}_{c+1}^2 (\underline{\gamma}^+)} \right]^2 \left( \frac{1}{\gamma_i} \right)^2 \left( \frac{1}{\gamma_j} \right)^2 (\tilde{b}_{i-i} \tilde{b}_{i-i}) (\tilde{b}_{j-j} \tilde{b}_{j-j}) \right\},$$

$$\text{if } \gamma_i \neq 0, \gamma_j \neq 0$$

$$\begin{aligned}
\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i} = & -\frac{1}{2} \left\{ -\left(\frac{1}{\gamma_i}\right)^2 \text{tr}[(I - T_{ii})^2] \right. \\
& + 2 \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \right] \left(\frac{1}{\gamma_i}\right)^3 \tilde{b}_{-i}' (I - T_{ii}) \tilde{b}_{-i} \\
& \left. - \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \right]^2 \left(\frac{1}{n-p^*}\right) \left(\frac{1}{\gamma_i}\right)^4 (\tilde{b}_{-i}' \tilde{b}_{-i})^2 \right\} \text{ if } \gamma_i \neq 0.
\end{aligned}$$

Define

$$\underline{k}(\underline{Y}^+) = \left[ \frac{\partial L_c}{\partial \gamma_i} \right] \quad (i = 1, \dots, c),$$

$$K(\underline{Y}^+) = \left[ \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} \right] \quad (i, j = 1, \dots, c).$$

When applied to the concentrated log-likelihood function  $L_c(\underline{Y}^+; \underline{y})$ , the (p+1)-st iterate of the traditional Newton-Raphson method is

$$\underline{Y}^{+(p+1)} = \underline{Y}^{+(p)} - K^{-1}(\underline{Y}^{+(p)}) \underline{k}(\underline{Y}^{+(p)}). \quad (3.52)$$

Corresponding to (p+1)-st iterate  $\underline{Y}^{+(p+1)}$  is the value  $\hat{\sigma}_{c+1}^2(\underline{Y}^{+(p+1)})$  at which  $L_1(\underline{Y}^{+(p+1)}, \sigma_{c+1}^2; \underline{y})$  attains its maximum. The vector  $\begin{bmatrix} \underline{Y}^{+(p+1)} \\ \hat{\sigma}_{c+1}^2(\underline{Y}^{+(p+1)}) \end{bmatrix}$

can thus be regarded as the  $(p+1)$ -st iterate of an algorithm for determining a REML estimate of  $\underline{\gamma}$ .

b. Alternative parameterization                      Define

$$\underline{\psi} \equiv (\psi_1, \dots, \psi_{c+1})'$$

and

$$\underline{\psi}^+ \equiv (\psi_1, \dots, \psi_c)',$$

where

$$\psi_i \equiv \begin{cases} \frac{1}{1 + \sum_{j=i}^c \lambda_{ji} \gamma_j} & \text{for } i = 1, \dots, c \\ \sigma_{c+1}^2 & \text{for } i = c+1. \end{cases}$$

Let

$$\underline{\psi}^* = \begin{pmatrix} \frac{1}{\psi_1} - 1 \\ \vdots \\ \frac{1}{\psi_c} - 1 \end{pmatrix}.$$

Then

$$\underline{\psi}^* = \Lambda'_{11} \underline{\gamma}^+,$$

where

$$\Lambda_{11} = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ \lambda_{21} & \lambda_{22} & & 0 \\ \vdots & \vdots & & \vdots \\ \lambda_{c1} & \lambda_{c2} & \dots & \lambda_{cc} \end{pmatrix},$$

so that

$$\underline{y}^+ = W'_{11} \underline{\psi}^*, \text{ and}$$

$$\gamma_{c+1} = \psi_{c+1},$$

where

$$W_{11} = \Lambda_{11}^{-1} = \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ w_{21} & w_{22} & & 0 \\ \vdots & \vdots & & \vdots \\ w_{c1} & w_{c2} & \dots & w_{cc} \end{pmatrix}.$$

Let

$$L_R^*(\underline{\psi}^+, \sigma_{c+1}^2; \underline{y}) = L_1(W'_{11} \underline{\psi}^*, \sigma_{c+1}^2; \underline{y}).$$

Then,  $L_R^*$  is (aside from an additive constant) a reparameterization of the log-likelihood function associated with  $A'\underline{y}$ . Let  $\Omega_R^*$  represent the set of  $\underline{\psi}$  values that correspond to the set  $\Omega_1^*$  of  $\underline{y}$  values. Take

$$L_R(\underline{\psi}^+; \underline{y}) = L_R^*(\underline{\psi}^+, \hat{\psi}_{c+1}(\underline{\psi}^+); \underline{y}),$$

where  $\hat{\psi}_{c+1}(\underline{\psi}^+)$  is, for an arbitrary (fixed) value of  $\underline{\psi}^+$ , the value of  $\psi_{c+1}$  at which the function  $L_R^*(\underline{\psi}^+, \sigma_{c+1}^2; \underline{y})$  assumes its maximum over the interval  $0 < \sigma_{c+1}^2 < \infty$ . Clearly,

$$\hat{\psi}_{c+1}(\underline{\psi}^+) = \hat{\sigma}_{c+1}^2(W'_{11}\underline{\psi}^*),$$

so that

$$L_R^*(\underline{\psi}^+, \hat{\psi}_{c+1}(\underline{\psi}^+); \underline{y}) = L_1(W'_{11}\underline{\psi}^*, \hat{\sigma}_{c+1}^2(W'_{11}\underline{\psi}^*); \underline{y})$$

$$= L_c(W'_{11}\underline{\psi}^*; \underline{y}) .$$

It follows from the chain rule of calculus that

$$\frac{\partial L_R}{\partial \psi_i} = \sum_{j=1}^c \frac{\partial \gamma_j}{\partial \psi_i} \frac{\partial L_c}{\partial \gamma_j} \quad (i = 1, \dots, c) .$$

for  $i, j \in \{1, \dots, c\}$ , we have that

$$\frac{\partial \gamma_j}{\partial \psi_i} = - \frac{w_{ij}}{\psi_i^2} .$$

Thus,

$$\frac{\partial L_R}{\partial \psi_i} = - \sum_{j=1}^c \left( \frac{w_{ij}}{\psi_i^2} \right) \frac{\partial L_c}{\partial \gamma_j} \quad (i = 1, \dots, c). \quad (3.53)$$

Further, for  $i, k \in \{1, \dots, c\}$ ,

$$\begin{aligned} \frac{\partial^2 L_R}{\partial \psi_i \partial \psi_k} &= \sum_{j=1}^c \left\{ \frac{\partial^2 \gamma_j}{\partial \psi_i \partial \psi_k} \frac{\partial L_c}{\partial \gamma_j} + \frac{\partial \gamma_j}{\partial \psi_i} \left[ \sum_{\ell=1}^c \frac{\partial \gamma_\ell}{\partial \psi_k} \frac{\partial^2 L_c}{\partial \gamma_j \partial \gamma_\ell} \right] \right\} \\ &= \begin{cases} 2 \left( \frac{1}{\psi_i} \right)^3 \sum_{j=1}^c w_{ij} \frac{\partial L_c}{\partial \gamma_j} + \left( \frac{1}{\psi_i} \right)^4 \sum_{j=1}^c w_{ij} \left[ \sum_{\ell=1}^c w_{i\ell} \frac{\partial^2 L_c}{\partial \gamma_j \partial \gamma_\ell} \right] & \text{if } i = k \\ \left( \frac{1}{\psi_i} \right)^2 \left( \frac{1}{\psi_k} \right)^2 \sum_{j=1}^c w_{ij} \left[ \sum_{\ell=1}^k w_{k\ell} \frac{\partial^2 L_c}{\partial \gamma_j \partial \gamma_\ell} \right], & \text{if } i \neq k. \end{cases} \end{aligned}$$

Expressions for  $\frac{\partial L_c}{\partial \gamma_i}$  and  $\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j}$  ( $i, j = 1, \dots, c$ ) are given in Section

III.B.3.a.

Define

$$\underline{m}(\underline{\psi}^+) = \left[ \left[ \frac{\partial L_R}{\partial \psi_i} \right] \right] \quad (i = 1, \dots, c)$$

and

$$\underline{M}(\underline{\psi}^+) = \left[ \left[ \frac{\partial^2 L_R}{\partial \psi_i \partial \psi_j} \right] \right] \quad (i, j = 1, \dots, c).$$

As applied to the maximization of the function  $L_R(\underline{\psi}^+; \underline{y})$ , the (p+1)-st iterate of the traditional Newton-Raphson method is

$$\underline{\psi}^{(p+1)} = \underline{\psi}^{+(p)} - M^{-1}(\underline{\psi}^{+(p)}) \underline{m}(\underline{\psi}^{+(p)}) . \quad (3.54)$$

Corresponding to the (p+1)-st iterate  $\underline{\psi}^{+(p+1)}$  is the value

$\hat{\sigma}_{c+1}^2(W_{11}^* \underline{\psi}^{+(p+1)})$  at which  $L_R^*(\underline{\psi}^{+(p)}, \sigma_{c+1}^2; \underline{y})$  attains its maximum with

respect to  $\sigma_{c+1}^2$ . The vector  $\begin{pmatrix} \underline{\psi}^{+(p+1)} \\ \hat{\sigma}_{c+1}^2(W_{11}^* \underline{\psi}^{+(p+1)}) \end{pmatrix}$  can thus be regarded as

the (p+1)-st iterate of an algorithm for determining a REML estimate of  $\underline{\psi}$ .

Since  $\psi_1, \dots, \psi_c$  are nonlinear functions of  $\gamma_1, \dots, \gamma_c$ , the sequence of iterates obtained by applying the Newton-Raphson method to the problem of maximizing  $L_R(\underline{\psi}^+; \underline{y})$  will not, in general, be the same as that obtained by applying it to the problem of maximizing  $L_c(\underline{\gamma}^+; \underline{y})$ .

#### 4. Linearization as applied to the concentrated log-likelihood function

In Section III.B.2, we attempted to improve on the performance of the Newton-Raphson method, as applied to  $L_1$  or, equivalently, to  $L_v$ , by "linearizing" the likelihood equations before applying Newton's method. In this section, we consider linearization as a device for improving on the performance of the Newton-Raphson method when this method is applied to the problem of maximizing  $L_R$  or  $L_c$ .

a. The  $\psi$ -parameterization

The iterates obtained by applying the traditional Newton-Raphson method to the problem of maximizing  $L_R$  are the same as those obtained by using Newton's method to solve the nonlinear system of equations  $\left[ \frac{\partial L_R}{\partial \psi_i} \right] = 0$  ( $i = 1, \dots, c$ ). The rate of convergence of these iterates depends, in part, on the degree of nonlinearity of these equations. We wish to speed up the rate of convergence by finding an equivalent system of equations that are more nearly linear than those in the original system.

Following essentially the same approach as in Section III.B.2, we begin by supposing that an ANOVA( $\sigma^2$ ) exists and that  $s = c+1$ . Then, since  $L_1^*(\underline{m}; \underline{y}) = L_s^*(\underline{m}; \underline{y}) + k_1$  for some scalar  $k_1$  that does not depend on  $\underline{m}$ ,

$$L_1^*(\underline{m}; \underline{y}) = k_2 - \sum_{i=1}^{c+1} \left( \frac{1}{2} \right) (r_i - 2) \log(m_i)$$

$$- \sum_{i=1}^{c+1} \left( \frac{1}{2} \right) \frac{1}{m_i} s_i$$

$$- \sum_{i=1}^{c+1} \log(m_i) ,$$

for some scalar  $k_2$  that does not depend on  $\underline{m}$ . Thus,

$$L_R^*(\psi_1, \dots, \psi_c, \sigma_{c+1}^2; \underline{y}) = L_1^*\left(\frac{\sigma_{c+1}^2}{\psi_1}, \dots, \frac{\sigma_{c+1}^2}{\psi_c}, \sigma_{c+1}^2; \underline{y}\right)$$

$$= k_2 - \sum_{i=1}^c \left(\frac{1}{2}\right) (r_i - 2) \log\left(\frac{\sigma_{c+1}^2}{\psi_i}\right)$$

$$- \sum_{i=1}^c \left(\frac{1}{2}\right) S_i\left(\frac{\psi_i}{\sigma_{c+1}^2}\right)$$

$$- \sum_{i=1}^c \log\left(\frac{\sigma_{c+1}^2}{\psi_i}\right)$$

$$- \left(\frac{1}{2}\right) (r_{c+1} - 2) \log(\sigma_{c+1}^2)$$

$$- \left(\frac{1}{2}\right) S_{c+1}\left(\frac{1}{\sigma_{c+1}^2}\right) - \log(\sigma_{c+1}^2)$$

$$= k_2 - \sum_{i=1}^c \left(\frac{1}{2}\right) (r_i - 2) \log(\sigma_{c+1}^2)$$

$$+ \sum_{i=1}^c \left(\frac{1}{2}\right) (r_i - 2) \log(\psi_i)$$

$$- \sum_{i=1}^c \left(\frac{1}{2}\right) S_i\left(\frac{\psi_i}{\sigma_{c+1}^2}\right) - \sum_{i=1}^c \log(\sigma_{c+1}^2)$$

$$\begin{aligned}
& + \sum_{i=1}^c \log(\psi_i) - \left(\frac{1}{2}\right)(r_{c+1}-2)\log(\sigma_{c+1}^2) \\
& - \left(\frac{1}{2}\right)s_{c+1}\left(\frac{1}{\sigma_{c+1}^2}\right) - \log(\sigma_{c+1}^2) \\
= & k_2 - [\log(\sigma_{c+1}^2)] \left[ \sum_{i=1}^c \left(\frac{1}{2}\right)(r_i-2) + c \right. \\
& \left. + \left(\frac{1}{2}\right)(r_{c+1}-2) + 1 \right] \\
& + \sum_{i=1}^c \left[ \left(\frac{1}{2}\right)(r_i-2) + 1 \right] [\log(\psi_i)] \\
& - \left(\frac{1}{2}\right) \sum_{i=1}^c s_i \left(\frac{\psi_i}{\sigma_{c+1}^2}\right) - \left(\frac{1}{2}\right) s_{c+1} \left(\frac{1}{\sigma_{c+1}^2}\right) \\
= & k_2 - \left(\frac{1}{2}\right)(n-p^*)\log(\sigma_{c+1}^2) \\
& + \sum_{i=1}^c [\log(\psi_i)] \left(\frac{r_i}{2}\right) \\
& - \left(\frac{1}{2}\right) \sum_{i=1}^c s_i \left(\frac{\psi_i}{\sigma_{c+1}^2}\right) - \left(\frac{1}{2}\right) s_{c+1} \left(\frac{1}{\sigma_{c+1}^2}\right) \\
& \quad \quad \quad \left[ \text{since } \sum_{i=1}^{c+1} r_i = n-p^* \right]
\end{aligned}$$

$$\begin{aligned}
&= k_2 - \left(\frac{1}{2}\right)(n-p^*) \log(\sigma_{c+1}^2) \\
&\quad + \left(\frac{1}{2}\right) \sum_{i=1}^c r_i \log(\psi_i) \\
&\quad - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right), \\
&\quad \underline{\psi} \in \Omega_R^* . \tag{3.55}
\end{aligned}$$

Recall that  $L_R(\underline{\psi}^+; \underline{y})$  can be obtained from  $L_R^*(\underline{\psi}; \underline{y})$  by replacing  $\psi_{c+1}$  by  $\hat{\psi}_{c+1}(\underline{\psi}^+)$ . Making use of expression (3.55), we find that

$$\begin{aligned}
\frac{\partial L_R^*}{\partial \sigma_{c+1}^2} &= - \left(\frac{1}{2}\right)(n-p^*) \left(\frac{1}{\sigma_{c+1}^2}\right) \\
&\quad + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right) \\
&= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (n-p^*) [\sigma_{c+1}^2 - \left(\frac{1}{n-p^*}\right) \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right)] ,
\end{aligned}$$

implying that

$$\hat{\psi}_{c+1}(\underline{\psi}^+) = \frac{1}{n-p^*} \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right) . \tag{3.56}$$

Thus,

$$\begin{aligned}
 L_R(\underline{\psi}^+; \underline{y}) &= L_R^*(\underline{\psi}^+, \hat{\psi}_{c+1}(\underline{\psi}^+); \underline{y}) \\
 &= k_2 - \left(\frac{1}{2}\right)(n-p^*) \log(\hat{\psi}_{c+1}(\underline{\psi}^+)) \\
 &\quad + \left(\frac{1}{2}\right) \sum_{i=1}^c r_i \log(\psi_i) \\
 &\quad - \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right] \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right).
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \frac{\partial L_R}{\partial \psi_j} &= - \left(\frac{1}{2}\right)(n-p^*) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right] \frac{\partial \hat{\psi}_{c+1}(\underline{\psi}^+)}{\partial \psi_j} + \left(\frac{1}{2}\right) \left( \frac{r_j}{\psi_j} \right) \\
 &\quad + \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right]^2 \frac{\partial \hat{\psi}_{c+1}(\underline{\psi}^+)}{\partial \psi_j} \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right) \\
 &\quad - \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right] s_j \quad (j = 1, \dots, c) .
 \end{aligned}$$

Moreover, using result (3.56), we find that

$$\frac{\partial \hat{\psi}_{c+1}(\underline{\psi}^+)}{\partial \psi_j} = \left(\frac{1}{n-p^*}\right) s_j \quad (j = 1, \dots, c)$$

so that

$$\begin{aligned} \frac{\partial L_R}{\partial \psi_j} &= - \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right] s_j + \left(\frac{1}{2}\right) \left(\frac{r_j}{\psi_j}\right) \\ &\quad + \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right]^2 \left(\frac{1}{n-p^*}\right) s_j [(n-p^*) \hat{\psi}_{c+1}(\underline{\psi}^+)] \\ &\quad - \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\psi}_{c+1}(\underline{\psi}^+)} \right] s_j \\ &= - \left(\frac{1}{2}\right) (n-p^*) \left\{ \frac{1}{\sum_{i=1}^c \psi_i s_i + s_{c+1}} \right\} s_j + \left(\frac{1}{2}\right) \left(\frac{r_j}{\psi_j}\right) \\ &= - \left(\frac{1}{2}\right) \left[ \frac{(n-p^*) s_j \psi_j - r_j \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right)}{\psi_j \left( \sum_{i=1}^c \psi_i s_i + s_{c+1} \right)} \right] \\ &= - \left(\frac{1}{2}\right) \left\{ \frac{(n-p^*) \left(\frac{s_j}{s_{c+1}}\right) \psi_j - r_j \left[ \sum_{i=1}^c \psi_i \left(\frac{s_i}{s_{c+1}}\right) + 1 \right]}{\psi_j \left[ \sum_{i=1}^c \psi_i \left(\frac{s_i}{s_{c+1}}\right) + 1 \right]} \right\} \quad (j=1, \dots, c) \end{aligned} \quad (3.57)$$

Setting the partial derivatives (3.57) to zero gives the equations

$$\frac{\partial L_R}{\partial \psi_j} = 0 \quad (j = 1, \dots, c). \quad (3.58)$$

Note that, if we multiply both sides of the  $j$ -th of these equations by

$$\psi_j \left[ \sum_{i=1}^c \psi_i \left( \frac{S_i}{S_{c+1}} \right) + 1 \right] \quad (j = 1, \dots, c),$$

we obtain the equivalent set of equations

$$\psi_j \left[ \sum_{i=1}^c \psi_i \left( \frac{S_i}{S_{c+1}} \right) + 1 \right] \frac{\partial L_R}{\partial \psi_j} = 0 \quad (j = 1, \dots, c). \quad (3.59)$$

The left-hand sides of equations (3.58) are nonlinear in  $\psi_1, \dots, \psi_c$ , while the left-hand sides of equations (3.59) are linear. Consequently, if Newton's method were applied to equations (3.59), it would converge in a single iteration. In contrast, if it were applied directly to equations (3.58), additional iterations would be required (and, even then, only an approximate solution would be obtained).

Let us now consider the general problem of computing a solution to the equations

$$\frac{\partial L_R}{\partial \psi_j} = 0 \quad (j = 1, \dots, c) \quad (3.60)$$

(dropping the supposition that an ANOVA( $\sigma^2$ ) exists). Instead of applying

Newton's method directly to these equations, that is, instead of applying the Newton-Raphson method directly to the problem of maximizing the function  $L_R$ , we can instead apply Newton's method to the equations

$$m_j^*(\underline{\psi}^+) = 0 \quad (j = 1, \dots, c), \quad (3.61)$$

where  $m_j^*(\underline{\psi}^+) = \psi_j \left[ \sum_{i=1}^c \psi_i \left( \frac{S_i}{S_{c+1}} \right) + 1 \right] \frac{\partial L_R}{\partial \psi_j}$ . Clearly, equations (3.60)

are equivalent to (3.61). Hopefully, the functions  $m_j^*(\underline{\psi}^+)$  ( $j = 1, \dots, c$ ) will be more nearly linear than the functions  $\frac{\partial L_R}{\partial \psi_j}$  ( $j = 1, \dots, c$ ).

Let

$$\underline{m}^*(\underline{\psi}^+) = \left[ m_i^*(\underline{\psi}^+) \right] \quad (i = 1, \dots, c), \text{ and}$$

$$M^*(\underline{\psi}^+) = \left[ \frac{\partial m_i^*}{\partial \psi_j} \right] \quad (i, j = 1, \dots, c).$$

We have, for  $j, k \in \{1, \dots, c\}$ , that

$$\frac{\partial m_j^*}{\partial \psi_k} = \begin{cases} m_j^*(\underline{\psi}^+) \frac{\partial^2 L_R}{\partial \psi_j \partial \psi_k} + \psi_j \left( \frac{S_k}{S_{c+1}} \right) \frac{\partial L_R}{\partial \psi_j}, & \text{if } j \neq k \\ m_j^*(\underline{\psi}^+) \frac{\partial^2 L_R}{\partial \psi_j^2} + \left\{ \left[ \sum_{i=1}^c \psi_i \left( \frac{S_i}{S_{c+1}} \right) + 1 \right] \right. \\ \quad \left. + \psi_j \left( \frac{S_j}{S_{c+1}} \right) \right\} \frac{\partial L_R}{\partial \psi_j}, & \text{if } j = k. \end{cases}$$

Expressions for  $\frac{\partial L_R}{\partial \psi_j}$  and  $\frac{\partial^2 L_R}{\partial \psi_j \partial \psi_k}$  are given in Section III.B.3.b. Let  $\underline{\psi}^{+(p)}$  represent the p-th iterate obtained by applying Newton's method to equations (3.61). Then,

$$\underline{\psi}^{+(p+1)} = \underline{\psi}^{+(p)} - M^{*-1}(\underline{\psi}^{+(p)}) \underline{m}^*(\underline{\psi}^{+(p)}). \quad (3.62)$$

#### b. The $\gamma$ -parameterization

The iterates obtained by

applying the traditional Newton-Raphson method to the problem of maximizing  $L_c$  are the same as those obtained by using Newton's method to solve the nonlinear system of equations  $\frac{\partial L_c}{\partial \gamma_j} = 0$  ( $j = 1, \dots, c$ ). We now consider whether these equations, like the equations  $\frac{\partial L_1^*}{\partial m_j} = 0$  ( $j = 1, \dots, c+1$ ) and  $\frac{\partial L_R}{\partial \psi_j} = 0$  ( $j = 1, \dots, c$ ), can be "linearized."

Let us begin by supposing that an ANOVA( $\sigma^2$ ) exists and that  $s = c+1$ . Then, letting  $g_i(\underline{\gamma}^+) = 1 + \sum_{j=i}^c \lambda_{ji} \gamma_j = \psi_i^{-1}$  ( $i = 1, \dots, c$ ),

$$L_c(\underline{\gamma}^+; \underline{y}) = L_R[g_1^{-1}(\underline{\gamma}^+), \dots, g_c^{-1}(\underline{\gamma}^+); \underline{y}]$$

$$= - \left(\frac{1}{2}\right) (n-p^*) \log(\hat{\sigma}_{c+1}^2(\underline{\gamma}^+))$$

$$- \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{\gamma}^+)} \right] \left( \sum_{i=1}^c \frac{s_i}{g_i(\underline{\gamma}^+)} + s_{c+1} \right)$$

$$- \left(\frac{1}{2}\right) \sum_{i=1}^c r_i \log(g_i(\underline{Y}^+))$$

so that

$$\begin{aligned} \frac{\partial L_c}{\partial \gamma_k} = & - \left(\frac{1}{2}\right) (n-p^*) \left[ \frac{1}{\sigma_{c+1}^2(\underline{Y}^+)} \right] \frac{\partial \hat{\sigma}_{c+1}^2(\underline{Y}^+)}{\partial \gamma_k} \\ & + \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \right]^2 \left\{ \sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1} \right\} \frac{\partial \hat{\sigma}_{c+1}^2(\underline{Y}^+)}{\partial \gamma_k} \\ & - \left(\frac{1}{2}\right) \left[ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \right] \left[ - \sum_{i=1}^c \frac{s_i}{[g_i(\underline{Y}^+)]^2} \frac{\partial g_i(\underline{Y}^+)}{\partial \gamma_k} \right] \\ & - \left(\frac{1}{2}\right) \sum_{i=1}^c r_i \left[ \frac{1}{g_i(\underline{Y}^+)} \right] \frac{\partial g_i(\underline{Y}^+)}{\partial \gamma_k} \quad (k = 1, \dots, c). \end{aligned}$$

Observe that

$$\frac{\partial g_i(\underline{Y}^+)}{\partial \gamma_k} = \lambda_{ki} \quad (i, k = 1, \dots, c)$$

and that

$$\hat{\sigma}_{c+1}^2(\underline{Y}^+) = \frac{1}{n-p^*} \left( \sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1} \right),$$

implying that

$$\frac{\partial \hat{\sigma}_{c+1}^2(\underline{Y}^+)}{\partial \gamma_k} = - \frac{1}{n-p^*} \left[ \sum_{i=1}^c \frac{s_i}{[g_i(\underline{Y}^+)]^2} (\lambda_{ki}) \right].$$

Thus,

$$\begin{aligned} \frac{\partial L_c}{\partial \gamma_k} = & - \left(\frac{1}{2}\right) (n-p^*) \left[ \frac{1}{\sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1}} \right] \left[ - \sum_{i=1}^c \frac{s_i}{[g_i(\underline{Y}^+)]^2} (\lambda_{ki}) \right] \\ & + \left\{ \left(\frac{1}{2}\right) \left[ \frac{n-p^*}{\sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1}} \right]^2 \left( \sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1} \right) \right. \\ & \quad \cdot \left. \left[ \frac{-1}{n-p^*} \right] \left[ \sum_{i=1}^c \frac{s_i}{[g_i(\underline{Y}^+)]^2} (\lambda_{ki}) \right] \right\} \\ & - \left(\frac{1}{2}\right) \left[ \frac{n-p^*}{\sum_{i=1}^c \frac{s_i}{g_i(\underline{Y}^+)} + s_{c+1}} \right] \left[ - \sum_{i=1}^c \frac{s_i}{[g_i(\underline{Y}^+)]^2} (\lambda_{ki}) \right] \\ & - \left(\frac{1}{2}\right) \sum_{i=1}^c r_i \left( \frac{1}{g_i(\underline{Y}^+)} \right) \lambda_{ki} \end{aligned}$$

$$\begin{aligned}
&= - \left(\frac{1}{2}\right) \left\{ \left[ \frac{\frac{n-p^*}{\sum_{i=1}^c \frac{S_i}{(1 + \sum_{j=i}^c \lambda_{ji} \gamma_j)} + S_{c+1}}}{\sum_{i=1}^c \frac{S_i \lambda_{ki}}{(1 + \sum_{j=i}^c \lambda_{ji} \gamma_j)^2}} \right] \right. \\
&\quad \left. + \sum_{i=1}^c \frac{r_i \lambda_{ki}}{(1 + \sum_{j=i}^c \lambda_{ji} \gamma_j)} \right\} \quad (k = 1, \dots, c). \quad (3.63)
\end{aligned}$$

Based on expression (3.63), it would seem that, in general, there is no convenient way to linearize the equations  $\frac{\partial L_c}{\partial \gamma_k} = 0 \quad (k = 1, \dots, c)$ .

However, in the special case  $c=1$ , these equations can be linearized as we now show. Letting

$$F \equiv \frac{r_2}{r_1} \frac{S_1}{S_2},$$

we find that, when  $c=1$ , expression (3.63) can be rewritten as follows:

$$\begin{aligned}
\frac{\partial L_c}{\partial \gamma_1} = - \left(\frac{1}{2}\right) \left\{ \left[ \frac{\frac{n-p^*}{\frac{S_1}{1 + \lambda_{11} \gamma_1} + S_2}}{- \frac{S_1 \lambda_{11}}{(1 + \lambda_{11} \gamma_1)^2}} \right] \right. \\
\quad \left. + \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\}
\end{aligned}$$

$$\begin{aligned}
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left\{ 1 - \frac{1}{r_1 \lambda_{11}} \left\{ \frac{s_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \right. \\
&\quad \cdot \left. \left[ \frac{(n-p^*)(1 + \lambda_{11} \gamma_1)}{s_1 + s_2(1 + \lambda_{11} \gamma_1)} \right] \right\} \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left[ 1 - \frac{1}{r_1} (n-p^*) \left\{ \frac{s_1}{s_1 + s_2 + \lambda_{11} s_2 \gamma_1} \right\} \right] \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left\{ 1 - \frac{1}{r_1} (n-p^*) \left[ \frac{\left(\frac{r_1}{r_2}\right)^F}{\left(\frac{r_1}{r_2}\right)^F + 1 + \lambda_{11} \gamma_1} \right] \right\} \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left\{ 1 - \frac{(r_1 + r_2) \left(\frac{1}{r_2}\right)^F}{[1 + \lambda_{11} \gamma_1 + \left(\frac{r_1}{r_2}\right)^F]} \right\} \\
&\quad \text{[since } r_1 + r_2 = n-p^*] \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left[ \frac{1 + \lambda_{11} \gamma_1 + \left(\frac{r_1}{r_2}\right)^F - \left(\frac{r_1}{r_2}\right)^F - F}{1 + \lambda_{11} \gamma_1 + \left(\frac{r_1}{r_2}\right)^F} \right] \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{r_1 \lambda_{11}}{1 + \lambda_{11} \gamma_1} \right\} \left[ \frac{1 + \lambda_{11} \gamma_1 - F}{1 + \lambda_{11} \gamma_1 + \left(\frac{r_1}{r_2}\right)^F} \right].
\end{aligned}$$

Thus, when  $c=1$ , we can linearize the equation  $\frac{\partial L_c}{\partial \gamma_1} = 0$  by multiplying both sides by the quantity  $(1 + \lambda_{11}\gamma_1) [1 + \lambda_{11}\gamma_1 + (\frac{r_1}{r_2})F]$ .

Let us now consider, for the special case  $c=1$ , the general problem of computing a solution to the equation  $\frac{\partial L_c}{\partial \gamma_1} = 0$  (dropping the supposition that an ANOVA( $\sigma^2$ ) exists). Instead of applying Newton's method directly to the equation  $\frac{\partial L_c}{\partial \gamma_1} = 0$ , that is, instead of applying the Newton-Raphson method directly to the problem of maximizing the function  $L_c$ , we can instead apply Newton's method to the "linearized" equation  $k^*(\gamma_1) = 0$ , where

$$k^*(\gamma_1) \equiv (1 + \lambda_{11}\gamma_1) [1 + \lambda_{11}\gamma_1 + (\frac{r_1}{r_2})F] \frac{\partial L_c}{\partial \gamma_1}.$$

Define  $K^*(\gamma_1) = \frac{\partial k^*}{\partial \gamma_1}$ . Then

$$\begin{aligned} K^*(\gamma_1) &= \lambda_{11} [1 + \lambda_{11}\gamma_1 + (\frac{r_1}{r_2})F] \frac{\partial L_c}{\partial \gamma_1} \\ &+ (1 + \lambda_{11}\gamma_1) \left\{ [1 + \lambda_{11}\gamma_1 + (\frac{r_1}{r_2})F] \frac{\partial^2 L_c}{\partial \gamma_1^2} + \lambda_{11} \frac{\partial L_c}{\partial \gamma_1} \right\} \\ &= [2 \lambda_{11} (1 + \lambda_{11}\gamma_1) + (\frac{r_1}{r_2}) \lambda_{11} F] \frac{\partial L_c}{\partial \gamma_1} \end{aligned}$$

$$+ (1 + \lambda_{11}\gamma_1) [1 + \lambda_{11}\gamma_1 + (\frac{r_1}{r_2})F] \frac{\partial^2 L_c}{\partial \gamma_1^2}.$$

Expressions for  $\frac{\partial L_c}{\partial \gamma_1}$  and  $\frac{\partial^2 L_c}{\partial \gamma_1^2}$  are given in Section III.B.3.a. Let  $\gamma_1^{(p)}$  represent the p-th iterate obtained by applying Newton's method to the equation  $k^*(\gamma_1) = 0$ . Then,

$$\gamma_1^{(p+1)} = \gamma_1^{(p)} - K^{*-1}(\gamma_1^{(p)}) k^*(\gamma_1^{(p)}). \quad (3.64)$$

### C. Method of Scoring

The Method of Scoring was described by Rao (1965) as a general gradient method that applies when the Hessian matrix of the function being maximized depends on observed values of random variables. The Method of Scoring is the same as the traditional Newton-Raphson method except that the second-order partial derivatives of the function are replaced by their expected values.

Let  $B^*(\underline{\gamma}) = \mathbb{E}[B(\underline{\gamma})]$  represent the  $(c+1) \times (c+1)$  matrix whose  $(i,j)$ -th element is  $\mathbb{E}(\frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j})$  and  $G^*(\underline{\sigma}) = \mathbb{E}[G(\underline{\sigma})]$  represent the  $(c+1) \times (c+1)$  matrix whose  $(i,j)$ -th element is  $\mathbb{E}(\frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2})$ . General expressions for the elements of  $B^*(\underline{\gamma})$  and  $G^*(\underline{\sigma})$  are given in Sections II.B and II.C, respectively. Note that  $-B^*(\underline{\gamma})$  and  $-G^*(\underline{\sigma})$  are the Fisher information matrices associated with  $L_1(\underline{\gamma}; \underline{y})$  and  $L_v(\underline{\sigma}; \underline{y})$ , respectively.

Letting  $\underline{\gamma}^{(p)}$  represent the  $p$ -th iterate of the Method of Scoring, as applied to the maximization of  $L_1$ , we have that

$$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} - B^{*-1}(\underline{\gamma}^{(p)}) \underline{b}(\underline{\gamma}^{(p)}). \quad (3.65)$$

Similarly, letting  $\underline{\sigma}^{(p)}$  represent the  $p$ -th iterate of the Method of Scoring, as applied to the maximization of  $L_v$ , we have

$$\underline{\sigma}^{(p+1)} = \underline{\sigma}^{(p)} - G^{*-1}(\underline{\sigma}^{(p)}) \underline{g}(\underline{\sigma}^{(p)}). \quad (3.66)$$

Harville (1977) showed that

$$-B^*(\underline{\gamma}) = \left(\frac{1}{2}\right) \sigma_{c+1}^4 \llbracket \text{tr}(Z_j' P Z_i Z_i' P Z_j) \rrbracket \quad (i, j = 1, \dots, c+1),$$

$$-G^*(\underline{\sigma}) = \left(\frac{1}{2}\right) \llbracket \text{tr}(Z_j' P Z_i Z_i' P Z_j) \rrbracket \quad (i, j = 1, \dots, c+1).$$

Jennrich and Sampson (1976) note that the matrices  $-B^*(\underline{\gamma})$  and  $-G^*(\underline{\sigma})$  are nonnegative definite for all  $\underline{\gamma} \in \Omega_1^*$  and that, except in certain degenerate cases, they are positive definite. Lemma 3.2 provides additional detail on this property. More, specifically, defining

$$J^* \equiv \llbracket \text{tr}(Z_j' P Z_i Z_i' P Z_j) \rrbracket \quad (i, j = 1, \dots, c+1),$$

we reach the following conclusion.

Lemma 3.2: For  $\gamma \in \Omega_1^*$ , the matrix  $J^*$  is nonnegative definite. Moreover, letting  $P^{\frac{1}{2}}$  represent a symmetric matrix such that  $P = P^{\frac{1}{2}}P^{\frac{1}{2}}$ ,  $J^*$  is a positive definite matrix if, and only if,  $P^{\frac{1}{2}}Z_1Z_1'P^{\frac{1}{2}}$ , ...,  $P^{\frac{1}{2}}Z_cZ_c'P^{\frac{1}{2}}$ ,  $P$  are linearly independent matrices [i.e., if, and only if,

$\sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}}Z_iZ_i'P^{\frac{1}{2}} = \emptyset$  implies that  $\delta_i = 0$  ( $i = 1, \dots, c$ )].

Proof: To prove this result, we adopt a proof of an analogous result given by Miller [(1973), Proposition 5.4.1].

Let  $\underline{\delta} = (\delta_1, \dots, \delta_{c+1})'$  represent an arbitrary  $(c+1) \times 1$  vector.

Observe that

$$\begin{aligned}
 \underline{\delta}' J^* \underline{\delta} &= \sum_{j=1}^{c+1} \left\{ \sum_{i=1}^{c+1} \delta_i \operatorname{tr}[P^{\frac{1}{2}}Z_iZ_i'PZ_jZ_j'P^{\frac{1}{2}}] \right\} \delta_j \\
 &= \sum_{j=1}^{c+1} \left\{ \operatorname{tr} \left[ \sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}}Z_iZ_i'PZ_jZ_j'P^{\frac{1}{2}} \right] \right\} \delta_j \\
 &= \sum_{j=1}^{c+1} \operatorname{tr} \left[ \delta_j P^{\frac{1}{2}}Z_jZ_j'P^{\frac{1}{2}} \left( \sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}}Z_iZ_i'P^{\frac{1}{2}} \right) \right] \\
 &= \operatorname{tr} \left[ \sum_{j=1}^{c+1} \delta_j P^{\frac{1}{2}}Z_jZ_j'P^{\frac{1}{2}} \left( \sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}}Z_iZ_i'P^{\frac{1}{2}} \right) \right] \\
 &= \operatorname{tr} \left[ \left( \sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}}Z_iZ_i'P^{\frac{1}{2}} \right)^2 \right].
 \end{aligned}$$

Thus,  $\delta' J^* \delta$  is nonnegative, and it is equal to zero if, and only if,

$$\sum_{i=1}^{c+1} \delta_i P^{\frac{1}{2}} Z_i Z_i' P^{\frac{1}{2}} = 0. \text{ We conclude that } J^* \text{ is nonnegative definite and}$$

that it is positive definite if, and only if,  $P^{\frac{1}{2}} Z_1 Z_1' P^{\frac{1}{2}}, \dots, P^{\frac{1}{2}} Z_c Z_c' P^{\frac{1}{2}}, P$  are linearly independent matrices.  $\square$

If  $-B^*(\underline{y}^{(p)})$  is positive definite, it follows from the discussion of Section III.B.1 that at least some increase in the function  $L_1$  can be achieved by moving in the direction  $-B^{*-1}(\underline{y}^{(p)}) \underline{b}(\underline{y}^{(p)})$  [provided  $\underline{b}(\underline{y}^{(p)}) \neq 0$ ]. This direction is precisely that taken by the Method of Scoring algorithm (3.65). Similarly, if  $-G^*(\underline{\sigma})$  is positive definite, at least some increase in  $L_v$  can be achieved by moving in the direction taken by algorithm (3.66).

Miller (1973, 1979) showed that, if closed form solutions to  $\frac{\partial L_v}{\partial \underline{\sigma}} = 0$  exist, and if  $-G^*(\underline{\sigma}^{(0)})$  is positive definite, then algorithm (3.66) will converge to a solution in one iteration from any starting value. This is not necessarily the case for algorithm (3.65).

As discussed by Vandaele and Chowdhury (1971), the Method of Scoring algorithms (3.65) and (3.66) may fail to converge, or they may converge to a stationary point that is not a local maximum. To assure convergence of algorithm (3.65) to a local maximum, they suggested that these algorithms be modified by varying the stepsize. In particular, they recommended taking

$$\underline{Y}^{(p+1)} = \underline{Y}^{(p)} - \alpha_p B^{*-1}(\underline{Y}^{(p)}) \underline{b}(\underline{Y}^{(p)}) \quad (p = 0, 1, 2, \dots, ) ,$$

where  $\alpha_p$  ( $\alpha_p > 0$ ) is a scalar chosen so that  $L_1(\underline{Y}^{(p+1)}; \underline{y}) > L_1(\underline{Y}^{(p)}; \underline{y})$ . The choice of  $\alpha_p$  could be based on Rule 1 (Section III.B.1). Vandaele and Chowdhury proposed an approximation to this rule which is less exact, but which is also less costly.

Jennrich and Sampson (1976) and Miller (1979) used unmodified Method of Scoring algorithms to compute maximum likelihood estimates of variance components for various data sets. They occasionally obtained iterates outside the parameter space. In Chapter IV, we consider remedies proposed by them for this problem and discuss various alternative strategies.

#### D. Method of Successive Approximations

Let  $f_i(\underline{z}) = 0$  ( $i = 1, \dots, n$ ) represent a system of  $n$  possibly non-linear equations in an  $n$ -dimensional real-valued vector  $\underline{z} = (z_1, \dots, z_n)'$  of unknowns. Like Newton's method, the Method of Successive Approximations is an iterative method for solving such a system. Unlike Newton's method, its use does not require the evaluation of the partial derivatives  $\frac{\partial f_i(\underline{z})}{\partial z_j}$  ( $i, j = 1, \dots, n$ ). To apply the Method of Successive Approximations, we must first re-express the equations  $[[f_i(\underline{z})]] = 0$  ( $i = 1, \dots, n$ ) in the form  $\underline{z} = [[\bar{g}_i(\underline{z})]]$  ( $i = 1, \dots, n$ ), for some functions  $\bar{g}_1, \dots, \bar{g}_n$ . Then, starting with an initial guess  $\underline{z}^{(0)}$ , the method generates a

sequence of iterates  $\underline{z}^{(1)}, \underline{z}^{(2)}, \dots$  in accordance with the formula

$$\underline{z}^{(p+1)} = [\underline{\bar{g}}_i(\underline{z}^{(p)})] \quad (i = 1, \dots, n ; p = 0, 1, 2, \dots).$$

In this section, we apply the Method of Successive Approximations to the likelihood equations  $\frac{\partial L_1}{\partial \underline{\gamma}} = \underline{0}$  and  $\frac{\partial L_v}{\partial \underline{\sigma}} = \underline{0}$ . We show that both of the EM algorithms  $\{(3.9), (3.13)\}$  and  $\{(3.19), (3.20)\}$  can alternatively be derived by the Method of Successive Approximations. In addition, we use this method to derive three other algorithms.

### 1. Applications

Subsequently, we define

$$h_i(\underline{\sigma}) \equiv \frac{\underline{\tilde{b}}_i' \underline{\tilde{b}}_i}{q_i - \text{tr}(T_{ii})} \quad (i = 1, \dots, c),$$

$$h_{c+1}(\underline{\sigma}) \equiv \frac{(\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}})'(\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}})}{n - p^* - q + \text{tr}(T)},$$

$$h_{c+1}^*(\underline{\sigma}) \equiv \frac{\underline{y}'(\underline{y} - X\underline{\tilde{\alpha}} - Z\underline{\tilde{b}})}{n - p^*}$$

$$\epsilon_i(\underline{\sigma}) \equiv \frac{\underline{\tilde{b}}_i' \underline{\tilde{b}}_i + \sigma_i^2 \text{tr}(T_{ii})}{q_i} \quad (i = 1, \dots, c),$$

$$\epsilon_{c+1}(\underline{\sigma}) \equiv \frac{(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) + \sigma_{c+1}^2[p^* + q - \text{tr}(T)]}{n},$$

$$\epsilon_{c+1}^*(\underline{\sigma}) \equiv \frac{(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) + \sigma_{c+1}^2[q - \text{tr}(T)]}{n-p^*}.$$

Take the parameter space for  $\underline{\sigma}$  to be  $\Omega_2' = \{\underline{\sigma} : \sigma_{c+1}^2 > 0, \sigma_i^2 > 0 \text{ (} i = 1, \dots, c)\}$ , and consider the likelihood equations

$$\frac{\partial L_v}{\partial \sigma_i^2} = 0 \quad (i = 1, \dots, c),$$

$$\frac{\partial L_v}{\partial \sigma_{c+1}^2} = 0.$$

Using expressions (2.33) and (2.34), these equations can be rewritten as

$$- \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{ii})] + \left(\frac{1}{2}\right) \tilde{v}_{-i}' \tilde{v}_{-i} = 0 \quad (i = 1, \dots, c),$$

$$- \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(T)]$$

$$+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) = 0$$

or, alternatively, as

$$\left. \begin{aligned} - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{ii})] + \left(\frac{1}{2}\right) \frac{1}{\sigma_i^4} \tilde{b}'_{-i-i} \tilde{b}_{-i-i} &= 0 \quad (i = 1, \dots, c), \\ - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n-p^*-q+\text{tr}(T)] + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y}-X\underline{\tilde{\alpha}}-Z\underline{\tilde{b}})'(\underline{y}-X\underline{\tilde{\alpha}}-Z\underline{\tilde{b}}) &= 0. \end{aligned} \right\} \quad (3.67)$$

To apply the Method of Successive Approximations to equations (3.67), we need to re-express them in the form  $\sigma_i^2 = \bar{g}_i(\underline{\sigma})$  ( $i = 1, \dots, c+1$ ).

There are several ways in which this can be done. Keeping in mind that  $\sigma_i^2 > 0$  ( $i = 1, \dots, c+1$ ), we see that the equations  $\frac{\partial L_v}{\partial \sigma_i^2} = 0$  ( $i = 1, \dots, c$ )

can be re-expressed in the form

$$\sigma_i^2 = h_i(\underline{\sigma}) \quad (i = 1, \dots, c), \quad (3.68)$$

or the form

$$\sigma_i^2 = \epsilon_i(\underline{\sigma}) \quad (i = 1, \dots, c) . \quad (3.69)$$

Similarly, the equation  $\frac{\partial L_v}{\partial \sigma_{c+1}^2} = 0$  can be re-expressed in the form

$$\sigma_{c+1}^2 = h_{c+1}(\underline{\sigma}) , \quad (3.70)$$

or the form

$$\sigma_{c+1}^2 = \epsilon_{c+1}^*(\underline{\sigma}) , \quad (3.71)$$

or the form

$$\sigma_{c+1}^2 = \epsilon_{c+1}(\underline{\sigma}) . \quad (3.72)$$

If we use (3.69) and (3.72) in applying the Method of Successive Approximations, we obtain the EM algorithm {(3.9), (3.13)}. Alternatively, if we use (3.69) and (3.71), we obtain the EM algorithm {(3.19), (3.20)}.

A third possibility is to use (3.68) and (3.70) in applying the Method of Successive Approximations, in which case, starting with the initial guess  $\underline{\sigma}^{(0)}$ , we obtain the sequence of iterates  $\underline{\sigma}^{(1)}, \underline{\sigma}^{(2)}, \dots$ , where

$$\underline{\sigma}^{(p+1)} = [[h_i(\underline{\sigma}^{(p)})]] \quad (i = 1, \dots, c+1) . \quad (3.73)$$

Adopting an approach analagous to that taken by Anderson (1973) in the maximum likelihood estimation of  $\underline{\sigma}$ , we see that the likelihood equations (3.67) can also be rewritten in the form

$$\underline{\sigma} = - G^{*-1}(\underline{\sigma}) \underline{d}(\underline{\sigma}) , \quad (3.74)$$

where

$$\underline{d}(\underline{\sigma}) = \left(\frac{1}{2}\right) \begin{bmatrix} \left[ \frac{1}{\sigma_i^4} \tilde{\mathbf{b}}_i' \tilde{\mathbf{b}}_i \right] \quad (i = 1, \dots, c) \\ \frac{1}{\sigma_{c+1}^4} \underline{\mathbf{y}}' (\underline{\mathbf{y}} - \mathbf{X}\tilde{\underline{\alpha}} - \mathbf{Z}\tilde{\underline{\mathbf{b}}}) - \frac{1}{\sigma_{c+1}^2} \sum_{i=1}^c \frac{1}{\sigma_i^2} \tilde{\mathbf{b}}_i' \tilde{\mathbf{b}}_i \end{bmatrix}$$

[provided, of course, that the matrix  $-\mathbf{G}^*(\underline{\sigma})$  is positive definite for  $\underline{\sigma} \in \Omega_2'$ ]. If we use (3.74) in applying the Method of Successive Approximations, then we obtain the Method of Scoring algorithm (3.66), as noted previously by J. N. K. Rao [cited in Miller (1979)], Hocking and Kutner (1975), Harville (1977), and Searle (1979).

Still other implementations of the Method of Successive Approximations can be obtained by re-expressing the equations  $\frac{\partial L_1}{\partial \gamma_i} = 0$  ( $i = 1, \dots, c+1$ ). Take the parameter space for  $\underline{\gamma}$  to be  $\Omega_1' = \{\underline{\gamma} : \gamma_{c+1} > 0, \gamma_i > 0 \text{ (} i = 1, \dots, c \text{)}\}$ . Then, using parts (i) and (ii) of Lemma 2.7, the equations  $\frac{\partial L_1}{\partial \gamma_i} = 0$  ( $i = 1, \dots, c+1$ ) can be expressed in the form

$$- \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(\mathbf{T}_{ii})] - \sigma_{c+1}^2 \tilde{\mathbf{y}}_i' \tilde{\mathbf{y}}_i \right\} = 0 \quad (i = 1, \dots, c),$$

$$- \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{\mathbf{y}}' \mathbf{S}(\underline{\mathbf{y}} - \mathbf{Z}\tilde{\underline{\mathbf{b}}})] = 0$$

or, recalling results (2.13) and (2.14) [which imply that  $\underline{\mathbf{y}}' \mathbf{S}(\underline{\mathbf{y}} - \mathbf{Z}\tilde{\underline{\mathbf{b}}}) = \frac{1}{\sigma_{c+1}^2} \underline{\mathbf{y}}' (\underline{\mathbf{y}} - \mathbf{X}\tilde{\underline{\alpha}} - \mathbf{Z}\tilde{\underline{\mathbf{b}}})$ ], in the form

$$\left. \begin{aligned} & - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii})] - \sigma_{c+1}^2 \left(\frac{1}{\sigma_i^2}\right) \tilde{b}_i' \tilde{b}_i \right\} \quad (i = 1, \dots, c) \\ & - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \frac{1}{\sigma_{c+1}^2} \underline{y}'(\underline{y} - X\tilde{\alpha} - Z\tilde{b})] = 0 . \end{aligned} \right\} \quad (3.75)$$

Equations (3.75) can be regarded as equations in  $\underline{\sigma}$ , instead of  $\underline{\gamma}$ . Moreover, the equations  $\frac{\partial L_1}{\partial \gamma_i} = 0$  ( $i = 1, \dots, c$ ) can be re-expressed in the form

$$\sigma_i^2 = h_i(\underline{\sigma}) \quad (i = 1, \dots, c) , \quad (3.76)$$

or the form

$$\sigma_i^2 = \epsilon_i(\underline{\sigma}) \quad (i = 1, \dots, c) . \quad (3.77)$$

Similarly, the equation  $\frac{\partial L_1}{\partial \sigma_{c+1}^2} = 0$  can be re-expressed in the form

$$\sigma_{c+1}^2 = h_{c+1}^*(\underline{\sigma}) . \quad (3.78)$$

Note that  $\left(\frac{1}{n-p^*}\right) \underline{y}'(\underline{y} - X\tilde{\alpha} - Z\tilde{b}) = \hat{\sigma}_{c+1}^2(\underline{\gamma}^+)$  [as is evident from result (2.14)].

If we use (3.76) and (3.78) in applying the Method of Successive Approximations, then, starting with the initial guess  $\underline{\sigma}^{(0)}$ , we obtain the sequence of iterates  $\underline{\sigma}^{(1)}$ ,  $\underline{\sigma}^{(2)}$ , ..., where

$$\underline{\sigma}^{(p+1)} = \begin{bmatrix} h_1(\underline{\sigma}^{(p)}) \\ \vdots \\ h_c(\underline{\sigma}^{(p)}) \\ h_{c+1}^*(\underline{\sigma}^{(p)}) \end{bmatrix}. \quad (3.79)$$

Alternatively, if we use (3.77) and (3.78), then the iterates are those generated by the recursive formula

$$\underline{\sigma}^{(p+1)} = \begin{bmatrix} \epsilon_1(\underline{\sigma}^{(p)}) \\ \vdots \\ \epsilon_c(\underline{\sigma}^{(p)}) \\ h_{c+1}^*(\underline{\sigma}^{(p)}) \end{bmatrix}. \quad (3.80)$$

Algorithm (3.80) can be regarded as the REML analog of Henderson's (1973) algorithm for determining maximum likelihood estimates of  $\underline{\sigma}$ , as discussed by Harville (1977, Section 6).

## 2. Nonnegativity constraints

Suppose that  $\text{rank}(X, Z_i) > p^*$  ( $i = 1, \dots, c$ ). Harville (1975, Lemma 1) showed that, if  $\sigma_i^2 > 0$ , then  $q_i - \text{tr}(T_{ii}) > 0$ . A similar argument reveals that, if  $\sigma_i^2 < 0$ , then  $q_i - \text{tr}(T_{ii}) < 0$ . Moreover, if  $\sigma_i^2 = 0$ , then  $T_{ii} = I_{q_i}$ , implying that  $q_i - \text{tr}(T_{ii}) = 0$ . Thus, if the  $i$ -th component of the  $p$ -th iterate of algorithm (3.73) or (3.79) is greater than (less than) zero, then the  $i$ -th component of the  $(p+1)$ -st iterate is greater than or equal to (less than or equal to) zero ( $p = 0, 1, 2, \dots$ ;  $i = 1, \dots, c$ ). If the  $i$ -th component of the  $p$ -th iterate of algorithm (3.73) or (3.79) equals zero, then the  $i$ -th component of the  $(p+1)$ -st iterate is undefined and could, following Harville (1977), arbitrarily be assigned the value zero ( $p = 0, 1, 2, \dots$ ;  $i = 1, \dots, c$ ). The  $(c+1)$ -st component of each iterate of algorithm (3.73) or (3.79) exceeds zero [unless  $\underline{y} \in \mathcal{C}(X)$ , in which case it equals zero].

It follows from the discussion of Section III.A.4 that, if the  $i$ -th component of the  $p$ -th iterate of algorithm (3.80) is greater than (equal to) zero, then the  $i$ -th component of the  $(p+1)$ -st iterate is greater than (equal to) zero ( $p = 0, 1, 2, \dots$ ;  $i = 1, \dots, c$ ). As in the case of algorithms (3.73) and (3.79), the  $(c+1)$ -st component of each iterate of algorithm (3.80) exceeds zero [unless  $\underline{y} \in \mathcal{C}(X)$ , in which case it equals zero].

## E. Summary Table

A total of fourteen algorithms have been presented in this chapter. They are listed, for easy reference, in Table 3.1.

Table 3.1. Fourteen iterative algorithms for computing REML estimates.

Algorithm Number	Equation Number	Description <sup>a</sup>	(p+1)-st iterate (p=0,1,2,...)
1	(3.27)	TNR, $L_1$	$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} - \underline{B}^{-1}(\underline{\gamma}^{(p)}) \underline{b}(\underline{\gamma}^{(p)})$
2	(3.29)	TNR, $L_v$	$\underline{\sigma}^{(p+1)} = \underline{\sigma}^{(p)} - \underline{G}^{-1}(\underline{\sigma}^{(p)}) \underline{g}(\underline{\sigma}^{(p)})$
3	(3.52)	TNR, $L_c$	$\underline{\gamma}^{+(p+1)} = \underline{\gamma}^{+(p)} - \underline{K}^{-1}(\underline{\gamma}^{+(p)}) \underline{k}(\underline{\gamma}^{+(p)})$ $\sigma_{c+1}^2(p+1) = \hat{\sigma}_{c+1}^2(\underline{\gamma}^{+(p+1)})$
4	(3.54)	TNR, $L_R$	$\underline{\psi}^{+(p+1)} = \underline{\psi}^{+(p)} - \underline{M}^{-1}(\underline{\psi}^{+(p)}) \underline{m}(\underline{\psi}^{+(p)})$ $\sigma_{c+1}^2(p+1) = \hat{\sigma}_{c+1}^2(W'_{11} \underline{\psi}^{+(p+1)})$
5	(3.64)	LNR, $L_c (c=1)$	$\gamma_1^{(p+1)} = \gamma_1^{(p)} - \underline{K}^{*-1}(\gamma_1^{(p)}) k^*(\gamma_1^{(p)})$ $\sigma_2^2(p+1) = \hat{\sigma}_2^2(\gamma_1^{(p+1)})$
6	(3.62)	LNR, $L_R$	$\underline{\psi}^{+(p+1)} = \underline{\psi}^{+(p)} - \underline{M}^{*-1}(\underline{\psi}^{+(p)}) \underline{m}^*(\underline{\psi}^{+(p)})$ $\sigma_{c+1}^2(p+1) = \hat{\sigma}_{c+1}^2(W'_{11} \underline{\psi}^{+(p+1)})$
7	(3.33)	LNR, $L_1^*$	$\underline{m}^{(p+1)} = \underline{m}^{(p)} - \underline{Q}^{*-1}(\underline{m}^{(p)}) \underline{q}^*(\underline{m}^{(p)})$
8	(3.65)	MOS, $L_1$	$\underline{\gamma}^{(p+1)} = \underline{\gamma}^{(p)} - \underline{B}^{*-1}(\underline{\gamma}^{(p)}) \underline{b}(\underline{\gamma}^{(p)})$
9	(3.66)	MOS, $L_v$	$\underline{\sigma}^{(p+1)} = \underline{\sigma}^{(p)} - \underline{G}^{*-1}(\underline{\sigma}^{(p)}) \underline{g}(\underline{\sigma}^{(p)})$
10	{(3.9), (3.13)}	EM, ( $\underline{b}, \underline{e}$ )	$\underline{\sigma}^{(p+1)} = [\epsilon_1(\underline{\sigma}^{(p)}), \dots, \epsilon_{c+1}(\underline{\sigma}^{(p)})]'$
11	{(3.19), (3.20)}	EM, ( $\underline{b}, A' \underline{e}$ )	$\underline{\sigma}^{(p+1)} = [\epsilon_1(\underline{\sigma}^{(p)}), \dots, \epsilon_c(\underline{\sigma}^{(p)}), \epsilon_{c+1}^*(\underline{\sigma}^{(p)})]'$
12	(3.73)	MOSA, $L_v$	$\underline{\sigma}^{(p+1)} = [h_1(\underline{\sigma}^{(p)}), \dots, h_{c+1}(\underline{\sigma}^{(p)})]'$
13	(3.80)	MOSA, $L_1$	$\underline{\sigma}^{(p+1)} = [\epsilon_1(\underline{\sigma}^{(p)}), \dots, \epsilon_c(\underline{\sigma}^{(p)}), h_{c+1}^*(\underline{\sigma}^{(p)})]'$
14	(3.79)	MOSA, $L_1$	$\underline{\sigma}^{(p+1)} = [h_1(\underline{\sigma}^{(p)}), \dots, h_c(\underline{\sigma}^{(p)}), h_{c+1}^*(\underline{\sigma}^{(p)})]'$

<sup>a</sup>TNR, L denotes the traditional Newton-Raphson method applied to the maximization of L ( $L = L_1, L_v, L_c, L_R$ );

LNR,L denotes the linearized Newton-Raphson method applied to the maximization of L [ $L = L_c$  (with  $c=1$ ),  $L_R$ ,  $L_1^*$ ];

MOS,L denotes the Method of Scoring applied to the maximization of L ( $L = L_1$ ,  $L_v$ );

EM,( $\underline{b}, \underline{e}$ ) denotes the EM algorithm when ( $\underline{b}, \underline{e}$ ) is viewed as the complete data;

EM,( $\underline{b}, A'\underline{e}$ ) denotes the EM algorithm when ( $\underline{b}, A'\underline{e}$ ) is viewed as the complete data;

MOSA,L denotes a version of the Method of Successive Approximations when applied to the likelihood equations associated with L ( $L = L_1$ ,  $L_v$ ).

#### F. Special Case: $c=1$

In this section, we present simplified results for the special case  $c=1$ . The derivations are relegated to Section B of the Appendix.

Define  $\bar{\Delta} = \frac{1}{r} \sum_{i=1}^r \Delta_i$  and let

$$\underline{\Delta}^{-\frac{1}{2}} \equiv \text{diag}\{\Delta_1^{-\frac{1}{2}}, \dots, \Delta_r^{-\frac{1}{2}}\},$$

$$\underline{q} \equiv Z'(I - P_X)\underline{y},$$

$$\tilde{t} \equiv (\tilde{t}_1, \dots, \tilde{t}_r)' \equiv \underline{\Delta}^{-\frac{1}{2}} R^{*'} \underline{q},$$

$$S_1 \equiv \underline{y}'(P_1 - P_0)\underline{y},$$

$$S_2 \equiv \underline{y}'(I - P_1)\underline{y},$$

$$M_1 \equiv \sum_{i=1}^r \frac{\Delta_i}{1 + \gamma_1 \Delta_i} ,$$

$$M_2 \equiv \sum_{i=1}^r \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} ,$$

$$M_3 \equiv \sum_{i=1}^r \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i} ,$$

$$M_4 \equiv \sum_{i=1}^r \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} ,$$

$$M_5 \equiv \sum_{i=1}^r \frac{\Delta_i^2 \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^3} .$$

According to Harville and Fenech (1985),

$$\bar{\Delta} = \frac{1}{r} \text{tr}(C) ,$$

$$M_1 = \text{tr}[(I + \gamma_1 C)^{-1} C],$$

$$M_2 = \text{tr}\{[(I + \gamma_1 C)^{-1} C]^2\}$$

$$M_3 = \underline{u}' \tilde{\underline{s}} ,$$

$$M_4 = \underline{u}' \underline{u} ,$$

$$M_5 = \underline{u}' [(I + \gamma_1 C)^{-1} C] \underline{u} ,$$

where  $r = \text{rank}(C)$ ,  $\underline{\tilde{s}}$  = any solution to  $C\underline{\tilde{s}} = \underline{q}$  , and  $\underline{u} = (I + \gamma_1 C)^{-1} \underline{q}$ .

Moreover, as shown in Section B of the Appendix,

$$\hat{\sigma}_2^2(\gamma_1) = (S_2 + M_3)/(n-p^*) . \quad (3.81)$$

The following simplified expressions are relevant to the Newton-Raphson method, as applied to  $L_1$ :

$$\frac{\partial L_1}{\partial \gamma_1} = - \left(\frac{1}{2}\right) (M_1 - \frac{1}{\sigma_2^2} M_4) , \quad (3.82)$$

$$\frac{\partial L_1}{\partial \sigma_2^2} = - \left(\frac{1}{2}\right) (n-p^*) \frac{1}{\sigma_2^4} [\sigma_2^2 - \hat{\sigma}_2^2(\gamma_1)] , \quad (3.83)$$

$$\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} = \left(\frac{1}{2}\right) [M_2 - 2\left(\frac{1}{\sigma_2^2}\right) M_5] , \quad (3.84)$$

$$\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} M_4 , \quad (3.85)$$

$$\frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2} = \left(\frac{1}{2}\right) (n-p^*) \frac{1}{\sigma_2^6} [\sigma_2^2 - 2 \hat{\sigma}_2^2(\gamma_1)] . \quad (3.86)$$

The following simplified expressions are [together with expressions (3.82) and (3.83) for  $\frac{\partial L_1}{\partial \gamma_1}$  and  $\frac{\partial L_1}{\partial \sigma_2^2}$ , respectively] relevant to the Method of Scoring, as applied to  $L_1$ :

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) = - \left(\frac{1}{2}\right) M_2 , \quad (3.87)$$

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right) = - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^2} M_1 , \quad (3.88)$$

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \sigma^2 \partial \sigma_2^2}\right) = - \left(\frac{1}{2}\right) (n-p^*) \frac{1}{\sigma_2^4} . \quad (3.89)$$

The following simplified expressions are relevant to the Newton-Raphson method, as applied to  $L_c$ :

$$\frac{\partial L_c}{\partial \gamma_1} = - \left(\frac{1}{2}\right) \left[ M_1 - \frac{1}{\hat{\sigma}_2^2(\gamma_1)} M_4 \right] , \quad (3.90)$$

$$\begin{aligned} \frac{\partial^2 L_c}{\partial \gamma \partial \gamma} &= \left(\frac{1}{2}\right) M_2 - \frac{1}{\hat{\sigma}_2^2(\gamma_1)} M_5 \\ &+ \left(\frac{1}{2}\right) (n-p^*) \left[ \frac{1}{\hat{\sigma}_2^2(\gamma_1)} \right]^2 (M_4)^2 . \end{aligned} \quad (3.91)$$

The following simplified expressions are [together with expressions (3.90) and (3.91) for  $\frac{\partial L_c}{\partial \gamma_1}$  and  $\frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1}$ , respectively] relevant to the

linearized version of the Newton-Raphson method, as applied to  $L_c$ :

$$k^*(\gamma_1) = (1 + \gamma_1 \bar{\Delta})(1 + \gamma_1 \bar{\Delta} + \frac{s_1}{s_2}) \frac{\partial L_c}{\partial \gamma_1}, \quad (3.92)$$

$$\begin{aligned} K^*(\gamma_1) = & [2 \bar{\Delta} (1 + \gamma_1 \bar{\Delta}) + \bar{\Delta} \frac{s_1}{s_2}] \frac{\partial L_c}{\partial \gamma_1} \\ & + (1 + \gamma_1 \bar{\Delta})(1 + \gamma_1 \bar{\Delta} + \frac{s_1}{s_2}) \frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1}. \end{aligned} \quad (3.93)$$

The following simplified expressions are [together with expressions (3.90) and (3.91) for  $\frac{\partial L_c}{\partial \gamma_1}$  and  $\frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1}$ , respectively] relevant to the Newton-Raphson method as applied to  $L_R$ :

$$\frac{\partial L_R}{\partial \psi_1} = - \left( \frac{1}{\bar{\Delta}} \right) \frac{1}{\psi_1^2} \frac{\partial L_c}{\partial \gamma_1}, \quad (3.94)$$

$$\begin{aligned} \frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1} = & \left( \frac{1}{\bar{\Delta}} \right) \frac{1}{\psi_1^4} \frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1} \\ & + 2 \left( \frac{1}{\bar{\Delta}} \right) \frac{1}{\psi_1^3} \frac{\partial L_c}{\partial \gamma_1}. \end{aligned} \quad (3.95)$$

The following simplified expressions are [together with expressions (3.82) and (3.84) for  $\frac{\partial L_1}{\partial \gamma_1}$  and  $\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}$ , respectively] relevant to the

Newton-Raphson method, as applied to  $L_v$ :

$$\frac{\partial L_v}{\partial \sigma_1^2} = \frac{1}{\sigma_2^2} \frac{\partial L_1}{\partial \gamma_1}, \quad (3.96)$$

$$\begin{aligned} \frac{\partial L_v}{\partial \sigma_2^2} = & - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} \{ \sigma_2^2 (n - p^* - \gamma_1 M_1) \\ & - [(n-p^*) \hat{\sigma}_2^2(\gamma_1) - \gamma_1 M_4] \}, \end{aligned} \quad (3.97)$$

$$\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} = \frac{1}{\sigma_2^4} \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}, \quad (3.98)$$

$$\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} = - \frac{1}{\sigma_2^4} \left[ \gamma_1 \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} - \left(\frac{1}{2}\right) M_1 + \frac{1}{\sigma_2^2} M_4 \right], \quad (3.99)$$

$$\begin{aligned} \frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2} = & \left(\frac{1}{2}\right) \frac{1}{\sigma_2^6} \{ \sigma_2^2 (n - p^* - 2 \gamma_1 M_1 + \gamma_1^2 M_2) \\ & - 2 [(n-p^*) \hat{\sigma}_2^2(\gamma_1) - 2 \gamma_1 M_4 + \gamma_1^2 M_5] \}. \end{aligned} \quad (3.100)$$

The following simplified expressions are [together with expressions (3.96) and (3.97) for  $\frac{\partial L_v}{\partial \sigma_1^2}$  and  $\frac{\partial L_v}{\partial \sigma_2^2}$ , respectively] relevant to the Method of Scoring, as applied to  $L_v$ :

$$\mathbb{E}\left(\frac{\partial^2 L_V}{\partial \sigma_1^2 \partial \sigma_1^2}\right) = \frac{1}{\sigma_2^4} \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) , \quad (3.101)$$

$$\mathbb{E}\left(\frac{\partial^2 L_V}{\partial \sigma_1^2 \partial \sigma_2^2}\right) = \frac{1}{\sigma_2^2} \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right) - \left(\frac{1}{\sigma_2^4}\right) \gamma_1 \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) , \quad (3.102)$$

$$\begin{aligned} \mathbb{E}\left(\frac{\partial^2 L_V}{\partial \sigma_2^2 \partial \sigma_2^2}\right) &= \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2}\right) - 2 \left(\frac{1}{\sigma_2^2}\right) \gamma_1 \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right) \\ &\quad + \left(\frac{1}{\sigma_2^4}\right) \gamma_1^2 \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) . \end{aligned} \quad (3.103)$$

The following simplified expressions are [together with expressions (3.96), (3.98), and (3.84) for  $\frac{\partial L_V}{\partial \sigma_1^2}$ ,  $\frac{\partial^2 L_V}{\partial \sigma_1^2 \partial \sigma_1^2}$ , and  $\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}$ , respectively] relevant to the linearized version of the Newton-Raphson method as applied to  $L_1^*$ :

$$q_1^*(\underline{m}) = m_1^2 \left(\frac{1}{\Delta}\right) \frac{\partial L_V}{\partial \sigma_1^2} , \quad (3.104)$$

$$\begin{aligned} q_2^*(\underline{m}) &= -\frac{1}{2} \{ \sigma_2^2 [n - p^* - \left(\frac{1}{\Delta}\right) (1 + \gamma_1 \bar{\Delta}) M_1] \\ &\quad - [(n-p^*) \hat{\sigma}_2^2 (\gamma_1) - \left(\frac{1}{\Delta}\right) (1 + \gamma_1 \bar{\Delta}) M_4] \} , \end{aligned} \quad (3.105)$$

$$\frac{\partial q_1^*}{\partial m_1} = m_1 \left(\frac{1}{\Delta}\right) \left[ m_1 \left(\frac{1}{\Delta}\right) \frac{\partial^2 L_V}{\partial \sigma_1^2 \partial \sigma_1^2} + 2 \frac{\partial L_V}{\partial \sigma_1^2} \right] , \quad (3.106)$$

$$\frac{\partial q_1^*}{\partial m_2} = - \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta})^2 \left[ - \left( \frac{1}{2} \right) M_1 + \frac{1}{\sigma_2^2} M_4 + \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \right], \quad (3.107)$$

$$\frac{\partial q_2^*}{\partial m_1} = \frac{1}{(1 + \gamma_1 \bar{\Delta})^2} \frac{\partial q_1^*}{\partial m_2}, \quad (3.108)$$

$$\begin{aligned} \frac{\partial q_2^*}{\partial m_2} = & - \left( \frac{1}{2} \right) \frac{1}{\sigma_2^2} \{ \sigma_2^2 [n - p^* - \left( \frac{1}{\bar{\Delta}} \right)^2 (1 + \gamma_1 \bar{\Delta})^2 M_2] \\ & - 2 \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) [M_4 - \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) M_5] \}. \end{aligned} \quad (3.109)$$

The following simplified expressions are [together with expressions (3.94) and (3.95) for  $\frac{\partial L_R}{\partial \psi_1}$  and  $\frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1}$ , respectively] relevant to the linearized version of the Newton-Raphson method as applied to  $L_R$ :

$$m_1^*(\psi_1) = \psi_1 \left( 1 + \psi_1 \frac{s_1}{s_2} \right) \frac{\partial L_R}{\partial \psi_1}, \quad (3.110)$$

$$\frac{\partial m_1^*}{\partial \psi_1} = \psi_1 \left( 1 + \psi_1 \frac{s_1}{s_2} \right) \frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1} + (1 + 2 \psi_1 \frac{s_1}{s_2}) \frac{\partial L_R}{\partial \psi_1}. \quad (3.111)$$

The following simplified expressions are relevant to the EM algorithms {(3.9), (3.13)} and {(3.19), (3.20)}, and to the Method of Successive Approximations algorithms (3.73), (3.79), and (3.80):

$$h_1(\underline{\sigma}) = \gamma_1 \frac{M_4}{M_1} , \quad (3.112)$$

$$h_2(\underline{\sigma}) = \frac{S_2 + M_3 - \gamma_1 M_4}{n - p^* - \gamma_1 M_1} , \quad (3.113)$$

$$h_2^*(\underline{\sigma}) = \frac{1}{n-p^*} (S_2 + M_3) , \quad (3.114)$$

$$\epsilon_1(\underline{\sigma}) = \frac{1}{q_1} [\gamma_1^2 M_4 + \sigma_1^2 (q_1 - \gamma_1 M_1)] , \quad (3.115)$$

$$\epsilon_2^*(\underline{\sigma}) = \frac{1}{n-p^*} [S_2 + M_3 - \gamma_1 M_4 + \sigma_1^2 M_1] , \quad (3.116)$$

$$\epsilon_2(\underline{\sigma}) = \frac{1}{n} [S_2 + M_3 - \gamma_1 M_4 + \sigma_2^2 (p^* + \gamma_1 M_1)] . \quad (3.117)$$

## IV. MODIFICATIONS TO ACCOMMODATE PARAMETRIC CONSTRAINTS

Some of the iterative algorithms presented, in Chapter III, for computing a REML estimate of  $\underline{\gamma}$  can produce iterates outside the parameter space. This feature is undesirable for at least two reasons. First, since by definition a REML estimate belongs to the parameter space, we want the final iterate to be in the parameter space. Second, if the  $p$ -th iterate  $\underline{\gamma}^{(p)}$  lies outside the parameter space, then the formula for  $\underline{\gamma}^{(p+1)}$  may be ill-conditioned or even undefined. For example, if  $\underline{\gamma}^{(p)} \notin \Omega_1$ , then the value of  $V$  at  $\underline{\gamma} = \underline{\gamma}^{(p)}$  may be singular or nearly singular. Miller (1979) encountered this problem and found that it rendered his Method of Scoring algorithm "unstable."

In Chapter III, we noted that if  $\underline{y} \notin \mathcal{C}(X)$  and the initial guess  $\underline{\gamma}^{(0)}$  belongs to  $\Omega_2$ , then the iterates produced by the EM algorithms 10-11 (in Table 3.1) belong to  $\Omega_2$ , as do the iterates produced by the Method of Successive Approximation algorithms 12-14 (unless  $\tilde{b}_i^{(p)} = 0$  for some  $i \in \{1, \dots, c\}$  and  $p \in \{0, 1, 2, \dots\}$ ). Unfortunately, the Newton-Raphson algorithms (algorithms 1-7 in Table 3.1) and the Method of Scoring algorithms (algorithms 8 and 9 in Table 3.1) do not have this property. Unless modified, these and other gradient methods can generate iterates outside the parameter space. In this chapter, we discuss some strategies for modifying algorithms 1-9 so that their iterates do not violate the parametric constraints. Several of these strategies have been considered previously in the statistical literature.

Hemmerle and Hartley (1973) applied the Newton-Raphson method to the problem of computing maximum likelihood estimates of variance

components under model (1.2). In doing so, they parameterized in terms of  $\sigma_{c+1}^2$  and the positive square roots  $\tau_1, \dots, \tau_c$  of  $\gamma_1, \dots, \gamma_c$ . They set  $\tau_i^{(p+1)}$  equal to zero if  $\tau_i^{(p)}$  were negative and if, in addition, a certain approximation to  $\tau_i$  was sufficiently small in magnitude. If  $\tau_i^{(p+1)}$  were set to zero, then subsequent iterates  $\tau_i^{(p+2)}, \tau_i^{(p+3)}, \dots$  were also set to zero. With this strategy, the algorithm may converge to a value of  $\underline{\gamma}$  that is not a local maximum, as acknowledged by Hartley and Rao themselves, and as illustrated by Bard [(1974), Section 6.3].

In using a Newton-Raphson algorithm to compute maximum likelihood estimates of  $\underline{\gamma}$  under model (1.2), Jennrich and Sampson (1976) employed a "partial stepping" strategy. This strategy consists of replacing the constant stepsize rule of the traditional Newton-Raphson method, with a rule that permits steps to, but not beyond, the boundary. Jennrich and Sampson (1976), and Vandaele and Chowdhury (1971) also used partial stepping in conjunction with Method of Scoring algorithms. Partial stepping is discussed further in Section IV.A.

In this chapter, we discuss, in the context of computing a REML estimate of  $\underline{\gamma}$ , various strategies for constrained optimization. As is customary, we single out the case of linear (inequality) constraints for special attention. The constraints on  $\underline{\gamma}$  are linear in the special case  $c=1$  but, in general, are possibly nonlinear.

In the special case  $c=1$ , we have considered the following three parameter spaces for  $\underline{\gamma}$ :

$$\Omega_1 = \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, \gamma_1 \geq 0\},$$

$$\Omega_1^* = \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, \gamma_1 > -\frac{1}{\lambda^*}\},$$

$$\Omega_3^* = \{(\gamma_1, \sigma_2^2)' : \sigma_2^2 > 0, \gamma_1 > -\frac{1}{\Delta^*}\}.$$

Note that  $\Omega_3^* \supset \Omega_1^* \supset \Omega_1$ . In all three spaces, the constraints on  $\underline{\gamma}$  are simple linear inequality constraints.

For  $c > 1$ , the constraints determined by  $\Omega_1$  are linear, but those determined by the extended parameter space  $\Omega_1^*$  are not. The set of  $\underline{\gamma}$  values belonging to  $\Omega_1^*$  are such that, in part, the matrix  $I_n + \sum_{i=1}^c \gamma_i Z_i Z_i'$  is positive definite. Let  $h_{ij}(\underline{\gamma}^+)$  represent the  $(i,j)$ -th element of the matrix  $H = I_n + \sum_{i=1}^c \gamma_i Z_i Z_i'$  [regarded as a function of the vector  $\underline{\gamma}^+ = (\gamma_1, \dots, \gamma_c)'$ ]. A necessary and sufficient condition for  $H$  to be a positive definite matrix is that

$$h_{11}(\underline{\gamma}^+) > 0, \quad \begin{vmatrix} h_{11}(\underline{\gamma}^+) & h_{12}(\underline{\gamma}^+) \\ h_{21}(\underline{\gamma}^+) & h_{22}(\underline{\gamma}^+) \end{vmatrix} > 0, \dots, |H| > 0 \quad (4.1)$$

[e.g., Graybill (1983), Theorem 12.2.2]. Aside from  $h_{11}(\underline{\gamma}^+) > 0$ , the remaining  $n-1$  constraints in (4.1) are, in general, nonlinear in  $\underline{\gamma}$ .

Following what is common practice in the numerical optimization literature, we consider the problem of minimizing  $-L_1(\underline{\gamma}; \underline{y})$  rather than the equivalent problem of maximizing  $L_1(\underline{\gamma}; \underline{y})$ . Also, in considering inequality constraints, we suppose that any strict inequality constraint  $a(\underline{\gamma}) > c_1$  [where  $a(\underline{\gamma})$  represents some function of  $\underline{\gamma}$  and  $c_1$  a fixed known constant] has been replaced by the approximation  $a(\underline{\gamma}) \geq c_1 + \epsilon$ , where  $\epsilon$  is some small positive constant.

Each of algorithms 1-9 in Table 4.1 can be regarded as an algorithm for maximizing a function  $L(\underline{\theta}; \underline{y})$  of an  $a \times 1$  vector  $\underline{\theta} = (\theta_1, \dots, \theta_a)'$  [or for minimizing  $-L(\underline{\theta}; \underline{y})$ ]. Table 4.1 lists the vector  $\underline{\theta}$ , the number of variables  $a$ , and the function  $L(\underline{\theta}; \underline{y})$  for each algorithm. Also, each algorithm is of the general form

$$\underline{\theta}^{(p+1)} = \underline{\theta}^{(p)} - N(\underline{\theta}^{(p)}) \underline{n}(\underline{\theta}^{(p)}),$$

where  $N(\underline{\theta})$  and  $\underline{n}(\underline{\theta})$  are an  $a \times a$  matrix and an  $a \times 1$  vector, respectively, whose elements are functions of  $\underline{\theta}$ , and  $\underline{\theta}^{(p)}$  represents the guess for  $\underline{\theta}$  generated on the  $p$ -th iteration. Table 4.1 gives  $N(\underline{\theta})$  and  $\underline{n}(\underline{\theta})$  for each algorithm.

Table 4.1. Key for Chapter IV.

Algorithm	$\underline{\theta}_{a \times 1}$	$a$	$L(\underline{\theta}; \underline{y})$	$N(\underline{\theta})$	$\underline{n}(\underline{\theta})$
1	$\underline{\gamma}$	$c+1$	$L_1(\underline{\gamma}; \underline{y})$	$B(\underline{\gamma})$	$\underline{b}(\underline{\gamma})$
2	$\underline{\sigma}$	$c+1$	$L_v(\underline{\sigma}; \underline{y})$	$G(\underline{\sigma})$	$\underline{g}(\underline{\sigma})$
3	$\underline{\gamma}^+$	$c$	$L_c(\underline{\gamma}^+; \underline{y})$	$K(\underline{\gamma}^+)$	$\underline{k}(\underline{\gamma}^+)$
4	$\underline{\psi}^+$	$c$	$L_R(\underline{\psi}^+; \underline{y})$	$M(\underline{\psi}^+)$	$\underline{m}(\underline{\psi}^+)$
5	$\gamma_1$	1	$L_c(\gamma_1; \underline{y})$	$K^*(\underline{\gamma}^+)$	$\underline{k}^*(\underline{\gamma}^+)$
6	$\underline{\psi}^+$	$c$	$L_R(\underline{\psi}^+; \underline{y})$	$M^*(\underline{\psi}^+)$	$\underline{m}^*(\underline{\psi}^+)$
7	$\underline{m}$	$c+1$	$L_1^*(\underline{m}; \underline{y})$	$Q^*(\underline{m})$	$\underline{q}^*(\underline{m})$
8	$\underline{\gamma}$	$c+1$	$L_1(\underline{\gamma}; \underline{y})$	$B^*(\underline{\gamma})$	$\underline{b}(\underline{\gamma})$
9	$\underline{\sigma}$	$c+1$	$L_v(\underline{\sigma}; \underline{y})$	$G^*(\underline{\sigma})$	$\underline{g}(\underline{\sigma})$

Define

$$\underline{\phi}(\underline{\theta}) \equiv \Pi - \frac{\partial L}{\partial \theta_i}$$

$$= [\phi_1(\underline{\theta}), \dots, \phi_a(\underline{\theta})]', \quad (i = 1, \dots, a)$$

$$\Phi(\underline{\theta}) \equiv \Pi \left[ - \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] \quad (i, j = 1, \dots, a) .$$

Note, for algorithms 1-4 in Table 4.1, that  $N(\underline{\theta}) = - \Phi(\underline{\theta})$  and

$$\underline{n}(\underline{\theta}) = - \underline{\phi}(\underline{\theta}) .$$

### A. Linear Inequality Constraints

In this section, we discuss techniques for minimizing  $-L(\underline{\theta}; \underline{y})$  subject to a set of linear inequality constraints. That is, we consider the problem:

$$\left. \begin{array}{l} \text{minimize } -L(\underline{\theta}; \underline{y}) \\ \text{with respect to } \underline{\theta} \\ \text{subject to } A_o \underline{\theta} \geq \underline{b}_o, \end{array} \right\} \quad (4.2)$$

where  $A_o$  is a given  $m \times a$  matrix whose  $i$ -th row, say  $\underline{a}'_{o,i}$ , contains the coefficients corresponding to the  $i$ -th constraint, and  $\underline{b}_o$  is a  $m \times 1$  known vector whose  $i$ -th element is  $b_{o,i}$  ( $i = 1, \dots, m$ ).

#### 1. Active set methods

a. Definitions and notation If the  $i$ -th constraint,  $\underline{a}'_{o,i} \underline{\theta} \geq b_{o,i}$ , is satisfied as an equality at a point  $\bar{\underline{\theta}}$  (i.e., if  $\underline{a}'_{o,i} \bar{\underline{\theta}} = b_{o,i}$ ), then the  $i$ -th constraint is said to be active at the point  $\bar{\underline{\theta}}$ . If  $\underline{a}'_{o,i} \bar{\underline{\theta}} > b_{o,i}$ , then the  $i$ -th constraint is said to be inactive at  $\bar{\underline{\theta}}$ .

Let  $\underline{\theta}^*$  represent a solution to problem (4.2). The active set of constraints for problem (4.2) is that subset of the  $m$  constraints determined by  $A_o \underline{\theta} \geq \underline{b}_o$  that are active at  $\underline{\theta}^*$ . We refer to the subset of the  $m$  constraints that are active at the  $p$ -th iterate  $\underline{\theta}^{(p)}$  as a working set of constraints.

Let  $t_p$  represent the number of constraints in the working set as of the  $p$ -th iteration, that is, the number of constraints that are active at  $\underline{\theta}^{(p)}$ ; let  $t^*$  represent the number of constraints in the active set [for problem (4.2)]; let  $\hat{A}_p$  represent the  $t_p \times a$  matrix whose  $i$ -th row is the row of  $A_0$  that corresponds to the  $i$ -th constraint in the working set as of the  $p$ -th iteration [assume that  $\hat{A}_p$  has full row rank]; let  $A^*$  represent the  $t^* \times a$  matrix whose  $i$ -th row, say  $\underline{a}_i^{*}$ , is the row of  $A_0$  that corresponds to the  $i$ -th constraint in the active set; let  $\hat{\underline{b}}_p$  represent the  $t_p \times 1$  vector whose  $i$ -th element is the element of  $\underline{b}_0$  that corresponds to the  $i$ -th constraint in the working set as of the  $p$ -th iteration; let  $\underline{b}^*$  represent the  $t^* \times 1$  vector whose  $i$ -th element is the element of  $\underline{b}_0$  that corresponds to the  $i$ -th constraint in the active set; let  $\hat{Z}_p$  represent an  $a \times (a - t_p)$  matrix whose columns form a basis for the set of vectors orthogonal to the rows of  $\hat{A}_p$ , and let  $Z^*$  represent an  $a \times (a - t^*)$  matrix whose columns form a basis for the set of vectors orthogonal to the rows of  $A^*$ .

b. Motivation and general strategy      Note that if  $\underline{\theta}^*$  is a point at which  $-L(\underline{\theta}; \underline{y})$  attains a local minimum in problem (4.2), then it is also a point at which  $-L(\underline{\theta}; \underline{y})$  attains a local minimum in the problem:

$$\left. \begin{array}{l} \text{minimize } -L(\underline{\theta}; \underline{y}) \\ \text{with respect to } \underline{\theta} \\ \text{subject to } A^* \underline{\theta} = \underline{b}^* . \end{array} \right\} \quad (4.3)$$

Thus, if the active set of constraints were known, it would suffice to solve problem (4.3). This would be advantageous, since problem (4.3) could easily be reduced to an unconstrained optimization problem.

[See, e.g., GMW (1981).]

Unfortunately, the active set of constraints is generally unknown and, hence, the available information is insufficient for solving problem (4.3). However, the relationship between problems (4.2) and (4.3) has led to a class of methods, known as active set methods, which are widely used in solving linearly constrained problems. In active set methods, the working set at each iteration is regarded as an approximation to the true active set. In what follows, we concentrate on feasible point active set methods, that is, active set methods in which the  $p$ -th iterate  $\underline{\theta}^{(p)}$  satisfies  $A_{O-}\underline{\theta}^{(p)} \geq \underline{b}_O$  ( $p = 1, 2, \dots$ ).

The general form of the  $(p+1)$ -st iterate in these methods is  $\underline{\theta}^{(p)} + \alpha_p \underline{\ell}_p$ , where the "search direction"  $\underline{\ell}_p$  and the "steplength"  $\alpha_p$  are chosen to satisfy the following two conditions:

$$\begin{aligned}
 & \text{(i) } \underline{\theta}^{(p)} + \underline{\ell}_p \text{ is a solution to the following problem:} \\
 & \left. \begin{aligned}
 & \text{minimize } -\tilde{F}(\underline{\theta}) \\
 & \text{with respect to } \underline{\theta} \\
 & \text{subject to } \hat{A}_{p-}\underline{\theta} = \hat{\underline{b}}_p,
 \end{aligned} \right\} \quad (4.4)
 \end{aligned}$$

where  $\tilde{F}(\underline{\theta})$  is the function  $L(\underline{\theta}; \underline{y})$  or some approximation to  $L(\underline{\theta}; \underline{y})$ ;

$$(ii) \quad A_o(\underline{\theta}^{(p)} + \alpha \underline{\ell}_{p-p}) \geq \underline{b}_o.$$

Note that  $\hat{A}_{p-p} \underline{\ell}_{p-p} = \underline{0}$ , implying that

$$\hat{A}_p(\underline{\theta}^{(p)} + \alpha \underline{\ell}_{p-p}) = \hat{\underline{b}}_p,$$

so that the working constraints are active at the point  $\underline{\theta}^{(p)} + \alpha \underline{\ell}_{p-p}$  as well as at the point  $\underline{\theta}^{(p)}$ .

c. Search direction A search direction of the Newton-Raphson type is obtained by taking the function  $\tilde{F}(\underline{\theta})$  in problem (4.4) to be the quadratic approximation to  $L_1(\underline{\theta}; \underline{y})$  given by

$$\begin{aligned} \tilde{F}(\underline{\theta}) = & L(\underline{\theta}^{(p)}; \underline{y}) + \underline{n}(\underline{\theta}^{(p)})'(\underline{\theta} - \underline{\theta}^{(p)}) \\ & + \frac{1}{2} (\underline{\theta} - \underline{\theta}^{(p)})' \underline{N}(\underline{\theta}^{(p)}) (\underline{\theta} - \underline{\theta}^{(p)}). \end{aligned} \quad (4.5)$$

For this choice of  $\tilde{F}(\underline{\theta})$ , the solution to problem (4.4) is

$$\underline{\theta}_p \equiv \underline{\theta}^{(p)} - \hat{\underline{z}}_p [\hat{\underline{z}}_p' \underline{N}(\underline{\theta}^{(p)}) \hat{\underline{z}}_p]^{-1} \hat{\underline{z}}_p' \underline{n}(\underline{\theta}^{(p)})$$

[GMW (1981)], provided that the matrix  $\hat{\underline{z}}_p' \underline{N}(\underline{\theta}^{(p)}) \hat{\underline{z}}_p$  is positive definite.

The search direction is then  $\underline{\ell}_p = \underline{\theta}_p - \underline{\theta}^{(p)}$ .

d. Steplength For sufficiently large values of  $\alpha$ , the vector  $\underline{\theta}^{(p)} + \alpha \underline{\ell}_p$  may not satisfy the inequality constraints  $A_0 \underline{\theta} \geq \underline{b}_0$ .

Let  $\bar{\alpha}_p$  represent the supremum of those values of  $\alpha$  for which  $A_0 (\underline{\theta}^{(p)} + \alpha \underline{\ell}_p) \geq \underline{b}_0$ . To determine  $\bar{\alpha}_p$ , note that if  $\underline{a}'_{0,i} \underline{\ell}_p \geq 0$ , then  $\underline{a}'_{0,i} (\underline{\theta}^{(p)} + \alpha \underline{\ell}_p) = \underline{a}'_{0,i} \underline{\theta}^{(p)} + \alpha \underline{a}'_{0,i} \underline{\ell}_p \geq \underline{b}_{0,i}$  for any nonnegative scalar  $\alpha$  [since  $\underline{\theta}^{(p)}$  is a feasible iterate, i.e., since  $A_0 \underline{\theta}^{(p)} \geq \underline{b}_0$ ]. If, however,  $\underline{a}'_{0,i} \underline{\ell}_p < 0$ , then

$$\underline{a}'_{0,i} (\underline{\theta}^{(p)} + \alpha \underline{\ell}_p) \begin{cases} = \underline{b}_{0,i} , & \text{if } \alpha = \xi_i \\ < \underline{b}_{0,i} , & \text{if } \alpha > \xi_i \\ > \underline{b}_{0,i} , & \text{if } \alpha < \xi_i , \end{cases}$$

where  $\xi_i = \frac{\underline{b}_{0,i} - \underline{a}'_{0,i} \underline{\theta}^{(p)}}{\underline{a}'_{0,i} \underline{\ell}_p}$ . Thus, the maximum nonnegative feasible

step along  $\underline{\ell}_p$  is defined by

$$\bar{\alpha}_p = \begin{cases} \text{minimum}_{i \in \mathcal{K}} \{ \xi_i \} , & \text{if } \mathcal{K} \neq \emptyset \\ +\infty , & \text{if } \mathcal{K} = \emptyset , \end{cases}$$

where  $\mathcal{K} = \{i \in \{1, \dots, m\} : \underline{a}'_{0,i} \underline{\ell}_p < 0\}$ . Note that if  $\underline{a}'_{0,i} \in \mathcal{R}(\hat{A}_p)$ , then  $i \notin \mathcal{K}$ . The stepsize  $\alpha_p$  can be chosen from the interval  $0 < \alpha_p \leq \bar{\alpha}_p$  by adopting one of the rules discussed in Section III.B.1.

e. Updating the working set In practice, some modifications to the algorithm set forth in Section IV.A.1.a - IV.A.1.d are desirable. If the stepsize  $\alpha_p$  is chosen to be the maximum feasible stepsize  $\bar{\alpha}_p$ , then one or more constraints is added to the working set. Generally, it is recommended that no more than one constraint be added. If the algorithm calls for adding more than one constraint to the working set, then it may be desirable to modify the algorithm so that only one is added. One such modification is described by Gill and Murray (1974). It is recommended that the algorithm be further modified to allow for deletions from the working set. Lenard (1979) evaluated several strategies for adding and deleting constraints from the working set.

f. Summary In summary, the  $(p+1)$ -st iteration of a feasible active set method for solving problem (4.2) might consist of the following steps [GMW (1981)]:

(1) Check whether  $\theta^{(p)}$  satisfies some user-supplied termination criteria; if they are not satisfied, then proceed to step (2).

(2) Determine which, if any, constraints are to be deleted from the working set; if a constraint is to be deleted, proceed to step (6); otherwise, proceed to step (3).

(3) Determine a feasible search direction  $\underline{\ell}_p$  by solving problem (4.4), taking  $\underline{\ell}_p$  to be the difference between this solution and  $\theta^{(p)}$ . For example, take  $\underline{\ell}_p = -\hat{Z}_p [\hat{Z}_p' N(\theta^{(p)}) \hat{Z}_p]^{-1} \hat{Z}_p' n(\theta^{(p)})$  when  $\hat{Z}_p' N(\theta^{(p)}) \hat{Z}_p$  is a positive definite matrix.

(4) Determine the maximum nonnegative feasible step  $\bar{\alpha}_p$  and choose the stepsize  $\alpha_p$  so that  $\alpha_p \leq \bar{\alpha}_p$  and so that  $-L(\underline{\theta}^{(p)} + \alpha_p \underline{l}_p)$  is "sufficiently" smaller than  $-L(\underline{\theta}^{(p)})$ . If  $\alpha_p < \bar{\alpha}_p$ , then proceed to step (7); otherwise, proceed to step (5).

(5) Add to the working set one of those constraints (outside the working set) that are active at the point  $\underline{\theta}^{(p)} + \alpha_p \underline{l}_p$ ; then proceed to step (7).

(6) Delete a constraint from the working set (and modify  $t_p$ ,  $\hat{A}_p$ ,  $\hat{b}_p$ , and  $\hat{Z}_p$  accordingly); then return to step (2).

(7) Set  $\underline{\theta}^{(p+1)} = \underline{\theta}^{(p)} + \alpha_p \underline{l}_p$  and increase by one the value of  $p$ ; then return to step (1).

We refer to this algorithm as Algorithm 4.1.

#### g. Variations

##### Powell's algorithm

There are numerous ways in which Algorithm 4.1 might be modified or implemented. Powell (1974) suggested that, in carrying out step (3) of Algorithm 4.1 [i.e., in solving problem (4.4)],  $\tilde{F}(\underline{\theta})$  be taken to be  $L(\underline{\theta}; \underline{y})$ . If  $\hat{A}_p = A^*$ , then, with  $\tilde{F}(\underline{\theta}) = L(\underline{\theta}; \underline{y})$ , problem (4.4) would reduce to problem (4.3). The drawback of choosing  $\tilde{F}(\underline{\theta}) = L(\underline{\theta}; \underline{y})$  is that problem (4.4) then requires an iterative solution (in general). Consequently, when  $\hat{A}_p \neq A^*$ , considerable computation may be expended on finding a minimum of  $-L(\underline{\theta}; \underline{y})$  over a subspace

that differs considerably from that determined by the equalities  $A^* \underline{\theta}^* = \underline{b}^*$ . Powell also omitted steps (2) and (6) of Algorithm 4.1. He, alternatively, recommended repeated applications of an algorithm comprised of steps (1), (3)-(5), and (7) of Algorithm 4.1 [with  $\tilde{F}(\underline{\theta}) = L(\underline{\theta}; \underline{y})$ ], where constraints were deleted from the working set only between applications (if at all).

Powell (1974) proposed that Algorithm 4.1 be terminated on the  $p$ -th iteration if, in step (1), the necessary conditions for  $\underline{\theta}^{(p)}$  to be a local minimum of  $-L(\underline{\theta}; \underline{y})$  subject to  $\hat{A}_p(\underline{\theta}) = \hat{\underline{b}}_p$  are satisfied. As given by GMW (1981), these conditions are:

$$\begin{aligned}
 (1) \quad & A_{p-} \underline{\theta}^{(p)} = \hat{\underline{b}}_p ; \\
 (2) \quad & \underline{\phi}(\underline{\theta}^{(p)}) = \hat{A}_p' \hat{\underline{\lambda}}_{-p} \quad \text{for some } t_p \times 1 \text{ vector} \\
 & \hat{\underline{\lambda}}_{-p} \equiv (\hat{\lambda}_{p,1}, \dots, \hat{\lambda}_{p,t_p})' ; \quad \text{and} \\
 (3) \quad & \text{the matrix } \hat{\underline{Z}}_p' \underline{\phi}(\underline{\theta}^{(p)}) \hat{\underline{Z}}_p \text{ is nonnegative definite.}
 \end{aligned} \tag{4.6}$$

Powell (1974) showed that, if  $\hat{\lambda}_{p,s} < 0$  for some  $s \in \{1, \dots, t_p\}$ , then a new feasible point  $\bar{\underline{\theta}}_p$  could be found (by deleting the  $s$ -th constraint in the working set) such that  $-L(\bar{\underline{\theta}}_p; \underline{y}) < -L(\underline{\theta}^{(p)}; \underline{y})$ . Thus, if conditions (4.6) were satisfied at  $\underline{\theta}^{(p)}$  and if  $\hat{\lambda}_{p,s} < 0$ , Powell deleted the  $s$ -th constraint from the working set and restarted his version of Algorithm 4.1, taking the modified working set to be the initial working set, and taking the initial value of  $\underline{\theta}$  to be the final iterate from the previous application of the algorithm.

Deleting constraints from the working set

Let  $\tilde{\theta}_{(p)}$  be a local minimum (or global minimum) of  $-L(\theta; \underline{y})$  subject to  $\hat{A}_p \theta = \hat{b}_p$ . Powell (1974) showed that there exists a  $t_p \times 1$  Lagrange multiplier vector  $\tilde{\lambda}_{-p}$  such that

$$\phi(\tilde{\theta}_{(p)}) = \hat{A}_p' \tilde{\lambda}_{-p}.$$

He argued that if an element of  $\tilde{\lambda}_{-p}$  had a negative sign, then the corresponding constraint should be deleted from the working set. He thus, tried to devise a method for computing  $\tilde{\theta}_{(p)}$  and  $\tilde{\lambda}_{-p}$  explicitly.

An alternative strategy is to use an estimate of  $\tilde{\lambda}_{-p}$  in deciding which, if any, constraints to delete. GMW (1981) gave two estimators:

$$(1) \quad \tilde{\lambda}_{-p}^{(1)} = (\hat{A}_p \hat{A}_p')^{-1} \hat{A}_p \phi(\tilde{\theta}_{(p)}), \text{ and}$$

$$(2) \quad \tilde{\lambda}_{-p}^{(2)} = (\hat{A}_p \hat{A}_p')^{-1} \hat{A}_p [\phi(\tilde{\theta}_{(p)}) + \phi(\tilde{\theta}_{(p)}) \hat{\ell}_{-p}],$$

where  $\hat{\ell}_{-p}$  is some estimate of  $\tilde{\theta}_{-p} - \theta_{(p)}$ . Unfortunately, unless  $\tilde{\theta}_{(p)}$  is sufficiently close to  $\tilde{\theta}_{(p)}$ , some elements of  $\tilde{\lambda}_{-p}^{(1)}$  and  $\tilde{\lambda}_{-p}^{(2)}$  may differ in sign from the corresponding elements of  $\tilde{\lambda}_{-p}$  [GMW (1981)]. Further, the use of estimates of  $\tilde{\lambda}_{-p}$  can cause the algorithm to "zigzag", i.e., to repeatedly add and delete the same constraint(s). Powell (1974) showed that this can result in slow progress and possibly in convergence to a nonoptimal point.

If more than one constraint in the working set has an associated Lagrange multiplier (or estimated Lagrange multiplier) that is negative, then, generally, only one of them (e.g., the one whose Lagrange multiplier is largest in magnitude) is deleted.

#### Steplength

Step (4) of Algorithm 4.1 is begun by computing  $\bar{\alpha}_p$  (the maximum nonnegative feasible step in the direction  $\underline{\ell}_p$ ). Since most stepsize rules (see Section III.B.1) require that  $L(\underline{\theta}; \underline{y})$  be evaluated at several points, computing  $\bar{\alpha}_p$  prevents the unnecessary evaluation of  $L(\underline{\theta}; \underline{y})$  at a point outside the parameter space. The partial stepping rule of Vandaele and Chowdhury (1971) could be improved upon by first computing  $\bar{\alpha}_p$ . Their rule begins with the evaluation of  $L(\underline{\theta}^{(p)} + \underline{\ell}_p; \underline{y})$ . If  $L(\underline{\theta}^{(p)} + \underline{\ell}_p; \underline{y}) > L(\underline{\theta}^{(p)}; \underline{y})$ , then it sets  $\alpha_p = 2^{v-1}$ , where  $v$  is the first positive integer such that

$$L(\underline{\theta}^{(p)} + 2^v \underline{\ell}_p; \underline{y}) < L(\underline{\theta}^{(p)} + 2^{v-1} \underline{\ell}_p; \underline{y}) .$$

The quantity  $\bar{\alpha}_p$  provides a useful upper bound for  $\alpha_p$ . If  $L(\underline{\theta}^{(p)} + \underline{\ell}_p; \underline{y}) \leq L(\underline{\theta}^{(p)}; \underline{y})$  or if  $\bar{\alpha}_p < 1$ , then  $\alpha_p$  could be set equal to  $\min\{\bar{\alpha}_p, (\frac{1}{2})^v\}$ , where  $v$  is the first positive integer such that  $L[\underline{\theta}^{(p)} + (\frac{1}{2})^v \underline{\ell}_p; \underline{y}] > L(\underline{\theta}^{(p)}; \underline{y})$ . [If the matrix  $-\hat{Z}_p' N(\underline{\theta}^{(p)}) \hat{Z}_p$  is positive definite, then such an integer  $v$  will necessarily exist.]

The algorithms 1-9 in Table 3.1 were constructed using a constant stepsize rule of  $\alpha_p = 1$ . A simple generalization of this rule is to take  $\alpha_p = \min\{1, \bar{\alpha}_p\}$ . This, however, does not insure that  $-L(\underline{\theta}^{(p)} + \alpha_p \underline{\ell}_p; \underline{y})$  is "sufficiently" less than  $-L(\underline{\theta}^{(p)}; \underline{y})$ .

#### Null space versus range space methods

Suppose that,

in step (3) of Algorithm 4.1, the search direction  $\underline{\ell}_p$  was taken to be  $\underline{\theta}_p - \underline{\theta}^{(p)}$ , where  $\underline{\theta}_p$  is the solution to problem (4.4) when choice (4.5) is made for  $\tilde{F}(\underline{\theta})$ . Then,

$$\underline{\ell}_p = - \hat{\underline{Z}}_p [\hat{\underline{Z}}_p' \underline{N}(\underline{\theta}^{(p)}) \hat{\underline{Z}}_p]^{-1} \hat{\underline{Z}}_p' \underline{n}(\underline{\theta}^{(p)}) . \quad (4.7)$$

Note that expression (4.7) depends on  $\hat{\underline{A}}_p$  only through the matrix  $\hat{\underline{Z}}_p$  whose columns span the null space of  $\hat{\underline{A}}_p$ . Consequently, the determination of  $\underline{\ell}_p$  from expression (4.7) is called a null-space method.

Range-space methods provide an alternative to null-space methods for choosing  $\underline{\ell}_p$ . In one such method,

$$\underline{\ell}_p = - \underline{N}^{-1}(\underline{\theta}^{(p)}) [\hat{\underline{A}}_p \underline{\lambda}_{p-p} + \underline{n}(\underline{\theta}^{(p)})] ,$$

where

$$\underline{\lambda}_{p-p} = - [\hat{\underline{A}}_p \underline{N}^{-1}(\underline{\theta}^{(p)}) \hat{\underline{A}}_p']^{-1} \hat{\underline{A}}_p \underline{N}^{-1}(\underline{\theta}^{(p)}) \underline{n}(\underline{\theta}^{(p)}) .$$

GMW (1981) describe other null-space methods.

Null space methods for choosing  $\underline{l}_p$  tend to be most successful when  $t_p$ , the number of constraints in the working set, is large, while range-space methods may be preferable when  $t_p$  is small.

h. A special case - simple bounds Suppose that in problem (4.2), the constraints  $A_o \underline{\theta} \geq \underline{b}_o$  are actually simple bounds  $\underline{l} \leq \underline{\theta} \leq \underline{u}$ , where for a particular problem, some elements of  $\underline{l}$  or  $\underline{u}$  may be omitted. For example, in mixed model (1.2),  $\gamma_i \geq 0$  ( $i = 1, \dots, c$ ) and  $\gamma_{c+1} > 0$  or, approximately,  $\gamma_{c+1} \geq \epsilon$ , where  $\epsilon$  is a small positive constant.

The constraints  $\underline{l} \leq \underline{\theta} \leq \underline{u}$  can be written in the form  $A_o \underline{\theta} \geq \underline{b}_o$  by taking

$$A_o = \begin{pmatrix} I_a \\ -I_a \end{pmatrix}, \quad \underline{b}_o = \begin{pmatrix} \underline{l} \\ -\underline{u} \end{pmatrix}. \quad (4.8)$$

When  $A_o$  and  $\underline{b}_o$  are of the form (4.8), the rows of the matrix  $\hat{A}_p$  consist of linearly independent signed rows of an  $a \times a$  identity matrix. Moreover, the number  $t_p$  of rows in  $\hat{A}_p$  is then the number of variables that are currently fixed at one of the bounds - these variables are called "fixed", while the other variables are called "free." The  $(a - t_p)$  columns of  $\hat{Z}_p$  are the columns of the identity matrix which correspond to the free variables. Thus, when  $A_o$  and  $\underline{b}_o$  are of the form (4.8) and when  $\tilde{F}(\underline{\theta})$  is given by (4.5), the solution to problem (4.4) reduces to

$$\underline{\theta}_p = \underline{\theta}^{(p)} - \hat{z}_p G_{FR}^{-1} \underline{g}_{FR},$$

where the elements of  $\underline{g}_{FR}$  are the elements of  $\underline{n}(\underline{\theta}^{(p)})$  that correspond to the free variables and where the rows and columns of  $G_{FR}$  consist of the rows and columns of  $N(\underline{\theta}^{(p)})$  that correspond to the free variables.

## 2. Bertsekas' method

Bertsekas (1982a) noted that active set methods "are quite efficient for problems of relatively small dimension." However, he expressed skepticism about the efficiency of these methods in solving problem (4.2) when  $a$  and  $m$  are extremely large. He noted that a large number of iterations may be required before the working set closely approximates the active set. He described an alternative to active set methods which (i) allows for the addition of up to  $a$  constraints to the "working set" in a single iteration, and (ii) circumvents the need to solve a quadratic programming subproblem at each iteration. In this section, we describe Bertsekas' (1982a) algorithm for three types of constraints: simple nonnegativity constraints ( $\underline{\theta} \geq \underline{0}$ ), simple bounds ( $\underline{b}_1^0 \leq \underline{\theta} \leq \underline{b}_2^0$ ), and general linear constraints ( $\underline{b}_1^0 \leq A_0 \underline{\theta} \leq \underline{b}_2^0$ ).

a. Simple nonnegativity constraints      Consider the problem

$$\left. \begin{array}{l} \text{minimize } -L(\underline{\theta}; \underline{y}) \\ \text{with respect to } \underline{\theta} \\ \text{subject to } \underline{\theta} \geq \underline{0} . \end{array} \right\} \quad (4.9)$$

Terminology and notation

The following definitions

and notation are useful in describing Bertsekas' methodology:

1. For any  $a \times 1$  vector  $\underline{\theta} = (\theta_1, \dots, \theta_a)'$ ,  
define

$$\underline{\theta}^{++} \equiv [\max\{0, \theta_i\}] \quad (i = 1, \dots, a).$$

2. A symmetric matrix  $\bar{D} = [d_{ij}]$  ( $i, j = 1, \dots, a$ ) is said to be "diagonal with respect to a subset of indices  $\bar{I} \subset \{1, \dots, a\}$ " if, for all  $i \in \bar{I}$ , the  $i$ -th row of  $\bar{D}$  is null except possibly for the  $(i, i)$ -th element.

3. A point  $\bar{\theta}$  which satisfies  $\bar{\theta} = [\bar{\theta} - \alpha \phi(\bar{\theta})]^{++}$  for all  $\alpha$  ( $\alpha > 0$ ) is said to be a critical point for problem (4.9).

The algorithm

Let  $\underline{\theta}^{(p)} = (\theta_1^{(p)}, \dots, \theta_a^{(p)})'$  represent the  $p$ -th iterate of Bertsekas' (1982a) method. Define

$$I_p^+ \equiv \{i \mid 0 \leq \theta_i^{(p)} \leq \epsilon_p, \phi_i(\underline{\theta}^{(p)}) > 0\}, \quad i \in \{1, \dots, a\},$$

where, for some small scalar  $\epsilon$  ( $\epsilon > 0$ ) and some diagonal positive definite matrix  $M$  (e.g., the identity matrix),

$$\epsilon_p \equiv \min\{\epsilon, \|\underline{\theta}^{(p)} - [\underline{\theta}^{(p)} - M\phi(\underline{\theta}^{(p)})]^{++}\|\},$$

where  $||\cdot||$  denotes Euclidean norm. Let  $D_p$  represent a positive definite symmetric matrix that is diagonal with respect to  $I_p^+$ . Define

$$\underline{\theta}^{(p)}(\alpha) \equiv [\underline{\theta}^{(p)} - \alpha D_p \underline{\phi}(\underline{\theta}^{(p)})]^{++}, \alpha \geq 0.$$

The  $(p+1)$ -st iterate of Bertsekas' (1982a) method is

$$\underline{\theta}^{(p+1)} = [\underline{\theta}^{(p)} - \alpha_p D_p \underline{\phi}(\underline{\theta}^{(p)})]^{++} = \underline{\theta}^{(p)}(\alpha_p)$$

( $p = 0, 1, 2, \dots$ ). Here,  $\alpha_p$  ( $\alpha_p \geq 0$ ) is a stepsize determined by the following Armijo-like rule: for a scalar  $\beta \in (0, 1)$  and a scalar  $\sigma \in (0, \frac{1}{2})$ , put

$$\alpha_p = \beta^{m_p},$$

where  $m_p$  is the first nonnegative integer  $m$  such that

$$\begin{aligned} & -L(\underline{\theta}^{(p)}; \underline{y}) - \{-L[\underline{\theta}^{(p)}(\beta^m); \underline{y}]\} \\ & \geq \sigma \left\{ \beta^m \sum_{i \in I_p^+} \phi_i(\underline{\theta}^{(p)}) \ell_{p,i} + \sum_{i \in I_p^+} \phi_i(\underline{\theta}^{(p)}) [\theta_i^{(p)} - \theta_i^{(p)}(\beta^m)] \right\}, \end{aligned}$$

where  $\ell_{p,i}$  is the  $i$ -th element of the vector  $\underline{\ell}_p \equiv D_p \underline{\phi}(\underline{\theta}^{(p)})$  and  $\theta_i^{(p)}(\beta^m)$  is the  $i$ -th element of  $\underline{\theta}^{(p)}(\beta^m) \equiv [\underline{\theta}^{(p)} - \beta^m D_p \underline{\phi}(\underline{\theta}^{(p)})]^{++}$  ( $i=1, \dots, a$ ).

Bertsekas (1982a) showed that  $\{-L(\underline{\theta}^{(p)}; \underline{y})\}$  forms a monotone decreasing sequence, and that each iterate  $\underline{\theta}^{(p)}$  satisfies the constraints  $\underline{\theta} \geq 0$ . Further, he showed that, under certain rather unrestrictive conditions,

(1) every limit point of the sequence  $\{\underline{\theta}^{(p)}\}$  is a critical point of problem (4.9) (assuming that  $\alpha_p$  is determined by the forementioned Armijo-like rule),

(2)  $\{\underline{\theta}^{(p)}\}$  converges to a local minimum  $\underline{\theta}^*$  of (4.9) if the sequence ever enters a sufficiently small neighborhood of  $\underline{\theta}^*$ , and

(3) if  $\phi_i(\underline{\theta}^*) > 0$  for all  $i$  belonging to the set of indices corresponding to the active constraints at  $\underline{\theta}^*$ , say  $B^*$ , then after a finite number of iterations,  $I_p^+$  becomes  $B^*$  and does not change after that.

Bertsekas' algorithm allows some flexibility in the choice of the matrix  $D_p$  at the  $(p+1)$ -st iteration. He shows that if the rows and columns of  $\Phi(\underline{\theta}^{(p)})$ , corresponding to the indices  $i \notin I_p^+$ , form a positive definite matrix, and if the corresponding portion of  $D_p$  is chosen to be the inverse of this "submatrix", then the algorithm displays a super-

linear rate of convergence, i.e.,  $\lim_{p \rightarrow \infty} \frac{\|\underline{\theta}^{(p+1)} - \underline{\theta}^*\|}{\|\underline{\theta}^{(p)} - \underline{\theta}^*\|} = 0$ . A generali-

zation of this, which would exploit the differences in algorithms 1-9 in Table 4.1, is to use the appropriate rows and columns from  $-N(\underline{\theta}^{(p)})$ , rather than from  $\Phi(\underline{\theta}^{(p)})$ .

b. Simple bounds Consider the problem

$$\left. \begin{array}{l} \text{minimize } -L(\underline{\theta}; \underline{y}) \\ \text{with respect to } \underline{\theta} \\ \text{subject to } \underline{b}_1^0 \leq \underline{\theta} \leq \underline{b}_2^0, \end{array} \right\} \quad (4.10)$$

where  $\underline{b}_1^0 = (b_{1,1}^0, \dots, b_{1,a}^0)'$  and  $\underline{b}_2^0 = (b_{2,1}^0, \dots, b_{2,a}^0)'$ . If  $\theta_i$  has only an upper (lower) bound then  $\theta_i \leq b_{2,i}^0$  ( $\geq b_{1,i}^0$ ) can be approximated by putting  $b_{1,i}^0$  ( $b_{2,i}^0$ ) equal to an arbitrarily large negative (positive) number. If  $\theta_i$  has no bound at all, then  $b_{1,i}^0$  and  $b_{2,i}^0$  can, together, be taken to be arbitrarily large negative and positive numbers, respectively.

The described algorithm in Section IV.A.2.a for solving problem (4.9) can be generalized to problem (4.10) by making the following substitutions:

1. Replace  $I_p^+$  by

$$I_p^\# = \{i | b_{1,i}^0 \leq \theta_i^{(p)} \leq b_{1,i}^0 + \epsilon_p \text{ and } \phi_i(\theta^{(p)}) > 0, \\ \text{or } b_{2,i}^0 - \epsilon_p \leq \theta_i^{(p)} \leq b_{2,i}^0 \text{ and } \phi_i(\theta^{(p)}) < 0\},$$

$$i \in \{1, \dots, a\}.$$

2. Redefine  $\underline{\theta}^{(p)}(\alpha)$  to be

$$\underline{\theta}^{(p)}(\alpha) \equiv [\underline{\theta}^{(p)} - \alpha D_p \phi(\underline{\theta}^{(p)})]^\# ,$$

where, for an arbitrary  $a \times 1$  vector  $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_a)'$ , the  $i$ -th element of  $[\bar{\theta}]^\#$  is defined to be

$$[\bar{\theta}_i]^\# \equiv \begin{cases} b_{2,i}^0 & , \text{ if } b_{2,i}^0 \leq \bar{\theta}_i \\ \bar{\theta}_i & , \text{ if } b_{1,i}^0 < \bar{\theta}_i < b_{2,i}^0 \\ b_{1,i}^0 & , \text{ if } \bar{\theta}_i \leq b_{1,i}^0 \end{cases} .$$

3. Redefine  $\epsilon_p$  by

$$\epsilon_p \equiv \min\{\epsilon, ||\bar{\theta}^{(p)} - [\bar{\theta}^{(p)} - M\bar{\theta}(\bar{\theta}^{(p)})]^\#||\} .$$

4. Replace  $[\theta_i^{(p)} - \theta_i^{(p)}(\beta^m)]^+$ , in the Armijo-like rule for determining  $\alpha_p$ , by  $[\theta_i^{(p)} - \theta_i^{(p)}(\beta^m)]^\#$ .

c. General linear constraints

The problem

Consider the problem

$$\left. \begin{array}{l} \text{minimize } -L(\underline{\theta}; \underline{y}) \\ \text{with respect to } \underline{\theta} \\ \text{subject to } \underline{b}_1^0 \leq A_0 \underline{\theta} \leq \underline{b}_2^0 , \end{array} \right\} \quad (4.11)$$

where  $A_0 = (\underline{a}_{0,1}, \dots, \underline{a}_{0,m})'$  and  $\underline{b}_i^0 = (b_{i,1}^0, \dots, b_{i,m}^0)'$  ( $i = 1, 2$ ).

Let  $\theta_i$  represent the  $i$ -th element of  $\underline{\theta}$ .

General strategy

If  $A_0$  is a square nonsingular matrix, then problem (4.11) can be transformed into a problem that has simple bounds for constraints. Consider re-expressing  $-L(\underline{\theta}; \underline{y})$  in terms of  $\underline{\xi} \equiv A_0 \underline{\theta}$ . Assuming  $A_0$  is nonsingular, problem (4.11) is transformed into the equivalent problem

$$\left. \begin{array}{l} \text{minimize } H(\underline{\xi}) \equiv -L(A_0^{-1}\underline{\xi}; \underline{y}) \\ \text{with respect to } \underline{\xi} \\ \text{subject to } \underline{b}_1^0 \leq \underline{\xi} \leq \underline{b}_2^0 . \end{array} \right\} \quad (4.12)$$

The latter problem is of the form considered in Section IV.A.2.b. If  $\underline{\xi}^*$  is a solution to problem (4.12), then  $\underline{\theta}^* = A_0^{-1}\underline{\xi}^*$  is a solution to problem (4.11).

Bertsekas (1982a) indicated how this approach can be extended to the case where  $A_0$  is not a square matrix.

The active generalized rectangle

The "active generalized rectangle" is to Bertsekas' (1982a) method what the working set is to active set methods. To introduce the concept of an active generalized rectangle, put

$$B^0(\underline{\theta}) \equiv \{j \mid b_{1,j}^0 = a'_{0,j}\underline{\theta} \text{ or } b_{2,j}^0 = a'_{0,j}\underline{\theta}\},$$

$$j \in \{1, \dots, m\} ,$$

so that  $B^0(\underline{\theta})$  is the set of indices corresponding to constraints which are active at  $\underline{\theta}$ . Assume that the set  $\{\underline{a}_j | j \in B^0(\underline{\theta})\}$  is a set of linearly independent vectors for all  $\underline{\theta} \in \mathbb{E}^a$ . To insure that this assumption is satisfied, it may be necessary to include among the constraints  $\underline{b}_1^0 \leq A_{o\underline{\theta}} \leq \underline{b}_2^0$  in problem (4.11) the trivial constraints  $-\infty \leq \theta_i \leq +\infty$  ( $i = 1, \dots, a$ ).

At the  $(p+1)$ -st iteration of the algorithm, select a subset  $B_p$  of the set  $\{1, \dots, m\}$  of size  $a$  such that  $B^0(\underline{\theta}^{(p)}) \subset B_p$ , and  $\{\underline{a}_j | j \in B_p\}$  forms a linearly independent set of vectors. For example,  $B_p$  could consist of the elements of the set  $B^0(\underline{\theta}^{(p)})$  together with indices corresponding to trivial constraints. Define

$$\bar{X}_p \equiv \{\underline{\theta} | \underline{b}_{1,j}^0 \leq \underline{a}_{o,j}' \underline{\theta} \leq \underline{b}_{2,j}^0, j \in B_p\}.$$

Bertsekas calls  $\bar{X}_p$  the "active generalized rectangle at iteration  $p$ ."

The role of  $\bar{X}_p$  in Bertsekas' algorithm is similar to that of the set  $\{\underline{\theta} | \hat{A}_p \underline{\theta} = \hat{b}_p\}$  in active set methods.

The set  $\bar{X}_p$  is the collection of all points  $\underline{\theta}$  that satisfy  $a$  linearly independent constraints, including those constraints that are active at  $\underline{\theta}^{(p)}$ . In contrast, the working set of an active set method consists only of those constraints that are active at  $\underline{\theta}^{(p)}$ .

On the  $(p+1)$ -st iteration of an active set method, the  $(p+1)$ -st iterate  $\underline{\theta}^{(p+1)}$  is chosen from those values of  $\underline{\theta}$  at which the working set

constraints are active, i.e., from the set  $\{\theta | \hat{A}_p \theta = \hat{b}_p\}$ . On the  $(p+1)$ -st iteration of Bertsekas' algorithm,  $\theta^{(p+1)}$  is chosen from the set  $\bar{X}_p$ .

On the  $(p+1)$ -st iteration of an active set method, the direction of search is restricted by the constraints in working set, while the stepsize is restricted by the constraints outside the working set. On the  $(p+1)$ -st iteration of Bertsekas' algorithm, the direction of search is restricted by the constraints corresponding to the indices in  $B_p$ , while the remaining constraints restrict the stepsize.

In active set methods, the working set is continually updated in an attempt to obtain a better approximation to the true active set. The set  $B_p$ , which determines the active generalized rectangle  $\bar{X}_p$ , is also updated on each iteration, in an attempt to include all of those constraints that are active at a solution to problem (4.11).

A major difference between the two types of algorithms is that, in active set methods, the working set is usually reduced, or enlarged, by no more than one constraint at a time. In contrast, the a constraints corresponding to the set of indices in  $B_p$  can change completely from one iteration to the next. Bertsekas (1982a) indicated that this feature is advantageous in solving problems for which the active set may be large.

#### Some specifics

Assume, without loss of generality,

that

$$B_p = \{1, 2, \dots, a\}$$

and partition  $A_0$  as

$$A_o = \begin{pmatrix} A_p^+ \\ A_p^- \end{pmatrix}$$

where  $A_p^+$  is an  $a \times a$  nonsingular matrix with rows  $\{a_j^+ | j \in B_p\}$ . Similarly, partition  $b_1^0$  and  $b_2^0$  as

$$b_i^0 = \begin{pmatrix} b_{i,p}^+ \\ b_{i,p}^- \end{pmatrix} \quad (i = 1, 2).$$

Define  $\underline{\xi} = A_p^+ \theta$ . In terms of  $\underline{\xi}$ , the set  $\bar{X}_p$  is

$$\bar{Y}_p = \{\underline{\xi} : b_{1,p}^+ \leq \underline{\xi} \leq b_{2,p}^+\}$$

and problem (4.11) becomes

$$\left. \begin{array}{l} \text{minimize } H_p(\underline{\xi}) \equiv -L[(A_p^+)^{-1}\underline{\xi}; \underline{y}] \\ \text{with respect to } \underline{\xi} \\ \text{subject to } \left\{ \begin{array}{l} \underline{\xi} \in \bar{Y}_p \\ b_{1,p}^- \leq A_p^- [(A_p^+)^{-1}\underline{\xi}] \leq b_{2,p}^- \end{array} \right. \end{array} \right\} \quad (4.13)$$

Let  $\underline{\xi}^{(p)} = A_{p-}^+ \theta^{(p)}$ . Using  $\underline{\xi}^{(p)}$  as the initial value, take  $\underline{\xi}^{(p+1)}$  to be the first iterate obtained by applying Bertsekas' algorithm for problem (4.10) to the problem

$$\text{minimize } H_p(\underline{\xi})$$

with respect to  $\underline{\xi}$

$$\text{subject to } \underline{b}_{1,p}^+ \leq \underline{\xi} \leq \underline{b}_{2,p}^+ \text{ (i.e., subject to } \underline{\xi} \in \bar{Y}_p \text{) ,}$$

but, in applying the algorithm, restrict  $\alpha_p$  to the set

$$\{\alpha \mid \underline{b}_{1,p}^- \leq A_p^- [(A_p^+)^{-1} \underline{\xi}^{(p)}(\alpha)] \leq \underline{b}_{2,p}^-, \}$$

where

$$\underline{\xi}^{(p)}(\alpha) = [\underline{\xi}_p - \alpha D_p [(A_p^+)^{-1}]^{-1} \phi(\theta^{(p)})]^\# (\alpha > 0) .$$

Then,

$$\underline{\xi}^{(p+1)} = \underline{\xi}^{(p)}(\alpha_p) .$$

By construction,  $\underline{\xi}^{(p+1)}$  satisfies the constraints of problem (4.13). In terms of  $\theta$ , the  $(p+1)$ -st iterate of Bertsekas' algorithm for solving problem (4.11) is

$$\theta^{(p+1)} = (A_p^+)^{-1} \underline{\xi}^{(p+1)}$$

or, equivalently,

$$\underline{\theta}^{(p+1)} = (A_p^+)^{-1} [A_p^+ \underline{\theta}^{(p)} - \alpha_p D_p [(A_p^+)' ]^{-1} \underline{\phi}(\underline{\theta}^{(p)})]^\# .$$

Bertsekas (1982a) indicated that the convergence rate of his algorithm is typically superlinear.

### 3. Transformations

Another approach to solving optimization problems with inequality constraints is to transform them to unconstrained problems, or, at least, to problems that are easier to solve. Box (1966) evaluated the usefulness of transformations and suggested that, if an appropriate transformation can be found, then it may greatly reduce the computational effort required to find a solution. He suggested that even the elimination of only a few constraints, via a transformation, could be a "worthwhile step forward."

Although transformations may sometimes be useful in constrained optimization, their appeal has diminished with the development of efficient constrained optimization procedures. GMW (1981) suggested that "it is virtually never worthwhile to transform ... problems [with simple bounds] ... and it is rarely appropriate to alter linearly constrained problems."

In this section, we consider the use of transformations in obtaining a REML estimate of  $\underline{\gamma}$  when the parameter space is  $\Omega_1$ .

a. Squared-variable transformation      Consider the REML problem:

maximize  $L_1(\gamma_1, \dots, \gamma_{c+1}; \underline{y})$

with respect to  $\underline{\gamma}$

subject to the constraints  $\underline{\gamma} \in \Omega_1$  [i.e., subject to the

constraints  $\gamma_i \geq 0$  ( $i = 1, \dots, c$ ),  $\gamma_{c+1} > 0$ ].

Consider the change of variables

$$\gamma_i = \eta_i^2 \quad (i = 1, \dots, c+1),$$

known as the squared-variable transformation. In terms of  $\eta_1, \dots, \eta_{c+1}$ , the REML problem is:

maximize  $L_1(\eta_1^2, \dots, \eta_{c+1}^2; \underline{y})$

with respect to  $\eta_1, \dots, \eta_{c+1}$

subject to the constraint  $\eta_{c+1} \neq 0$ .

Since the constraint  $\eta_{c+1} \neq 0$  rarely comes into play, the transformed problem is essentially an unconstrained optimization problem to which algorithms 1-9 in Table 3.1 can be applied. Suppose that the vector  $(\eta_1^*, \dots, \eta_{c+1}^*)'$  is a solution to this problem. Then the vector  $\underline{\gamma}^*$ , whose  $i$ -th element is  $\gamma_i^* = \eta_i^{*2}$  ( $i = 1, \dots, c+1$ ), is a REML estimate of  $\underline{\gamma}$ .

As noted by Harville (1977), a possible difficulty with this approach is that additional stationary points are introduced. That is, the partial derivative of  $L_1(\eta_1^2, \dots, \eta_{c+1}^2; \underline{y})$ ,

$$\frac{\partial L_1(\eta_1^2, \dots, \eta_{c+1}^2; \underline{y})}{\partial \eta_i} = 2\eta_i \frac{\partial L_1(\underline{\gamma}; \underline{y})}{\partial \gamma_i},$$

equals zero if  $\eta_i = 0$  or  $\frac{\partial L_1(\underline{\gamma}; \underline{y})}{\partial \gamma_i} = 0$  ( $i = 1, \dots, c+1$ ). To avoid convergence to a point at which  $\gamma_i = 0$  but  $\frac{\partial L_1}{\partial \gamma_i} > 0$ , Harville (1977) suggested that this squared-variable transformation be used "only in conjunction with algorithms that guarantee at least some increase in the value of the objective function on each iteration."

GWM (1981) give examples where, after the squared-variable transformation, the objective function has a discontinuous first-order partial derivative or a Hessian matrix that is ill-conditioned in the vicinity of the solution. These features can adversely affect the performance of unconstrained optimization algorithms.

b. Drawbacks More generally, GMW (1981) listed the following undesirable consequences that may result when an optimization problem with linear or nonlinear constraints is transformed to an unconstrained optimization problem:

- (1) the desired maximum may be inadvertently excluded;
- (2) the degree of nonlinearity of the constraint functions may be significantly increased;
- (3) the transformed function may contain discontinuities not present in the original function;
- (4) the scaling of the problem may be adversely affected, and
- (5) the function may be periodic in the new variables.

GMW (1981) showed, in terms of an example where a trigonometric transformation was used to eliminate nonlinear constraints, how difficulties (3) and (5) can arise.

#### B. Nonlinear Constraints

Minimization of  $-L(\underline{\theta}; \underline{y})$  subject to general nonlinear constraints poses a significantly more difficult problem than that of linearly constrained optimization. The primary complication is that, if a nonlinear constraint holds as an equality at a point  $\underline{\theta}^{(p)}$ , then the constraint may be violated if movement (in any direction) is made away from  $\underline{\theta}^{(p)}$ . In this section, we present some algorithms for nonlinear constrained optimization. We restrict attention to algorithms whose iterates are all feasible (i.e., satisfy the constraints), since, as discussed in the first paragraph of Chapter IV, algorithms that can produce infeasible iterates seem undesirable for the REML estimation of  $\underline{\gamma}$ .

### 1. Barrier function methods

Consider the following problem

$$\begin{aligned}
 &\text{minimize } -L(\underline{\theta}; \underline{y}) \\
 &\text{with respect to } \underline{\theta} \\
 &\text{subject to the constraints } c_i(\underline{\theta}) \geq 0 \quad (i = 1, \dots, m),
 \end{aligned} \tag{4.14}$$

where  $c_i(\underline{\theta})$  ( $i = 1, \dots, m$ ) are possibly nonlinear, twice-continuously differentiable functions. One class of methods for solving problem (4.14) is the class of "barrier function methods." In these methods, a "barrier term" is added to  $-L(\underline{\theta}; \underline{y})$ . This term is, for values of  $\underline{\theta}$  in a sufficiently small neighborhood of a constraint boundary, arbitrarily large. Thus, it serves to deflect a minimization algorithm back into the interior of the feasible region. Two examples of barrier functions for problem (4.14) are

$$\begin{aligned}
 B_1(\underline{\theta}, r') &= -L(\underline{\theta}; \underline{y}) + r' \sum_{\ell=1}^m [c_{\ell}(\underline{\theta})]^{-1}, \text{ and} \\
 B_2(\underline{\theta}, r') &= -L(\underline{\theta}; \underline{y}) - r' \sum_{\ell=1}^m \log[c_{\ell}(\underline{\theta})],
 \end{aligned}$$

where  $r'(r' > 0)$  is a constant that is called the barrier parameter.

To apply a barrier function method, a sequence of barrier parameters  $\{r'_p\}$  is chosen so that  $\lim_{p \rightarrow \infty} r'_p = 0$  and  $r'_p > r'_{p+1}$  ( $p = 1, 2, \dots$ ). As the barrier parameter approaches zero, the influence of the barrier term, for  $\underline{\theta}$  in the interior of the feasible region, is reduced. Let  $\underline{\theta}(r'_p)$  denote a local unconstrained minimum of  $B_k(\underline{\theta}, r'_p)$  ( $k = 1, 2$ ). Under mild conditions, the iterative algorithm whose  $p$ -th iterate is  $\underline{\theta}(r'_p)$  will converge to a solution  $\underline{\theta}^*$  of problem (4.14). [See, e.g., GMW (1981)]. Moreover, each iterate will lie in the feasible region.

One approach to the problem of finding a local minimum of  $B_1(\underline{\theta}, r')$  or  $B_2(\underline{\theta}, r')$  is to apply the traditional Newton-Raphson algorithm or the Method of Scoring algorithm. However, to ensure that such an algorithm does not produce an iterate outside the parameter space, the algorithm could be modified to include a partial stepping procedure. To apply either algorithm, we require the first-order partial derivatives of  $B_1$  or  $B_2$  and also the second-order partial derivatives or their expected values. For  $i, j = 1, \dots, a$  and  $\underline{\theta}$  such that  $c_k(\underline{\theta}) \geq 0$  ( $k = 1, \dots, m$ ), we have that

$$\frac{\partial B_1}{\partial \theta_i} = - \frac{\partial L(\underline{\theta}; \underline{y})}{\partial \theta_i} + r' \sum_{\ell=1}^m \left\{ -[c_\ell(\underline{\theta})]^{-2} \frac{\partial c_\ell(\underline{\theta})}{\partial \theta_i} \right\},$$

$$\frac{\partial B_2}{\partial \theta_i} = - \frac{\partial L(\underline{\theta}; \underline{y})}{\partial \theta_i} - r' \sum_{\ell=1}^m \left\{ [c_\ell(\underline{\theta})]^{-1} \frac{\partial c_\ell(\underline{\theta})}{\partial \theta_i} \right\},$$

$$\frac{\partial^2 B_1}{\partial \theta_i \partial \theta_j} = - \frac{\partial^2 L(\underline{\theta}; \underline{y})}{\partial \theta_i \partial \theta_j} + r' \sum_{\ell=1}^m \left\{ 2[c_\ell(\underline{\theta})]^{-3} \frac{\partial c_\ell(\underline{\theta})}{\partial \theta_j} \frac{\partial c_\ell(\underline{\theta})}{\partial \theta_i} \right\}$$

$$- [c_{\ell}(\underline{\theta})]^{-2} \frac{\partial^2 c_{\ell}(\underline{\theta})}{\partial \theta_i \partial \theta_j} \Bigg\} ,$$

$$\begin{aligned} \frac{\partial^2 B_2}{\partial \theta_i \partial \theta_j} = & - \frac{\partial^2 L(\underline{\theta}; \underline{y})}{\partial \theta_i \partial \theta_j} - \mathbf{r}' \sum_{\ell=1}^m \left\{ -[c_{\ell}(\underline{\theta})]^{-2} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_j} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_i} \right. \\ & \left. + [c_{\ell}(\underline{\theta})]^{-1} \frac{\partial^2 c_{\ell}(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} , \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left( \frac{\partial^2 B_1}{\partial \theta_i \partial \theta_j} \right) = & - \mathbb{E} \left( \frac{\partial^2 L(\underline{\theta}; \underline{y})}{\partial \theta_i \partial \theta_j} \right) + \mathbf{r}' \sum_{\ell=1}^m \left\{ 2[c_{\ell}(\underline{\theta})]^{-3} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_j} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_i} \right. \\ & \left. - [c_{\ell}(\underline{\theta})]^{-2} \frac{\partial^2 c_{\ell}(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} , \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left( \frac{\partial^2 B_2}{\partial \theta_i \partial \theta_j} \right) = & - \mathbb{E} \left( \frac{\partial^2 L(\underline{\theta}; \underline{y})}{\partial \theta_i \partial \theta_j} \right) - \mathbf{r}' \sum_{\ell=1}^m \left\{ -[c_{\ell}(\underline{\theta})]^{-2} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_j} \frac{\partial c_{\ell}(\underline{\theta})}{\partial \theta_i} \right. \\ & \left. + [c_{\ell}(\underline{\theta})]^{-1} \frac{\partial^2 c_{\ell}(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right\} . \end{aligned}$$

One problem with the barrier function approach is that the Hessian matrix of  $B_1(\underline{\theta}, \mathbf{r}'_p)$  or  $B_2(\underline{\theta}, \mathbf{r}'_p)$  becomes increasingly ill-conditioned as  $\mathbf{r}'_p \rightarrow 0$  [GMW (1981)].

## 2. Sequential quadratic programming (SQP) methods

We now consider another approach to problem (4.14). Suppose that  $t^*$  constraints are active at a solution  $\underline{\theta}^*$  to this problem. Let  $c_i^*(\underline{\theta}) \geq 0$  ( $i = 1, \dots, t^*$ ) represent the constraints that are active at  $\underline{\theta}^*$ . Define  $\underline{c}^*(\underline{\theta}) = (c_1^*(\underline{\theta}), \dots, c_{t^*}^*(\underline{\theta}))'$  and take  $A^*(\underline{\theta})$  to be a  $t^* \times a$  matrix whose  $i$ -th row is  $[\frac{\partial c_i^*(\underline{\theta})}{\partial \underline{\theta}}]'$ . Further, assume that  $A^*(\underline{\theta})$  has full row rank. Let  $Z^*(\underline{\theta})$  represent an  $a \times (a - t^*)$  matrix whose columns form a basis for the orthogonal complement of the row space of  $A^*(\underline{\theta})$ .

GMW (1981) showed that necessary conditions for a point  $\underline{\theta}^*$  to be a solution to (4.14) are:

$$(1) \quad c_i(\underline{\theta}^*) \geq 0 \quad (i = 1, \dots, m) ;$$

$$(2) \quad \underline{\phi}(\underline{\theta}^*) = [A^*(\underline{\theta}^*)]' \underline{\lambda}^* \text{ for some } t^* \times 1 \text{ vector } \underline{\lambda}^* = (\lambda_1^*, \dots, \lambda_{t^*}^*)',$$

i.e.,  $\underline{\theta}^*$  is a stationary point of the Lagrangian function

$$LF(\underline{\theta}, \underline{\lambda}) = -L(\underline{\theta}; \underline{y}) - \underline{\lambda}' \underline{c}^*(\underline{\theta}) ,$$

where  $\underline{\lambda} = (\lambda_1, \dots, \lambda_{t^*})'$  ;

$$(3) \quad \lambda_i^* \geq 0 \quad (i = 1, \dots, t^*); \text{ and}$$

$$(4) \quad [Z^*(\underline{\theta}^*)]' W(\underline{\theta}^*, \underline{\lambda}^*) [Z^*(\underline{\theta}^*)] \text{ is a nonnegative definite matrix,}$$

where  $W(\underline{\theta}; \underline{\lambda})$  denotes the Hessian matrix of the Lagrangian function, i.e.,

$$W(\underline{\theta}, \underline{\lambda}) = \Phi(\underline{\theta}) - \sum_{i=1}^{t^*} \lambda_i \left[ \left[ \frac{\partial^2 c_i^*(\underline{\theta})}{\partial \theta_j \partial \theta_k} \right] \right] \quad (j, k = 1, \dots, a) .$$

SQP methods are iterative, with the (p+1)-st iteration consisting of the following four steps:

(1) approximate the problem (4.14) by a problem whose objective function is a quadratic approximation to  $-L(\underline{\theta}; \underline{y})$ , and whose m constraints are linear approximations to  $c_i(\underline{\theta})$  ( $i = 1, \dots, m$ );

(2) apply a method for linearly constrained optimization [discussed in Section IV.A] to the problem in step (1), and determine a search direction  $\underline{\ell}_p$ ;

(3) determine an appropriate stepsize  $\alpha_p$ ; and

(4) put  $\underline{\theta}^{(p+1)} = \underline{\theta}^{(p)} + \alpha_p \underline{\ell}_p$ .

A limit point  $\underline{\theta}^*$  of the sequence  $\underline{\theta}^{(p)}$  is regarded as a solution to problem (4.14).

In nonlinear constrained optimization, it is difficult to determine a maximum nonnegative feasible step  $\bar{\alpha}_p$ . Consequently, if an unconstrained univariate minimization algorithm is used to determine  $\alpha_p$  in step (3), an infeasible iterate may result. A possible solution to this problem is to substitute a barrier function for  $-L(\underline{\theta}; \underline{y})$  in step (3) [GMW (1981)].

It is common to take the quadratic approximation to  $-L(\underline{\theta}; \underline{y})$  in step (1) to be:

$$- \left\{ L(\underline{\theta}^{(p)}; \underline{y}) + [\underline{\phi}(\underline{\theta}^{(p)})]'(\underline{\theta} - \underline{\theta}^{(p)}) + \left(\frac{1}{2}\right)(\underline{\theta} - \underline{\theta}^{(p)})' \mathbf{B}_p (\underline{\theta} - \underline{\theta}^{(p)}) \right\},$$

where

$$B_p = \begin{cases} W(\underline{\theta}^{(p)}, \underline{\lambda}^{(p)}), \text{ where } \underline{\lambda}^{(p)} \text{ is some estimate of } \underline{\lambda}^* \\ \text{[provided } W(\underline{\theta}^{(p)}, \underline{\lambda}^{(p)}) \text{ is a positive definite matrix]} \\ \\ \text{a positive definite approximation to the matrix} \\ W(\underline{\theta}^{(p)}, \underline{\lambda}^{(p)}) \text{ [otherwise].} \end{cases}$$

The linear approximation to  $c_i(\underline{\theta})$  in step (1) is generally taken to be

$$c_i(\underline{\theta}^{(p)}) + \left[ \frac{\partial c_i(\underline{\theta})}{\partial \underline{\theta}} \bigg|_{\underline{\theta} = \underline{\theta}^{(p)}} \right]^{-1} (\underline{\theta} - \underline{\theta}^{(p)}) .$$

## V. EXTENSIONS TO CORRELATED RANDOM EFFECTS

In this chapter, we extend the results of Chapters II-IV to the more general version of model (1.1) obtained by replacing the assumption that  $\text{Var}(\underline{y}) = V$  with the assumption that  $\text{Var}(\underline{y}) = V_A$ , where

$$V_A = \sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i Z_i A_i Z_i'] \text{ and } A_i \text{ (} i = 1, \dots, c \text{) are } q_i \times q_i \text{ known}$$

matrices. We consider the following two special cases separately:

- (1) the matrices  $A_1, \dots, A_c$  are positive definite; and
- (2) the matrices  $A_1, \dots, A_c$  are nonnegative definite.

A. Case 1:  $A_1, \dots, A_c$  Positive Definite1. Derivatives of  $L_1(\underline{\gamma}; \underline{y})$ 

Since the matrix  $A_i$  is positive definite, there exists a  $q_i \times q_i$  nonsingular matrix  $\Gamma_i$  such that  $\Gamma_i' \Gamma_i = A_i$  ( $i = 1, \dots, c$ ). Thus, model (1.1) can be re-expressed as

$$\underline{y} \stackrel{d}{\sim} N_n(X\underline{\alpha}, V_A),$$

where

$$\begin{aligned} V_A &= \sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i Z_i A_i Z_i'] \\ &= \sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i (Z_i \Gamma_i') (Z_i \Gamma_i')'] . \end{aligned}$$

Define

$$Z_i^\# \equiv Z_i \Gamma_i' \quad (i = 1, \dots, c) ,$$

$$Z^\# \equiv (Z_1^\#, \dots, Z_c^\#) = Z \operatorname{diag}\{\Gamma_1', \dots, \Gamma_c'\} ,$$

$$D_A \equiv \sigma_{c+1}^2 \operatorname{diag}\{\gamma_1 A_1, \dots, \gamma_c A_c\} ,$$

and note that

$$V_A = \sigma_{c+1}^2 [I_n + \sum_{i=1}^c \gamma_i Z_i^\# Z_i^{\#'}]$$

$$= R + Z^\# D Z^{\#'} ,$$

$$= R + Z D_A Z' .$$

Consequently, all of our results on REML estimation for model (1.1), that hold for  $\underline{\gamma} \in \Omega_1^*$ , can be extended to case where  $\operatorname{Var}(\underline{y}) = V_A$  and

the set of possible  $\underline{\gamma}$  values is  $\Omega_{1,A}^* = \{\underline{\gamma}: \sigma_{c+1}^2 > 0, \sigma_{c+1}^2 [I_n + \sum_{j=1}^c$

$\gamma_j Z_j A_j Z_j']$  is positive definite ( $i = 1, \dots, c$ ) by simply substituting

$Z^\# = (Z_1^\#, \dots, Z_c^\#)$  for  $Z = (Z_1, \dots, Z_c)$ .

Lemma 5.1: For  $\underline{\gamma} \in \Omega_{1,A}^*$ , the matrices

$$I + Z'SZD_A$$

and

$$I + Z'R^{-1}ZD_A$$

are nonsingular.

Proof: Observe that

$$\begin{aligned} |I + Z^{\#'}SZ^{\#}D| &= |I + Z^{\#'}SZ^{\#} \text{diag}\{(\Gamma'_1)^{-1}, \dots, (\Gamma'_c)^{-1}\} \\ &\quad \cdot D_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}| \\ &= |I + \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\} Z^{\#'}SZ^{\#} \\ &\quad \cdot \text{diag}\{(\Gamma'_1)^{-1}, \dots, (\Gamma'_c)^{-1}\}D_A| \\ &= |I + Z'SZD_A| \end{aligned}$$

and similarly that

$$|I + Z^{\#'}R^{-1}Z^{\#}D| = |I + Z'R^{-1}ZD_A| .$$

Since, according to Lemma 2.3,  $I + Z^{\#'} S Z^{\#} D$  and  $I + Z^{\#'} R^{-1} Z^{\#} D$  are nonsingular for  $\underline{y} \in \Omega_{1,A}^*$ , it follows that  $I + Z' S Z D_A$  and  $I + Z' R^{-1} Z D_A$  are nonsingular.  $\square$

Define

$$V^{\#} \equiv R + Z^{\#} D Z^{\#'},$$

$$\underline{\tilde{\alpha}}^{\#} \equiv \text{any solution to the system of equations } X' V^{\#-1} X \underline{\tilde{\alpha}}^{\#} = X' V^{\#-1} \underline{y},$$

$$\underline{\tilde{\alpha}}_A \equiv \text{any solution to the system of equations } X' V_A^{-1} X \underline{\tilde{\alpha}}_A = X' V_A^{-1} \underline{y},$$

$$T_A^{\#} \equiv (I + Z^{\#'} S Z^{\#} D)^{-1} \equiv [[T_{ij}^{\#}]] \quad (i, j = 1, \dots, c),$$

$$T_A \equiv (I + Z' S Z D_A)^{-1} \equiv [[T_{A,ij}]] \quad (i, j = 1, \dots, c),$$

$$\begin{aligned} \underline{\tilde{v}}^{\#} &\equiv (I + Z^{\#'} R^{-1} Z^{\#} D)^{-1} Z^{\#'} R^{-1} (\underline{y} - X \underline{\tilde{\alpha}}^{\#}) \\ &\equiv (\tilde{v}_1^{\#'}, \dots, \tilde{v}_c^{\#'})', \end{aligned}$$

$$\begin{aligned} \underline{\tilde{v}}_A &\equiv (I + Z' R^{-1} Z D_A)^{-1} Z' R^{-1} (\underline{y} - X \underline{\tilde{\alpha}}_A) \\ &\equiv (\tilde{v}'_{A,1}, \dots, \tilde{v}'_{A,c})', \end{aligned}$$

$$\underline{\tilde{b}}^{\#} \equiv D \underline{\tilde{v}}^{\#} \equiv (\tilde{b}_1^{\#'}, \dots, \tilde{b}_c^{\#'})',$$

$$\underline{\tilde{b}}_A \equiv D \underline{\tilde{v}}_A \equiv (\tilde{b}'_{A,1}, \dots, \tilde{b}'_{A,c})'.$$

The extension of the results of Chapters II-IV to the case where  $\text{Var}(\underline{y}) = \underline{V}_A$  is facilitated by the following two lemmas:

Lemma 5.2: For  $\underline{\gamma} \in \Omega_{1,A}^*$ ,

$$(i) \quad \underline{V}^\# = \underline{V}_A,$$

$$(ii) \quad (\underline{I} + \underline{Z}^\#{}' \underline{R}^{-1} \underline{Z}^\# \underline{D})^{-1} \\ = \text{diag}\{\Gamma_1, \dots, \Gamma_c\} (\underline{I} + \underline{Z}' \underline{R}^{-1} \underline{Z} \underline{D}_A)^{-1} \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\},$$

$$(iii) \quad \underline{V}^{\#-1} = \underline{R}^{-1} - \underline{R}^{-1} \underline{Z} \underline{D}_A (\underline{I} + \underline{Z}' \underline{R}^{-1} \underline{Z} \underline{D}_A)^{-1} \underline{Z}' \underline{R},$$

$$(iv) \quad \underline{X} \tilde{\underline{\alpha}}^\# = \underline{X} \tilde{\underline{\alpha}}_{-A},$$

$$(v) \quad \tilde{\underline{y}}^\# = \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \tilde{\underline{y}}_{-A},$$

$$(vi) \quad \tilde{\underline{b}}^\# = \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} \tilde{\underline{b}}_A,$$

$$(vii) \quad \underline{T}^\# = \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \underline{T}_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}.$$

Proof:

$$(i) \quad \underline{V}^\# = \underline{R} + \underline{Z}^\# \underline{D} \underline{Z}^\#{}'$$

$$= \underline{R} + \underline{Z} \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \underline{D} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \underline{Z}'$$

$$= \underline{R} + \underline{Z} \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\}$$

$$\cdot \underline{D}_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \underline{Z}'$$

$$= R + Z D_A Z'$$

$$= V_A ;$$

$$(ii) \quad (I + Z^{\#'} R^{-1} Z^{\#} D)^{-1}$$

$$= [I + \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' R^{-1} Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\}]$$

$$\cdot \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} D_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}]^{-1}$$

$$= [\text{diag}\{\Gamma_1, \dots, \Gamma_c\} (I + Z' R^{-1} Z D_A) \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}]^{-1}$$

$$= \text{diag}\{\Gamma_1, \dots, \Gamma_c\} (I + Z' R^{-1} Z D_A)^{-1} \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\};$$

$$(iii) \quad V^{\#-1} = R^{-1} - R^{-1} Z^{\#} D (I + Z^{\#'} R^{-1} Z^{\#} D)^{-1} Z^{\#'} R^{-1} \text{ [using (2.10)]}$$

$$= R^{-1} - R^{-1} Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\}$$

$$\cdot D_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\} (I + Z^{\#'} R^{-1} Z^{\#} D)^{-1}$$

$$\cdot \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' R^{-1}$$

$$= R^{-1} - R^{-1} Z D_A (I + Z' R^{-1} Z D_A)^{-1} Z' R^{-1} \text{ [using part (ii)]};$$

(iv) Since  $V^{\#} = V_A$ ,  $\tilde{\alpha}^{\#}$  and  $\tilde{\alpha}_A$  are both solutions to the linear system  $X' V^{\#-1} \tilde{\alpha} = X' V^{\#-1} \underline{y}$ . Since the value of the vector  $X \tilde{\alpha}$  is the same for any solution to this system,  $X \tilde{\alpha}^{\#} = X \tilde{\alpha}_A$  ;

$$(v) \quad \tilde{\underline{y}}^{\#} = (I + Z^{\#'} R^{-1} Z^{\#} D)^{-1} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' R^{-1} (\underline{y} - X \tilde{\underline{\alpha}}^{\#})$$

$$= \text{diag}\{\Gamma_1, \dots, \Gamma_c\} (I + Z' R^{-1} Z D_A)^{-1} Z' R^{-1} (\underline{y} - X \tilde{\underline{\alpha}}_A)$$

[using parts (ii) and (iv)]

$$= \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \tilde{\underline{y}}_A ;$$

$$(vi) \quad \tilde{\underline{b}}^{\#} = D \tilde{\underline{y}}^{\#}$$

$$= \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} D_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\} \tilde{\underline{y}}^{\#}$$

$$= \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} D_A \tilde{\underline{y}}_A \text{ [using part (v)]}$$

$$= \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} \tilde{\underline{b}}_A ;$$

$$(vii) \quad T^{\#} = [I + \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' S Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\}]$$

$$\cdot \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} D_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}]^{-1}$$

$$= [\text{diag}\{\Gamma_1, \dots, \Gamma_c\} (I + Z' S Z D_A) \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}]^{-1}$$

$$= \text{diag}\{\Gamma_1, \dots, \Gamma_c\} (I + Z' S Z D_A)^{-1} \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\}$$

$$= \text{diag}\{\Gamma_1, \dots, \Gamma_c\} T_A \text{diag}\{\Gamma_1^{-1}, \dots, \Gamma_c^{-1}\} \quad \square$$

Lemma 5.3: For  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \text{tr}(T_{ii}^{\#}) = \text{tr}(T_{A,ii}) ,$$

$$(ii) \quad Z^{\#} \tilde{\underline{b}}^{\#} = Z \tilde{\underline{b}}_{-A} ,$$

$$(iii) \quad \tilde{\underline{v}}_{-i}^{\#'} \tilde{\underline{v}}_{-i}^{\#} = \tilde{\underline{v}}_{-A,i}' A_i \tilde{\underline{v}}_{-A,i} = \frac{1}{\sigma_i^4} \tilde{\underline{b}}_{-A,i}' A_i^{-1} \tilde{\underline{b}}_{-A,i} ,$$

$$(iv) \quad \text{tr}(T_{ij}^{\#} T_{ji}^{\#}) = \text{tr}(T_{A,ij} T_{A,ji}) ,$$

$$(v) \quad \tilde{\underline{v}}_{-i}^{\#'} T_{ij-j}^{\#} \tilde{\underline{v}}_{-j}^{\#} = \tilde{\underline{v}}_{-A,i}' A_i T_{A,ij} \tilde{\underline{v}}_{-A,j} ,$$

$$(vi) \quad \text{tr}[(I - T_{ii}^{\#})^2] = \text{tr}[(I - T_{A,ii})^2] ,$$

$$(vii) \quad \tilde{\underline{v}}_{-i}^{\#'} (I - T_{ii}^{\#}) \tilde{\underline{v}}_{-i}^{\#} = \tilde{\underline{v}}_{-A,i}' A_i (I - T_{A,ii}) \tilde{\underline{v}}_{-A,i} ,$$

$$(viii) \quad \text{tr}(T_A^{\#2}) = \text{tr}(T_A^2) .$$

Proof:

$$(i) \quad \text{tr}(T_{ii}^{\#}) = \text{tr}(\Gamma_i' T_{A,ii} \Gamma_i^{-1}) = \text{tr}(T_{A,ii}) ;$$

$$\begin{aligned} (ii) \quad Z^{\#} \tilde{\underline{b}}^{\#} &= Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \text{diag}\{(\Gamma_1')^{-1}, \dots, (\Gamma_c')^{-1}\} \tilde{\underline{b}}_{-A} \\ &= Z \tilde{\underline{b}}_{-A} ; \end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad \tilde{\mathbf{v}}_{-i}^{\#'} \tilde{\mathbf{v}}_{-i}^{\#} &= (\Gamma_{i-A,i} \tilde{\mathbf{v}}_{-A,i})' (\Gamma_{i-A,i} \tilde{\mathbf{v}}_{-A,i}) = \tilde{\mathbf{v}}_{-A,i}' \mathbf{A}_i \tilde{\mathbf{v}}_{-A,i} \\
&= \frac{1}{\sigma_i^2} \tilde{\mathbf{v}}_{-A,i}' \tilde{\mathbf{b}}_{-A,i} \\
&= \frac{1}{\sigma_i^4} \tilde{\mathbf{b}}_{-A,i}' \mathbf{A}_i^{-1} \tilde{\mathbf{b}}_{-A,i} ;
\end{aligned}$$

$$\begin{aligned}
\text{(iv)} \quad \text{tr}(\mathbf{T}_{ij}^{\#} \mathbf{T}_{ji}^{\#}) &= \text{tr}[(\Gamma_i \mathbf{T}_{A,ij} \Gamma_j^{-1}) (\Gamma_j \mathbf{T}_{A,ji} \Gamma_i^{-1})] \\
&= \text{tr}(\mathbf{T}_{A,ij} \mathbf{T}_{A,ji}) ;
\end{aligned}$$

$$\begin{aligned}
\text{(v)} \quad \tilde{\mathbf{v}}_{-i}^{\#'} \mathbf{T}_{ij}^{\#} \tilde{\mathbf{v}}_{-j}^{\#} &= (\Gamma_i \tilde{\mathbf{v}}_{-A,i})' (\Gamma_i \mathbf{T}_{A,ij} \Gamma_j^{-1}) (\Gamma_j \tilde{\mathbf{v}}_{-A,j}) \\
&= \tilde{\mathbf{v}}_{-A,i}' \mathbf{A}_i \mathbf{T}_{A,ij} \tilde{\mathbf{v}}_{-A,j} ;
\end{aligned}$$

$$\begin{aligned}
\text{(vi)} \quad \text{tr}[(\mathbf{I} - \mathbf{T}_{ii}^{\#})^2] &= \text{tr}[(\mathbf{I} - \Gamma_i \mathbf{T}_{A,ii} \Gamma_i^{-1})^2] \\
&= \text{tr}(\mathbf{I} - 2\Gamma_i \mathbf{T}_{A,ii} \Gamma_i^{-1} + \Gamma_i \mathbf{T}_{A,ii}^2 \Gamma_i^{-1}) \\
&= \text{tr}(\mathbf{I} - 2\mathbf{T}_{A,ii} + \mathbf{T}_{A,ii}^2) \\
&= \text{tr}[(\mathbf{I} - \mathbf{T}_{A,ii})^2] ;
\end{aligned}$$

$$\begin{aligned}
\text{(vii)} \quad \tilde{\mathbf{v}}_{-i}^{\#'} (\mathbf{I} - \mathbf{T}_{ii}^{\#}) \tilde{\mathbf{v}}_{-i}^{\#} &= (\Gamma_i \tilde{\mathbf{v}}_{-A,i})' (\mathbf{I} - \Gamma_i \mathbf{T}_{A,ii} \Gamma_i^{-1}) (\Gamma_i \tilde{\mathbf{v}}_{-A,i}) \\
&= \tilde{\mathbf{v}}_{-A,i}' (\mathbf{A}_i - \mathbf{A}_i \mathbf{T}_{A,ii}) \tilde{\mathbf{v}}_{-A,i}
\end{aligned}$$

$$= \tilde{v}_{A,i} A_i (I - T_{A,ii}) \tilde{v}_{A,i} ;$$

$$(viii) \quad \text{tr}(T^{\#2}) = \text{tr}[(\Gamma_i T_{A,ii} \Gamma_i^{-1})^2]$$

$$= \text{tr}(\Gamma_i T_{A,ii}^2 \Gamma_i^{-1})$$

$$= \text{tr}(T_{A,ii}^2) .$$

□

The following lemma extends the results of Lemma 2.7 to the case where  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\gamma} \in \Omega_{1,A}^*$ .

Lemma 5.4: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\gamma} \in \Omega_{1,A}^*$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - Z \tilde{b}_A)] ,$$

$$(ii) \quad \frac{\partial L_1}{\partial \gamma_i} = - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \sigma_{c+1}^2 \tilde{v}_{A,i}' A_i \tilde{v}_{A,i} \right\}$$

$$= - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \left(\frac{1}{\sigma_i^2}\right) \frac{1}{\gamma_i} \tilde{b}_{A,i}' A_i^{-1} \tilde{b}_{A,i} \right\} ,$$

if  $\gamma_i \neq 0$ ,

$$(iii) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \tilde{z}_{-A}^b)],$$

$$(iv) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} = \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i},$$

$$(v) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji})$$

$$- 2 \sigma_{c+1}^2 \left( \frac{1}{\gamma_j} \right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j},$$

$$\text{if } \gamma_i \neq 0, \gamma_j \neq 0,$$

$$(vi) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} = - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2]$$

$$+ 2 \sigma_{c+1}^2 \left( \frac{1}{\gamma_i} \right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i},$$

$$\text{if } \gamma_i \neq 0,$$

$$(vii) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] = \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2], \text{ if } \gamma_i \neq 0,$$

$$(viii) \quad (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right) = \frac{1}{\sigma_{c+1}^2} \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] ,$$

if  $\gamma_i \neq 0$  ,

$$(ix) \quad (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right) = \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) ,$$

if  $\gamma_i \neq 0, \gamma_j \neq 0$  ,

$$(x) \quad (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) = \frac{1}{\sigma_{c+1}^4} (n-p^*)$$

Proof: Substituting  $Z^\#$  for  $Z$  in parts (i)-(x) of Lemma 2.7 and then making use of Lemmas 5.2 and 5.3, we find that

$$(i) \quad \frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - Z^\# \tilde{\underline{b}}^\#)]$$

$$= - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - Z \tilde{\underline{b}}_A)] ;$$

$$(ii) \quad \frac{\partial L_1}{\partial \gamma_i} = - \left( \frac{1}{2} \right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii}^\#)] - \sigma_{c+1}^2 \tilde{\gamma}_i^\# \tilde{\gamma}_i^\# \right\}$$

$$= - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \sigma_{c+1}^2 \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} \right\}$$

and

$$\sigma_{c+1}^2 \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} = \frac{1}{\sigma_i^2} \frac{1}{\gamma_i} \tilde{b}'_{-A,i} A_i^{-1} \tilde{b}_{-A,i} ;$$

$$(iii) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - Z^{\#} \underline{b}^{\#})]$$

$$= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - Z \tilde{\underline{b}}_A)] ;$$

$$(iv) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} = \tilde{v}_{-i}^{\#'} \tilde{v}_{-i}^{\#} = \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} ;$$

$$(v) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij}^{\#} T_{ji}^{\#})$$

$$- 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_j}\right) \tilde{v}_{-i}^{\#'} T_{ij}^{\#} \tilde{v}_{-j}^{\#}$$

$$= - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji})$$

$$- 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_j}\right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} ;$$

$$(vi) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} = - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{ii}^\#)^2]$$

$$+ 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_i}\right) \tilde{v}_{-i}^{\#'} (I - T_{ii}^\#) \tilde{v}_{-i}^\#$$

$$= - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2]$$

$$+ 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_i}\right) \tilde{v}_{-A,i}^{\#'} A_i (I - T_{A,ii}) \tilde{v}_{-A,i}^\# ;$$

$$(vii) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] = \frac{1}{\gamma_i^2} \text{tr}[(I - T_{ii}^\#)^2]$$

$$= \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] ;$$

$$(viii) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right] = \frac{1}{\sigma_{c+1}^2} \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii}^\#)]$$

$$= \frac{1}{\sigma_{c+1}^2} \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] ;$$

$$\begin{aligned}
(\text{ix}) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] &= \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij}^\# T_{ji}^\#) \\
&= \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) ;
\end{aligned}$$

$$(\text{x}) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} (n-p^*) . \quad \square$$

## 2. Derivatives of $L_v(\underline{\sigma}; \underline{y})$

The following lemma extends the results of Section II.C.2 to the case where  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\sigma} \in \{\underline{\sigma} : \sigma_{c+1}^2 > 0, [\sigma_{c+1}^2 I_n + \sum_{j=i}^c \sigma_j^2 Z_{jA} Z_{jA}'] \text{ is positive definite } (i = 1, \dots, c)\}$  [Let  $\Omega_{2,A}^* \equiv \{\underline{\sigma} : \sigma_{c+1}^2 > 0, [\sigma_{c+1}^2 I_n + \sum_{j=i}^c \sigma_j^2 Z_{jA} Z_{jA}'] \text{ is positive definite } (i = 1, \dots, c)\}$ .]

Lemma 5.5: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\sigma} \in \Omega_{2,A}^*$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$\begin{aligned}
(\text{i}) \quad \frac{\partial L_v}{\partial \sigma_{c+1}^2} &= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(T_A)] \\
&\quad + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{x}_{\underline{A}} - \underline{z}_{\underline{A}})' (\underline{y} - \underline{x}_{\underline{A}} - \underline{z}_{\underline{A}}) ,
\end{aligned}$$

$$\begin{aligned}
(ii) \quad \frac{\partial L_v}{\partial \sigma_i^2} &= - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \tilde{v}'_{-A,i} A_{i-A,i} \tilde{v}_{-A,i} \\
&= - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \frac{1}{\sigma_i^4} \tilde{b}'_{-A,i} A_{i-A,i}^{-1} \tilde{b}_{-A,i} ,
\end{aligned}$$

$$\text{if } \sigma_i^2 \neq 0 ,$$

$$\begin{aligned}
(iii) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - Z \tilde{b}_A)] \\
&\quad - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{A,jk} T_{A,jk})] \\
&\quad + 2 \sigma_{j-A,j}^2 \tilde{v}'_j A_j^T T_{A,jk} \tilde{v}_{k-A,k}] \\
&\quad + 2 \left(\frac{1}{\sigma_{c+1}^4}\right) \sum_{j=1}^c \left[\left(\frac{1}{2}\right) q_j - \sigma_{j-A,j}^2 \tilde{v}'_j A_j^T \tilde{v}_{A,j}\right],
\end{aligned}$$

$$\begin{aligned}
(iv) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} &= - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{A,ii}) \\
&\quad + 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \frac{1}{\sigma_i^2} \sum_{j=1}^c \left[\left(\frac{1}{2}\right) \text{tr}(T_{A,ij} T_{A,ji})\right]
\end{aligned}$$

$$+ \sigma_{i-A,i}^2 \tilde{V}'_{-A,i} A_i T_{A,ij} \tilde{V}_{-A,j} ] , \text{ if } \sigma_i^2 \neq 0 ,$$

$$(v) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} = - \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji})$$

$$- 2 \left( \frac{1}{\sigma_j^2} \right) \tilde{V}'_{-A,i} A_i T_{A,ij} \tilde{V}_{-A,j} , \text{ if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0 ,$$

$$(vi) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} = - \frac{1}{\sigma_i^4} \text{tr}[(I - T_{A,ii})^2]$$

$$+ 2 \left( \frac{1}{\sigma_i^2} \right) \tilde{V}'_{-A,i} A_i (I - T_{A,ii}) \tilde{V}_{-A,i} , \text{ if } \sigma_i^2 \neq 0 ,$$

$$(vii) \quad (-2) \mathfrak{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_i^4} \text{tr}[(I - T_{A,ii})^2] , \text{ if } \sigma_i^2 \neq 0 ,$$

$$(viii) \quad (-2) \mathfrak{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [\text{tr}(T_{A,ii}) - \sum_{j=1}^c \text{tr}(T_{A,ij} T_{A,ji})],$$

$$\text{if } \sigma_i^2 \neq 0 ,$$

$$(ix) \quad (-2) \mathfrak{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}) ,$$

$$(x) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_V}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} [n - p^* - q + \text{tr}(T_A^2)] .$$

Proof: Substitution of  $Z^\#$  for  $Z$  in the results of Section II.C.2, and then making use of Lemmas 5.2 and 5.3, yields the above identities. □

### 3. Derivatives of $L_c(\underline{Y}^+; \underline{y})$

Define

$$H^\# \equiv \frac{1}{\sigma_{c+1}^2} V^\# = I_n + \sum_{i=1}^c \gamma_i Z_i^\# Z_i^{\#'} ,$$

$$H_A \equiv \frac{1}{\sigma_{c+1}^2} V_A = I_n + \sum_{i=1}^c \gamma_i Z_i^{A} Z_i' ,$$

$$P^{*\#} \equiv A(A'H^\#A)^{-1}A' ,$$

$$P_A^* \equiv A(A'H_A A)^{-1}A' ,$$

$$\hat{\sigma}_{c+1}^{2\#}(\underline{Y}^+) \equiv \left( \frac{1}{n-p^*} \right) \underline{y}' P^{*\#} \underline{y} ,$$

$$\hat{\sigma}_{c+1,A}^2(\underline{Y}^+) \equiv \left( \frac{1}{n-p^*} \right) \underline{y}' P_A^* \underline{y} .$$

Since  $V^{\#} = V_A$ , we have that

$$H^{\#} = H_A , \quad (5.1)$$

$$P^{*\#} = P_A^* , \quad (5.2)$$

and

$$\hat{\sigma}_{c+1}^{2\#}(\underline{Y}^+) = \hat{\sigma}_{c+1,A}^2(\underline{Y}^+) . \quad (5.3)$$

We now generalize expressions (3.44) and (3.45) to the case where  $\text{Var}(\underline{y}) = V_A$ .

Lemma 5.6: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_c}{\partial \gamma_i} = - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right] \right. \\ \left. \cdot \frac{1}{\gamma_i^2} \tilde{b}_{A,i}^t A_i^{-1} \tilde{b}_{A,i} \right\}, \text{ if } \gamma_i \neq 0 ,$$

$$(ii) \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \right.$$

$$\begin{aligned}
& - 2 \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right] \frac{1}{\gamma_j^2} \frac{1}{\gamma_i} \tilde{b}'_{-A,i} T_{A,ij} A_{-A,j}^{-1} \tilde{b}_{-A,j} \\
& - \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right]^2 \left( \frac{1}{n-p^*} \right) \frac{1}{\gamma_i^2} \frac{1}{\gamma_j^2} (\tilde{b}'_{-A,i} A_{-A,i}^{-1} \tilde{b}_{-A,i}) \\
& \quad \cdot (\tilde{b}'_{-A,j} A_{-A,j}^{-1} \tilde{b}_{-A,j}) \left. \right\} , \text{ if } \gamma_i \neq 0 , \gamma_j \neq 0 , \\
\\
\text{(iii)} \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i} &= - \left( \frac{1}{2} \right) \left\{ - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \right. \\
& + 2 \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right] \frac{1}{\gamma_i^3} \tilde{b}'_{-A,i} (I - T_{A,ii}) A_{-A,i}^{-1} \tilde{b}_{-A,i} \\
& \quad \left. - \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right]^2 \left( \frac{1}{n-p^*} \right) \frac{1}{\gamma_i^4} (\tilde{b}'_{-A,i} A_{-A,i}^{-1} \tilde{b}_{-A,i})^2 \right\} , \\
& \quad \text{if } \gamma_i \neq 0 .
\end{aligned}$$

Proof: Substituting  $Z^\#$  for  $Z$  in the expressions for  $\frac{\partial L_c}{\partial \gamma_i}$ ,  $\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j}$  and  $\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i}$  given on pages 116-117, and making use of identities (5.1)-(5.3) as well as Lemmas 5.2 and 5.3, we find that

$$\begin{aligned}
(i) \quad \frac{\partial L_c}{\partial \gamma_i} &= - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{ii}^\#)] \right. \\
&\quad \left. - \left[ \frac{1}{\hat{\sigma}_{c+1}^2(Y^+)} \right] \frac{1}{\gamma_i^2} \tilde{b}_i^{\#'} \tilde{b}_i^\# \right\} \\
&= - \left(\frac{1}{2}\right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] \right. \\
&\quad \left. - \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(Y^+)} \right] \frac{1}{\gamma_i^2} \tilde{b}_{A,i}^{\#'} A_i^{-1} \tilde{b}_{A,i} \right\} \\
&\quad [\text{since } \tilde{b}_i^{\#'} \tilde{b}_i^\# = [(\Gamma_i')^{-1} \tilde{b}_{A,i}]' [(\Gamma_i')^{-1} \tilde{b}_{A,i}] \\
&\quad = \tilde{b}_{A,i}^{\#'} A_i^{-1} \tilde{b}_{A,i}] ;
\end{aligned}$$

$$\begin{aligned}
(ii) \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} &= - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{ij}^\# T_{ji}^\#) \right. \\
&\quad \left. - 2 \left[ \frac{1}{\hat{\sigma}_{c+1}^2(Y^+)} \right] \frac{1}{\gamma_j^2} \frac{1}{\gamma_i} \tilde{b}_i^{\#'} T_{ij}^\# \tilde{b}_j^\# \right. \\
&\quad \left. - \left[ \frac{1}{\hat{\sigma}_{c+1}^2(Y^+)} \right]^2 \left(\frac{1}{n-p^*}\right) \frac{1}{\gamma_i^2} \frac{1}{\gamma_j^2} (\tilde{b}_i^{\#'} \tilde{b}_j^\#) (\tilde{b}_j^{\#'} \tilde{b}_i^\#) \right\}
\end{aligned}$$

$$\begin{aligned}
&= - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \operatorname{tr}(T_{A,ij} T_{A,ji}) \right. \\
&\quad - 2 \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(Y^+)} \right) \frac{1}{\gamma_j^2} \frac{1}{\gamma_i} \tilde{b}_{A,i}' T_{A,ij} A_j^{-1} \tilde{b}_{A,j} \\
&\quad - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(Y^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \frac{1}{\gamma_i^2} \frac{1}{\gamma_j^2} \\
&\quad \cdot (\tilde{b}_{A,i}' A_i^{-1} \tilde{b}_{A,i}) (\tilde{b}_{A,j}' A_j^{-1} \tilde{b}_{A,j}) \left. \right\}
\end{aligned}$$

$$\begin{aligned}
&[\text{since } \tilde{b}_i^{\#'} T_{ij}^{\#} \tilde{b}_j^{\#} \\
&= [(\Gamma_i')^{-1} \tilde{b}_{A,i}]' (\Gamma_i T_{A,ij} A_j^{-1}) [(\Gamma_j')^{-1} \tilde{b}_{A,j}] \\
&= \tilde{b}_{A,i}' T_{A,ij} A_j^{-1} \tilde{b}_{A,j}
\end{aligned}$$

$$\text{and } \tilde{b}_i^{\#'} \tilde{b}_i^{\#} = \tilde{b}_{A,i}' A_i^{-1} \tilde{b}_{A,i} ;$$

$$\begin{aligned}
\text{(iii)} \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i} &= - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i^2} \operatorname{tr}[(I - T_{ii}^{\#})^2] \right. \\
&\quad + 2 \left( \frac{1}{\hat{\sigma}_{c+1}^{\#}(Y^+)} \right) \frac{1}{\gamma_i^3} \tilde{b}_i^{\#'} (I - T_{ii}^{\#}) \tilde{b}_i^{\#}
\end{aligned}$$

$$\begin{aligned}
& - \left\{ \frac{1}{\hat{\sigma}_{c+1}^2(\underline{Y}^+)} \right\}^2 \left( \frac{1}{n-p^*} \right) \frac{1}{\gamma_i^4} (\tilde{\underline{b}}_i^{\#'} \tilde{\underline{b}}_i^{\#})^2 \Big\} \\
& = - \left( \frac{1}{2} \right) \left\{ - \frac{1}{\gamma_i^2} \text{tr}[(\mathbf{I} - \mathbf{T}_{A,ii})^2] \right. \\
& \quad + 2 \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \frac{1}{\gamma_i^3} \tilde{\underline{b}}_{A,i} (\mathbf{I} - \mathbf{T}_{A,ii}) \mathbf{A}_i^{-1} \tilde{\underline{b}}_{A,i} \\
& \quad \left. - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \frac{1}{\gamma_i^4} (\tilde{\underline{b}}_{A,i}^{\#'} \mathbf{A}_i^{-1} \tilde{\underline{b}}_{A,i}^{\#}) \right\} \\
& \quad [\text{since } \tilde{\underline{b}}_i^{\#'} (\mathbf{I} - \mathbf{T}_{ii}^{\#}) \tilde{\underline{b}}_i^{\#} \\
& = [(\Gamma_i')^{-1} \tilde{\underline{b}}_{A,i}]' (\mathbf{I} - \Gamma_i \mathbf{T}_{A,ii} \Gamma_i^{-1}) [(\Gamma_i')^{-1} \tilde{\underline{b}}_{A,i}] \\
& = \tilde{\underline{b}}_{A,i}^{\#'} (\mathbf{A}_i^{-1} - \mathbf{T}_{A,ii} \mathbf{A}_i^{-1}) \tilde{\underline{b}}_{A,i} \\
& = \tilde{\underline{b}}_{A,i}^{\#'} (\mathbf{I} - \mathbf{T}_{A,ii}) \mathbf{A}_i^{-1} \tilde{\underline{b}}_{A,i} \\
& \text{and } \tilde{\underline{b}}_i^{\#'} \tilde{\underline{b}}_i^{\#} = \tilde{\underline{b}}_{A,i}^{\#'} \mathbf{A}_i^{-1} \tilde{\underline{b}}_{A,i}^{\#} .
\end{aligned}$$

□

#### 4. Modification of algorithms

In this section, we extend the fourteen algorithms summarized in Table 3.1 to the case where  $\text{Var}(\underline{y}) = V_A$ .

a. The Newton-Raphson and Method of Scoring algorithms To extend algorithms 1 through 9 in Table 3.1 (the Newton-Raphson and Method of Scoring algorithms) to the case where  $\text{Var}(\underline{y}) = V_A$ , we need only substitute the expressions for the first-order partial derivatives and for the second-order partial derivatives (or their expected values) given Sections V.A.1-V.A.3 for those given in Chapter II. [The relevant derivatives for algorithms 4 and 6 are not given explicitly in Sections V.B.1-V.B.3, however Section III.B.3.b indicates how they can be obtained, via the chain rule, from the derivatives of  $L_c(\underline{Y}^+; \underline{y})$ .]

b. EM algorithm In this section, we extend the results of Section III.A to the more general version of model (1.2) obtained by replacing the assumption that  $\text{Var}(\underline{b}) = D$  with the assumption that  $\text{Var}(\underline{b}) = D_A$ . Under this new assumption, the model equation for model (1.2) can be re-expressed as

$$\underline{y} = X\underline{\alpha} + Z^{\#}\underline{b}^{\#} + \underline{e} = X\underline{\alpha} + \sum_{i=1}^c Z_{i-}^{\#} \underline{b}_{i-}^{\#} + \underline{e} ,$$

where

$$\underline{b}_i^{\#} = (\Gamma_i')^{-1} \underline{b}_i \quad (i = 1, \dots, c),$$

$$\underline{b}^{\#} = (\underline{b}_1^{\#}, \dots, \underline{b}_c^{\#})' \stackrel{d}{\sim} N_q(\underline{0}, D),$$

$$\underline{e} \stackrel{d}{\sim} N_n(\underline{0}, R),$$

$$\underline{y} \stackrel{d}{\sim} N_n(X\underline{\alpha}, V^{\#}) \text{ or, equivalently, } \underline{y} \stackrel{d}{\sim} N_n(X\underline{\alpha}, V_A).$$

Consequently, all of our results on REML estimation for model (1.2) can be extended to the case where  $\text{Var}(\underline{b}) = D_A$  simply by substituting  $Z^* = (Z_1^*, \dots, Z_c^*)$  for  $Z = (Z_1, \dots, Z_c)$ .

Suppose that  $\underline{b}$  and  $\underline{e}$  or, equivalently,  $\underline{b}^{\#}$  and  $\underline{e}$ , are regarded as the complete data in the generalized version of model (1.2). Then, substituting  $Z^{\#}$  for  $Z$  in algorithm 10 of Table 3.1, we find that the components of the  $(p+1)$ -st iterate of the EM algorithm are given by

$$\sigma_i^{2(p+1)} = \frac{\frac{\tilde{\underline{b}}_i^{\#(p)'} \tilde{\underline{b}}_i^{\#(p)}}{\underline{b}_i} + \sigma_i^{2(p)} \text{tr}[T_{ii}^{\#(p)}]}{q_i} \quad (i = 1, \dots, c),$$

$$\sigma_{c+1}^{2(p+1)} = \left( \frac{1}{n} \right) \left\{ (\underline{y} - X\tilde{\underline{\alpha}}^{\#(p)} - Z^{\#} \tilde{\underline{b}}^{\#(p)})' (\underline{y} - X\tilde{\underline{\alpha}}^{\#(p)} - Z^{\#} \tilde{\underline{b}}^{\#(p)}) \right. \\ \left. + \sigma_{c+1}^{2(p)} [p^* + q - \text{tr}(T^{\#(p)})] \right\},$$

where  $\underline{\sigma}^{(p)} = (\sigma_1^{2(p)}, \dots, \sigma_{c+1}^{2(p)})'$  represents the  $p$ -th iterate of the

algorithm, and where a superscript (p) indicates that a scalar or vector that is functionally dependent on  $\underline{g}$  is to be evaluated at  $\underline{g} = \underline{g}^{(p)}$ .

Since

$$\tilde{\underline{b}}_{-i}^{\#'} \tilde{\underline{b}}_{-i}^{\#} = \tilde{\underline{b}}_{-A,i}^{\#'} \tilde{\underline{b}}_{-A,i}^{\#} \tilde{\underline{b}}_{-A,i}^{-1} \tilde{\underline{b}}_{-A,i}^{\#} ,$$

$$\text{tr}(\underline{T}_{ii}^{\#}) = \text{tr}(\underline{T}_{A,ii}) \quad (i = 1, \dots, c) ,$$

and

$$\underline{Z}^{\#'} \tilde{\underline{b}}^{\#} = \tilde{\underline{Z}}_{-A}^{\#} \tilde{\underline{b}}^{\#} ,$$

we obtain the following alternative expressions for the components of the (p+1)-st iterate of the EM algorithm:

$$\sigma_i^{2(p+1)} = \frac{\tilde{\underline{b}}_{-A,i}^{(p)'} \tilde{\underline{b}}_{-A,i}^{-1} \tilde{\underline{b}}_{-A,i}^{(p)} + \sigma_i^{2(p)} \text{tr}(\underline{T}_{A,ii}^{(p)})}{q_i} \quad (i = 1, \dots, c) , \quad (5.4)$$

$$\sigma_{c+1}^{2(p+1)} = \frac{(\underline{y} - \underline{X}\tilde{\underline{\alpha}}_A^{(p)} - \tilde{\underline{Z}}_{-A} \tilde{\underline{b}}_{-A}^{(p)})' (\underline{y} - \underline{X}\tilde{\underline{\alpha}}_A^{(p)} - \tilde{\underline{Z}}_{-A} \tilde{\underline{b}}_{-A}^{(p)}) + \sigma_{c+1}^{2(p)} [p+q - \text{tr}(\underline{T}_A^{(p)})]}{n} .$$

If  $\underline{b}$  and  $A'\underline{e}$  or, equivalently,  $\underline{b}^{\#}$  and  $A'\underline{e}$  (rather than  $\underline{b}^{\#}$  and  $\underline{e}$ ), are regarded as the complete data for model (1.2) in the case  $\text{Var}(\underline{b}) = D_A$ , then the first c components of the (p+1)-st iterate of the EM algorithm are given by expression (5.4) but the (c+1)-st component is given by

$$\begin{aligned}
\sigma_{c+1}^2(p+1) &= \left(\frac{1}{n-p^*}\right) \left\{ (\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#(p)} - \underline{Z}^{\# \sim \#(p)} \underline{\tilde{b}}_{\underline{A}})^{\prime} (\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#(p)} - \underline{Z}^{\# \sim \#(p)} \underline{\tilde{b}}_{\underline{A}}) \right. \\
&\quad \left. + \sigma_{c+1}^2(p) [q - \text{tr}(\underline{T}^{\#(p)})] \right\} \\
&= \frac{(\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#(p)} - \underline{Z}^{\# \sim \#(p)} \underline{\tilde{b}}_{\underline{A}})^{\prime} (\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#(p)} - \underline{Z}^{\# \sim \#(p)} \underline{\tilde{b}}_{\underline{A}}) + \sigma_{c+1}^2(p) [q - \text{tr}(\underline{T}_{\underline{A}}^{\#(p)})]}{n-p^*} .
\end{aligned}$$

c. Method of Successive Approximations      Substituting  $\underline{Z}^{\#}$  for  $\underline{Z}$  in the expressions given in Section III.D.1 for  $h_i(\underline{\sigma})$ ,  $h_{c+1}^*(\underline{\sigma})$ , and  $\epsilon_i(\underline{\sigma})$  ( $i = 1, \dots, c$ ), we find that

$$h_i(\underline{\sigma}) = \frac{\underline{\tilde{b}}_{\underline{A},i}^{\prime} \underline{\tilde{b}}_{\underline{A},i}^{\sim \#}}{q_i - \text{tr}(\underline{T}_{ii}^{\#})} = \frac{\underline{\tilde{b}}_{\underline{A},i}^{\prime} \underline{A}_{\underline{A},i}^{-1} \underline{\tilde{b}}_{\underline{A},i}^{\sim}}{q_i - \text{tr}(\underline{T}_{\underline{A},ii})} ,$$

$$h_{c+1}^*(\underline{\sigma}) = \frac{\underline{y}^{\prime} (\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#} - \underline{Z}^{\# \sim \#} \underline{\tilde{b}}_{\underline{A}})}{n-p^*} = \frac{\underline{y}^{\prime} (\underline{y} - \underline{\tilde{x}}_{\underline{A}}^{\#} - \underline{Z}^{\# \sim} \underline{\tilde{b}}_{\underline{A}})}{n-p^*} ,$$

$$\epsilon_i(\underline{\sigma}) = \frac{\underline{\tilde{b}}_{\underline{A},i}^{\prime} \underline{\tilde{b}}_{\underline{A},i}^{\sim \#} + \sigma_i^2 \text{tr}(\underline{T}_{ii}^{\#})}{q_i} = \frac{\underline{\tilde{b}}_{\underline{A},i}^{\prime} \underline{A}_{\underline{A},i}^{-1} \underline{\tilde{b}}_{\underline{A},i}^{\sim} + \sigma_i^2 \text{tr}(\underline{T}_{\underline{A},ii})}{q_i} .$$

To extend algorithms 12 through 14 of Table 3.1 (three versions of the Method of Successive Approximations) to the case  $\text{Var}(\underline{b}) = \underline{D}_{\underline{A}}$ , we substitute these expressions for  $h_i(\underline{\sigma})$ ,  $h_{c+1}^*(\underline{\sigma})$ , and  $\epsilon_i(\underline{\sigma})$  ( $i = 1, \dots, c$ ) for those given in Section III.D.1.

B. Case 2:  $A_1, \dots, A_c$  Nonnegative Definite

Let us now relax the assumption, imposed in Section V.A, that the matrices  $A_1, \dots, A_c$  are positive definite. When  $A_1$  is nonnegative definite, it is still the case that there exists a  $q_1 \times q_1$  matrix  $\Gamma_1$  such that  $\Gamma_1' \Gamma_1 = A_1$ , however, since  $\text{rank}(\Gamma_1) = \text{rank}(A_1)$ ,  $\Gamma_1$  is nonsingular only if  $A_1$  is positive definite.

Define

$$D_{A,i} \equiv \sigma_{c+1}^2 \text{diag}\{\emptyset, \dots, \emptyset, \gamma_i A_i, \dots, \gamma_c A_c\} \quad (i = 1, \dots, c),$$

$$D_A \equiv D_{A,1},$$

$$V_{A,i} \equiv \sigma_{c+1}^2 \left[ I_n + \sum_{j=1}^c \gamma_j Z_j A_j Z_j' \right] \quad (i = 1, \dots, c),$$

so that

$$V_{A,i} = R + Z D_{A,i} Z' \quad (i = 1, \dots, c),$$

$$V_A = V_{A,1} = R + Z D_A Z'.$$

Let  $\Omega_{1,A}^* = \{\underline{\gamma} : \sigma_{c+1}^2 > 0, V_{A,i} \text{ is a positive definite matrix } (i = 1, \dots, c)\}$  represent the parameter space for  $\underline{\gamma}$  associated with the more general version of model (1.1) that is obtained by replacing the assumption  $\text{Var}(\underline{y}) = V$  by the assumption that  $\text{Var}(\underline{y}) = V_A$ .

Lemma 5.7: For  $\underline{y} \in \Omega_{1,A}^*$ , the matrices

$$I + Z'SZD_A$$

and

$$I + Z'R^{-1}ZD_A$$

are nonsingular.

Proof: Substitution of  $\Omega_{1,A}^*$ ,  $V_{A,i}$ , and  $D_{A,i}$  for  $\Omega_1^*$ ,  $V_i$ , and  $D_i$ , respectively, in the proof of Lemma 2.3 yields the above results.  $\square$

For  $\underline{y} \in \Omega_{1,A}^*$ , let us adopt the following additional notation:

$$Z_i^\# \equiv Z_i \Gamma_i' \quad (i = 1, \dots, c),$$

$$Z^\# \equiv (Z_1^\#, \dots, Z_c^\#),$$

$$T_A \equiv (I + Z'SZD_A)^{-1} \equiv [[T_{A,ij}]] \quad (i, j = 1, \dots, c),$$

$$\tilde{\alpha}_{-A} \equiv \text{any solution to the system of equations } X'V_A^{-1}X\tilde{\alpha}_{-A} = X'V_A^{-1}\underline{y},$$

$$\tilde{\nu}_{-A} \equiv (I + Z'R^{-1}ZD_A)^{-1}Z'R^{-1}(\underline{y} - X\tilde{\alpha}_{-A})$$

$$\equiv (\tilde{\nu}_{-A,1}', \dots, \tilde{\nu}_{-A,c}')',$$

$$\tilde{\underline{b}}_{-A} \equiv D_{A-A} \tilde{\underline{y}} \equiv (\tilde{b}'_{-A,1}, \dots, \tilde{b}'_{-A,c})',$$

$$P_A \equiv V_A^{-1} - V_A^{-1} X (X' V_A^{-1} X)^{-1} X' V_A^{-1}.$$

It should be pointed out that, for  $\underline{\gamma} \in \Omega_{1,A}^*$ , the results of Section II.A remain valid if  $D_A$  is substituted for  $D$  (as is evident from the proofs of these results).

#### 1. Derivatives of $L_1(\underline{\gamma}; \underline{y})$

Lemma 5.8 extends the expressions given in Lemma 2.7 for the first- and second-order partial derivatives of  $L_1(\underline{\gamma}; \underline{y})$  and for the expected values of the second-order partial derivatives to the case where  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\gamma} \in \Omega_{1,A}^*$  in model (1.1). Note that these extensions are essentially the same as those obtained in Section V.A under the assumption that the matrices  $A_1, \dots, A_c$  are all positive definite.

Lemma 5.8: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\gamma} \in \Omega_{1,A}^*$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - Z \tilde{\underline{b}}_{-A})],$$

$$(ii) \quad \frac{\partial L_1}{\partial \gamma_i} = - \left(\frac{1}{2}\right) \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})]$$

$$+ \left(\frac{1}{2}\right) \sigma_{c+1}^2 \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} , \text{ if } \gamma_i \neq 0 ,$$

$$(iii) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \tilde{z}_{-A}^b)] ,$$

$$(iv) \quad (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} = \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} ,$$

$$(v) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji})$$

$$- 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_j}\right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} ,$$

$$\text{if } \gamma_i \neq 0, \gamma_j \neq 0 ,$$

$$(vi) \quad (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} = - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2]$$

$$+ 2 \sigma_{c+1}^2 \left(\frac{1}{\gamma_i}\right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} ,$$

$$\text{if } \gamma_i \neq 0 ,$$

$$(vii) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] = \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] , \text{ if } \gamma_i \neq 0 ,$$

$$(viii) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right] = \frac{1}{\gamma_i} \frac{1}{\sigma_{c+1}^2} [q_i - \text{tr}(T_{A,ii})] , \text{ if } \gamma_i \neq 0 ,$$

$$(ix) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] = \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) , \text{ if } \gamma_i \neq 0, \gamma_j \neq 0 ,$$

$$(x) \quad (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} (n-p^*) .$$

Proof:

(i) Applying standard results on matrix differentiation, we find [following Harville (1977, p. 326)] that

$$\begin{aligned} \frac{\partial L_1}{\partial \gamma_i} = & - \left( \frac{1}{2} \right) \text{tr} \left( P_A \frac{\partial V_A}{\partial \gamma_i} \right) \\ & + \left( \frac{1}{2} \right) (\underline{y} - \underline{X} \tilde{\underline{\alpha}}_A)' V_A^{-1} \left( \frac{\partial V_A}{\partial \gamma_i} \right) V_A^{-1} (\underline{y} - \underline{X} \tilde{\underline{\alpha}}_A) \quad (i = 1, \dots, c+1). \end{aligned} \quad (5.5)$$

Moreover,

$$\frac{\partial V_A}{\partial \gamma_i} = \begin{cases} \sigma_{c+1}^2 Z_i A_i Z_i' & , \text{ if } i = 1, \dots, c \\ I_n + \sum_{i=1}^c \gamma_i Z_i A_i Z_i' = \frac{1}{\sigma_{c+1}^2} V_A & , \text{ if } i = c+1 . \end{cases} \quad (5.6)$$

Thus,

$$\frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \text{tr}(P_A V_A) + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} (\underline{y} - \underline{X\tilde{\alpha}}_A)' V_A^{-1} (\underline{y} - \underline{X\tilde{\alpha}}_A) .$$

we find that

$$\begin{aligned} \text{tr}(P_A V_A) &= \text{tr}[I - V_A^{-1} X(X' V_A^{-1} X)^{-1} X'] \text{ [using (2.2)]} \\ &= n - \text{tr}[V_A^{-1} X(X' V_A^{-1} X)^{-1} X'] \\ &= n - \text{tr}[(X' V_A^{-1} X)^{-1} X' X_A^{-1} X] \\ &= n - \text{rank}(X' V_A^{-1} X) \\ &= n - \text{rank}(X) \\ &= n - p^* \end{aligned} \tag{5.7}$$

and that

$$(\underline{y} - \underline{X\tilde{\alpha}}_A)' V_A^{-1} (\underline{y} - \underline{X\tilde{\alpha}}_A) = \underline{y}' S (\underline{y} - \underline{Z\tilde{b}}_A) \text{ [using (2.13)]} .$$

We conclude that

$$\frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S (\underline{y} - \underline{Z\tilde{b}}_A)] . \tag{5.8}$$

(ii) It follows from results (5.5) and (5.6) that, for  
 $i = 1, \dots, c$ ,

$$\begin{aligned}
 \frac{\partial L_1}{\partial \gamma_i} &= - \left(\frac{1}{2}\right) \sigma_{c+1}^2 \operatorname{tr}(P_A Z_i A_i Z_i') \\
 &\quad + \left(\frac{1}{2}\right) \sigma_{c+1}^2 (\underline{y} - \underline{x}_{-A})' V_A^{-1} Z_i A_i Z_i' V_A^{-1} (\underline{y} - \underline{x}_{-A}) \\
 &= - \left(\frac{1}{2}\right) \sigma_{c+1}^2 \operatorname{tr}(P_A Z_i A_i Z_i') \\
 &\quad + \left(\frac{1}{2}\right) \sigma_{c+1}^2 \tilde{y}_{-A,i}' A_{i-A,i} \tilde{y}_{-A,i} .
 \end{aligned} \tag{5.9}$$

Recalling the notation

$$\underline{\Delta}_i \equiv \operatorname{diag}\{\emptyset_{q_1 \times q_1}, \dots, \emptyset_{q_{i-1} \times q_{i-1}}, I_{q_i}, \emptyset_{q_{i+1} \times q_{i+1}}, \dots, \emptyset_{q_c \times q_c}\},$$

we find that

$$\begin{aligned}
 \operatorname{tr}(P_A Z_i A_i Z_i') &= \operatorname{tr}[P_A (Z_i \Gamma_i') (\Gamma_i Z_i')] \\
 &= \operatorname{tr}(\Gamma_i Z_i' P_A Z_i \Gamma_i') \\
 &= \operatorname{tr}(\underline{\Delta}_i Z_i^{\#'} P_A Z_i^{\#} \underline{\Delta}_i)
 \end{aligned}$$

$$\begin{aligned}
&= \text{tr}[\underline{\Delta}_i \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' P_A Z \\
&\quad \cdot \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} \underline{\Delta}_i] \\
&= \text{tr}[\underline{\Delta}_i \text{diag}\{\Gamma_1, \dots, \Gamma_c\} T_A Z' S Z \\
&\quad \cdot \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} \underline{\Delta}_i] \quad [\text{using (2.27) and part (ii)} \\
&\quad \text{of Lemma 2.4}] \\
&= \text{tr}[\Gamma_i (\sum_{j=1}^c T_{A,ij} Z'_j S Z_i) \Gamma'_i] \\
&= \text{tr}(\sum_{j=1}^c T_{A,ij} Z'_j S Z_i A_i) . \tag{5.10}
\end{aligned}$$

Since, for  $i, k \in \{1, \dots, c\}$  with  $i \neq k$ ,

$$T_{A,ii} + \sigma_i^2 \sum_{j=1}^c T_{A,ij} Z'_j S Z_i A_i = I_{q_i}$$

$$T_{A,ik} + \sigma_k^2 \sum_{j=1}^c T_{A,ij} Z'_j S Z_k A_k = \emptyset ,$$

we have that, for  $i, k \in \{1, \dots, c\}$  with  $i \neq k$ ,

$$\left. \begin{aligned} \sum_{j=1}^c T_{A,kj} Z_j' S Z_k A_k &= \frac{1}{\sigma_k^2} (I_{q_k} - T_{A,kk}) \\ \sum_{j=1}^c T_{A,ij} Z_j' S Z_k A_k &= -\frac{1}{\sigma_k^2} T_{A,ik} . \end{aligned} \right\} \quad (5.11)$$

Together, results (5.10) and (5.11) imply that

$$\begin{aligned} \text{tr}(P_{Z_i A_i Z_i'}) &= \frac{1}{\sigma_i^2} \text{tr}(I_{q_i} - T_{A,ii}) \\ &= \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] . \end{aligned} \quad (5.12)$$

Substituting from expression (5.12) into expression (5.9), we obtain

$$\frac{\partial L_1}{\partial \gamma_i} = - \left(\frac{1}{2}\right) \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \sigma_{c+1-A,i}^2 \tilde{y}'_{c+1-A,i} \tilde{y}_{c+1-A,i} \quad (i = 1, \dots, c).$$

(iii) Result (5.8), together with part (iv) of Lemma 2.4, implies that

$$\frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' P_{A^c} \underline{y}] . \quad (5.13)$$

Thus,

$$\begin{aligned}
\frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (n-p^*) - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} \underline{y}' P_A \underline{y} \\
&+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \underline{y}' \left( \frac{\partial P_A}{\partial \sigma_{c+1}^2} \right) \underline{y} .
\end{aligned} \tag{5.14}$$

Moreover, for  $i = 1, \dots, c+1$ ,

$$\begin{aligned}
\frac{\partial P_A}{\partial \gamma_i} &= \frac{\partial [A(A'V_A A)^{-1}A']}{\partial \gamma_i} \text{ [using part (i) of Lemma 2.1]} \\
&= A \frac{\partial [(A'V_A A)^{-1}]}{\partial \gamma_i} A' \\
&= A[-(A'V_A A)^{-1} (A' \frac{\partial V_A}{\partial \gamma_i} A) (A'V_A A)^{-1}] A' \\
&= -P_A \frac{\partial V_A}{\partial \gamma_i} P_A \text{ [again, using part (i) of Lemma 2.1]}.
\end{aligned} \tag{5.15}$$

In particular,

$$\frac{\partial P_A}{\partial \sigma_{c+1}^2} = -\frac{1}{\sigma_{c+1}^2} P_A V_A P_A$$

$$= - \frac{1}{\sigma_{c+1}^2} \text{ [using part (iv) of Lemma 2.1] } .$$

Substituting this expression into expression (5.14), we obtain

$$(-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{y}' P_A \underline{y}] \quad (5.16)$$

$$= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{y}' S(\underline{y} - \tilde{Z} \tilde{b}_A)]$$

[using part (iv) of Lemma 2.4].

(iv) Starting with expression (5.13), we find that, for  
 $i = 1, \dots, c$ ,

$$\begin{aligned} (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} &= - \frac{1}{\sigma_{c+1}^2} \underline{y}' \frac{\partial P_A}{\partial \gamma_i} \underline{y} \\ &= \frac{1}{\sigma_{c+1}^2} \underline{y}' P_A \frac{\partial V_A}{\partial \gamma_i} P_A \underline{y} \quad \text{[using (5.15)]} \\ &= \underline{y}' P_A Z_i A_i Z_i' P_A \underline{y} \\ &= \tilde{V}_{-A,i}' A_i \tilde{V}_{-A,i} . \end{aligned} \quad (5.17)$$

(v) Result (5.9) can be re-expressed as

$$\begin{aligned} \frac{\partial L_1}{\partial \gamma_i} = & - \left(\frac{1}{2}\right) \sigma_{c+1}^2 \operatorname{tr}(Z_i A_i Z_i' P_A) \\ & + \left(\frac{1}{2}\right) \sigma_{c+1}^2 \underline{y}' P_A Z_i A_i Z_i' P_A \underline{y} \end{aligned}$$

Therefore, for  $i, j \in \{1, \dots, c\}$ ,

$$\begin{aligned} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = & - \left(\frac{1}{2}\right) \sigma_{c+1}^2 \operatorname{tr}[Z_i A_i Z_i' (-P_A \frac{\partial V_A}{\partial \gamma_j} P_A)] \\ & + \left(\frac{1}{2}\right) \sigma_{c+1}^2 [\underline{y}' (-P_A \frac{\partial V_A}{\partial \gamma_j} P_A) Z_i A_i Z_i' P_A \underline{y}] \\ & + \underline{y}' P_A Z_i A_i Z_i' (-P_A \frac{\partial V_A}{\partial \gamma_j} P_A) \underline{y} \text{ [using (5.15)]}. \end{aligned} \quad (5.18)$$

If  $i \neq j$ , then

$$\begin{aligned} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} = & \left(\frac{1}{2}\right) \sigma_{c+1}^4 [\operatorname{tr}(Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A) \\ & - 2(\underline{y}' P_A Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A \underline{y})] \text{ [using (5.6)]}. \end{aligned} \quad (5.19)$$

Moreover,

$$\begin{aligned}
& \text{tr}(Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A) \\
&= \text{tr}[P_A (Z_i \Gamma_i' (\Gamma_i Z_i') P_A (Z_j \Gamma_j') (\Gamma_j Z_j'))] \\
&= \text{tr}(Z_i^{\#'} P_A Z_j^{\#} Z_j^{\#'} P_A Z_i^{\#}) \\
&= \text{tr}(\Delta_{-i} Z_i^{\#'} P_A Z_j^{\#} Z_j^{\#'} P_A Z_{-i}^{\#}) \\
&= \text{tr}(Z_j^{\#'} P_A Z_{-i}^{\#} \Delta_{-i} Z_i^{\#'} P_A Z_j^{\#}) \\
&= \text{tr}(\Delta_{-j} Z_j^{\#'} P_A Z_{-i}^{\#} \Delta_{-i} Z_i^{\#'} P_A Z_{-j}^{\#}) \\
&= \text{tr}[\Delta_{-j} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' P_A Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \\
&\quad \cdot \Delta_{-i} \Delta_{-i} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' P_A Z \\
&\quad \cdot \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \Delta_{-j}] \\
&= \text{tr} \left[ \begin{aligned} & [\Delta_{-i} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} T_A Z' S Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \Delta_{-i}] \\ & \cdot [\Delta_{-i} \text{diag}\{\Gamma_1, \dots, \Gamma_c\} T_A Z' S Z \text{diag}\{\Gamma_1', \dots, \Gamma_c'\} \Delta_{-i}] \end{aligned} \right] \\
&\quad [\text{using (2.27) and part (ii) of Lemma 2.4}]
\end{aligned}$$

$$\begin{aligned}
&= \text{tr}\{[\Gamma_j (\sum_{k=1}^c T_{A,jk} Z_k' S Z_i) \Gamma_i'] [\Gamma_i (\sum_{k=1}^c T_{A,ik} Z_k' S Z_j) \Gamma_j']\} \\
&= \text{tr}[(\sum_{k=1}^c T_{A,jk} Z_k' S Z_i A_i) (\sum_{k=1}^c T_{A,ik} Z_k' S Z_j A_j)] \\
&= \text{tr}\left[(-\frac{1}{\sigma_i^2} T_{A,ji}) (-\frac{1}{\sigma_j^2} T_{A,ij})\right] \text{ [using (5.11)]} \\
&= \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}) , \tag{5.20}
\end{aligned}$$

and

$$\begin{aligned}
&\underline{y}' P_A Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A \underline{y} \\
&= \text{tr}(\underline{y}' P_A Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A \underline{y}) \\
&= \text{tr}(\underline{y}' P_A Z_i^{\#} Z_i^{\#'} P_A Z_j^{\#} Z_j^{\#'} P_A \underline{y}) \\
&= \text{tr}(Z_i^{\#'} P_A Z_j^{\#} Z_j^{\#'} P_A \underline{y} \underline{y}' P_A Z_i^{\#}) \\
&= \text{tr}(\underline{\Delta}_i Z_i^{\#'} P_A Z_j^{\#} Z_j^{\#'} P_A \underline{y} \underline{y}' P_A Z_i^{\#} \underline{\Delta}_i) \\
&= \text{tr}(\underline{\Delta}_j Z_j^{\#'} P_A \underline{y} \underline{y}' P_A Z_i^{\#} \underline{\Delta}_i \underline{\Delta}_i' Z_i^{\#'} P_A Z_j^{\#} \underline{\Delta}_j)
\end{aligned}$$

$$\begin{aligned}
&= \underline{y}' P_A Z' \underline{\Delta}_i Z' P_A Z' \underline{\Delta}_j Z' P_A \underline{y} \\
&= \underline{y}' P_A Z' \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} \underline{\Delta}_i \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \\
&\quad \cdot Z' P_A Z' \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} \underline{\Delta}_j \\
&\quad \cdot \text{diag}\{\Gamma_1, \dots, \Gamma_c\} Z' P_A \underline{y} \\
&= \tilde{v}'_{-A} \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} [\underline{\Delta}_i \text{diag}\{\Gamma_1, \dots, \Gamma_c\} T_A Z' S Z \\
&\quad \cdot \text{diag}\{\Gamma'_1, \dots, \Gamma'_c\} \underline{\Delta}_j] \text{diag}\{\Gamma_1, \dots, \Gamma_c\} \tilde{v}_{-A} \\
&= \tilde{v}'_{-A,i} \Gamma'_i \Gamma_i \left( \sum_{\ell=1}^c T_{A,i\ell} Z'_\ell S Z_j \right) \Gamma'_j \Gamma_j \tilde{v}_{-A,j} \\
&= \tilde{v}'_{-A,i} A_i \left( \sum_{\ell=1}^c T_{A,i\ell} Z'_\ell S Z_j A_j \right) \tilde{v}_{-A,j} \\
&= \tilde{v}'_{-A,i} A_i \left( -\frac{1}{\sigma_j^2} T_{A,ij} \right) \tilde{v}_{-A,j} \text{ [using (5.11)]} \\
&= -\frac{1}{\sigma_j^2} \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} . \tag{5.21}
\end{aligned}$$

Substituting expressions (5.20) and (5.21) into expression (5.19), we obtain

$$\begin{aligned}
(-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} &= - \sigma_{c+1}^4 \left[ \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}) \right] \\
&\quad + 2 \left( \frac{1}{\sigma_j^2} \right) \tilde{v}'_{-A,i} T_{A,ij} \tilde{v}_{-A,j} \\
&= - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ji} T_{A,ij}) \\
&\quad - 2 \sigma_{c+1}^2 \left( \frac{1}{\gamma_j} \right) \tilde{v}'_{-A,i} T_{A,ij} \tilde{v}_{-A,j} .
\end{aligned}$$

(vi) Setting  $i=j$  in expression (5.18) and using result (5.6), we have that

$$\begin{aligned}
\frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} &= \left( \frac{1}{2} \right) \sigma_{c+1}^4 \left[ \text{tr}(Z_{iA} Z_{iA}' P_A Z_{iA} Z_{iA}' P_A) \right. \\
&\quad \left. - 2 \text{tr}(\underline{y}' P_A Z_{iA} Z_{iA}' P_A Z_{iA} Z_{iA}' P_A \underline{y}) \right] .
\end{aligned} \tag{5.22}$$

Proceeding as in the derivation of results (5.20) and (5.21), we find that

$$\text{tr}(Z_{iA} Z_{iA}' P_A Z_{iA} Z_{iA}' P_A)$$

$$\begin{aligned}
&= \text{tr}[(\sum_{k=1}^c T_{A,ik} Z_k' S Z_{i,i})^2] \\
&= \text{tr}[\frac{1}{\sigma_i^4} (I_{q_i} - T_{A,ii})^2] \text{ [using (5.11)]} \tag{5.23}
\end{aligned}$$

and

$$\begin{aligned}
&\text{tr}(\underline{y}' P_{A,i} Z_{i,i} Z_{i,i}' P_{A,i} Z_{i,i} Z_{i,i}' P_{A,i} \underline{y}) \\
&= \tilde{y}_{-A,i}' A_i (\sum_{\ell=1}^c T_{A,i\ell} Z_{\ell}' S Z_{i,i}) \tilde{y}_{-A,i} \\
&= \tilde{y}_{-A,i}' A_i (\frac{1}{\sigma_i^2}) (I_{q_i} - T_{A,ii}) \tilde{y}_{-A,i} \tag{5.24}
\end{aligned}$$

[again, using (5.11)].

Substituting expressions (5.23) and (5.24) into expression (5.22), we obtain

$$\begin{aligned}
(-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} &= -\frac{1}{\gamma_i^2} \text{tr}[(I_{q_i} - T_{A,ii})^2] \\
&+ 2 \sigma_{c+1}^2 (\frac{1}{\gamma_i}) \tilde{y}_{-A,i}' A_i (I_{q_i} - T_{A,ii}) \tilde{y}_{-A,i} .
\end{aligned}$$

(vii) Together, equation (5.22) and part (i) of Lemma 2.1 imply that

$$\begin{aligned}
 (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} &= -\sigma_{c+1}^4 \{ \text{tr}(Z_i A_i Z_i' P_A Z_i A_i Z_i' P_A) \\
 &\quad - 2 \text{tr}[\underline{y}' A (A' V_A A)^{-1} A' Z_i A_i Z_i' P_A Z_i A_i Z_i' A (A' V_A A)^{-1} A' \underline{y}] \}.
 \end{aligned}$$

Thus,

$$\begin{aligned}
 (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] &= -\sigma_{c+1}^4 \text{tr}(Z_i A_i Z_i' P_A Z_i A_i Z_i' P_A) \\
 &\quad + 2 \sigma_{c+1}^4 \mathbb{E} \{ \text{tr}[(A' V_A A)^{-1} A' Z_i A_i Z_i' \\
 &\quad \cdot P_A Z_i A_i Z_i' A (A' V_A A)^{-1} (A' \underline{y})(A' \underline{y})'] \} .
 \end{aligned}$$

Moreover,

$$\begin{aligned}
 &\mathbb{E} \{ \text{tr}[(A' V_A A)^{-1} A' Z_i A_i Z_i' P_A Z_i A_i Z_i' A (A' V_A A)^{-1} (A' \underline{y})(A' \underline{y})'] \} \\
 &= \text{tr} \{ (A' V_A A)^{-1} A' Z_i A_i Z_i' P_A Z_i A_i Z_i' A (A' V_A A)^{-1} \mathbb{E}[(A' \underline{y})(A' \underline{y})'] \} \\
 &= \text{tr}[(A' V_A A)^{-1} A' Z_i A_i Z_i' P_A Z_i A_i Z_i' A (A' V_A A)^{-1} A' V_A A]
 \end{aligned}$$

$$= \text{tr}(Z_{i i} A_i Z_{i i}' P_A Z_{i i} A_i Z_{i i}' P_A) \text{ [using part (i) of Lemma 2.1].}$$

Therefore,

$$\begin{aligned} (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] &= \sigma_{c+1}^4 \text{tr}(Z_{i i} A_i Z_{i i}' P_A Z_{i i} A_i Z_{i i}' P_A) \\ &= \frac{1}{\gamma_i^2} \text{tr}[(I_{q_i} - T_{A, ii})^2] \text{ [using (5.23)].} \end{aligned}$$

(viii) It follows from result (5.17) that

$$\begin{aligned} (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right] &= \mathbb{E}[\text{tr}(\underline{y}' P_A Z_{i i} A_i Z_{i i}' P_A \underline{y})] \\ &= \mathbb{E}\{\text{tr}[\underline{y}' A (A' V_A A)^{-1} A' Z_{i i} A_i Z_{i i}' A (A' V_A A)^{-1} A' \underline{y}]\} \\ &= \text{tr}\{(A' V_A A)^{-1} A' Z_{i i} A_i Z_{i i}' A (A' V_A A)^{-1} \mathbb{E}[(A' \underline{y})(A' \underline{y})']\} \\ &= \text{tr}[(A' V_A A)^{-1} A' Z_{i i} A_i Z_{i i}' A (A' V_A A)^{-1} A' V_A A] \\ &= \text{tr}(Z_{i i} A_i Z_{i i}' P_A) \end{aligned}$$

$$= \frac{1}{\gamma_i} \left( \frac{1}{\sigma_{c+1}^2} \right) [q_i - \text{tr}(T_{A,ii})] \text{ [using (5.12)]}.$$

(ix) Starting with expression (5.19) and proceeding as in the derivation of part (vii), we obtain

$$\begin{aligned} (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] &= \sigma_{c+1}^4 \text{tr}(Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A) \\ &= \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \text{ [using (5.20)]}. \end{aligned}$$

(x) According to expression (5.16),

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \mathbb{E}(\underline{y}' P_A \underline{y})] .$$

Since

$$\mathbb{E}(\underline{y}' P_A \underline{y}) = (X \underline{\alpha})' P_A (X \underline{\alpha}) + \text{tr}(P_A V_A)$$

$$= \text{tr}(P_A V_A)$$

$$= n - p^* \text{ [using (5.7)] ,}$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} (n - p^*) .$$

□

## 2. Derivatives of $L_v(\underline{\sigma}; \underline{y})$

Lemma 5.9: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$  and  $\underline{y} \in \Omega_{1,A}^*$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_v}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(T_A)] \\ + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{x}\tilde{\alpha}_{-A} - \underline{z}\tilde{b}_{-A})' (\underline{y} - \underline{x}\tilde{\alpha}_{-A} - \underline{z}\tilde{b}_{-A}),$$

$$(ii) \quad \frac{\partial L_v}{\partial \sigma_i^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \tilde{v}_{-A,i}'^A \tilde{v}_{-A,i},$$

if  $\sigma_i^2 \neq 0$ ,

$$(iii) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{y}' S(\underline{y} - \underline{z}\tilde{b}_{-A})] \\ - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{A,jk} T_{A,kj}) \\ + 2\sigma_{j-A,j}^2 \tilde{v}_{j-A,j}'^A T_{A,jk} \tilde{v}_{-A,k}] \\ + 2 \left(\frac{1}{\sigma_{c+1}^4}\right) \sum_{j=1}^c \left[\left(\frac{1}{2}\right) q_j - \sigma_{j-A,j}^2 \tilde{v}_{j-A,j}'^A \tilde{v}_{-A,j}\right],$$

$$(iv) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_1^2} = - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_1^2} \operatorname{tr}(T_{A,ii})$$

$$+ 2 \left( \frac{1}{\sigma_{c+1}^2} \right) \frac{1}{\sigma_1^2} \sum_{j=1}^c \left[ \left( \frac{1}{2} \right) \operatorname{tr}(T_{A,ij} T_{A,ji}) \right. \\ \left. + \sigma_{i-A, i}^2 T_{A,ij} \tilde{V}_{-A,j} \right] ,$$

$$\text{if } \sigma_1^2 \neq 0 ,$$

$$(v) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_j^2} = - \frac{1}{\sigma_1^2} \frac{1}{\sigma_j^2} \operatorname{tr}(T_{A,ij} T_{A,ji})$$

$$- 2 \left( \frac{1}{\sigma_j^2} \right) \tilde{V}_{-A,i}^i T_{A,ij} \tilde{V}_{-A,j}^i ,$$

$$\text{if } \sigma_1^2 \neq 0, \sigma_j^2 \neq 0 ,$$

$$(vi) \quad (-2) \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} = - \frac{1}{\sigma_1^4} \operatorname{tr}[(I - T_{A,ii})^2]$$

$$+ 2 \left( \frac{1}{\sigma_1^2} \right) \tilde{V}_{-A,i}^i T_{A,ii} (I - T_{A,ii}) \tilde{V}_{-A,i}^i ,$$

$$\text{if } \sigma_1^2 \neq 0 ,$$

$$(vii) \quad (-2) \quad \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_i^4} \text{tr}[(I - T_{A,ii})^2], \text{ if } \sigma_i^2 \neq 0,$$

$$(viii) \quad (-2) \quad \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [\text{tr}(T_{A,ii}) - \sum_{j=1}^c \text{tr}(T_{A,ij} T_{A,ji})],$$

$$\text{if } \sigma_i^2 \neq 0,$$

$$(ix) \quad (-2) \quad \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}),$$

$$\text{if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0,$$

$$(x) \quad (-2) \quad \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right] = \frac{1}{\sigma_{c+1}^4} [n - p^* - q + \text{tr}(T_A^2)].$$

Proof: We use the chain rule to derive the results of Lemma 5.9 from the results of Lemma 5.8.

(i) Using result (9.1) and parts (i) and (ii) of Lemma 5.8, we find that

$$\frac{\partial L_v}{\partial \sigma_{c+1}^2} = - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S(\underline{y} - \tilde{Z}_{-A} \tilde{b}_{-A})]$$

$$\begin{aligned}
& - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \left\{ - \left( \frac{1}{2} \right) [q_j - \text{tr}(T_{A,jj})] + \left( \frac{1}{2} \right) \gamma_j \sigma_{c+1}^2 \tilde{v}_{j-A,j}'^A \tilde{v}_{j-A,j} \right\} \\
& = - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} (n-p^*) + \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} \underline{y}' S (\underline{y} - \tilde{z}_{\underline{b}_A}) \\
& \quad + \left( \frac{1}{2} \right) \left( \frac{1}{\sigma_{c+1}^2} \right) q - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \text{tr}(T_{A,jj}) \\
& \quad - \left( \frac{1}{2} \right) \sum_{j=1}^c \gamma_j \tilde{v}_{j-A,j}'^A \tilde{v}_{j-A,j} \\
& = - \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(T_A)] \\
& \quad + \left( \frac{1}{2} \right) \frac{1}{\sigma_{c+1}^2} [\underline{y}' S (\underline{y} - \tilde{z}_{\underline{b}_A}) - \sigma_{c+1}^2 \sum_{j=1}^c \gamma_j \tilde{v}_{j-A,j}'^A \tilde{v}_{j-A,j}] .
\end{aligned}$$

Observing that

$$\begin{aligned}
& \underline{y}' S (\underline{y} - \tilde{z}_{\underline{b}_A}) - \sigma_{c+1}^2 \sum_{j=1}^c \gamma_j \tilde{v}_{j-A,j}'^A \tilde{v}_{j-A,j} \\
& = (\underline{y} - \underline{x} \tilde{\alpha}_{\underline{A}})' \underline{V}_A^{-1} (\underline{y} - \underline{x} \tilde{\alpha}_{\underline{A}}) - \tilde{\underline{v}}_{\underline{A}}' \underline{D}_{\underline{A}-\underline{A}} \tilde{\underline{v}}_{\underline{A}} \quad [\text{using (2.13)}]
\end{aligned}$$

$$= (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A})' \underline{v}_A^{-1} (\underline{y} - \underline{x\tilde{\alpha}}_{-A})$$

$$+ \underline{\tilde{b}}_{-A}' \underline{z}' \underline{v}_A^{-1} (\underline{y} - \underline{x\tilde{\alpha}}_{-A}) - \underline{\tilde{v}}_{-A}' \underline{D}_{A-A} \underline{\tilde{v}}_{-A}$$

$$= (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A})' \underline{v}_A^{-1} (\underline{y} - \underline{x\tilde{\alpha}}_{-A})$$

$$+ \underline{\tilde{b}}_{-A-A}' \underline{\tilde{v}}_{-A} - \underline{\tilde{b}}_{-A-A}' \underline{\tilde{v}}_{-A}$$

$$= (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A})' \underline{R}^{-1} (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A})$$

[using part (v) of Lemma 2.4] ,

we conclude that

$$\frac{\partial L_v}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(\underline{T}_A)]$$

$$+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A})' (\underline{y} - \underline{x\tilde{\alpha}}_{-A} - \underline{z\tilde{b}}_{-A}) .$$

(ii) Part (ii) of Lemma 5.8, together with result (9.2), implies that

$$\frac{\partial L_v}{\partial \sigma_i^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i}.$$

(iii) Result (9.1) implies that

$$\begin{aligned} \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= \left[ \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_{c+1}} \right) \right] + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \gamma_j \frac{\partial L_1}{\partial \gamma_j} \\ &\quad + \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \left( \frac{1}{\sigma_{c+1}^4} \right) \frac{\partial L_1}{\partial \gamma_j} \\ &\quad - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \gamma_j \left[ \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_j} \right) \right] \\ &= \frac{\partial^2 L_1}{\partial \gamma_{c+1} \partial \gamma_{c+1}} - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \gamma_j \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} \\ &\quad + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \gamma_j \frac{\partial L_1}{\partial \gamma_j} + \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \left( \frac{1}{\sigma_{c+1}^4} \right) \frac{\partial L_1}{\partial \gamma_j} \\ &\quad - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \gamma_j \left\{ \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_{c+1}} - \frac{1}{\sigma_{c+1}^2} \sum_{k=1}^c \gamma_k \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} \\
&\quad + 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
&\quad + \left( \frac{1}{\sigma_{c+1}^8} \right) \sum_{j=1}^c \sigma_j^2 \left[ \sum_{k=1}^c \sigma_k^2 \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right]. \tag{5.25}
\end{aligned}$$

Substituting from parts (ii)-(vi) of Lemma 5.8, we obtain

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \\
&\quad - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \left[ (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} \right] \\
&\quad - 4 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
&\quad + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \left[ \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left\{ (-2) \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right\} \right]
\end{aligned}$$

$$\begin{aligned}
&= (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \left[ (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} \right] \\
&\quad - 4 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
&\quad + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \left\{ \left[ \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right] \right] \right. \\
&\quad \left. + \sigma_j^4 \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_j} \right] \right\}
\end{aligned}$$

$$= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{y}' S(\underline{y} - \underline{Z}\underline{b}_A)]$$

$$- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{v}'_{-A,j} \tilde{v}_{-A,j}^{A,j}$$

$$- 4 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \left\{ - \left( \frac{1}{2} \right) \frac{1}{\gamma_j} [q_j - \text{tr}(T_{A,jj})] \right.$$

$$\left. + \left( \frac{1}{2} \right) \sigma_{c+1}^2 \tilde{v}'_{-A,j} \tilde{v}_{-A,j}^{A,j} \right\}$$

$$\begin{aligned}
& + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \left\{ \left[ \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 \left[ -\sigma_{c+1}^4 \left( \frac{1}{\sigma_j^2} \right) \left( \frac{1}{\sigma_k^2} \right) \text{tr}(T_{A,jk} T_{A,kj}) \right. \right. \right. \\
& \quad \left. \left. - 2 \sigma_{c+1}^4 \left( \frac{1}{\sigma_k^2} \right) \tilde{y}'_{-A,j} A_j T_{A,jk} \tilde{y}_{-A,k} \right] \right. \\
& \quad + \sigma_j^4 \left[ -\sigma_{c+1}^4 \left( \frac{1}{\sigma_j^4} \right) \text{tr}[(I - T_{A,jj})^2] \right. \\
& \quad \left. \left. + 2 \sigma_{c+1}^4 \left( \frac{1}{\sigma_j^2} \right) \tilde{y}'_{-A,j} A_j (I - T_{A,jj}) \tilde{y}_{-A,j} \right] \right\} \\
& = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \tilde{z} \underline{b}_A)] \\
& \quad - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}'_{-A,j} A_j \tilde{y}_{-A,j} \\
& \quad - 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \left\{ -\left( \frac{1}{2} \right) \frac{1}{\sigma_j^2} [q_j - \text{tr}(T_{A,jj})] + \left( \frac{1}{2} \right) \tilde{y}'_{-A,j} A_j \tilde{y}_{-A,j} \right\} \\
& \quad + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \left\{ \left[ \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \left[ -\frac{1}{\sigma_j^2} \text{tr}(T_{A,jk} T_{A,kj}) \right. \right. \right.
\end{aligned}$$

$$\begin{aligned}
& - 2 \tilde{\mathbf{v}}'_{-A,j} \mathbf{A}_j \mathbf{T}_{A,jk} \tilde{\mathbf{v}}_{-A,k} \Big] - \text{tr}[(\mathbf{I} - \mathbf{T}_{A,jj})^2] \\
& + 2 \sigma_j^2 \tilde{\mathbf{v}}'_{-A,j} \mathbf{A}_j \tilde{\mathbf{v}}_{-A,j} \\
& - 2 \sigma_j^2 \tilde{\mathbf{v}}_{-A,j} \mathbf{A}_j \mathbf{T}_{A,jj} \tilde{\mathbf{v}}_{-A,j} \Big\} \\
= & - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\mathbf{y}'\mathbf{S}(\mathbf{y} - \mathbf{Z}\tilde{\mathbf{b}}_A)] \\
& - 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \left\{ -\left(\frac{1}{2}\right) \frac{1}{\sigma_j^2} [q_j - \text{tr}(\mathbf{T}_{A,jj})] + \left(\frac{1}{2}\right) \tilde{\mathbf{v}}'_{-A,j} \mathbf{A}_j \tilde{\mathbf{v}}_{-A,j} \right\} \\
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(\mathbf{T}_{A,jk} \mathbf{T}_{A,kj}) + 2\sigma_j^2 \tilde{\mathbf{v}}'_{-A,j} \mathbf{A}_j \mathbf{T}_{A,jk} \tilde{\mathbf{v}}_{-A,k}] \\
& + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [-q_j + 2\text{tr}(\mathbf{T}_{A,jj})] \\
= & - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2\mathbf{y}'\mathbf{S}(\mathbf{y} - \mathbf{Z}\tilde{\mathbf{b}}_A)]
\end{aligned}$$

$$\begin{aligned}
& + 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c [q_j - \text{tr}(T_{A,jj})] \\
& - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{v}'_{-A,j} A_j \tilde{v}_{-A,j} \\
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{A,jk} T_{A,kj}) + 2 \sigma_j^2 \tilde{v}'_{-A,j} A_j T_{A,jk} \tilde{v}_{-A,k}] \\
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [q_j - 2 \text{tr}(T_{A,jj})] \\
& = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - Z \tilde{b}_{-A})] \\
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{A,jk} T_{A,kj}) + 2 \sigma_j^2 \tilde{v}'_{-A,j} A_j T_{A,jk} \tilde{v}_{-A,k}] \\
& + 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \left[ \left( \frac{1}{2} \right) q_j - \sigma_j^2 \tilde{v}'_{-A,j} A_j \tilde{v}_{-A,j} \right] .
\end{aligned}$$

(iv) Using result (9.1) and substituting from parts (ii) and (iv)-(vi) of Lemma 5.8, we obtain

$$\begin{aligned}
\frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} &= -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i} + \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \\
&= -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i} + \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \gamma_j \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] \\
&= -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i} + \frac{1}{\sigma_{c+1}^2} \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \\
&\quad - \frac{1}{\sigma_{c+1}^4} \sum_{\substack{j=1 \\ j \neq i}}^c \gamma_j \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} - \frac{1}{\sigma_{c+1}^4} (\gamma_i) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \quad (5.26)
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} \tilde{v}_{-A,i} \\
&\quad - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} \tilde{v}_{-A,i} - \frac{1}{\sigma_{c+1}^4} \sum_{\substack{j=1 \\ j \neq i}}^c \gamma_j \left[ \left(\frac{1}{2}\right) \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \right. \\
&\quad \left. + \sigma_{c+1}^2 \left(\frac{1}{\gamma_j}\right) \tilde{v}'_{-A,i} \tilde{v}_{-A,ij} \right] \\
&\quad - \frac{1}{\sigma_{c+1}^4} (\gamma_i) \left\{ \left(\frac{1}{2}\right) \left(\frac{1}{\gamma_i^2}\right) \text{tr}[(I - T_{A,ii})^2] \right. \\
&\quad \left. - \sigma_{c+1}^2 \left(\frac{1}{\gamma_i}\right) \tilde{v}'_{-A,i} (I - T_{A,ii}) \tilde{v}_{-A,i} \right\}
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] - \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} \\
&\quad - \frac{1}{\sigma_{c+1}^2} \sum_{\substack{j=1 \\ j \neq i}}^c \left[ \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} \text{tr}(T_{A,ij} T_{A,ji}) + \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} \right] \\
&\quad - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(I_{q_i} - 2T_{A,ii} + T_{A,ii}^2) \\
&\quad + \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i} - \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} A_i T_{A,ii} \tilde{v}_{-A,i} \\
&= \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{A,ii}) - \frac{1}{\sigma_{c+1}^2} \sum_{\substack{j=1 \\ j \neq i}}^c \left[ \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} \text{tr}(T_{A,ij} T_{A,ji}) \right. \\
&\quad \left. + \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} \right] \\
&\quad - \frac{1}{\sigma_{c+1}^2} \tilde{v}'_{-A,i} A_i T_{A,ii} \tilde{v}_{-A,i} - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{A,ii}^2) .
\end{aligned}$$

Therefore,

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} &= - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{A,ii}) \\
&+ 2 \left( \frac{1}{\sigma_{c+1}^2} \right) \frac{1}{\sigma_i^2} \sum_{j=1}^c \left[ \left( \frac{1}{2} \right) \text{tr}(T_{A,ij} T_{A,ji}) \right. \\
&\quad \left. + \sigma_i^2 \tilde{v}_{-A,i}^{A_i} T_{A,ij} \tilde{v}_{-A,j} \right] .
\end{aligned}$$

(v) It follows from result (9.2) that

$$\begin{aligned}
\frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial}{\partial \sigma_j^2} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \\
&= \frac{1}{\sigma_{c+1}^2} \sum_{k=1}^{c+1} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_k} \frac{\partial \gamma_k}{\partial \sigma_j^2} \\
&= \frac{1}{\sigma_{c+1}^4} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} .
\end{aligned} \tag{5.27}$$

Thus, in light of part (v) of Lemma 5.8, we have that

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{\sigma_{c+1}^4} \left[ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \right]$$

$$\begin{aligned}
& - 2 \sigma_{c+1}^2 \left( \frac{1}{\gamma_j} \right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} \\
& = - \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}) \\
& - 2 \left( \frac{1}{\sigma_j^2} \right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} .
\end{aligned}$$

(vi) Similarly,

$$\frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} = \frac{1}{\sigma_{c+1}^4} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \quad (5.28)$$

so that, in light of part (vi) of Lemma 5.8,

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} & = \frac{1}{\sigma_{c+1}^4} \left\{ - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \right. \\
& + 2 \sigma_{c+1}^2 \left( \frac{1}{\gamma_i} \right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} \} \\
& = - \frac{1}{\sigma_i^4} \text{tr}[(I - T_{A,ii})^2] \\
& + 2 \left( \frac{1}{\sigma_i^2} \right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} .
\end{aligned}$$

(vii) Using result (5.28) and substituting from part (vii) of Lemma 5.8, we obtain

$$\begin{aligned}
 (-2) \mathfrak{E} \left( \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right) &= \frac{1}{\sigma_{c+1}^4} \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right) \right] \\
 &= \frac{1}{\sigma_{c+1}^4} \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \\
 &= \frac{1}{\sigma_i^4} \text{tr}[(I - T_{A,ii})^2] .
 \end{aligned}$$

(viii) Result (5.26) implies that

$$\begin{aligned}
 (-2) \mathfrak{E} \left( \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_i^2} \right) &= 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \mathfrak{E} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \\
 &\quad + \frac{1}{\sigma_{c+1}^2} \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right) \right] \\
 &\quad - \frac{1}{\sigma_{c+1}^4} \sum_{\substack{j=1 \\ j \neq i}}^c \gamma_j \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right) \right] \\
 &\quad - \frac{1}{\sigma_{c+1}^4} \gamma_i \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right) \right] .
 \end{aligned}$$

According to parts (ii), (iv), and (vii) of Lemma 5.8,

$$\begin{aligned}
\mathbb{E} \left( \frac{\partial L_1}{\partial \gamma_i} \right) &= - \left( \frac{1}{2} \right) \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] \\
&+ \left( \frac{1}{2} \right) \sigma_{c+1}^2 \mathbb{E}(\tilde{y}'_{-A,i} A_i \tilde{y}_{-A,i}) \\
&= - \left( \frac{1}{2} \right) \sigma_{c+1}^2 [(-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right) ] \\
&+ \left( \frac{1}{2} \right) \sigma_{c+1}^2 [(-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right) ] \\
&= 0 .
\end{aligned} \tag{5.29}$$

Based on parts (vii)-(ix) of Lemma 5.8, we conclude that

$$\begin{aligned}
(-2) \mathbb{E} \left( \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} \right) &= \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] \\
&- \frac{1}{\sigma_{c+1}^4} \sum_{\substack{j=1 \\ j \neq i}}^c \gamma_i \left[ \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \right] \\
&- \frac{1}{\sigma_{c+1}^4} (\gamma_i) \frac{1}{\gamma_i^2} \text{tr}(I - 2T_{A,ii} + T_{A,ii}^2)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \left\{ q_i - \text{tr}(T_{A,ii}) \right. \\
&\quad - \left[ \sum_{\substack{j=1 \\ j \neq i}}^c \text{tr}(T_{A,ij} T_{A,ji}) \right] - q_i \\
&\quad \left. + 2 \text{tr}(T_{A,ii}) - \text{tr}(T_{A,ii}^2) \right\} \\
&= \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \left[ \text{tr}(T_{A,ii}) - \sum_{j=1}^c \text{tr}(T_{A,ij} T_{A,ji}) \right].
\end{aligned}$$

(ix) Using result (5.27), and substituting from part (ix) of Lemma 5.8, we obtain

$$\begin{aligned}
(-2) \mathbb{E} \left( \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right) &= \frac{1}{\sigma_{c+1}^4} \left[ (-2) \mathbb{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right) \right] \\
&= \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{A,ij} T_{A,ji}) .
\end{aligned}$$

(x) Using results (5.25) and (5.29), and substituting from parts (vii)-(x) of Lemma 5.8, we find that

$$\begin{aligned}
(-2) \mathfrak{E} \left( \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) &= [(-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right)] \\
&- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 [(-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} \right) \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \left\{ \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 [(-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right)] \right. \\
&\quad \left. + \sigma_j^4 [(-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_j} \right)] \right\} \\
&= \frac{1}{\sigma_{c+1}^4} (n-p^*) - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c [q_j - \text{tr}(T_{A,jj})] \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \left[ \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 \left( \frac{1}{\gamma_j} \right) \frac{1}{\gamma_k} \text{tr}(T_{A,jk} T_{A,kj}) \right. \\
&\quad \left. + \sigma_j^4 \left( \frac{1}{\gamma_j^2} \right) \text{tr}(I - 2T_{A,jj} + T_{A,jj}^2) \right] \\
&= \frac{1}{\sigma_{c+1}^4} (n-p^*)
\end{aligned}$$

$$- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) q + 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \text{tr}(T_A)$$

$$+ \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c \text{tr}(T_{A,jk} T_{A,kj})$$

$$+ \left( \frac{1}{\sigma_{c+1}^4} \right) q - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \text{tr}(T_A)$$

$$= \frac{1}{\sigma_{c+1}^4} [n - p^* - 2q + \sum_{j=1}^c \sum_{k=1}^c \text{tr}(T_{A,jk} T_{A,kj}) + q]$$

$$= \frac{1}{\sigma_{c+1}^4} [n - p^* - q + \text{tr}(T_A^2)] .$$

□

### 3. Derivatives of $L_c(\underline{Y}^+; \underline{y})$

Define

$$H_A \equiv I_n + \sum_{j=1}^c \gamma_j Z_j A_j Z_j' = \frac{1}{\sigma_{c+1}^2} V_A ,$$

$$P_A^* \equiv H_A^{-1} - H_A^{-1} X (X' H_A^{-1} X)^{-1} X' H_A^{-1} = \sigma_{c+1}^2 P_A ,$$

$$\begin{aligned} \hat{\sigma}_{c+1,A}^2(\underline{Y}^+) &\equiv \left( \frac{1}{n-p^*} \right) (\underline{y} - X \tilde{\alpha}_A)' H_A^{-1} (\underline{y} - X \tilde{\alpha}_A) \\ &= \left( \frac{1}{n-p^*} \right) \underline{y}' P_A^* \underline{y} . \end{aligned}$$

Following the approach taken in Section III.B.3.a, the result of part (i) of Lemma 5.8 can be re-expressed as

$$\frac{\partial L_1}{\partial \sigma_{c+1}^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (n-p^*) [\sigma_{c+1}^2 - \hat{\sigma}_{c+1,A}^2(\underline{Y}^+)] .$$

Thus, for an arbitrary (fixed) value of  $\underline{Y}^+$ , the function  $L_1(\underline{Y}; \underline{y})$  attains its maximum, over the interval  $0 < \sigma_{c+1}^2 < \infty$ , uniquely at the value  $\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)$ . Accordingly, the concentrated log-likelihood function (up to an additive constant) associated with  $A'\underline{y}$  is

$$L_c(\underline{Y}^+; \underline{y}) \equiv L_1(\underline{Y}^+, \hat{\sigma}_{c+1,A}^2(\underline{Y}^+); \underline{y}) .$$

We have that

$$\begin{aligned} L_1(\underline{Y}^+, \sigma_{c+1}^2; \underline{y}) &= - \left(\frac{1}{2}\right) \log |\sigma_{c+1}^2 H_A| \\ &\quad - \left(\frac{1}{2}\right) \left| \log X^{*'} \frac{1}{\sigma_{c+1}^2} H_A^{-1} X^* \right| - \left(\frac{1}{2}\right) \underline{y}' \frac{1}{\sigma_{c+1}^2} P_A^* \underline{y} \end{aligned}$$

and, hence, that

$$L_c(\underline{Y}^+; \underline{y}) = - \left(\frac{n}{2}\right) \log \hat{\sigma}_{c+1,A}^2(\underline{Y}^+) - \left(\frac{1}{2}\right) \log |H_A|$$

$$+ \left(\frac{1}{2}\right) p^* \log \hat{\sigma}_{c+1,A}^2 (\underline{Y}^+) - \left(\frac{1}{2}\right) |\log X^{*'} H_A^{-1} X^*|$$

$$- \left( \frac{1}{\hat{\sigma}_{c+1,A}^2 (\underline{Y}^+)} \right) \left(\frac{1}{2}\right) \underline{y}' P_A^* \underline{y} .$$

Since

$$\begin{aligned} \underline{y}' P_A^* \underline{y} &= \underline{y}' (\sigma_{c+1}^2 P_A) \underline{y} \\ &= \sigma_{c+1}^2 (\underline{y} - X \tilde{\alpha}_A)' V_A^{-1} (\underline{y} - X \tilde{\alpha}_A) \quad [\text{using result (2.5)}] \\ &= (\underline{y} - X \tilde{\alpha}_A)' H_A^{-1} (\underline{y} - X \tilde{\alpha}_A) \\ &= (n-p^*) \hat{\sigma}_{c+1,A}^2 (\underline{Y}^+) , \end{aligned} \tag{5.30}$$

$L_c(\underline{Y}^+; \underline{y})$  can be re-expressed in the form

$$\begin{aligned} L_c(\underline{Y}^+; \underline{y}) &= - \left(\frac{1}{2}\right) (n-p^*) [1 + \log \hat{\sigma}_{c+1,A}^2 (\underline{Y}^+)] \\ &\quad - \left(\frac{1}{2}\right) \log |H_A| - \left(\frac{1}{2}\right) \log |X^{*'} H_A^{-1} X^*| , \quad \underline{y} \in \Omega_{1,A}^* , \end{aligned} \tag{5.31}$$

In the special case  $A_i = I_{q_i}$  ( $i = 1, \dots, c$ ), expression (5.31) reduces to (3.40).

Lemma 5.10 gives the first- and second-order partial derivatives of  $L_c(\underline{\gamma}^+; \underline{y})$ .

Lemma 5.10: Under the modified version of model (1.1) in which  $\text{Var}(\underline{y}) = V_A$  and  $\underline{\gamma} \in \Omega_{1,A}^*$ , and for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$(i) \quad \frac{\partial L_c}{\partial \gamma_i} = - \left( \frac{1}{2} \right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{\gamma}^+)} \right) \sigma_{c+1}^4 \tilde{v}'_{A,i} \tilde{v}_{A,i} \right\}, \quad \text{if } \gamma_i \neq 0,$$

$$(ii) \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = - \left( \frac{1}{2} \right) \left\{ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) - 2 \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{\gamma}^+)} \right) \sigma_{c+1}^4 \left( \frac{1}{\gamma_j} \right) \tilde{v}'_{A,i} \tilde{v}_{A,i} T_{A,ij} \tilde{v}_{A,j} - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{\gamma}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \sigma_{c+1}^8 (\tilde{v}'_{A,j} \tilde{v}_{A,j}) (\tilde{v}'_{A,i} \tilde{v}_{A,i}) \right\},$$

if  $\gamma_i \neq 0, \gamma_j \neq 0$ ,

$$\begin{aligned}
\text{(iii)} \quad \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i} = & - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \right. \\
& + 2 \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \sigma_{c+1}^4 \left( \frac{1}{\gamma_i} \right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} \\
& \left. - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \sigma_{c+1}^8 (\tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i})^2 \right\}, \\
& \text{if } \gamma_i \neq 0.
\end{aligned}$$

Proof:

(i) It follows from expression (5.31) that

$$\begin{aligned}
\frac{\partial L_c}{\partial \gamma_i} = & - \left(\frac{1}{2}\right) (n-p^*) \frac{\partial \log[\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)]}{\partial \gamma} \\
& - \left(\frac{1}{2}\right) \left[ \frac{\partial \log |H_A|}{\partial \gamma_i} + \frac{\partial \log |X^{*'} H_A^{-1} X^*|}{\partial \gamma_i} \right] \quad (5.32)
\end{aligned}$$

Using result (5.6) and proceeding as in Section III.B.3.a, we find that

$$\frac{\partial \log |X^{*'} H_A^{-1} X^*|}{\partial \gamma_i} = - \text{tr}[(X^{*'} H_A^{-1} X^*)^{-1} X^{*'} H_A^{-1} \left( \frac{\partial H_A}{\partial \gamma_i} \right) H_A^{-1} X^*]$$

$$\begin{aligned}
&= - \operatorname{tr}[(X^{*'} H_A^{-1} X^*)^{-1} X^{*'} H_A^{-1} Z_{i i} Z_{i i}' H_A^{-1} X^*] \\
&= - \operatorname{tr}[H_A^{-1} X^* (X^{*'} H_A^{-1} X^*)^{-1} X^{*'} H_A^{-1} Z_{i i} Z_{i i}']
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \log |H_A|}{\partial \gamma_i} &= \operatorname{tr}(H_A^{-1} \frac{\partial H_A}{\partial \gamma_i}) \\
&= \operatorname{tr}(H_A^{-1} Z_{i i} A_{i i} Z_{i i}') .
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{\partial \log |H_A|}{\partial \gamma_i} + \frac{\partial \log |X^{*'} H_A^{-1} X^*|}{\partial \gamma_i} \\
&= \operatorname{tr}\{[H_A^{-1} - H_A^{-1} X^* (X^{*'} H_A^{-1} X^*)^{-1} X^{*'} H_A^{-1}] Z_{i i} A_{i i} Z_{i i}'\} \\
&= \operatorname{tr}(P_{A_{i i}}^* Z_{i i} A_{i i} Z_{i i}') \quad [\text{using part (ii) of Lemma 2.1}] \quad (5.33) \\
&= \sigma_{c+1}^2 \operatorname{tr}(P_{A_{i i}} Z_{i i} A_{i i} Z_{i i}') \\
&= \frac{1}{\gamma_i} [q_i - \operatorname{tr}(T_{A, ii})] \quad [\text{using result (5.12)}]. \quad (5.34)
\end{aligned}$$

In addition,

$$\frac{\partial \log[\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)]}{\partial \gamma_i} = \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \left( \frac{1}{n-p^*} \right) \underline{y}' \frac{\partial P_A^*}{\partial \gamma_i} \underline{y} ,$$

and, hence, since

$$\frac{\partial P_A^*}{\partial \gamma_i} = - P_A^* \frac{\partial H_A}{\partial \gamma_i} P_A^* \quad [\text{using (5.15)}]$$

$$= - P_A^* Z_{i,A} Z_{i,A}' P_A^* \quad [\text{using (5.6)}] ,$$

$$\frac{\partial \log[\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)]}{\partial \gamma_i} = - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \left( \frac{1}{n-p^*} \right) \underline{y}' P_A^* Z_{i,A} Z_{i,A}' P_A^* \underline{y} \quad (5.35)$$

$$= - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \left( \frac{1}{n-p^*} \right) \sigma_{c+1}^4 \underline{y}' P_A Z_{i,A} Z_{i,A}' P_A \underline{y}$$

$$= - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) \left( \frac{1}{n-p^*} \right) \sigma_{c+1-A,i}^4 \tilde{V}_{i-A,i}' A_{i-A,i} \tilde{V}_{i-A,i} . \quad (5.36)$$

Substituting expressions (5.34) and (5.36) into expression (5.32), we obtain

$$\frac{\partial L_c}{\partial \gamma_i} = - \left( \frac{1}{2} \right) \left\{ \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] \right. \\ \left. - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\gamma^+)} \right) \sigma_{c+1}^4 \tilde{y}'_{A,i} \tilde{y}_{-A,i}^A \right\}.$$

(ii) Substituting expressions (5.33) and (5.35) into expression (5.32) gives

$$\frac{\partial L_c}{\partial \gamma_i} = - \left( \frac{1}{2} \right) \left[ \text{tr}(P_{A,i}^* Z_i A_i Z_i') \right. \\ \left. - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\gamma^+)} \right) (y' P_{A,i}^* Z_i A_i Z_i' P_{A,i}^* y) \right] \\ = - \left( \frac{1}{2} \right) \left[ \sigma_{c+1}^2 \text{tr}(P_{A,i} Z_i A_i Z_i') \right. \\ \left. - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\gamma^+)} \right) \sigma_{c+1}^4 y' P_{A,i} Z_i A_i Z_i' P y \right].$$

Differentiating this expression with respect to  $\gamma_j$ , we find that

$$\begin{aligned}
\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = & - \left( \frac{1}{2} \right) \left\{ \sigma_{c+1}^2 \operatorname{tr} \left( \frac{\partial P_A}{\partial \gamma_j} Z_{iA} Z_i' \right) \right. \\
& - \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\gamma^+)} \right] \frac{\partial (\underline{y}' P_A Z_{iA} Z_i' P_A \underline{y})}{\partial \gamma_j} (\sigma_{c+1}^4) \\
& \left. - \left[ \frac{\partial}{\partial \gamma_j} \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(\gamma^+)} \right] \right] \sigma_{c+1}^4 \underline{y}' P_A Z_{iA} Z_i' P_A \underline{y} \right\} . \quad (5.37)
\end{aligned}$$

Observe that, according to result (5.15) ,

$$\frac{\partial P_A}{\partial \gamma_j} = - P_A \sigma_{c+1}^2 Z_{jA} Z_j' P_A$$

implying, in light of result (5.20) that, for  $i \neq j$  ,

$$\begin{aligned}
\operatorname{tr} \left( \frac{\partial P_A}{\partial \gamma_j} Z_{iA} Z_i' \right) &= - \sigma_{c+1}^2 \operatorname{tr} (P_A Z_{jA} Z_j' P_A Z_{iA} Z_i') \\
&= - \left( \frac{1}{\sigma_{c+1}^2} \right) \frac{1}{\gamma_i} \frac{1}{\gamma_j} \operatorname{tr} (T_{A,ij} T_{A,ji}) .
\end{aligned}$$

Further, for  $i \neq j$  ,

$$\frac{\partial (\underline{y}' P_A Z_{iA} Z_i' P_A \underline{y})}{\partial \gamma_j} = \underline{y}' P_A Z_{iA} \frac{\partial (Z_i' P_A \underline{y})}{\partial \gamma_j} + \frac{\partial (\underline{y}' P_A Z_{iA})}{\partial \gamma_j} Z_i' P_A \underline{y}$$

$$\begin{aligned}
&= \underline{y}' P_A Z_i A_i Z_i' \frac{\partial P_A}{\partial \gamma_j} \underline{y} \\
&\quad + \underline{y}' \frac{\partial P_A}{\partial \gamma_j} Z_i A_i Z_i' P_A \underline{y} \\
&= -2 \sigma_{c+1}^2 \underline{y}' P_A Z_i A_i Z_i' P_A Z_j A_j Z_j' P_A \underline{y} \\
&= 2 \left( \frac{1}{\gamma_j} \right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} \quad [\text{using (5.21)}]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial \gamma_j} \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right) &= - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \frac{\partial [\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)]}{\partial \gamma_j} \\
&= - \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \underline{y}' \frac{\partial P_A^*}{\partial \gamma_j} \underline{y} \\
&\quad [\text{using (5.30)}] \\
&= \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \sigma_{c+1}^2 \underline{y}' P_A Z_j A_j Z_j' P_A \\
&= \left( \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right)^2 \left( \frac{1}{n-p^*} \right) \sigma_{c+1}^4 \tilde{v}'_{-A,j} A_j \tilde{v}_{-A,j} \quad (5.38)
\end{aligned}$$

Substituting into expression (5.37), we find that, for  $i \neq j$ ,

$$\begin{aligned} \frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_j} = & - \left(\frac{1}{2}\right) \left\{ - \frac{1}{\gamma_i} \frac{1}{\gamma_j} \text{tr}(T_{A,ij} T_{A,ji}) \right. \\ & - 2 \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(Y^+)} \right] \sigma_{c+1}^4 \left(\frac{1}{\gamma_j}\right) \tilde{v}'_{-A,i} A_i T_{A,ij} \tilde{v}_{-A,j} \\ & \left. - \left(\frac{1}{n-p^*}\right) \left[ \frac{1}{\hat{\sigma}_{c+1,A}^2(Y^+)} \right]^2 \sigma_{c+1}^8 (\tilde{v}'_{A,j} A_j \tilde{v}_{-A,j}) (\tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i}) \right\}. \end{aligned}$$

(iii) Observing that

$$\begin{aligned} \text{tr} \left( \frac{\partial P_A}{\partial \gamma_i} Z_{iA} Z_i' \right) &= - \sigma_{c+1}^2 \text{tr}(P_A Z_{iA} Z_i' P_A Z_{iA} Z_i') \\ &= - \frac{1}{\sigma_{c+1}^2} \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \text{ [using (5.23)]} \end{aligned}$$

and

$$\begin{aligned} \frac{(\underline{y}' P_A Z_{iA} Z_i' P_A \underline{y})}{\partial \gamma_i} &= - 2 \sigma_{c+1}^2 \underline{y}' P_A Z_{iA} Z_i' P_A Z_{iA} Z_i' P_A \underline{y} \\ &= - 2 \left(\frac{1}{\gamma_i}\right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} \text{ [using (5.24)]} \end{aligned}$$

and recalling result (5.38), we see that

$$\begin{aligned}
\frac{\partial^2 L_c}{\partial \gamma_i \partial \gamma_i} = & - \left( \frac{1}{2} \right) \left\{ - \frac{1}{\gamma_i^2} \text{tr}[(I - T_{A,ii})^2] \right. \\
& + 2 \left\{ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right\} \sigma_{c+1}^4 \left( \frac{1}{\gamma_i} \right) \tilde{v}'_{-A,i} A_i (I - T_{A,ii}) \tilde{v}_{-A,i} \\
& \left. - \left( \frac{1}{n-p^*} \right) \left\{ \frac{1}{\hat{\sigma}_{c+1,A}^2(\underline{Y}^+)} \right\}^2 \sigma_{c+1}^8 (\tilde{v}'_{-A,i} A_i \tilde{v}_{-A,i})^2 \right\} . \quad \square
\end{aligned}$$

#### 4. Algorithm modifications

In this section, we extend the fourteen algorithms summarized in Table 3.1 to the case where, in model (1.1),  $\text{Var}(\underline{y}) = V_A$ , while assuming only that  $A_1, \dots, A_c$  are nonnegative definite matrices.

a. The Newton-Raphson and Method of Scoring algorithms To extend algorithms 1-9 in Table 3.1 to the case where  $\text{Var}(\underline{y}) = V_A$  and  $\underline{y} \in \Omega_{1,A}^*$ , we need only substitute the expressions for the first-order partial derivatives and for the second-order partial derivatives (or their expected values) given in Sections V.B.1-V.B.2 for those given in Chapter II. [The relevant derivatives for algorithms 4 and 6 are not given explicitly in Sections V.B.1-V.B.2, however results on page 121 indicate how they can be obtained, via the chain rule, from the derivatives of  $L_c(\underline{Y}^+; \underline{y})$ .]

b. EM algorithm In this section, we extend the results of Section III.A to the more general version of model (1.2) obtained by replacing the assumption that  $\text{Var}(\underline{b}) = D$  with the assumption that  $\text{Var}(\underline{b}) = D_A$ .

The construction of the EM algorithm in Section III.A presumed the existence of a probability density function for the vector of complete data. In the case where  $A_i$  is a nonnegative definite matrix (rather than a positive definite matrix), however, the random vector  $\underline{b}_i$  does not have a probability density function. We now propose a generalization of the EM algorithm described in Section III.A which can accommodate  $A_1, \dots, A_c$  being nonnegative definite matrices.

Let  $q_i^* = \text{rank}(A_i)$ . Define  $Q_i = (Q_{i1}, Q_{i2})$  to be a  $q_i \times q_i$  orthogonal matrix such that

$$Q_i' A_i Q_i = \begin{pmatrix} D_i & \emptyset \\ \emptyset & \emptyset \end{pmatrix},$$

where  $Q_{i1}$  has dimension  $q_i \times q_i^*$  and  $D_i$  is a  $q_i^* \times q_i^*$  positive definite, diagonal matrix ( $i = 1, \dots, c$ ). Then, for  $i = 1, \dots, c$ ,

$$\begin{aligned} Z_i \underline{b}_i &= Z_i Q_i Q_i' \underline{b}_i \\ &= Z_i (Q_{i1} Q_{i1}' + Q_{i2} Q_{i2}') \underline{b}_i \end{aligned}$$

$$= \underline{Z}_i Q_{i1} Q'_{i1} \underline{b}_{i1} \quad (\text{with probability one})$$

$$[\text{since } E(Q'_{i2} \underline{b}_i) = \underline{0} \text{ and}$$

$$\text{Var}(Q'_{i2} \underline{b}_i) = \sigma_i^2 Q'_{i2} A_i Q_{i2} = \emptyset]$$

$$= \underline{Z}_i^* \underline{b}_i^* ,$$

with  $\underline{Z}_i^* = \underline{Z}_i Q_{i1}$  and  $\underline{b}_i^* = Q'_{i1} \underline{b}_i$ . Note that  $E(\underline{b}_i^*) = \underline{0}$  and  $\text{Var}(\underline{b}_i^*) = \sigma_i^2 \underline{D}_i$

so that  $\underline{b}_i^*$  (unlike  $\underline{b}_i$  itself) does have a probability density function.

Subsequently, we define

$$\underline{Z}^* \equiv (\underline{Z}_1^*, \dots, \underline{Z}_c^*) = \underline{Z} \text{diag} \{Q_{11}, \dots, Q_{c1}\} ,$$

$$\underline{b}^* \equiv (\underline{b}_1^{*'}, \dots, \underline{b}_c^{*'}) = \text{diag}\{Q'_{11}, \dots, Q'_{c1}\} \underline{b} ,$$

$$\underline{D} \equiv \text{diag}\{\sigma_1^2 \underline{D}_1, \dots, \sigma_c^2 \underline{D}_c\}$$

$$\underline{T}^* \equiv (\underline{I} + \underline{Z}^{*'} \underline{S} \underline{Z}^* \underline{D})^{-1} \quad [\underline{T}_{ij}^*] \quad (i, j = 1, \dots, c) ,$$

$$\underline{V}^* \equiv \sigma_{c+1}^2 [\underline{I}_n + \sum_{i=1}^c \gamma_i \underline{Z}_i^* \underline{D}_i \underline{Z}_i^{*'}] ,$$

$$\underline{P}^* \equiv \underline{V}^{*-1} - \underline{V}^{*-1} \underline{X} (\underline{X}' \underline{V}^{*-1} \underline{X})^{-1} \underline{X}' \underline{V}^{*-1} ,$$

$\underline{\tilde{\alpha}}^* \equiv$  any solution to the system of equations  $\underline{X}' \underline{V}^{*-1} \underline{X}' \underline{\tilde{\alpha}}^* = \underline{X}' \underline{V}^{*-1} \underline{y}$ ,

$$\underline{\tilde{v}}_{-A,i}^* \equiv \underline{z}_{-i}^{*'} \underline{V}^{*-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}^*) ,$$

$$\underline{\tilde{v}}_{-A}^* \equiv (\underline{\tilde{v}}_{-A,1}^{*'}, \dots, \underline{\tilde{v}}_{-A,c}^{*'})' = \underline{Z}^{*'} \underline{V}^{*-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}^*) ,$$

$$\underline{\tilde{b}}_{-A,i}^* \equiv \sigma_i^2 \underline{D}_{-i} \underline{\tilde{v}}_{-A,i}^* ,$$

$$\underline{\tilde{b}}_{-A}^* \equiv (\underline{\tilde{b}}_{-A,1}^{*'}, \dots, \underline{\tilde{b}}_{-A,c}^{*'})' = \underline{D} \underline{\tilde{v}}_{-A}^* .$$

The model equation for model (1.2) can then be re-expressed as

$$\underline{y} = \underline{X} \underline{\alpha} + \sum_{i=1}^c \underline{z}_{-i}^{*'} \underline{\tilde{b}}_{-i}^* + \underline{e} \quad (\text{with probability one})$$

$$= \underline{X} \underline{\alpha} + \underline{Z}^{*'} \underline{\tilde{b}}^* + \underline{e} ,$$

where

$$\underline{y} \stackrel{d}{\sim} N(\underline{X} \underline{\alpha}, \underline{V}^*) ,$$

$$\underline{\tilde{b}}^* \stackrel{d}{\sim} N(\underline{0}, \underline{D}) ,$$

$$\underline{e} \stackrel{d}{\sim} N(\underline{0}, \underline{R}) .$$

Note that

$$\underline{V}^* = \underline{V}_A$$

since  $\underline{Z}_i^* \underline{D}_i \underline{Z}_i^{*'} = \underline{Z}_i \underline{Q}_{i1} \underline{D}_i \underline{Q}_{i1}' \underline{Z}_i' = \underline{Z}_i \underline{A}_i \underline{Z}_i'$  ( $i = 1, \dots, c$ ). It follows,

since  $\underline{D}_1, \dots, \underline{D}_c$  are positive definite matrices, that the EM algorithms in Section V.A.4.b can be generalized to accommodate  $\underline{A}_1, \dots, \underline{A}_c$  being nonnegative definite by simply substituting  $\underline{Z}^*$  and  $\underline{D}$  for  $\underline{Z}$  and  $\underline{D}_A$ , respectively.

The identities in Lemma 5.11 will facilitate this generalization of the results in Section IV.A.4.b.

Lemma 5.11: For  $i = 1, \dots, c$ ,

$$(i) \quad \underline{X\tilde{\alpha}}^* = \underline{X\tilde{\alpha}}_A,$$

$$(ii) \quad \underline{\tilde{V}}_{A,i}^* = \underline{Q}_{i1}' \underline{\tilde{V}}_{A,i},$$

$$(iii) \quad \underline{\tilde{b}}_{A,i}^* = \underline{Q}_{i1}' \underline{\tilde{b}}_{A,i},$$

$$(iv) \quad \underline{\tilde{b}}_{A,i}^{*'} \underline{D}_{-i}^{-1} \underline{\tilde{b}}_{A,i}^* = \underline{\tilde{b}}_{A,i}' \underline{A}_{-i} \underline{\tilde{b}}_{A,i},$$

$$(v) \quad \underline{Z}_{-i}^* \underline{\tilde{b}}_{A,i}^* = \underline{Z}_{-i} \underline{\tilde{b}}_{A,i},$$

$$(vi) \quad \text{tr}(\underline{T}_{ii}^*) = q_i^* - q_i + \text{tr}(\underline{T}_{A,ii}).$$

Proof:

(i) Since  $\underline{v}^* = \underline{v}_A$ ,  $\underline{\tilde{\alpha}}^*$  and  $\underline{\alpha}_A$  are both solutions to the linear system  $X'\underline{v}^{*-1}\underline{x\tilde{\alpha}} = X'\underline{v}^{*-1}\underline{y}$ . Since the value of the vector  $\underline{x\tilde{\alpha}}$  is the same for any solution to this system,  $\underline{x\tilde{\alpha}}^* = \underline{x\tilde{\alpha}}_A$ .

$$(ii) \quad \tilde{v}_{-A,i}^* = Q'_{i1} Z'_i \underline{v}^{*-1} (\underline{y} - \underline{x\tilde{\alpha}}^*)$$

$$= Q'_{i1} Z'_i \underline{v}_A^{-1} (\underline{y} - \underline{x\tilde{\alpha}}_A) \quad [\text{using part (i)}]$$

$$= Q'_{i1-A,i} \tilde{v}_{-A,i}.$$

$$(iii) \quad \tilde{b}_{-A,i}^* = \sigma^2 D_{i-i} \tilde{v}_{-A,i}^*$$

$$= \sigma^2 D_{i-i} Q'_{i1-A,i} \tilde{v}_{-A,i}$$

$$= \sigma^2 Q'_{i1} Q_{i1-i} D_{i-i} Q'_{i1-A,i} \tilde{v}_{-A,i}$$

$$= \sigma^2 Q'_{i1} A_{i-i} \tilde{v}_{-A,i} \quad [\text{since } Q_{i1-i} D_{i-i} Q'_{i1} = A_{i-i}]$$

$$= Q'_{i1} D_{A,i-i} \tilde{v}_{-A,i}$$

$$= Q'_{i1-A,i} \tilde{b}_{-A,i}.$$

$$(iv) \quad \tilde{b}_{-A,i}^{*'} D_{-i}^{-1} \tilde{b}_{-A,i}^{**} = \tilde{b}_{-A,i}' Q_{i1} D_i^{-1} Q_{i1}' \tilde{b}_{i1-A,i} \quad [\text{using part (iii)}]$$

$$= \tilde{b}_{-A,i}' A_{i-A,i}^{+} \tilde{b}_{i-A,i} \quad [\text{where } A_i^{+} \text{ represents the} \\ \text{Moore-Penrose inverse of } A_i]$$

$$= \sigma_{i-A,i}^4 \tilde{v}_{i-A,i}' A_i A_i^{+} A_i \tilde{v}_{i-A,i} \quad [\text{since } \tilde{b}_{-A,i} = \sigma_{i-A,i}^2 A_i \tilde{v}_{i-A,i}]$$

$$= \sigma_{i-A,i}^4 \tilde{v}_{i-A,i}' A_i A_i^{-} A_i \tilde{v}_{i-A,i}$$

$$= \tilde{b}_{-A,i} A_i^{-} \tilde{b}_{-A,i} .$$

$$(v) \quad Z_{i-A,i}^{**} \tilde{b}_{i-A,i}^{**} = Z_i Q_{i1} Q_{i1}' \tilde{b}_{i1-A,i} \quad [\text{using part (iii)}]$$

$$= Z_i (Q_{i1} Q_{i1}' + Q_{i2} Q_{i2}') \tilde{b}_{-A,i} \quad [\text{since } Q_{i2}' A_i = \emptyset]$$

$$= Z_i Q_i Q_i' \tilde{b}_{i-A,i}$$

$$= Z_{i-A,i} \tilde{b}_{i-A,i} .$$

(vi) It follows from the definition of  $\underline{T}^*$  that

$$\underline{T}^{*-1} - I = \underline{Z}^{*'} \underline{S} \underline{Z}^* \underline{D}$$

so that

$$I_{q_i}^* - \underline{T}^* \underline{Z}^{*'} \underline{S} \underline{Z}^* \underline{D} = \underline{T}^* .$$

Thus,

$$\text{tr}(\underline{T}_{ii}^*) = q_i^* - \text{tr}(\underline{\Delta}_i \underline{T}^* \underline{Z}^{*'} \underline{S} \underline{Z}^* \underline{D} \underline{\Delta}_i) .$$

Further,

$$\underline{T}^* \underline{Z}^{*'} \underline{S} = \underline{Z}^{*'} \underline{P}^* \quad [\text{using part (ii) of Lemma 2.4 and expression (2.27)}]$$

$$= \underline{Z}^{*'} \underline{P}_A \quad [\text{since } \underline{V}^* = \underline{V}_A]$$

$$= \text{diag}\{Q'_{11}, \dots, Q'_{c1}\} \underline{Z}' \underline{P}_A$$

$$= \text{diag}\{Q'_{11}, \dots, Q'_{c1}\} \underline{T}_A \underline{Z}' \underline{S}$$

and

$$\underline{Z}^* \underline{D} = \underline{Z} \text{diag}\{Q_{11}, \dots, Q_{c1}\} \underline{D}$$

$$\cdot \text{diag}\{Q'_{11}, \dots, Q'_{c1}\} \text{diag}\{Q_{11}, \dots, Q_{c1}\}$$

$$= \underline{Z} \underline{D}_A \text{diag}\{Q_{11}, \dots, Q_{c1}\}$$

so that

$$\begin{aligned}
 \text{tr}(\underline{T}_{-ii}^*) &= q_i^* - \text{tr}(\underline{\Delta}_i \underline{T}^* \underline{Z}^{**'} \underline{S} \underline{Z}^* \underline{D} \underline{\Delta}_i) \\
 &= q_i^* - \text{tr}[\underline{\Delta}_i \text{diag}\{Q'_{11}, \dots, Q'_{c1}\} \underline{T}_A \underline{Z}' \underline{S} \\
 &\quad \cdot \underline{Z} \underline{D}_A \text{diag}\{Q_{11}, \dots, Q_{c1}\} \underline{\Delta}_i] \\
 &= q_i^* - \text{tr}[\underline{T}_A \underline{Z}' \underline{S} \underline{Z} \text{diag}\{\sigma_1^2 A_1 Q_{11}, \dots, \sigma_c^2 A_c Q_{c1}\} \\
 &\quad \cdot \underline{\Delta}_i \underline{\Delta}_i \text{diag}\{Q'_{11}, \dots, Q'_{c1}\}] \\
 &= q_i^* - \text{tr}[\underline{T}_A \underline{Z}' \underline{S} \underline{Z} \text{diag}\{\emptyset, \dots, \emptyset, \sigma_i^2 A_i Q_{i1} Q'_{i1}, \emptyset, \dots, \emptyset\}] \\
 &= q_i^* - \text{tr}[\underline{T}_A \underline{Z}' \underline{S} \underline{Z} \text{diag}\{\emptyset, \dots, \emptyset, \sigma_i^2 A_i, \emptyset, \dots, \emptyset\}] \\
 &\quad [\text{since } A_i Q_{i1} Q'_{i1} = A_i (Q_{i1} Q'_{i1} + Q_{i2} Q'_{i2}) = A_i] \\
 &= q_i^* - \text{tr}(\underline{T}_A \underline{Z}' \underline{S} \underline{Z} \underline{D}_A \underline{\Delta}_i) \\
 &= q_i^* - \text{tr}[(\underline{I} - \underline{T}_A) \underline{\Delta}_i] \\
 &= q_i^* - q_i + \text{tr}(\underline{T}_A \underline{\Delta}_i) \\
 &= q_i^* - q_i + \text{tr}(\underline{T}_{A,ii}) .
 \end{aligned}$$

□

Substituting  $\underline{Z}^*$  and  $\underline{D}$  for  $Z$  and  $D_A$ , respectively, in the EM iterates of Section V.A.4.b, and using results from Lemma 5.11, we find that the  $(p+1)$ -st iterate of EM algorithm 10 in Table 3.1 becomes

$$\sigma_i^{2(p+1)} = \frac{\tilde{b}_{-A,i}^{(p)'} \tilde{b}_{i-A,i}^{(p)} + \sigma_i^{2(p)} [q_i^* - q_i + \text{tr}(T_{A,ii}^{(p)})]}{q_i} \quad (i = 1, \dots, c), \quad (5.40)$$

$$\begin{aligned} \sigma_{c+1}^{2(p+1)} = & \frac{1}{n} \{ (\underline{y} - \tilde{x}_{-A}^{(p)} - \tilde{z}_{-A}^{(p)})' (\underline{y} - \tilde{x}_{-A}^{(p)} - \tilde{z}_{-A}^{(p)}) \\ & + \sigma_{c+1}^{2(p)} [p^* + 2q - \sum_{i=1}^c q_i^* - \text{tr}(T_A^{(p)})] \} , \end{aligned}$$

and that the  $\sigma_{c+1}^2$ -component of the  $(p+1)$ -st iterate of EM algorithm 11 in Table 3.1 is given by

$$\begin{aligned} \sigma_{c+1}^{2(p+1)} = & \frac{1}{n-p^*} \{ (\underline{y} - \tilde{x}_{-A}^{(p)} - \tilde{z}_{-A}^{(p)})' (\underline{y} - \tilde{x}_{-A}^{(p)} - \tilde{z}_{-A}^{(p)}) \\ & + \sigma_{c+1}^{2(p)} [2q - \sum_{i=1}^c q_i^* - \text{tr}(T_A^{(p)})] \} . \end{aligned}$$

[Like algorithm 10, the first  $c$  components of the  $(p+1)$ -st iterate of algorithm 11 are described by (5.40).]

#### c. Method of Successive Approximations

Using parts (i) and

(ii) of Lemma 5.9, the REML equations  $\frac{\partial L_v}{\partial \sigma_i^2} = 0$  ( $i = 1, \dots, c+1$ ) are expressible as

$$\begin{aligned}
& - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \text{tr}(T_A)] \\
& + \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - X\tilde{\alpha}_A - Z\tilde{b}_A)' (\underline{y} - X\tilde{\alpha}_A - Z\tilde{b}_A) = 0, \quad (5.41)
\end{aligned}$$

$$\begin{aligned}
& - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \tilde{v}_{A,i}' A_{i-A,i} \tilde{v}_{A,i} = 0 \\
& (i = 1, \dots, c) . \quad (5.42)
\end{aligned}$$

As discussed for the special case  $A_i = I_{q_i}$  ( $i = 1, \dots, c$ ), in Section III.D.1, equations (5.42) are equivalent to the equations  $\sigma_i^2 = h_i(\underline{\sigma})$  ( $i = 1, \dots, c$ ) and also to the equations  $\sigma_i^2 = \epsilon_i(\underline{\sigma})$  ( $i = 1, \dots, c$ ), while equation (5.41) is equivalent to the equation  $\sigma_{c+1}^2 = h_{c+1}(\underline{\sigma})$ , to the equation  $\sigma_{c+1}^2 = \epsilon_{c+1}^*(\underline{\sigma})$ , and to the equation  $\sigma_{c+1}^2 = \epsilon_{c+1}(\underline{\sigma})$ , where

$$\left. \begin{aligned}
h_i(\underline{\sigma}) &= \frac{\sigma_{i-A,i}^4 \tilde{v}_{i-A,i}' A_{i-A,i} \tilde{v}_{i-A,i}}{q_i - \text{tr}(T_{A,ii})} \\
&= \frac{\tilde{b}_{A,i}' A_{i-A,i} \tilde{b}_{A,i}}{q_i - \text{tr}(T_{A,ii})} \quad (i = 1, \dots, c) ,
\end{aligned} \right\} \quad (5.43)$$

$$\begin{aligned}
\epsilon_i(\sigma) &= \frac{\sigma^4 \tilde{v}'_{i-A,i} A_{i-A,i} \tilde{v}_{i-A,i} + \sigma_i^2 \text{tr}(T_{A,ii})}{q_i} \\
&= \frac{\tilde{b}'_{A,i} A_{i-A,i} \tilde{b}_{i-A,i} + \sigma_i^2 \text{tr}(T_{A,ii})}{q_i} \quad (i = 1, \dots, c),
\end{aligned}
\tag{5.44}$$

$$h_{(c+1)}(\underline{\sigma}) = \frac{(\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A})' (\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A})}{n - p^* - q + \text{tr}(T_A)}, \tag{5.45}$$

$$\epsilon_{c+1}^*(\underline{\sigma}) = \frac{(\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A})' (\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A}) + \sigma_{c+1}^2 [q - \text{tr}(T_A)]}{n - p^*}, \tag{5.46}$$

and

$$\epsilon_{c+1}(\underline{\sigma}) = \frac{(\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A})' (\underline{y} - X_{-A} \tilde{\alpha}_{-A} - Z_{-A} \tilde{b}_{-A}) + \sigma_{c+1}^2 [p^* + q - \text{tr}(T_A)]}{n}. \tag{5.47}$$

Similarly, using parts (i) and (ii) of Lemma 5.8, the REML equations

$$\frac{\partial L_1}{\partial \gamma_i} = 0 \quad (i = 1, \dots, c+1) \text{ are expressible as}$$

$$\begin{aligned}
- \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S (\underline{y} - Z_{-A} \tilde{b}_{-A})] &= 0, \\
- \left(\frac{1}{2}\right) \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \sigma_{c+1}^2 \tilde{v}'_{i-A,i} A_{i-A,i} \tilde{v}_{i-A,i} &= 0 \quad (i = 1, \dots, c),
\end{aligned}$$

or, alternatively, according to results (2.13) and (2.14), as

$$- \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \left(\frac{1}{\sigma_{c+1}^2}\right) \underline{y}'(\underline{y} - \underline{X}\tilde{\alpha}_A - \underline{Z}\tilde{b}_A)] = 0, \quad (5.48)$$

$$- \left(\frac{1}{2}\right) \frac{1}{\gamma_i} [q_i - \text{tr}(T_{A,ii})] + \left(\frac{1}{2}\right) \sigma_{c+1-A,i}^2 \tilde{v}'_{1-A,i} \tilde{v}_{1-A,i} = 0 \quad (5.49)$$

$$(i + 1, \dots, c).$$

Equations (5.49), like equations (5.42), are equivalent to the equations

$$\sigma_i^2 = h_i(\underline{\sigma}) \quad (i = 1, \dots, c) \text{ and also to the equations } \sigma_i^2 = \epsilon_i(\underline{\sigma})$$

( $i = 1, \dots, c$ ), while equation (5.48) is equivalent to the equation

$$\sigma_{(c+1)}^2 = h_{c+1}^*(\underline{\sigma}), \text{ where}$$

$$h_{c+1}^*(\underline{\sigma}) = \frac{\underline{y}'(\underline{y} - \underline{X}\tilde{\alpha}_A - \underline{Z}\tilde{b}_A)}{n-p^*}. \quad (5.50)$$

To extend the five Method of Successive Approximation algorithms in Table 3.1 (algorithms 10 through 14) to the case where  $\text{Var}(\underline{y}) = V_A$ , it suffices to adopt the more general definitions of the functions  $h_1(\underline{\sigma}), \dots, h_{c+1}(\underline{\sigma}), h_{c+1}^*(\underline{\sigma}), \epsilon_1(\underline{\sigma}), \dots, \epsilon_{c+1}(\underline{\sigma})$ , and  $\epsilon_{c+1}^*(\underline{\sigma})$  given by expressions (5.43)–(5.47) and (5.50).

## VI. NUMERICAL RESULTS

In this chapter, we use the fourteen algorithms summarized in Table 3.1 to compute a REML estimate of  $\underline{\gamma}$  from each of four data sets. For each data set, the model is taken to be of the form (1.1), with  $c = 1$ , and the parameter space is taken to be the extended parameter space  $\Omega_j^*$ . Each algorithm is tried with two different starting values.

Each algorithm begins with a preliminary step consisting of the computation of quantities that do not depend on  $\underline{\gamma}$  and, hence, that need to be computed only once. We will consider two alternatives for the preliminary steps. They differ in whether or not they incorporate a diagonalization of the matrix  $C$ .

In Section VI.A, we review the advantages and disadvantages of diagonalizing  $C$ . In Section VI.B, we describe a general strategy for implementing the iterative algorithms. We introduce the four data sets in Section VI.C, and the performances of the various iterative algorithms is reported in Section VI.D.

### A. Diagonalization Versus Inversion

On each iteration of the algorithms summarized in Table 3.1, the matrix  $T = (I + \frac{1}{\sigma^2_{c+1}} CD)^{-1}$  must be evaluated for a new value of  $\underline{\gamma}$ .

Suppose now that  $c = 1$ , in which case  $T = (I + \gamma_1 C)^{-1}$ . According to Lemma 2.9, the matrix  $I + \gamma_1 C$  is symmetric and positive definite.

Moreover, the Cholesky decomposition provides an efficient way of inverting  $I + \gamma_1 C$  for any particular value of  $\gamma_1$  [see, e.g., Kennedy and Gentle (1980)]. However, as discussed in Section II.C.1, even this procedure can become quite costly if  $q = q_1$  is large and if  $I + \gamma_1 C$  must be inverted for a large number of  $\gamma_1$  values. As an alternative,  $T$  can be computed from the formula

$$T = R^* (I + \gamma_1 \underline{\Delta})^{-1} R^{*'} + U^* U^{*'}$$

where, recall,

$\Delta_1, \dots, \Delta_r$  are the nonzero (and, hence, positive) characteristic roots of  $C$ ,

$$\underline{\Delta} = \text{diag}\{\Delta_1, \dots, \Delta_r\},$$

$R^*$  = a  $q_1 \times r$  matrix whose columns are orthonormal characteristic vectors associated with  $\Delta_1, \dots, \Delta_r$ , and

$U^*$  = a  $q_1 \times (q_1 - r)$  matrix whose columns are orthonormal characteristic vectors associated with the zero roots of  $C$ .

The use of this formula can be advantageous if  $T$  is to be computed for a large number of  $\gamma_1$  values, since  $R^*$ ,  $U^*$ , and  $\underline{\Delta}$  do not depend on  $\gamma$

and, hence, need be computed only once. Moreover, as can be seen from the results of Section III.F, the matrix  $T$  appears in algorithms 1-14 in the form  $TC$ , that is, what needs to be computed is not  $T$  itself, but rather quantities like  $TC$ . Since  $U^{*'}C = \emptyset$ , the computing formula for  $TC$  based on the diagonalization of  $C$  is then

$$TC = R^*(I + \gamma_1 \underline{\Delta})^{-1} R^{*'} C = R^*(I + \gamma_1 \underline{\Delta})^{-1} \underline{\Delta} R^{*'} \quad (6.1)$$

which does not require the computation of  $U^*$ . Further, the matrix  $I + \gamma_1 \underline{\Delta}$  is diagonal and, hence, is easy to invert. Note, however, that the computation of  $R^*$  and  $\underline{\Delta}$  can be quite costly if  $q = q_1$  is large. Thus, the use of formula (6.1) is likely to be advantageous only when the number of different  $\gamma_1$  values for which  $T$  must be evaluated is sufficiently large.

#### B. Algorithm Implementation

Table 6.1 lists various quantities that must be evaluated in computing each iterate. Two sets of formulas are given - the first set is to be used if  $C$  has been diagonalized, and the second set is to be used if  $C$  has not been diagonalized.

The matrix  $C$  and the vector  $\underline{q}$  can be formed by first solving the linear system

$$(X'X)(\underline{H}, \underline{a}) \equiv (X'Z, X'y) \quad (6.2)$$

Table 6.1. Representations of quantities needed in algorithms 1-14 in Table 3.1.

With the diagonalization of C	Without the diagonalization of C
$r$ = number of nonzero characteristic roots of C	$r$ = number of linearly independent columns of C
$S_1 = \tilde{\underline{x}}' \tilde{\underline{x}}$ , where $\tilde{\underline{x}} = (R^* \underline{\Delta}^{-1/2})' \underline{q}$	$S_1 = \tilde{\underline{s}}' \underline{q}$ , where $\tilde{\underline{s}}$ is any solution to $C \tilde{\underline{s}} = \underline{q}$
$S_2 = \underline{y}' (I - P_X) \underline{y} - S_1$	$S_2 = \underline{y}' (I - P_X) \underline{y} - S_1$
$\bar{\Delta} = \frac{1}{r} \sum_{i=1}^r \Delta_i$	$\bar{\Delta} = \frac{1}{r} \text{tr}(C)$
$\sum_{i=1}^r \frac{\Delta_i}{1 + \gamma_1 \Delta_i}$	$\text{tr}[(I + \gamma_1 C)^{-1} C]$
$\sum_{i=1}^r \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2}$	$\text{tr}\{[(I + \gamma_1 C)^{-1} C]^2\}$
$\sum_{i=1}^r \frac{\tilde{x}_i^2}{1 + \gamma_1 \Delta_i}$	$\underline{u}' \tilde{\underline{s}}$ , where $\underline{u}$ is any solution to $(I + \gamma_1 C) \underline{u} = \underline{q}$
$\sum_{i=1}^r \frac{\Delta_i \tilde{x}_i^2}{(1 + \gamma_1 \Delta_i)^2}$	$\underline{u}' \underline{u}$
$\sum_{i=1}^r \frac{\Delta_i^2 \tilde{x}_i^2}{(1 + \gamma_1 \Delta_i)^3}$	$\underline{u}' [(I + \gamma_1 C)^{-1} C] \underline{u}$

for  $H$  and  $\underline{a}$ , and by then putting

$$C = Z'Z - (Z'X)\underline{H}.$$

and

$$\underline{q} = Z'\underline{y} - (Z'X)\underline{a},$$

as observed by Harville and Fenech (1985). Further, the quantity  $\underline{y}'(I - P_X)\underline{y}$  can be computed from the formula

$$\underline{y}'(I - P_X)\underline{y} = \underline{y}'\underline{y} - \underline{a}'X'\underline{y}.$$

One way to solve non-full-rank linear systems, like (6.2) and  $C\underline{\xi} = \underline{q}$ , is to multiply both sides of the system by orthogonal Householder transformation matrices chosen to triangularize the coefficient matrix. More specifically, by following the prescription given by Kennedy and Gentle (1980), we could choose Householder transformation matrices so as to obtain an orthogonal matrix  $HQ$  such that  $HQ(X'X)$  is an upper-triangular matrix. If the linearly dependent rows and columns of  $HQ(X'X)$  (those rows and columns that have a zero element on the matrix diagonal) are deleted, as well as the corresponding rows in  $HQ(X'Z, X'y)$ , then a full-rank linear system is obtained. By augmenting the solution to this subsystem with rows of zeroes (corresponding to dependencies), we can obtain a solution  $(\underline{H}, \underline{a})$  to (6.2). Note that, in this approach to solving a linear system, the rank of the coefficient matrix is obtained as a by-product.

The process of implementing each iterative algorithm consists of the following five stages:

Stage 0. Determine  $n$ ,  $p$ ,  $q$ ,  $X$ ,  $Z$ ,  $\underline{y}$  ;

Stage 1. Form  $X'X$ ,  $X'Z$ ,  $X'\underline{y}$ ,  $Z'Z$ ,  $Z'\underline{y}$ , and  $\underline{y}'\underline{y}$  ;

Stage 2. Solve  $(X'X)(\underline{H}, \underline{a}) = (X'Z, X'\underline{y})$  for  $\underline{H}$  and  $\underline{a}$ , and then compute  $p^*$ ,  $\underline{y}'\underline{y} - \underline{a}'X'\underline{y}$ ,  $C$ , and  $q$  ;

Stage 3. Carry out the preliminary step, consisting of the computation of any of the quantities  $n-p^*$ ,  $r$ ,  $\bar{\Delta}$ ,  $S_1$ ,  $S_2$ , and  $\frac{S_1}{S_2}$  required by the algorithm and, if  $C$  is to be diagonalized, of the diagonalization of  $C$ ;

Stage 4. Carry out the iteration.

In conjunction with Stage 3, note that algorithms 1-3 and 8-14 do not require  $\bar{\Delta}$ , while algorithms 1-4 and 7-14 do not require  $\frac{S_1}{S_2}$  [see Section III.F].

In applying each algorithm to the four data sets, each algorithm was terminated as soon as 50 consecutive iterates were obtained that were identical up to 10 significant digits or, in any case, after 3000 iterations. If an iterate was obtained that was outside the parameter space, it was changed to a point near the boundary.

All computations were carried out in double precision, using the WATFIV compiler on the IBM AS/6 mainframe computer. CPU times were

measured by a FORTRAN subroutine written for the AS/6 computer at Iowa State University. This routine is influenced slightly by factors external to the task being executed. Thus, small fluctuations in time are possible over repetitions of the same task.

### C. The Data Sets

The four data sets were as follows:

Data set 1: [Harville and Fenech (1985, Table 1)]

These data consist of birth weights of 62 single-birth male lambs. They come from 5 distinct population lines of sheep (2 control lines and 3 selected lines). Each lamb is the progeny of one of 23 rams (i.e., sires), and each lamb had a different dam. Age of dam is recorded as belonging to one of three categories, numbered 1(1-2 years), 2(2-3 years), and 3(over 3 years).

Let  $y_{ijkd}$  represent the weights of the d-th of those lambs that are the offspring of the k-th sire in the j-th population line and of a dam belonging to the i-th age category. The assumed model for  $y_{ijkd}$  is the mixed linear model

$$y_{ijkd} = \mu + \delta_i + \pi_j + s_{jk} + e_{ijkd} ,$$

where the age effects ( $\delta_1, \delta_2, \delta_3$ ) and the line effects ( $\pi_1, \dots, \pi_5$ ) are fixed effects, where the sire (within line) effects ( $s_{11}, s_{12}, \dots, s_{58}$ ) are random effects that are distributed independently as  $N(0, \sigma_s^2)$ ,

and where the random errors ( $e_{1111}, e_{1121}, \dots, e_{3582}$ ) are distributed as  $N(0, \sigma_e^2)$  independently of each other and of the sire effects.

Some properties of the data set which are used by algorithms 1-14 are given below:

$$n = 62 ,$$

$$p = 9 ,$$

$$q = 23 ,$$

$$p^* = 7 ,$$

$$r = 18 ,$$

$$s_1 = 80.296 ,$$

$$s_2 = 102.235 ,$$

$$\bar{\Delta} = 2.2118 .$$

The individual characteristic values  $\Delta_1, \dots, \Delta_{18}$  of C, and the corresponding observed values of  $\tilde{t}_1, \dots, \tilde{t}_{18}$ , are

$\Delta$	$\tilde{t}$	$\Delta$	$\tilde{t}$	$\Delta$	$\tilde{t}$
.8400	2.9062	1.4078	1.2882	2.7482	-2.5893
.9027	.4505	1.7077	-2.5006	3.1505	-2.4773
1.0000	-4.8083	1.9329	-2.1365	3.3236	-3.0049
1.0750	-.7319	2.0000	1.1294	3.5644	1.5521
1.1644	-.7361	2.0000	-1.4106	4.2340	-1.8835
1.3456	1.2924	2.3293	1.2735	5.0875	.7676

Data set 2: [Dempster et al., (1984, Table 1)]

These data are taken from a study on rats whose purpose was to assess the effects of an experimental compound on pup weights. Thirty female rats were randomly allocated into 3 equal size treatment groups: control, low dose (of the experimental compound), and high dose. In the high dose group, one female failed to conceive, one cannibalized her litter, and one delivered one still-birth. Consequently, data from only 7 litters in the high dose group were available for analysis. Since litter size and the sex ratio differed from litter to litter, and since these two factors were thought to influence pup weight, Dempster et al., included them as covariates in modeling the data.

Let  $y_{ijp}$  represent the weight of the  $p$ -th of those pups that are offspring of the  $j$ -th dam in the  $i$ -th treatment group. The assumed model for  $y_{ijp}$  is the mixed linear model

$$y_{ijp} = \mu + \tau_i + d_{j(i)} + \beta_1 \ell + \beta_2 s + e_{ijp} ,$$

where the treatment effects ( $\tau_1, \tau_2, \tau_3$ ) are fixed effects, litter size ( $\ell$ ) and pup sex ( $s$ ) [recorded as 0 or 1] are treated as continuous variables, where the dam (with treatment) effects ( $d_{1(1)}, \dots, d_{7(3)}$ ) are random effects that are distributed independently as  $N(0, \sigma_d^2)$ , and where the random errors ( $e_{111}, \dots, e_{379}$ ) are distributed as  $N(0, \sigma_e^2)$  independently of each other and of the dam (within treatment) effects.

Some properties of the data set which are used by algorithms 1-14 are given below:

$$n = 322,$$

$$p = 7,$$

$$q = 27,$$

$$p^* = 5 ,$$

$$r = 23 ,$$

$$s_1 = 31.4180 ,$$

$$s_2 = 48.5761 ,$$

$$\bar{\Delta} = 11.8394 .$$

The individual characteristic values  $\Delta_1, \dots, \Delta_{23}$  of  $C$ , and the corresponding observed values of  $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_{23}$ , are

$\Delta$	$\tilde{\epsilon}$	$\Delta$	$\tilde{\epsilon}$	$\Delta$	$\tilde{\epsilon}$
2.6288	-.0608	11.7324	-.4176	14.0893	2.2226
3.5998	-.9469	12.1105	-2.0488	14.7507	1.0756
6.3169	-.0868	12.2723	.2039	14.7914	1.3732
8.2746	-.7431	12.9182	-.7243	15.4862	1.0530
8.8718	-1.2331	13.0000	.7218	16.0000	.8998
9.5766	-1.5266	13.3045	-.0329	16.8801	1.9158
9.9161	.9528	13.4899	.7820	17.5676	2.2703
10.7349	.5436	13.9927	-.3256		

Data set 3: [Davies and Goldsmith (1972, Table 6.3)]

These data come from an investigation to determine the intrabatch correlation between yields of a dyestuff. Six batches of an intermediate product were randomly chosen and 5 preparations of a dyestuff were made from each sample. The data points consisted of the "equivalent yields" of the 30 preparations.

Let  $y_{ij}$  represent the yield of the  $j$ -th dyestuff preparation from the  $i$ -th batch. The assumed model for  $y_{ij}$  is the random linear model

$$y_{ij} = \mu + b_i + e_{ij} ,$$

where the batch effects ( $b_1, \dots, b_6$ ) are random effects that are distributed independently as  $N(0, \sigma_b^2)$ , and where the random errors ( $e_{11}, \dots, e_{65}$ ) are distributed as  $N(0, \sigma_e^2)$  independently of each other and of the batch effects.

The equivalent yield of each preparation as grams of standard color are presented below:

Batch					
1	2	3	4	5	6
1545	1540	1595	1445	1595	1520
1440	1555	1550	1440	1630	1455
1440	1490	1605	1595	1515	1450
1520	1560	1510	1465	1635	1480
1580	1495	1560	1545	1625	1445

Some properties of the data set which are used by algorithms 1-14 are given below:

$$n = 30 ,$$

$$p = 1 ,$$

$$q = 6 ,$$

$$p^* = 1 ,$$

$$r = 5 ,$$

$$s_1 = 56357.00$$

$$s_2 = 58830.00$$

$$\bar{\Delta} = 5.00 .$$

The individual characteristic values  $\Delta_1, \dots, \Delta_5$  of C, and the corresponding observed values of  $\tilde{t}_1, \dots, \tilde{t}_5$ , are:

$\Delta$	$\tilde{t}$
5.00	109.1947
5.00	152.5000
5.00	36.3662
5.00	-140.8457
5.00	-4.2164

Data set 4: This data set is a modified version of Data Set 3, obtained by subtracting 50 from each yield in batch 3 and subtracting 100 from each yield in batch 5. The model for these data was taken to be the same as for Data Set 3.

After modifying Data Set 3 as indicated above, the resulting data set (Data Set 4) is given by:

Batch					
1	2	3	4	5	6
1545	1540	1545	1445	1495	1520
1440	1555	1500	1440	1530	1455
1440	1490	1555	1595	1415	1450
1520	1560	1460	1465	1535	1480
1580	1495	1510	1545	1525	1445

Some properties of the data set which are used by algorithms 1-14 are given below:

$$n = 30 ,$$

$$p = 1 ,$$

$$q = 6 ,$$

$$p^* = 1 ,$$

$$r = 5 ,$$

$$S_1 = 9357.50 ,$$

$$S_2 = 58830.00 ,$$

$$\bar{\Delta} = 5.00 .$$

The individual characteristic values  $\Delta_1, \dots, \Delta_5$  of  $C$ , and the corresponding observed values of  $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_5$  are the same as for Data Set 3.

#### D. Numerical Results

Tables 6.2a-6.2b list the first ten  $\gamma$ -iterates produced by each algorithm when applied to Data Set 1, using the starting value  $\gamma^{(0)} = (.01, 1.)'$ . The number of iterations that each algorithm required to achieve an accuracy of 1, 2, 3, 4, and 5 significant digits is given in Table 6.2c. Tables 6.3a-6.3c give the same information as Tables 6.2a-6.2c, but for a different starting value, namely, the starting value  $\gamma^{(0)} = (1., 1.)'$ . Information similar to that given for Data Set 1 in Tables 6.2 and 6.3 is given for Data Sets 2, 3, and 4 in Tables 6.4-6.5, 6.6-6.7, and 6.8-6.9, respectively.

Table 6.10 provides information on the computing time required by each algorithm when applied to each Data Set. This information consists of (i) the CPU time required to execute the preliminary step, and (ii) the CPU time required to complete one iteration, and is given both for the case where  $C$  is diagonalized and for the case that  $C$  is not diagonalized. The times reported are actually average times: the preliminary step times are averages of 10 or more observed times and the per-iteration times are averages over a sequence of at least 100 iterations.

Table 6.2a.  $\gamma_1$ -iterates for data set 1, using the initial value  $\underline{\gamma}^{(0)} = (.01, 1.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.099	.018	.100	.225	.139	.139	.070	.326	.113	.003	.003	.012	.003	.012
2	.191	.030	.158	.178	.174	.174	.165	.217	.156	.003	.003	.014	.003	.014
3	.252	.052	.174	.175	.175	.175	.175	.186	.169	.003	.003	.017	.003	.017
4	.236	.088	.175					.178	.173	.003	.003	.020	.003	.020
5	.185	.137						.175	.174	.003	.003	.024	.003	.024
6	.175	.169							.174	.003	.003	.028	.003	.028
7		.174							.175	.003	.003	.033	.003	.033
8		.175								.003	.003	.039	.003	.038
9										.003	.003	.045	.003	.044
10										.003	.003	.051	.003	.050

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.17459.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Table 6.2b.  $\sigma_2^2$ -iterates for data set 1, using the initial value  $\gamma^{(0)} = (.01, 1.)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.366	1.401	3.086	2.894	3.017	3.017	3.157	3.068	3.068	3.013	3.270	3.286	3.290	3.290
2	1.806	1.890	2.987	2.956	2.963	2.963	2.974	2.912	2.991	3.273	3.308	3.279	3.310	3.285
3	2.270	2.402	2.963	2.962	2.962	2.962	2.962	2.945	2.970	3.304	3.308	3.272	3.310	3.279
4	2.671	2.798	2.962					2.957	2.964	3.307	3.308	3.264	3.310	3.271
5	2.906	2.955						2.960	2.962	3.308	3.308	3.254	3.310	3.263
6	2.959	2.964						2.961		3.308	3.308	3.243	3.310	3.253
7	2.962	2.962						2.961		3.308	3.308	3.231	3.310	3.243
8								2.962		3.308	3.308	3.218	3.310	3.231
9										3.308	3.308	3.204	3.310	3.218
10										3.308	3.308	3.188	3.310	3.204

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2.9616.

<sup>b</sup>If after achieving accuracy to 4 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Table 6.2c. Number of iterations required, in the case of data set 1, to achieve accuracy to N significant digits, using the initial value  $\underline{\gamma}^{(0)} = (.01, 1.)'$ .<sup>a</sup>

Algorithm														
N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4(4)	6(4)	2(1)	1(1)	2(1)	2(1)	2(1)	2(1)	2(1)	985(1)	986(1)	25(1)	1001(1)	26(1)
2	6(6)	6(5)	3(2)	3(2)	2(1)	2(1)	3(2)	6(4)	3(2)	1015(949)	1015(949)	30(19)	1032(968)	32(21)
3	6(6)	8(6)	4(3)	3(2)	3(2)	3(2)	3(3)	6(4)	7(4)	1140(1050)	1140(1051)	56(38)	1167(1076)	60(42)
4	7(7)	8(7)	4(4)	3(3)	3(3)	3(3)	4(3)	8(8)	7(5)	1160(1086)	1160(1086)	60(45)	1189(1114)	64(50)
5	7(7)	8(7)	4(4)	4(3)	3(3)	3(3)	4(3)	11(8)	9(8)	1203(1161)	1203(1161)	68(61)	1235(1196)	74(66)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

Table 6.3a.  $\gamma_1$ -iterates for data set 1, using the initial value  $\gamma^{(0)} = (1., 1.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.956	.821	-.196 <sup>c</sup>	.521	.210	.210	-.146	-.196 <sup>c</sup>	.305	.585	.548	.571	.523	.628
2	.608	.585	-.196	.280	.174	.174	-.095	-.156	.207	.474	.452	.418	.453	.467
3	.331	.352	-.195	.189	.175	.175	.033	-.091	.184	.422	.409	.340	.418	.380
4	.224	.219	-.194	.175			.153	.023	.177	.390	.381	.294	.392	.326
5	.181	.179	-.193				.174	.113	.175	.366	.359	.264	.370	.290
6	.175	.175	-.191				.175	.155		.346	.341	.243	.351	.264
7			-.188					.169		.330	.325	.228	.336	.246
8			-.183					.173		.316	.312	.217	.322	.232
9			-.172					.174		.304	.301	.209	.311	.221
10			-.145					.174		.294	.291	.202	.300	.212

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.17459.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\gamma_1$ -iterate produced by the algorithm was less than  $-\frac{1}{\Delta^*} = - .197$ .

Table 6.3b.  $\sigma_2^2$ -iterates for data set 1, using the initial value  $\underline{y}^{(0)} = (1., 1.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.373	1.397	8.439	2.627	2.913	3.182	2.825	2.825	2.825	2.154	2.300	2.650	2.409	2.409
2	1.834	1.880	6.973	2.829	2.962	3.780	4.583	2.918	2.918	2.567	2.656	2.720	2.626	2.564
3	2.330	2.386	6.110	2.941		3.266	3.685	2.949	2.949	2.706	2.745	2.776	2.674	2.664
4	2.727	2.778	5.576	2.961		2.993	3.251	2.958	2.958	2.758	2.775	2.818	2.701	2.732
5	2.925	2.941	5.226	2.962		2.962	3.055	2.961	2.961	2.782	2.792	2.849	2.722	2.782
6	2.961	2.961	4.979				2.991	2.961	2.961	2.798	2.805	2.872	2.741	2.819
7	2.962	2.962	4.784				2.970	2.962	2.962	2.811	2.816	2.890	2.758	2.846
8			4.599				2.964			2.821	2.826	2.904	2.772	2.868
9			4.375				2.962			2.831	2.835	2.915	2.785	2.885
10			4.045							2.839	2.843	2.923	2.797	2.899

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2.9616.

<sup>b</sup>If after achieving accuracy to 4 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Table 6.3c. Number of iterations required, in the case of data set 1, to achieve accuracy to N significant digits, using the initial value  $\underline{\gamma}^{(0)} = (1., 1.)'$ .<sup>a</sup>

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4(4)	4(4)	14(12)	3(1)	1(1)	1(1)	4(3)	6(4)	2(1)	17(2)	17(2)	6(1)	18(2)	7(2)
2	6(6)	6(6)	15(14)	4(4)	2(2)	2(2)	5(4)	7(6)	6(4)	142(61)	141(60)	32(17)	153(70)	36(20)
3	6(6)	6(6)	16(15)	4(4)	3(2)	3(2)	6(5)	11(8)	7(4)	142(75)	141(74)	32(20)	153(86)	36(23)
4	7(7)	7(7)	16(16)	5(5)	3(2)	3(2)	6(5)	11(9)	8(7)	194(186)	194(185)	43(42)	210(206)	48(48)
5	7(7)	7(7)	16(16)	5(5)	3(3)	3(3)	6(6)	13(12)	11(8)	297(205)	296(205)	64(46)	321(266)	71(53)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

Table 6.4a.  $\gamma_1$ -iterates for data set 2, using the initial value  $\underline{\gamma}^{(0)} = (1., 1.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2.415	.942	-.056 <sup>c</sup>	.703	.584	.584	.587	.930	.577	1.315	1.388	.617	1.899	.633
2	5.038	.925	-.056	.600			.584	.578	.584	.856	.863	.588	.914	.592
3	10.23	.918	-.054	.584				.584		.672	.673	.584	.688	.585
4	20.58	.915	-.051							.614	.614		.620	.584
5	41.28	.914	-.047							.595	.595		.597	
6	82.67	.913	-.038							.588	.588		.589	
7	165.5	.913	-.022							.585	.585		.588	
8	331.0	.912	.005							.584	.584		.586	
9	662.1		.047										.585	
10	1324		.108										.584	

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.58384.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\gamma_1$ -iterate produced by the algorithm was less than  $-\frac{1}{\Delta^*} = -.057$ .

Table 6.4b.  $\sigma_2^2$ -iterates for data set 2, using the initial value  $\gamma^{(0)} = (1., 1.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2.250	2.247	1.654	.163	.165	.165	.165	.165	.165	.232	.220	.165	.161	.161
2	4.686	4.687	1.153	.165						.170	.168		.157	.164
3	9.540	9.551	.840							.165	.165		.161	.165
4	19.24	19.27	.629										.163	
5	38.64	38.72	.479										.164	
6	77.43	77.60	.371										.165	
7	155.0	155.4	.296											
8	310.2	310.9	.246											
9	620.5	621.9	.214											
10	1241	1244	.194											

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 0.16506.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Table 6.4c. Number of iterations required, in the case of data set 2, to achieve accuracy to N significant digits, using the initial value  $\underline{\gamma}^{(0)} = (1., 1.)'$ .<sup>a</sup>

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	* <sup>b</sup>	*	15(8)	2(1)	1(1)	1(1)	1(1)	2(1)	1(1)	4(1)	4(1)	1(1)	4(1)	1(1)
2	*	*	16(12)	3(3)	1(1)	1(1)	2(1)	2(1)	1(1)	8(2)	8(6)	3(3)	8(8)	4(4)
3	*	*	17(15)	3(3)	1(1)	1(1)	2(1)	3(1)	2(1)	8(3)	8(6)	3(3)	9(8)	4(4)
4	*	*	17(16)	4(3)	2(2)	2(2)	2(2)	4(1)	3(1)	12(8)	12(8)	5(4)	13(10)	6(5)
5	*	*	17(16)	4(4)	2(2)	2(2)	2(2)	4(3)	3(3)	13(8)	13(9)	6(4)	13(11)	7(6)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

<sup>b</sup>A \* indicates that the algorithm failed to converge.

Table 6.5a.  $\gamma_1$ -iterates for data set 2, using the initial value  $\gamma^{(0)} = (.2561172, .251014)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.575	$1.0 \times 10^6$ <sup>c</sup>	.372	$8.4 \times 10^5$ <sup>d</sup>	.573	.573	.579	.480	.597	.441	.444	.500	.434	.476
2	.730	$7.3 \times 10^5$	.484	$4.2 \times 10^5$	.584	.584	.584	.587	.584	.529	.530	.571	.520	.561
3	.904	$4.9 \times 10^5$	.559	$2.1 \times 10^5$				.584		.563	.564	.582	.559	.579
4	1.072	$3.3 \times 10^5$	.582	$1.0 \times 10^5$						.576	.576	.584	.574	.583
5	1.177	$2.2 \times 10^5$	.584	52790.						.581	.581		.580	.584
6	1.116	$1.4 \times 10^5$		26395.						.583	.583		.582	
7	.871	96965.		13198.						.583	.583		.583	
8	.691	64644.		6599.						.584	.584		.584	
9	.611	43096.		3300.										
10	.586	28731.		1650.										

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.58384.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\sigma_2^2$ -iterate produced by the algorithm was less than 0.

<sup>d</sup>The first  $\psi_1$ -iterate produced by the algorithm was less than 0.

Table 6.5b.  $\sigma_2^2$ -iterates for data set 2, using the initial value  $\underline{\gamma}^{(0)} = (.2561172, .251014)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.013	$1.0 \times 10^{-7}$ <sup>c</sup>	.170	.153	.165	.165	.165	.165	.165	.174	.172	.169	.176	.176
2	.020	$1.5 \times 10^{-7}$	.167	.153						.166	.166	.165	.168	.167
3	.029	$2.2 \times 10^{-7}$	.166	.153						.165	.165		.166	.165
4	.041	$3.4 \times 10^{-7}$	.165	.153									.166	
5	.059	$5.1 \times 10^{-7}$		.153									.165	
6	.082	$7.6 \times 10^{-7}$		.153										
7	.109	$1.1 \times 10^{-6}$		.153										
8	.137	$1.7 \times 10^{-6}$		.153										
9	.157	$2.6 \times 10^{-6}$		.153										
10	.164	$3.8 \times 10^{-6}$		.153										

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 0.16506.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\sigma_2^2$ -iterate produced by the algorithm was less than 0.

Table 6.5c. Number of iterations required, in the case of data set 2, to achieve accuracy to N significant digits, using the initial value  $\underline{\gamma}^{(0)} = (.2561172, .251014)'$ .<sup>a</sup>

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	9(9)	38(38)	3(1)	23(1)	1(1)	1(1)	1(1)	2(1)	1(1)	3(1)	3(1)	2(1)	3(1)	2(1)
2	11(11)	40(40)	4(1)	25(24)	2(1)	2(1)	1(1)	3(2)	2(2)	4(1)	4(1)	2(1)	5(2)	3(2)
3	11(11)	40(40)	5(4)	25(24)	2(1)	2(1)	2(1)	3(2)	2(2)	8(3)	8(3)	4(2)	8(5)	5(3)
4	12(12)	41(41)	5(4)	25(25)	2(2)	2(2)	2(2)	4(2)	3(2)	9(4)	9(4)	5(3)	9(6)	6(4)
5	12(12)	41(41)	6(5)	25(25)	2(2)	2(2)	2(2)	4(3)	4(2)	12(6)	12(6)	6(4)	13(9)	8(6)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

Table 6.6a.  $\gamma_1$ -iterates for data set 3, using the initial value  $\gamma^{(0)} = (5., 1000.)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	31.01	19.74	11.94	2.689	.720	.720	.720	-.199 <sup>c</sup>	.720	1.266	1.243	.890	1.298	1.031
2	64.89	31.85	25.39	1.547				.720		.911	.905	.755	.948	.805
3	131.8	53.81	52.17	1.003						.803	.801	.728	.827	.748
4	265.0	98.75	105.7	.779						.759	.758	.722	.774	.729
5	531.4	194.2	212.7	.724						.739	.739	.720	.748	.723
6	1064.	389.1	426.7	.720						.729	.729		.734	.721
7	2130.	779.4	854.8							.724	.724		.728	.720
8	4260.	1560.	1711.							.722	.722		.724	
9	8522.	3121.	3423.							.721	.721		.722	
10	17046	6243.	6848.							.720	.720		.721	

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.71965.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\gamma_1$ -iterate produced by the algorithm was less than  $-\frac{1}{\Delta^*} = -.200$ .

Table 6.6b.  $\sigma_2^2$ -iterates for data set 3, using the initial value  
 $\gamma^{(0)} = (5., 1000.)$ .<sup>a, b</sup>

Iteration	Algorithm						
	1	2	3	4	5	6	7
1	1227.48	1371.88	2060.64	2163.14	2451.25	2451.25	2451.25
2	1571.51	1791.28	2043.81	2251.12			
3	1859.02	2171.26	2036.04	2351.77			
4	2002.04	2393.84	2032.29	2425.45			
5	2027.89	2448.62	2030.45	2449.44			
6	2028.62	2451.24	2029.53	2451.24			
7	2028.62	2451.25	2029.08	2451.25			
8	2028.62		2028.85				
9	2028.62		2028.73				
10	2028.62		2028.68				

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2451.250.

<sup>b</sup>If after achieving accuracy to 6 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Algorithm						
8	9	10	11	12	13	14
2451.25	2451.25	2157.37	2197.28	2435.21	2103.37	2103.37
		2378.39	2392.29	2437.14	2288.06	2344.48
		2426.20	2430.11	2447.43	2367.15	2415.26
		2440.28	2441.45	2450.28	2407.09	2438.76
		2445.92	2446.34	2451.19	2427.84	2446.87
		2448.57	2448.76	2451.23	2438.77	2449.71
		2449.90	2449.99	2451.25	2444.57	2450.71
		2450.57	2450.61		2447.67	2451.06
		2450.91	2450.93		2449.33	2451.18
		2451.08	2451.09		2450.22	2451.23

Table 6.6c. Number of iterations required, in the case of data set 3, to achieve accuracy to N significant digits, using the initial value  $\gamma^{(0)} = (5., 1000.)'$ .<sup>a</sup>

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	* <sup>b</sup>	*	*	5(1)	1(1)	1(1)	1(1)	2(1)	1(1)	5(1)	5(1)	3(1)	5(1)	3(1)
2	*	*	*	5(6)	1(1)	1(1)	1(1)	2(1)	1(1)	7(8)	7(8)	4(4)	8(10)	5(7)
3	*	*	*	6(6)	1(1)	1(1)	1(1)	2(1)	1(1)	10(8)	10(8)	5(4)	11(10)	7(7)
4	*	*	*	6(6)	1(1)	1(1)	1(1)	2(1)	1(1)	13(8)	13(8)	7(5)	15(11)	9(7)
5	*	*	*	7(8)	1(1)	1(1)	1(1)	2(1)	1(1)	19(11)	19(11)	10(6)	21(14)	13(9)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

<sup>b</sup>A \* indicates that the algorithm failed to converge.

Table 6.7a.  $\gamma_1$ -iterates for data set 3, using the initial value  $\gamma^{(0)} = (.1, 1000.)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.183	.120	.243	$1.\times 10^7$ <sup>c</sup>	.720	.720	.720	1.619	.720	.097	.095	.245	.085	.226
2	.285	.155	.412	$1.\times 10^6$				.720		.132	.130	.464	.111	.407
3	.404	.215	.575	$5.\times 10^5$						.193	.190	.628	.154	.566
4	.533	.317	.683	$2.5\times 10^5$						.283	.279	.694	.217	.657
5	.649	.463	.717	$1.2\times 10^5$						.396	.392	.713	.301	.696
6	.709	.607	.720	62500.						.508	.505	.718	.400	.711
7	.719	.695		31250.						.595	.594	.719	.499	.717
8	.720	.718		15625.						.652	.651	.720	.579	.719
9		.720		7812.9						.684	.683		.636	.719
10				3906.6						.701	.701		.672	.720

<sup>a</sup>The REML estimate of  $\gamma_1$  is 0.71965.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\gamma_1$ -iterate produced by the algorithm exceeded  $1.\times 10^7$ .

Table 6.7b.  $\sigma_2^2$ -iterates for data set 3, using the initial value  
 $\gamma^{(0)} = (.1, 1000.)'$ .<sup>a, b</sup>

Iteration	Algorithm						
	1	2	3	4	5	6	7
1	1347.69	1371.88	2905.51	2028.62	2451.25	2451.25	2451.25
2	1753.30	1791.28	2663.85	2028.62			
3	2148.42	2171.26	2530.10	2028.62			
4	2408.35	2393.84	2468.66	2028.62			
5	2467.75	2448.62	2452.42	2028.62			
6	2454.10	2451.24	2451.26	2028.63			
7	2451.31	2451.25	2451.25	2028.63			
8	2451.25			2028.65			
9				2028.70			
10				2028.72			

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2451.250.

<sup>b</sup>If after achieving accuracy to 6 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Algorithm						
8	9	10	11	12	13	14
2451.25	2451.25	2884.81	2949.81	3068.70	3324.20	3324.20
		3064.26	3084.45	2675.29	3394.69	2940.89
		2946.61	2951.99	2507.05	3277.94	2669.25
		2768.64	2787.70	2464.20	3126.83	2535.93
		2647.80	2646.95	2454.39	2961.79	2482.21
		2553.86	2552.40	2452.03	2804.50	2462.31
		2501.38	2500.17	2451.45	2675.91	2455.16
		2475.23	2474.49	2451.30	2584.90	2452.63
		2462.78	2462.38	2451.26	2527.28	2451.74
		2456.86	2456.66	2451.25	2493.38	2451.42

Table 6.7c. Number of iterations required, in the case of data set 3, to achieve accuracy to N significant digits, using the initial value  $\gamma^{(0)} = (.1, 1000.)'$ .<sup>a</sup>

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	6(2)	7(2)	4(4)	25(1)	1(1)	1(1)	1(1)	2(1)	1(1)	8(8)	8(8)	4(4)	10(10)	4(5)
2	7(5)	8(6)	5(5)	25(25)	1(1)	1(1)	1(1)	2(1)	1(1)	13(8)	13(8)	6(4)	14(10)	7(5)
3	8(7)	9(6)	6(5)	26(25)	1(1)	1(1)	1(1)	2(1)	1(1)	17(11)	17(11)	8(5)	20(14)	10(8)
4	8(7)	10(6)	7(6)	26(25)	1(1)	1(1)	1(1)	2(1)	1(1)	23(15)	17(15)	11(7)	26(19)	14(10)
5	8(9)	10(6)	7(8)	27(26)	1(1)	1(1)	1(1)	2(1)	1(1)	23(16)	22(16)	11(8)	26(20)	14(11)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

Table 6.8a.  $\gamma_1$ -iterates for data set 4, using the initial value  $\underline{\gamma}^{(0)} = (4., 1000.)'$ .<sup>a,b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	8.81	6.26	8.40	1.946	-.047	-.047	-.047	-.199	-.047	.515	.505	.147	.542	.174
2	18.19	9.92	17.19	.920				-.047		.261	.261	.069	.284	.079
3	36.84	16.67	34.77	.408						.181	.182	.041	.192	.047
4	74.09	30.56	69.92	.153						.140	.141	.028	.146	.031
5	148.6	60.10	140.2	.029						.115	.115	.020	.119	.022
6	297.5	120.4	280.8	-.027						.098	.098	.014	.101	.016
7	595.4	241.2	562.0	-.045						.086	.086	.011	.088	.012
8	1191.	482.8	1124.	-.047						.077	.077	.008	.078	.009
9	2383.	965.9	2249.							.069	.069	.006	.070	.007
10	4766.	1932.	4499.							.063	.063	.005	.064	.005

<sup>a</sup>The REML estimate of  $\gamma_1$  is -0.047302.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Table 6.8b.  $\sigma_2^2$ -iterates for data set 4, using the initial value  
 $\gamma^{(0)} = (4., 1000.)'$ .<sup>a, b</sup>

Iteration	Algorithm						
	1	2	3	4	5	6	7
1	1332.39	1371.88	2036.13	2058.69	2451.25	2451.25	2451.25
2	1670.90	1791.28	2032.33	2086.24			
3	1920.60	2171.26	2030.47	2134.84			
4	2017.53	2393.84	2029.54	2211.51			
5	2028.49	2448.62	2029.08	2310.61			
6	2028.62	2451.24	2028.85	2400.79			
7	2028.62	2451.25	2028.74	2444.42			
8	2028.62		2028.68	2451.12			
9	2028.62		2028.65	2451.25			
10	2028.62		2028.64	2451.25			

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2451.250.

<sup>b</sup>If after achieving accuracy to 6 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Algorithm						
8	9	10	11	12	13	14
2451.25	2451.25	2153.77	2193.56	2428.05	2043.99	2043.99
		2315.72	2325.52	2304.10	2115.57	2200.99
		2315.44	2316.34	2309.16	2162.06	2259.75
		2307.48	2307.31	2318.63	2193.45	2289.90
		2304.21	2304.10	2326.34	2215.22	2307.87
		2303.75	2303.72	2332.16	2231.11	2319.55
		2304.54	2304.55	2336.53	2243.25	2327.56
		2305.85	2305.89	2339.84	2252.87	2333.28
		2307.37	2307.41	2342.36	2260.70	2337.46
		2308.92	2308.97	2344.29	2267.22	2340.59

Table 6.8c. Number of iterations required, in the case of data set 4, to achieve accuracy to N significant digits, using the initial value  $\gamma^{(0)} = (4., 1000.)'$ .<sup>a</sup>

N	Algorithm										11 <sup>b</sup>	12 <sup>b</sup>	13 <sup>b</sup>	14 <sup>b</sup>
	1	2	3	4	5	6	7	8	9	10 <sup>b</sup>				
1	* <sup>c</sup>	*	*	5(1)	1(1)	1(1)	1(1)	2(1)	1(1)	14(1)	14(1)	3(1)	14(1)	3(1)
2	*	*	*	8(8)	1(1)	1(1)	1(1)	2(1)	1(1)	198(1019)	198(1019)	10(18)	199(1399)	11(19)
3	*	*	*	8(8)	1(1)	1(1)	1(1)	2(1)	1(1)	2279(1919)	2279(1019)	20(18)	2282(1399)	20(19)
4	*	*	*	8(8)	1(1)	1(1)	1(1)	2(1)	1(1)	>3000(1699)	>3000(1679)	30(20)	>3000(2317)	30(21)
5	*	*	*	9(9)	1(1)	1(1)	1(1)	2(1)	1(1)	>3000(>3000)	>3000(>3000)	40(33)	>3000(>3000)	41(34)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

<sup>b</sup>The algorithm appeared to be converging to the  $\gamma$ -value (0., 2351.29)'.  
<sup>c</sup>A \* indicates that the algorithm failed to converge.

Table 6.9a.  $\gamma_1$ -iterates for data set 4, using the initial value  $\gamma^{(0)} = (.1, 1000.)'$ .<sup>a, b</sup>

Iteration	Algorithm													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.058	.055	-.199 <sup>c</sup>	.004	-.047	-.047	-.047	-.199 <sup>c</sup>	-.047	.049	.048	.054	.048	.056
2	-.002	.007	-.198	-.036				-.047		.044	.044	.034	.044	.036
3	-.031	-.028	-.196	-.046						.042	.042	.024	.042	.025
4	-.044	-.044	-.193	-.047						.040	.040	.017	.040	.018
5	-.047	-.047	-.187							.038	.038	.013	.038	.013
6			-.187							.036	.036	.009	.037	.010
7			-.164							.035	.035	.007	.035	.008
8			-.144							.033	.033	.006	.034	.006
9			-.119							.032	.032	.004	.032	.005
10			-.090							.031	.031	.003	.031	.004

<sup>a</sup>The REML estimate of  $\gamma_1$  is -0.047302.

<sup>b</sup>If after achieving accuracy to 3 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

<sup>c</sup>The first  $\gamma_1$ -iterate produced by the algorithm was less than  $-\frac{1}{\Delta^*} = -.200$ .

Table 6.9b.  $\sigma_2^2$ -iterates for data set 4, using the initial value  
 $\gamma^{(0)} = (.1, 1000.)'$ .<sup>a, b</sup>

Iteration	Algorithm						
	1	2	3	4	5	6	7
1	1365.21	1371.88	66563.	2344.61	2451.25	2451.25	2451.25
2	1780.73	1791.28	34883.	2421.72			
3	2158.59	2171.26	19020.	2448.88			
4	2387.17	2393.84	11051.	2451.23			
5	2447.84	2448.62	7010.08	2451.25			
6	2451.24	2451.24	4918.87				
7	2451.25	2451.25	3800.22				
8			3177.97				
9			2821.02				
10			2615.84				

<sup>a</sup>The REML estimate of  $\sigma_2^2$  is 2451.250.

<sup>b</sup>If after achieving accuracy to 6 significant digits, an algorithm maintained that accuracy over the remainder of the first 10 iterations, the entries for those iterates are omitted.

Algorithm						
8	9	10	11	12	13	14
2451.25	2451.25	2188.52	2229.50	2304.47	2243.74	2243.74
		2307.03	2312.90	2313.43	2289.22	2281.10
		2316.82	2317.41	2322.26	2292.84	2302.42
		2318.48	2318.57	2329.09	2295.32	2315.92
		2319.49	2319.55	2334.23	2297.57	2325.04
		2320.41	2320.46	2338.09	2299.61	2331.46
		2321.28	2321.32	2341.02	2301.49	2336.12
		2322.09	2322.13	2343.27	2303.23	2339.58
		2322.85	2322.90	2345.00	2304.83	2342.19
		2323.58	2323.62	2346.34	2306.32	2344.18

Table 6.9c. Number of iterations required, in the case of data set 4, to achieve accuracy to N significant digits, using the initial value  $\gamma^{(0)} = (.1, 1000.)'$ <sup>a</sup>.

N	Algorithm													
	1	2	3	4	5	6	7	8	9	10 <sup>b</sup>	11 <sup>b</sup>	12 <sup>b</sup>	13 <sup>b</sup>	14 <sup>b</sup>
1	5(2)	5(2)	12(12)	3(1)	1(1)	1(1)	1(1)	2(1)	1(1)	1(1)	1(1)	2(1)	1(1)	2(1)
2	5(6)	5(6)	13(12)	4(4)	1(1)	1(1)	1(1)	2(1)	1(1)	185(1005)	185(1005)	9(16)	186(1386)	9(18)
3	6(6)	6(6)	14(13)	4(4)	1(1)	1(1)	1(1)	2(1)	1(1)	2266(1005)	2266(1005)	19(16)	2268(1368)	19(18)
4	6(6)	6(6)	14(14)	5(4)	1(1)	1(1)	1(1)	2(1)	1(1)	>3000(1685)	>3000(1685)	29(19)	>3000(2303)	29(20)
5	6(6)	6(6)	15(14)	5(4)	1(1)	1(1)	1(1)	2(1)	1(1)	>3000(>3000)	>3000(>3000)	39(26)	>3000(>3000)	39(27)

<sup>a</sup>The first number of each entry represents the number of iterations required in the case of the  $\gamma_1$ -iterates, while the second (parenthesized) number represents the number of iterations required in the case of the  $\sigma_2^2$ -iterates.

<sup>b</sup>The algorithm appeared to be converging to the  $\gamma$ -value  $(0., 2351.29)'$ .

Table 6.10. Computing times (in  $\frac{1}{100} \times$  CPU seconds) with diagonalization (WD) and without diagonalization (WOD).

Algorithm	Data Set 1				Data Set 2				Data Sets 3 and 4			
	Preliminary Step		Per Iteration		Preliminary Step		Per Iteration		Preliminary Step		Per Iteration	
	WD	WOD	WD	WOD	WD	WOD	WD	WOD	WD	WOD	WD	WOD
1	353	87	.86	150	636	155	1.07	228	8.0	4.8	.57	6.9
2	353	87	.89	147	636	155	1.08	229	8.0	4.8	.62	6.9
3	353	87	.48	149	636	155	.58	228	8.0	4.8	.18	6.5
4	353	87	.48	148	636	155	.60	228	8.0	4.8	.19	6.6
5	353	87	.48	148	636	155	.59	227	8.0	4.8	.20	6.6
6	353	87	.49	148	636	155	.59	228	8.0	4.8	.20	6.6
7	353	87	.90	148	636	155	1.00	228	8.0	4.8	.59	7.1
8	353	87	.79	144	636	155	.88	224	8.0	4.8	.56	6.1
9	353	87	.96	144	636	155	1.06	221	8.0	4.8	.76	6.1
10	353	87	.37	92	636	155	.46	141	8.0	4.8	.15	4.6
11	353	87	.37	92	636	155	.45	140	8.0	4.8	.15	4.5
12	353	87	.36	93	636	155	.44	141	8.0	4.8	.15	4.4
13	353	87	.36	93	636	155	.44	142	8.0	4.8	.14	4.5
14	353	87	.36	93	636	155	.44	140	8.0	4.8	.14	4.5

## E. Discussion

In discussing the results reported in Section VI.D, we say that an algorithm has converged once its iterates agree to five significant digits with the REML estimate of  $\underline{\gamma}$ .

Data Sets 3 and 4 are such that an  $\text{ANOVA}(\sigma^2)$  exists with  $s = c+1$ . Consequently, when applied to Data Sets 3 and 4, the linearized Newton-Raphson methods (algorithms 5-7 in Table 3.1) converge to the REML estimate of  $\underline{\gamma}$  in a single iteration from any starting value. Likewise, when applied to these two data sets, the Method of Scoring algorithm 9 converges in a single iteration. By comparison, an  $\text{ANOVA}(\sigma^2)$  exists for Data Sets 1 and 2 but with  $s > c+1$ , so that one-step convergence cannot be expected from algorithms 5-7 or 9 when applied to Data Sets 1 and 2.

The one-step convergence of the linearized Newton-Raphson methods and the Method of Scoring, as applied to Data Sets 3 and 4, is in sharp contrast to the performance of the traditional Newton-Raphson methods (algorithms 1-4). For these data sets, algorithms 1-3 do not converge (Tables 6.6a-6.6c and 6.8a-6.8c), unless the starting value is relatively close to the REML estimate. Even with rather good starting values, algorithms 1-4 may require 5 or more iterations to achieve an accuracy of just one significant digit.

Algorithm 7 is a linearized version of algorithm 2 [the traditional Newton-Raphson method applied to  $L_{\underline{v}}(\underline{\sigma}; \underline{y})$ ]. When applied to Data Set 3, algorithm 7 achieved an accuracy of one significant digit in a single iteration from both starting values (Tables 6.4a-6.5c). In contrast,

algorithm 2 failed to converge from one starting value, and it required 38 iterations to achieve an accuracy of one significant digit from the other starting value. Moreover, after only 2 iterations, the accuracy of algorithm 7 had improved to five significant digits.

Algorithm 5 is a linearized version of algorithm 3 [the traditional Newton-Raphson method applied to  $L_c(\underline{\gamma}^+; \underline{y})$ ], and algorithm 6 is a linearized version of algorithm 4 [the traditional Newton-Raphson method applied to  $L_R(\underline{\psi}^+; \underline{y})$ ]. Algorithms 5 and 6 were consistently faster in achieving a given accuracy than were algorithms 3 and 4, respectively (Tables 6.2a-6.5c). Algorithms 5 and 6 consistently converged in 3 iterations or less, while algorithms 3 and 4 sometimes required 14 or more iterations to achieve an accuracy of one significant digit.

In summary, the linearization technique introduced in Section III.B.2 shows promise in improving the rate of convergence of the traditional Newton-Raphson method and also in improving the method's robustness to poor starting values. Moreover, once near the REML estimate, the linearized Newton-Raphson methods appear to converge just as rapidly, if not more rapidly, than the traditional versions.

In algorithm 1, the traditional Newton-Raphson method is applied to  $L_1(\underline{\gamma}; \underline{y})$ , while in algorithm 3, it is applied to  $L_c(\underline{\gamma}; \underline{y})$ . Both algorithms 1 and 3 sometimes diverged when applied to Data Sets 3 and 4 (Tables 6.6a-6.6c and 6.8a-6.8c), though, in one case in which algorithm 1 diverged when applied to Data Set 2, algorithm 3 converged in 17 iterations (Tables 6.4a-6.4c). In several cases, algorithm 3 achieved an accuracy of one significant digit at least 2 iterations prior to

algorithm 1 (Tables 6.2a-6.2c, 6.5a-6.5c, and 6.7a-6.7c). In these same cases, both algorithms took about 3 additional iterations to converge after achieving an accuracy of one significant digit. However, in some other cases, algorithm 1 achieved an accuracy of one significant digit and converged more rapidly than algorithm 3 (Tables 6.3a-6.3c and 6.9a-6.9c). In these latter two instances, the progress of algorithm 3 was deterred by a large first step that took it to the boundary of the parameter space.

The performance of the traditional Newton-Raphson method, as applied to the log-likelihood function, did not vary much with the parameterization. [Refer to the performance of algorithms 1 and 2.] Similarly, the performance of the linearized Newton-Raphson method as applied to the concentrated log-likelihood function did not vary much with the parameterization. [Refer to the performance of algorithms 5 and 6.]

In applying the traditional Newton-Raphson method to the concentrated log-likelihood function, the function was parameterized in terms of  $\gamma$  (algorithm 3) or in terms of  $\psi$  (algorithm 4). Algorithm 4 converged in two cases in which algorithm 3 failed to converge (Tables 6.6a-6.6c and 6.8a-6.8c). Further, in three other cases, algorithm 4 converged in significantly fewer iterations than algorithm 3 (Tables 6.3a-6.3c, 6.4a-6.4c, and 6.9a-6.9c). However, in various other cases, algorithm 3 was much faster to achieve an accuracy of one significant digit and to converge (Tables 6.5a-6.5c and 6.7a-6.7c). Overall, the performance of algorithm 4 seemed superior to that of algorithm 3.

In general, the Method of Scoring (algorithms 8 and 9) performed better when the likelihood function was parameterized in terms of  $\sigma$  (algorithm 9) than when parameterized in terms of  $\gamma$  (algorithm 8), though the difference in performance was not substantial.

Algorithms 10-14 represent five different versions of the Method of Successive Approximations, though two of them (algorithms 10 and 11) can also be derived from the EM algorithm. In every case, algorithms 10, 11, and 13 behaved similarly, as did algorithms 12 and 14. Algorithms 12 and 14 consistently required fewer iterations to achieve an accuracy of one significant digit and, subsequently, to converge than algorithms 10, 11, and 13.

As discussed in Chapter III, a common property of algorithms 10-14 is that if the  $\gamma_1$ -component of the starting value  $\gamma^{(0)}$  is positive, then the  $\gamma_1$ -component of every iterate is greater than or equal to zero. Thus, when applied to Data Set 4, for which the REML estimate  $-0.047$  is negative, the  $\gamma_1$ -component of the iterates generated by these algorithms approached zero rather than the REML estimate.

Overall, algorithms 5-9 (the linearized Newton-Raphson methods and the Method of Scoring algorithms) tended to achieve an accuracy of one significant digit in the fewest number of iterations. In general, the Method of Successive Approximations algorithms (including the EM algorithms) made relatively slow progress (e.g., Tables 6.2a-6.2c), though in some cases they fared better than the traditional Newton-Raphson methods (Tables 6.4a-6.6c and 6.8a-6.9c). Also, the Method of Scoring tended to make faster progress from poor starting values than the traditional Newton-Raphson method.

When judged on the basis of the number of iterations they required to converge, once an accuracy of one significant digit had been attained, algorithms 1-7 (the traditional and linearized Newton-Raphson methods) were all found to be very satisfactory. The Method of Scoring algorithms were only slightly less satisfactory in this regard. In contrast, the Method of Successive Approximations algorithms were very slow to converge.

While algorithms 1-9 generally required fewer iterations to converge than algorithms 10-14, they tended to require more computing time per iteration. The point at which the additional computing time per iteration offsets the faster convergence depends somewhat on whether or not the matrix  $C$  is diagonalized in the preliminary step.

If  $C$  was diagonalized in the preliminary step, then algorithms 3-6 and 10-14 required significantly less CPU time per iteration than algorithms 1, 2, 7, 8, and 9. When  $C$  was not diagonalized in the preliminary step, algorithms 10-14 required significantly less CPU time per iteration than algorithms 3-6, as well as algorithms 1-2 and 7-9.

We now quantify the effect of diagonalizing the matrix  $C$  in the preliminary step. When  $C$  is diagonalized, the CPU time required by the preliminary step is increased, while the CPU per iteration is decreased (Table 6.10). For each of the four data sets, Table 6.11 lists (in terms of the number of iterations) the point at which the savings per iteration from diagonalization is sufficient to offset the additional time expended in the preliminary step. For the cases considered here, this point was the same for all fourteen algorithms.

Table 6.11. Minimum number of iterations for diagonalization to be efficient.

Data Set	q	Number of Iterations
3,4	6	1
1	23	2
2	27	3

As expected, the minimum number of iterations for diagonalization to be efficient increases with  $q$ .

## VII. BIBLIOGRAPHY

- Anderson, T. W. 1973. Asymptotically Efficient Estimation of Covariance Matrices with Linear Structure. *Annals of Statistics* 1(1):135-141.
- Bard, Y. 1970. Comparison of Gradient Methods for the Solution of Nonlinear Parameter Estimation Problems. *SIAM Journal on Numerical Analysis* 7(1):157-186.
- Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York. 341 pp.
- Bertsekas, D. P. 1982a. Projected Newton Methods for Optimization Problems with Simple Constraints. *SIAM Journal on Control and Optimization* 20(2):221-246.
- Bertsekas, D. P. 1982b. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York. 395 pp.
- Box, M. J. 1966. A Comparison of Several Current Optimization Methods, and the Use of Transformations in Constrained Problems. *Computer Journal* 9:67-77.
- Brodlike, K. W. 1977. Unconstrained Minimization. Pages 229-268 in D. Jacobs, ed. *The State of the Art in Numerical Analysis*. Academic Press, New York. 978 pp.
- Brown, K. G. 1984. On Analysis of Variance in the Mixed Model. *Annals of Statistics* 12(4):1488-1499.
- Broyles, R. A. 1983. On the Convergence of the EM Algorithm. *Journal of the Royal Statistical Society, Series B* 45(1):47-50.
- Conte, S. D., and C. de Boor. 1972. *Elementary Numerical Analysis*. 2nd ed. McGraw-Hill Book Co., New York. 396 pp.
- Corbeil, R. R., and S. R. Searle. 1976a. Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. *Technometrics* 18(1):31-38.
- Corbeil, R. R., and S. R. Searle. 1976b. A Comparison of Variance Component Estimators. *Biometrics* 32:779-791.
- Davies, O. L., and P. L. Goldsmith. 1972. *Statistical Methods in Research and Production*. 4th ed. Hafner Publishing Co., New York. 478 pp.

- Dempster, A. P., N. M. Laird and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* 39:1-38.
- Dempster, A. P., M. R. Selwyn, C. M. Patel, and A. J. Roth. 1984. Statistical and Computational Aspects of Mixed Model Analysis. *Applied Statistics* 33(2):203-214.
- Dixon, L. C. W. 1972. Quasi-Newton Algorithms Generate Identical Points. *Mathematical Programming* 2:383-387.
- Dixon, L. C. W., J. Gomulka, and G. P. Szegö. 1975. Towards a Global Optimisation Technique. Pages 29-54 in L. C. W. Dixon, and G. P. Szegö, eds. *Towards Global Optimisation*. North-Holland Publishing Co., Amsterdam. 472 pp.
- Falconer, D. S. 1976. *Introduction to Quantitative Genetics*. 8th ed. The Ronald Press Co., New York. 365 pp.
- Fenech, A. P., and D. A. Harville. 1985. Confidence Regions for Variance Components in Unbalanced Mixed Linear Models. Unpublished manuscript.
- Ferguson, T. S. 1967. *Mathematical Statistics - A Decision Theoretic Approach*. Academic Press, New York. 396 pp.
- Fisher, R. A. 1925. *Statistical Methods for Research Workers*. 1st ed. Oliver and Boyd, Edinburgh. 239 pp.
- Fletcher, R. 1980. *Practical Methods of Optimization, Volume 1: Unconstrained Optimization*. John-Wiley and Sons, New York.
- Fletcher, R., and M. J. D. Powell. 1963. A Rapidly Convergent Descent Method for Minimization. *Computer Journal* 6:163-168.
- Gill, P. E., and W. Murray. 1974. Newton-type Methods for Linearly Constrained Optimization. Pages 39-66 in P. E. Gill, and W. Murray, eds. *Numerical Methods for Constrained Optimization*. Academic Press, New York. 283 pp.
- Gill, P. E., W. Murray, and M. H. Wright. 1981. *Practical Optimization*. Academic Press, New York. 401 pp.
- Graybill, F. A. 1983. *Matrices with Applications in Statistics*. 2nd ed. Wadsworth Publishing Co., Belmont, California. 461 pp.
- Greenstadt, J. L. 1967. On the Relative Efficiencies of Gradient Methods. *Mathematics of Computation* 21:360-367.

- Hartley, H. O., and J. N. K. Rao. 1967. Maximum Likelihood Estimation for the Mixed Analysis of Variance Model. *Biometrika* 54(1):93-108.
- Harville, D. A. 1974. Bayesian Inference for Variance Components Using Only Error Contrasts. *Biometrika* 61:383-385.
- Harville, D. A. 1975. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Technical Report No. 75-0175. (Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio.)
- Harville, D. A. 1976. Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects. *Annals of Statistics* 4(2):384-395.
- Harville, D. A. 1977. Maximum likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association* 72(358):320-340.
- Harville, D. A., and A. P. Fenech. 1985. Confidence Intervals for a Variance Ratio, or for Heritability, in an Unbalanced Mixed Linear Model. *Biometrics* 41(1):137-152.
- Hemmerle, W. J., and H. O. Hartley. 1973. Computing Maximum Likelihood Estimates for the Mixed A.O.V. Model Using the W. Transformation. *Technometrics* 15(4):819-831.
- Henderson, C. R. 1953. Estimation of Variance and Covariance Components. *Biometrics* 9:226-252.
- Henderson, C. R. 1963. Selection Index and Expected Genetic Advance. Pages 141-163 in *Statistical Genetics and Plant Breeding*. National Academy of Sciences - National Research Council Publication No. 982.
- Henderson, C. R. 1973. Sire Evaluation and Genetic Trends. Pages 10-41 in *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, Champaign, Illinois: American Society of Animal Science. 104 pp.
- Henderson, C. R. 1980. A Simple Method for Unbiased Estimation of Variance Components in the Mixed Model. Mimeo of talk presented at 72nd annual meeting of the American Society of Animal Science, July 27-30, 1980, Cornell University, Ithaca, New York. Its abstract appears as Abstract No. 34 on p. 119 in *Journal of Animal Science*, Vol. 52, Supplement 1 (Abstracts).
- Hocking, R. R., and M. H. Kutner. 1975. Some Analytical and Numerical Comparisons of Estimators for the Mixed A.O.V. Model. *Biometrics* 31:19-27.

- Jennrich, R. I., and P. F. Sampson. 1976. Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation. *Technometrics* 18(1):11-17.
- Kempthorne, O. 1973. *An Introduction to Genetic Statistics*. 3rd ed. The Iowa State University Press, Ames, Iowa. 545 pp.
- Kennedy, W. J., and J. E. Gentle. 1980. *Statistical Computing*. Marcel Dekker, New York. 591 pp.
- Laird, N. M. 1982. Computation of Variance Components Using the EM Algorithm. *Journal of Statistical Computation and Simulation* 14:295-303.
- Lenard, M. L. 1979. A Computation Study of Active Set Strategies in Nonlinear Programming with Linear Constraints. *Mathematical Programming* 16:81-97.
- Levenberg, K. 1944. A Method for the Solution of Certain Problems in Least Squares. *Quarterly of Applied Mathematics* 2(2):164-168.
- Marquardt, D. 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics* 11(2):431-441.
- Miller, J. J. 1973. Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance. Technical Report No. 12. (Department of Statistics, Stanford University, Stanford, California.)
- Miller, J. J. 1979. Maximum Likelihood Estimation of Variance Components - A Monte Carlo Study. *Journal of Statistical Computation and Simulation* 8:175-190.
- Murray, W. 1972. Second Derivative Methods. Pages 57-71 in W. Murray, ed. *Numerical Methods for Unconstrained Optimization*. Academic Press, New York. 144 pp.
- Patterson, H. D., and R. Thompson. 1971. Recovery of Interblock Information When Block Sizes are Unequal. *Biometrika* 58(3):545-554.
- Powell, M. J. D. 1974. Introduction to Constrained Optimization. Pages 1-29 in P. E. Gill, and W. Murray, eds. *Numerical Methods for Constrained Optimization*. Academic Press, New York. 283 pp.
- Rao, C. R. 1965. *Linear Statistical Inference*. John Wiley and Sons, New York. 522 pp.

- Rao, C. R. 1970. Estimation of Heteroscedastic Variances in a Linear Model. *Journal of the American Statistical Association* 65(329):161-172.
- Rao, C. R. 1971. Minimum Variance Quadratic Unbiased Estimation of Variance Components. *Journal of Multivariate Analysis* 1:257-275.
- Schnabel, R. B. 1982. Unconstrained Optimization in 1981. Pages 3-16 in M. J. D. Powell, ed. *Nonlinear Optimization 1981*. Academic Press, New York. 559 pp.
- Searle, S. R. 1971. *Linear Models*. John Wiley and Sons, New York. 532 pp.
- Searle, S. R. 1974. Prediction, Mixed Models, and Variance Components. Pages 229-266 in F. Proschan and R. J. Sofling, eds. *Reliability and Biometry*. Society for Industrial and Applied Mathematics, Philadelphia. 815 pp.
- Searle, S. R. 1979. Notes on Variance Component Estimation - A Detailed Account of Maximum Likelihood and Kindred Methodology. Biometrics Unit, New York State College of Agriculture and Life Sciences, Cornell University, Ithaca, New York.
- Snedecor, G. W., and W. G. Cochran. 1967. *Statistical Methods*. 6th ed. The Iowa State University Press, Ames, Iowa. 593 pp.
- Sundberg, R. 1974. Maximum Likelihood Theory for Incomplete Data from an Exponential Family. *Scandinavian Journal of Statistics* 1:49-58.
- Sundberg, R. 1976. An Iterative Method for Solution of the Likelihood Equations for Incomplete Data from Exponential Families. *Communications in Statistics: Simulation and Computation* B5(1):55-64.
- Thompson, W. A., Jr. 1962. The Problem of Negative Estimates of Variance Components. *Annals of Mathematical Statistics* 33:273-289.
- Vandaele, W. H., and S. R. Chowdhury. 1971. A Revised Method of Scoring. *Statistica Neerlandica* 25:101-112.
- Westlake, J. R. 1968. *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations*. John Wiley and Sons, New York. 171 pp.
- Wu, C. F. Jeff. 1983. On the Convergence Properties of the EM Algorithm. *Annals of Statistics* 11(1):95-103.

## VIII. ACKNOWLEDGEMENTS

I wish to express my extreme gratitude to my major professor, Dr. David A. Harville, for his patient guidance over the last 2½ years. His insights contributed greatly to the content of this dissertation, and his demand for a precise and concise writing style has improved its presentation.

On a more general note, I would like to thank the entire faculty of the Department of Statistics at Iowa State University for providing me the opportunity to pursue an advanced degree in statistics, and for making my stay here a very rewarding experience. Particular mention is due Dr. David F. Cox, my supervisor in the Agriculture Experiment Station. I learned a great deal from him about statistics and consulting. Moreover, it was just a pleasure to know him.

I would also like to acknowledge my fellow classmates who provided an ear to listen when times were tough, and with whom I shared many good moments. My friends, Jan van Schaik and Daniel R. Jeske in particular, helped me to laugh and to keep things in proper perspective. I won't forget them.

To my mom, dad, brothers, and sisters, I owe a special "thank you" for their letters and calls of support and encouragement. I needed them.

To my wife (Annie) and my son (Ryan), I dedicate this dissertation. Their unfailing love and support allowed me to work long hours while they endured sacrifices at home.

Lastly, the devotion of my typist, Mrs. Darlene Wicks, exceeded all my expectations. Her excellent work is greatly appreciated.

## IX. APPENDIX

A. Derivation of the Partial Derivatives of the  
Log-Likelihood Function With Respect to  
 $\sigma_1^2, \dots, \sigma_{c+1}^2$

1. General case

According to result (2.31), we have that

$$\begin{aligned} \frac{\partial L_v}{\partial \sigma_{c+1}^2} &= \sum_{j=1}^{c+1} \frac{\partial L_1}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \sigma_{c+1}^2} \\ &= \frac{\partial L_1}{\partial \sigma_{c+1}^2} - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \frac{L_1}{\partial \gamma_j} \end{aligned} \quad (9.1)$$

$$= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \underline{y}' S (\underline{y} - \underline{z} \tilde{\underline{b}})]$$

$$- \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \left[ - \left(\frac{1}{2}\right) \text{tr}(G_{jj}) + \left(\frac{1}{2}\right) \sigma_{c+1}^2 \tilde{\underline{v}}_j' \tilde{\underline{v}}_j \right]$$

[using parts (i) and (ii) of Lemma 2.5]

$$= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \text{tr}(G_{jj})]$$

$$+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [\underline{y}' S (\underline{y} - \underline{z} \tilde{\underline{b}}) - \sum_{j=1}^c \sigma_j^2 \tilde{\underline{v}}_j' \tilde{\underline{v}}_j] .$$

Further,

$$\begin{aligned}
\underline{y}'S(\underline{y} - \underline{z}\tilde{\underline{b}}) &= \sum_{j=1}^c \sigma_j^2 \tilde{\underline{v}}_j' \tilde{\underline{v}}_j \\
&= \underline{y}'S(\underline{y} - \underline{z}\tilde{\underline{b}}) - \tilde{\underline{v}}'D\tilde{\underline{v}} \\
&= \underline{y}'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) - \tilde{\underline{v}}'D\tilde{\underline{v}} \text{ [using (2.14)]} \\
&= (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) \\
&\quad + \tilde{\underline{\alpha}}'X'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) \\
&\quad + \tilde{\underline{b}}'Z'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) - \tilde{\underline{v}}'D\tilde{\underline{v}} \\
&= (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) \\
&\quad + (\tilde{\underline{\alpha}}'X' + \tilde{\underline{b}}'Z')V^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}}) - \tilde{\underline{v}}'D\tilde{\underline{v}} \\
&\quad \text{[using part (v) of Lemma 2.4]} \\
&= (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'R^{-1}(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})
\end{aligned}$$

$$\begin{aligned}
& + \underline{\tilde{\alpha}}' \underline{X}' \underline{V}^{-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}) \\
& + \underline{\tilde{b}}' \underline{Z}' \underline{V}^{-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}) - \underline{\tilde{v}}' \underline{D} \underline{\tilde{v}} \\
& = (\underline{y} - \underline{X} \underline{\tilde{\alpha}} - \underline{Z} \underline{\tilde{b}})' \underline{R}^{-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}} - \underline{Z} \underline{\tilde{b}}) \\
& \quad [\text{since } \underline{\tilde{\alpha}}' \underline{X}' \underline{V}^{-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}) = 0, \text{ and} \\
& \quad \underline{\tilde{b}}' \underline{Z}' \underline{V}^{-1} (\underline{y} - \underline{X} \underline{\tilde{\alpha}}) = \underline{\tilde{b}}' \underline{\tilde{v}} - (\underline{D} \underline{\tilde{v}})' \underline{\tilde{v}} = \underline{\tilde{v}}' \underline{D} \underline{\tilde{v}}].
\end{aligned}$$

Thus ,

$$\begin{aligned}
\frac{\partial L_v}{\partial \sigma_{c+1}^2} &= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \text{tr}(G_{jj})] \\
&+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - \underline{X} \underline{\tilde{\alpha}} - \underline{Z} \underline{\tilde{b}})' (\underline{y} - \underline{X} \underline{\tilde{\alpha}} - \underline{Z} \underline{\tilde{b}}) .
\end{aligned}$$

Also, for  $i = 1, \dots, c$  ,

$$\begin{aligned}
\frac{L_v}{\partial \sigma_i^2} &= \sum_{j=1}^{c+1} \frac{L_1}{\partial \gamma_j} \frac{\partial \gamma_j}{\partial \sigma_i^2} [\text{using (2.31)}] \\
&= \frac{1}{\sigma_{c+1}^2} \frac{L_1}{\partial \gamma_i} \\
&= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [\text{tr}(G_{ii}) - \sigma_{c+1}^2 \underline{\tilde{v}}_i' \underline{\tilde{v}}_i] \\
& \quad [\text{using part (ii) of Lemma 2.5}].
\end{aligned} \tag{9.2}$$

As a further consequence of result (9.2),

$$\begin{aligned}
 \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_k^2} &= \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial}{\partial \sigma_k^2} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \\
 &= \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^{c+1} \left[ \frac{\partial}{\partial \gamma_j} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \frac{\partial \gamma_j}{\partial \sigma_k^2} \\
 &\quad \text{[using the chain rule of calculus]} \\
 &= \frac{1}{\sigma_{c+1}^4} \left[ \frac{\partial}{\partial \gamma_k} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \quad \text{[using (2.32)]} \\
 &= \frac{1}{\sigma_{c+1}^4} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_k} \quad (i, k = 1, \dots, c) . \tag{9.3}
 \end{aligned}$$

Thus, for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ ,

$$\begin{aligned}
 (-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} &= \frac{1}{\sigma_{c+1}^4} \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] \\
 &= \frac{1}{\sigma_{c+1}^4} \left[ -\text{tr}(G_{ii}^2) + 2 \sigma_{c+1}^2 \tilde{v}_i' G_{ii} \tilde{v}_i \right] , \\
 (-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{1}{\sigma_{c+1}^4} \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right]
 \end{aligned}$$

$$= \frac{1}{\sigma_{c+1}^4} \left[ -\operatorname{tr}(G_{ij} G_{ji}) + 2 \sigma_{c+1}^2 \tilde{y}_i' G_{ij} \tilde{y}_j \right],$$

$$(-2)\mathfrak{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_{c+1}^4} \left[ (-2)\mathfrak{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_i} \right] \right]$$

$$= \frac{1}{\sigma_{c+1}^4} \operatorname{tr}(G_{ii}^2),$$

$$(-2)\mathfrak{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right] = \frac{1}{\sigma_{c+1}^4} \left[ (-2)\mathfrak{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] \right] \quad (9.4)$$

$$= \frac{1}{\sigma_{c+1}^4} \operatorname{tr}(G_{ij} G_{ji}).$$

Further, using (9.2),

$$\frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} = -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i} + \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right]$$

$$= -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i}$$

$$+ \frac{1}{\sigma_{c+1}^2} \left[ \frac{\partial}{\partial \gamma_{c+1}} \left( \frac{\partial L_1}{\partial \gamma_i} \right) - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \left( \frac{1}{\sigma_{c+1}^2} \right) \left[ \frac{\partial}{\partial \delta_j} \left( \frac{\partial L_1}{\partial \gamma_i} \right) \right] \right]$$

$$= -\frac{1}{\sigma_{c+1}^4} \frac{\partial L_1}{\partial \gamma_i} + \frac{1}{\sigma_{c+1}^2} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_{c+1}} - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \frac{\sigma_j^2}{\sigma_{c+1}^2} \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j}$$

$$(i = 1, \dots, c) . \quad (9.5)$$

Substituting from parts (ii) and (iv)-(vi) of Lemma 2.5 into expression (9.5), we find that

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} = -\frac{1}{\sigma_{c+1}^4} [\text{tr}(G_{ii}) - 2 \sigma_{c+1}^2 \tilde{v}_i' \tilde{v}_i]$$

$$- \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \left[ \frac{1}{\sigma_{c+1}^2} \right] \left[ -\text{tr}(G_{ij} G_{ji}) \right.$$

$$\left. + 2 \sigma_{c+1}^2 \tilde{v}_i' G_{ij} \tilde{v}_j \right]$$

$$(i = 1, \dots, c) .$$

Also, making use of (9.5), we find that,

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} \right] = 2 \left[ \frac{1}{\sigma_{c+1}^4} \right] \left[ \mathbb{E} \left[ \frac{\partial L_1}{\partial \gamma_i} \right] \right]$$

$$+ \frac{1}{\sigma_{c+1}^2} \left[ (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \sigma_{c+1}^2} \right] \right]$$

$$- \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \sigma_j^2 \left[ \frac{1}{\sigma_{c+1}^2} \right] \left[ (-2) \mathbb{E} \left[ \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right] \right] .$$

Since

$$\mathbb{E} \left[ \frac{\partial L_1}{\partial \gamma_i} \right] = \mathbb{E} \left[ - \left( \frac{1}{2} \right) \text{tr}(G_{ii}) + \left( \frac{1}{2} \right) \sigma_{c+1}^2 \tilde{\gamma}'_{-i} \tilde{\gamma}_{-i} \right]$$

$$= - \left( \frac{1}{2} \right) \text{tr}(G_{ii}) + \left( \frac{1}{2} \right) \sigma_{c+1}^2 \mathbb{E}(\tilde{\gamma}'_{-i} \tilde{\gamma}_{-i})$$

$$= - \left( \frac{1}{2} \right) \text{tr}(G_{ii}) + \left( \frac{1}{2} \right) \sigma_{c+1}^2 \mathbb{E} \left[ (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_i} \right]$$

[using part (iv) of Lemma 2.5]

$$= - \left( \frac{1}{2} \right) \text{tr}(G_{ii}) + \left( \frac{1}{2} \right) \sigma_{c+1}^2 \left[ \frac{1}{\sigma_{c+1}^2} \text{tr}(G_{ii}) \right]$$

[using part (viii) of Lemma 2.5]

$$= 0 \quad (i = 1, \dots, c) ,$$

it follows that, for  $i = 1, \dots, c$  ,

$$\begin{aligned}
(-2) \mathfrak{E} \left( \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) &= \frac{1}{\sigma_{c+1}^2} \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_{c+1}^2} \right) \right] \\
&\quad - \frac{1}{\sigma_{c+1}^6} \sum_{j=1}^c \sigma_j^2 \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_j} \right) \right] \\
&= \frac{1}{\sigma_{c+1}^4} \text{tr}(G_{ii}) - \frac{1}{\sigma_{c+1}^6} \sum_{j=1}^c \sigma_j^2 \text{tr}(G_{ij} G_{ji}) .
\end{aligned} \tag{9.6}$$

Since

$$\frac{\partial L_v}{\partial \sigma_{c+1}^2} = \frac{\partial L_1}{\partial \sigma_{c+1}^2} - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} ,$$

we find, using the chain rule, that

$$\begin{aligned}
\frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_{c+1}} \right) + 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
&\quad - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \left[ \frac{\partial}{\partial \sigma_{c+1}^2} \left( \frac{\partial L_1}{\partial \gamma_j} \right) \right] \\
&= \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \frac{\partial^2 L_1}{\partial \gamma_j \partial \sigma_{c+1}^2}
\end{aligned}$$

$$\begin{aligned}
& + 2 \left[ \frac{1}{\sigma_{c+1}^6} \right] \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sigma_j^2 \left\{ \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \gamma_j} - \frac{1}{\sigma_{c+1}^4} \sum_{k=1}^c \sigma_k^2 \frac{\partial^2 L_1}{\partial \gamma_k \partial \gamma_j} \right\} \\
& = \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} - 2 \left[ \frac{1}{\sigma_{c+1}^4} \right] \sum_{j=1}^c \sigma_j^2 \frac{\partial^2 L_1}{\partial \gamma_j \partial \sigma_{c+1}^2} \\
& + 2 \left[ \frac{1}{\sigma_{c+1}^6} \right] \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j} \\
& + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \left[ \sum_{k=1}^c \sigma_k^2 \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right]
\end{aligned} \tag{9.7}$$

Hence,

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= (-2) \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \\
&- 2 \left[ \frac{1}{\sigma_{c+1}^4} \right] \sum_{j=1}^c \sigma_j^2 \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_j \partial \sigma_{c+1}^2} \right] \\
&- 4 \left[ \frac{1}{\sigma_{c+1}^6} \right] \sum_{j=1}^c \sigma_j^2 \frac{\partial L_1}{\partial \gamma_j}
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left[ (-2) \frac{\partial^2 L_1}{\partial \gamma_j \partial \gamma_k} \right] \\
& = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \underline{Z} \underline{\tilde{b}})] \\
& \quad - 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}_{-j}' \tilde{y}_{-j} \\
& \quad - 4 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \left[ -\left(\frac{1}{2}\right) \text{tr}(G_{jj}) + \left(\frac{1}{2}\right) \sigma_{c+1}^2 \tilde{y}_{-j}' \tilde{y}_{-j} \right] \\
& \quad + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left[ -\text{tr}(G_{jk} G_{kj}) \right. \\
& \quad \quad \left. + 2 \sigma_{c+1}^2 \tilde{y}_{-j}' G_{jk} \tilde{y}_{-k} \right] \\
& \quad \quad \quad \text{[using parts (ii)-(v) of Lemma 2.5]}
\end{aligned}$$

$$\begin{aligned}
& = - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \underline{Z} \underline{\tilde{b}})] \\
& \quad - 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}_{-j}' \tilde{y}_{-j}
\end{aligned}$$

$$\begin{aligned}
& + 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \operatorname{tr}(G_{jj}) \\
& + \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left[ -\operatorname{tr}(G_{jk} G_{kj}) + 2 \sigma_{c+1}^2 \tilde{v}_j' G_{jk} \tilde{v}_k \right].
\end{aligned}$$

Further,

$$\begin{aligned}
(-2) \mathfrak{E} \left( \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) &= (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) \\
&- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_j \partial \sigma_{c+1}^2} \right) \right] \\
&+ 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 \left[ (-2) \mathfrak{E} \left( \frac{\partial L_1}{\partial \gamma_j} \right) \right] \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \left[ (-2) \mathfrak{E} \left( \frac{\partial^2 L_1}{\partial \gamma_i \partial \gamma_k} \right) \right]. \quad (9.8)
\end{aligned}$$

Recalling that  $\mathfrak{E} \left( \frac{\partial L_1}{\partial \gamma_i} \right) = 0$  ( $i = 1, \dots, c$ ), and making use of parts

(vii)-(x) of Lemma 2.5, we find that

$$\begin{aligned}
(-2)\mathfrak{E}\left(\frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2}\right) &= \frac{1}{\sigma_{c+1}^4} (n-p^*) - 2 \left(\frac{1}{\sigma_{c+1}^6}\right) \sum_{j=1}^c \sigma_j^2 \operatorname{tr}(G_{jj}) \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{k=1}^c \sigma_k^2 \operatorname{tr}(G_{jk} G_{kj}).
\end{aligned}$$

## 2. Special case

Substituting expressions (2.43) into formulas (2.33), (2.34), (2.38), and (2.39), we find that,

$$\begin{aligned}
\frac{\partial L_v}{\partial \sigma_{c+1}^2} &= - \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^2} [n - p^* - q + \operatorname{tr}(T)] \\
&+ \left(\frac{1}{2}\right) \frac{1}{\sigma_{c+1}^4} (\underline{y} - X\underline{\alpha} - Z\underline{b})' (\underline{y} - X\underline{\alpha} - Z\underline{b})
\end{aligned}$$

$$[\text{since } \sum_{j=1}^c \operatorname{tr}(T_{jj}) = \operatorname{tr}(T)] ,$$

$$\frac{\partial L_v}{\partial \sigma_i^2} = - \left(\frac{1}{2}\right) \frac{1}{\sigma_i^2} [q_i - \operatorname{tr}(T_{ii})] + \left(\frac{1}{2}\right) \underline{\tilde{y}}_i' \underline{\tilde{y}}_i, \text{ if } \sigma_i^2 \neq 0 ,$$

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} = \frac{1}{\sigma_{c+1}^4} \left\{ - \operatorname{tr}[\sigma_{c+1}^2 \left(\frac{1}{\sigma_i^2}\right) (I - T_{ii})]^2 \right.$$

$$\begin{aligned}
& + 2 \sigma_{c+1}^2 \tilde{v}_i' \sigma_{c+1}^2 \left(\frac{1}{\sigma_i^2}\right) (I - T_{ii}) \tilde{v}_i \left. \vphantom{\frac{1}{\sigma_i^2}} \right\} \\
& = - \frac{1}{\sigma_i^4} \text{tr}[(I - T_{ii})^2] + 2 \left(\frac{1}{\sigma_i^2}\right) \tilde{v}_i' (I - T_{ii}) \tilde{v}_i,
\end{aligned}$$

$$\text{if } \sigma_i^2 \neq 0 ,$$

$$(-2) \mathbb{E} \left[ \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_i^2} \right] = \frac{1}{\sigma_j^4} \text{tr}[(I - T_{ii})^2] , \text{ if } \sigma_i^2 \neq 0 .$$

Together, results (2.37), (2.41), and (2.43) imply, for  $i, j \in \{1, \dots, c\}$  with  $i \neq j$ , that

$$\begin{aligned}
(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{1}{\sigma_{c+1}^4} \left[ - \sigma_{c+1}^4 \left(\frac{1}{\sigma_i^2}\right) \left(\frac{1}{\sigma_j^2}\right) \text{tr}(T_{ij} T_{ji}) \right. \\
&\quad \left. - 2 \sigma_{c+1}^2 \tilde{v}_i' \sigma_{c+1}^2 \left(\frac{1}{\sigma_j^2}\right) T_{ij} \tilde{v}_j \right] \\
&= - \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{ij} T_{ji}) - 2 \left(\frac{1}{\sigma_j^2}\right) \tilde{v}_i' T_{ij} \tilde{v}_j ,
\end{aligned}$$

$$\text{if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0 ,$$

$$(-2) \mathbb{E} \left( \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_j^2} \right) = \frac{1}{\sigma_i^2} \frac{1}{\sigma_j^2} \text{tr}(T_{ij} T_{ji}) \quad , \quad \text{if } \sigma_i^2 \neq 0, \sigma_j^2 \neq 0 \quad .$$

Similarly, for  $i \in \{1, \dots, c\}$  ,

$$(-2) \frac{\partial^2 L_v}{\partial \sigma_i^2 \partial \sigma_{c+1}^2} = - \left( \frac{1}{\sigma_{c+1}^4} \right) \sigma_{c+1}^2 \left( \frac{1}{\sigma_i^2} \right) [q_i - \text{tr}(T_{ii})]$$

$$+ 2 \left( \frac{1}{\sigma_{c+1}^2} \right) \tilde{v}'_{-i} \tilde{v}_{-i}$$

$$+ 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{\substack{j=1 \\ j \neq i}}^c \sigma_j^2 \left( \frac{1}{\sigma_{c+1}^2} \right) \left[ \left( \frac{1}{2} \right) \sigma_{c+1}^4 \left( \frac{1}{\sigma_i^2} \right) \left( \frac{1}{\sigma_j^2} \right) \text{tr}(T_{ij} T_{ji}) \right.$$

$$\left. + \sigma_{c+1}^4 \left( \frac{1}{\sigma_j^2} \right) \tilde{v}'_{-i} T_{ij} \tilde{v}_{-j} \right]$$

$$+ \left( \frac{1}{\sigma_{c+1}^4} \right) \sigma_i^2 \left( \frac{1}{\sigma_{c+1}^2} \right) \sigma_{c+1}^4 \left( \frac{1}{\sigma_i^4} \right) \text{tr}[(I - T_{ii})^2]$$

$$- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sigma_i^2 \left( \frac{1}{\sigma_{c+1}^2} \right) \sigma_{c+1}^2 \tilde{v}'_{-i} \sigma_{c+1}^2 \left( \frac{1}{\sigma_i^2} \right) (I - T_{ii}) \tilde{v}_{-i}$$

$$\begin{aligned}
&= -\frac{1}{\sigma_{c+1}^2} \left(\frac{1}{\sigma_i^2}\right) q_i + \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{ii}) \\
&\quad + 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \tilde{v}'_{-i} \tilde{v}_{-i} \\
&\quad + 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \frac{1}{\sigma_i^2} \sum_{\substack{j=1 \\ j \neq i}}^c \left[\left(\frac{1}{2}\right) \text{tr}(T_{ij} T_{ji}) + \sigma_i^2 \tilde{v}'_{-i} T_{ij} \tilde{v}_{-j}\right] \\
&\quad + \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}[(I - T_{ii})^2] \\
&\quad - 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \tilde{v}'_{-i} \tilde{v}_{-i} + 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \tilde{v}'_{-i} T_{ii} \tilde{v}_{-i} \\
&= -\frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{ii}) \\
&\quad + 2 \left(\frac{1}{\sigma_{c+1}^2}\right) \frac{1}{\sigma_i^2} \sum_{j=1}^c \left[\left(\frac{1}{2}\right) \text{tr}(T_{ij} T_{ji}) + \sigma_i^2 \tilde{v}'_{-i} T_{ij} \tilde{v}_{-j}\right], \\
&\quad \text{if } \sigma_i^2 \neq 0,
\end{aligned}$$

and,

$$\begin{aligned}
(-2)\mathfrak{E}\left(\frac{\partial^2 L_V}{\partial \sigma_i^2 \partial \sigma_{c+1}^2}\right) &= \left(\frac{1}{\sigma_{c+1}^4}\right) \sigma_{c+1}^2 \left(\frac{1}{\sigma_i^2}\right) [q_i - \text{tr}(T_{ii})] \\
&\quad - \frac{1}{\sigma_{c+1}^6} \sum_{\substack{j=1 \\ j \neq i}}^c \sigma_j^2 \text{tr}[\sigma_{c+1}^4 \left(\frac{1}{\sigma_i^2}\right) \left(\frac{1}{\sigma_j^2}\right) T_{ij} T_{ji}] \\
&\quad - \left(\frac{1}{\sigma_{c+1}^6}\right) \sigma_i^2 \text{tr}[\sigma_{c+1}^4 \left(\frac{1}{\sigma_i^4}\right) (I - 2 T_{ii} + T_{ii}^2)] \\
&= \frac{1}{\sigma_{c+1}^2} \left(\frac{1}{\sigma_i^2}\right) q_i - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{ii}) \\
&\quad - \frac{1}{\sigma_{c+1}^2} \sum_{\substack{j=1 \\ j \neq i}}^c \frac{1}{\sigma_i^2} \text{tr}(T_{ij} T_{ji}) \\
&\quad - \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} [q_i - 2 \text{tr}(T_{ii}) + \text{tr}(T_{ii}^2)] \\
&= \frac{1}{\sigma_{c+1}^2} \frac{1}{\sigma_i^2} \text{tr}(T_{ii}) \\
&\quad - \frac{1}{\sigma_{c+1}^2} \sum_{j=1}^c \frac{1}{\sigma_i^2} \text{tr}(T_{ij} T_{ji}), \text{ if } \sigma_i^2 \neq 0.
\end{aligned}$$

Also,

$$\begin{aligned}
 (-2) \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} &= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \underline{Z} \tilde{\underline{b}})] \\
 &- 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}'_{-j} \tilde{y}_{-j} \\
 &+ 2 \left( \frac{1}{\sigma_{c+1}^6} \right) \sum_{j=1}^c \sigma_j^2 [\sigma_{c+1}^2 \left( \frac{1}{\sigma_j^2} \right) [q_j - \text{tr}(T_{jj})]] \\
 &+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 [-\sigma_{c+1}^4 \left( \frac{1}{\sigma_j^2} \right) \left( \frac{1}{\sigma_k^2} \right) \text{tr}(T_{jk} T_{kj})] \\
 &+ 2 \sigma_{c+1}^2 \tilde{y}'_{-j} (-\sigma_{c+1}^2) \left( \frac{1}{\sigma_k^2} \right) T_{jk} \tilde{y}_{-k}] \\
 &+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^4 [-\sigma_{c+1}^4 \left( \frac{1}{\sigma_j^4} \right) \text{tr}(I - 2 T_{jj} + T_{jj}^2)] \\
 &+ 2 \sigma_{c+1}^2 \tilde{y}'_{-j} \sigma_{c+1}^2 \left( \frac{1}{\sigma_j^2} \right) (I - T_{jj}) \tilde{y}_{-j}] \\
 &= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \underline{y}' S(\underline{y} - \underline{Z} \tilde{\underline{b}})]
 \end{aligned}$$

$$- 4 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}_j' \tilde{y}_j$$

$$+ 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c [q_j - \text{tr}(T_{jj})]$$

$$+ \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{\substack{k=1 \\ k \neq j}}^c [-\text{tr}(T_{jk} T_{kj}) - 2 \sigma_j^2 \tilde{y}_j' T_{jk} \tilde{y}_k]$$

$$+ \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [-q_j + 2 \text{tr}(T_{jj}) - \text{tr}(T_{jj}^2) + 2 \sigma_j^2 \tilde{y}_j' \tilde{y}_j$$

$$- 2 \sigma_j^2 \tilde{y}_j' T_{jj} \tilde{y}_j]$$

$$= - \frac{1}{\sigma_{c+1}^4} [n - p^* - 2 \tilde{y}' S(\tilde{y} - Z \tilde{b})]$$

$$- 2 \left( \frac{1}{\sigma_{c+1}^4} \right) \sum_{j=1}^c \sigma_j^2 \tilde{y}_j' \tilde{y}_j$$

$$+ \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [2q_j - 2 \text{tr}(T_{jj}) - q_j + 2 \text{tr}(T_{jj})]$$

$$- \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{jk} T_{kj}) + 2 \sigma_j^2 \tilde{y}_j' T_{jk} \tilde{y}_k]$$

$$\begin{aligned}
&= -\frac{1}{\sigma_{c+1}^4} [n - p^* - 2\underline{y}'s(\underline{y} - Z\underline{b})] \\
&+ 2 \left(\frac{1}{\sigma_{c+1}^4}\right) \sum_{j=1}^c \left(\frac{1}{2} q_j - \sigma_j^2 \tilde{v}_j' \tilde{v}_j\right) \\
&- \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c [\text{tr}(T_{jk} T_{kj}) + 2 \sigma_j^2 \tilde{v}_j' T_{jk} \tilde{v}_k] ,
\end{aligned}$$

and

$$\begin{aligned}
(-2) \mathbb{E} \left( \frac{\partial^2 L_v}{\partial \sigma_{c+1}^2 \partial \sigma_{c+1}^2} \right) &= \frac{1}{\sigma_{c+1}^4} (n-p^*) \\
&- 2 \left(\frac{1}{\sigma_{c+1}^6}\right) \sum_{j=1}^c \sigma_j^2 [\sigma_{c+1}^2 \left(\frac{1}{\sigma_j^2}\right) [q_j - \text{tr}(T_{jj})]] \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^2 \sum_{\substack{k=1 \\ k \neq j}}^c \sigma_k^2 \text{tr}[\sigma_{c+1}^4 \left(\frac{1}{\sigma_j^2}\right) \left(\frac{1}{\sigma_k^2}\right) T_{jk} T_{kj}] \\
&+ \frac{1}{\sigma_{c+1}^8} \sum_{j=1}^c \sigma_j^4 \sigma_{c+1}^4 \left(\frac{1}{\sigma_j^4}\right) [q_j - 2 \text{tr}(T_{jj}) + \text{tr}(T_{jj}^2)] \\
&= \frac{1}{\sigma_{c+1}^4} (n-p^*)
\end{aligned}$$

$$\begin{aligned}
& - \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [2 q_j - 2 \operatorname{tr}(T_{jj})] \\
& + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{\substack{k=1 \\ k \neq j}}^c \operatorname{tr}(T_{jk} T_{kj}) \\
& + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c [q_j - 2 \operatorname{tr}(T_{jj}) + \operatorname{tr}(T_{jj}^2)] \\
& = \frac{1}{\sigma_{c+1}^4} (n - p^*) + \frac{1}{\sigma_{c+1}^4} (-q) + \frac{1}{\sigma_{c+1}^4} \sum_{j=1}^c \sum_{k=1}^c \operatorname{tr}(T_{jk} T_{kj}) \\
& = \frac{1}{\sigma_{c+1}^4} [n - p^* - q + \operatorname{tr}(T^2)] \\
& \quad [\text{since } \sum_{j=1}^c \sum_{k=1}^c \operatorname{tr}(T_{jk} T_{kj}) = \operatorname{tr}(T^2)].
\end{aligned}$$

B. Derivation of Simplified Expressions  
for the Special Case  $c=1$

In this section, we derive the simplified results given in Section III.F for the special case  $c=1$ . In doing so, we take the parameter space for  $\underline{y}$  to be  $\Omega_3^*$ .

Let

$$\tilde{\underline{e}} \equiv (\underline{I} - \underline{P}_X)\underline{y} - (\underline{I} - \underline{P}_X)\underline{Z}\tilde{\underline{s}}$$

$$= \underline{M}^* \underline{y} ,$$

where  $\underline{M}^* = (\underline{I} - \underline{P}_X) - (\underline{I} - \underline{P}_X)\underline{Z}\underline{C}^-\underline{Z}'(\underline{I} - \underline{P}_X)$ . Since  $(\underline{I} - \underline{P}_X)$  and  $(\underline{I} - \underline{P}_X)\underline{Z}\underline{C}^-\underline{Z}'(\underline{I} - \underline{P}_X)$  are symmetric idempotent matrices, the matrix  $\underline{M}^*$  is symmetric idempotent. Further, the matrix  $\underline{M}^*$  has rank  $r_2$ , as we now show:

$$\begin{aligned} \text{rank}(\underline{M}^*) &= \text{tr}(\underline{M}^*) \quad [\text{since } \underline{M}^* \text{ is idempotent}] \\ &= \text{tr}(\underline{I} - \underline{P}_X) - \text{tr}[(\underline{I} - \underline{P}_X)\underline{Z}\underline{C}^-\underline{Z}'(\underline{I} - \underline{P}_X)] \\ &= n - p^* - \text{tr}(\underline{C}^-\underline{C}) \\ &= n - p^* - \text{rank}(\underline{C}^-\underline{C}) \quad [\text{since } \underline{C}^-\underline{C} \text{ is idempotent}] \\ &= n - p^* - \text{rank}(\underline{C}) \end{aligned}$$

$$= n - p^* - r$$

$$= n - p^* - [\text{rank}(X, Z) - p^*]$$

$$= n - \text{rank}(X, Z)$$

$$= r_2 .$$

Thus, there exists an orthogonal matrix  $W^* = (W_1^*, W_2^*)$  such that

$$\begin{aligned} \begin{pmatrix} W_1^{*'} M^* W_1^* & W_1^{*'} M^* W_2^* \\ W_2^{*'} M^* W_1^* & W_2^{*'} M^* W_2^* \end{pmatrix} &= W^{*'} M^* W^* \\ &= \begin{pmatrix} I_{r_2} & \emptyset \\ \emptyset & \emptyset \end{pmatrix}, \end{aligned} \quad (9.9)$$

where  $W_1^*$  has dimensions  $n \times r_2$ . Since  $M^{*'} M^* = M^*$ , expression (9.9) implies that  $(M^* W_2^*)' (M^* W_2^*) = \emptyset$  and, hence, that  $M^* W_2^* = \emptyset$ . Define

$$\underline{z}^* \equiv (z_1^*, \dots, z_{r_2}^*)' \equiv W_1^{*'} \tilde{\underline{e}} = W_1^{*'} M^* \underline{y} .$$

Observing that

$$\mathbb{E}(\underline{q}) = \underline{0} , \text{ and}$$

$$\begin{aligned}
\text{Var}(\underline{q}) &= \underline{Z}'(\underline{I} - \underline{P}_X)V(\underline{I} - \underline{P}_X)\underline{Z} \\
&= \underline{Z}'(\underline{I} - \underline{P}_X) (\sigma_2^2 \underline{I}_n + \sigma_1^2 \underline{Z}\underline{Z}') (\underline{I} - \underline{P}_X)\underline{Z} \\
&= \sigma_2^2 (\underline{C} + \gamma_1 \underline{C}^2) ,
\end{aligned}$$

note that

$$\mathbb{E}(\underline{\tilde{t}}) = \underline{0} ,$$

$$\begin{aligned}
\text{Var}(\underline{\tilde{t}}) &= \underline{\Delta}^{-\frac{1}{2}} \underline{R}^{*'} \text{Var}(\underline{q}) \underline{R}^* \underline{\Delta}^{-\frac{1}{2}} \\
&= \sigma_2^2 (\underline{\Delta}^{-\frac{1}{2}} \underline{R}^{*'} \underline{C} \underline{R}^* \underline{\Delta}^{-\frac{1}{2}} + \gamma_1 \underline{\Delta}^{-\frac{1}{2}} \underline{R}^{*'} \underline{C}^2 \underline{R}^* \underline{\Delta}^{-\frac{1}{2}}) \\
&= \sigma_2^2 (\underline{I} + \gamma_1 \underline{\Delta}) \quad [\text{using (2.44)}] \\
&= \sigma_2^2 \text{diag}\{1 + \gamma_1 \Delta_1, \dots, 1 + \gamma_1 \Delta_r\} ,
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\underline{\tilde{t}}, \underline{z}^*) &= \underline{\Delta}^{-\frac{1}{2}} \underline{R}^{*'} \underline{Z}' (\underline{I} - \underline{P}_X) \underline{V} \underline{M}^* \underline{W}_1^* \\
&= \sigma_2^2 \underline{\Delta}^{-\frac{1}{2}} \underline{R}^{*'} \underline{Z}' (\underline{I} - \underline{P}_X) (\underline{I} + \gamma_1 \underline{Z}\underline{Z}') \underline{M}^* \underline{W}_1^* \\
&= \emptyset \quad [\text{since } \underline{Z}' (\underline{I} - \underline{P}_X) \underline{M}^* = \underline{Z}' \underline{M}^* = \emptyset] ,
\end{aligned}$$

$$E(\underline{z}^*) = \underline{0} ,$$

$$\begin{aligned} \text{Var}(\underline{z}^*) &= W_1^{*'} M^* V M^* W_1^* \\ &= \sigma_2^2 W_1^{*'} M^* (I + \gamma_1 Z Z') M^* W_1^* \\ &= \sigma_2^2 W_1^{*'} M^* W_1^* \\ &= \sigma_2^2 I_{r_2} , \end{aligned}$$

and also that

$$\begin{aligned} \underline{z}^{*'} \underline{z}^* &= \underline{y}' M^* W_1^* W_1^{*'} M^* \underline{y} \\ &= \underline{y}' M^* (I - W_2^* W_2^{*'}) M^* \underline{y} \\ &= \underline{y}' M^{*2} \underline{y} \quad [\text{since } M^* W_2^* = \emptyset] \\ &= \underline{y}' M^* \underline{y} \\ &= \underline{y}' (I - P_X) \underline{y} - \underline{y}' (I - P_X) Z C^{-1} Z' (I - P_X) \underline{y} \\ &= \underline{y}' \underline{y} - \underline{y}' P_X \underline{y} - \tilde{s}' \underline{q} \\ &= S_2 . \end{aligned}$$

We conclude (since  $r + r_2 = n - p^*$ ) that  $\tilde{t}_1, \dots, \tilde{t}_r, z_1^*, \dots, z_{r_2}^*$  are  $n - p^*$  linearly independent error contrasts whose joint probability density function is

$$\begin{aligned}
 h(\underline{\tilde{t}}, \underline{z}^*; \underline{\gamma}) &= (2\pi)^{-\frac{1}{2}(n-p^*)} [\sigma_2^2 (n-p^*)]^{-\frac{1}{2}} \prod_{i=1}^r (1 + \gamma_1 \Delta_i)^{-\frac{1}{2}} \\
 &\quad \cdot \exp\left\{-\frac{1}{2} \left[ \frac{1}{\sigma_2^2} S_2 + \sum_{i=1}^r \frac{\tilde{t}_i^2}{\sigma_2^2 (1 + \gamma_1 \Delta_i)} \right]\right\} \\
 &= (2\pi\sigma_2^2)^{-\frac{1}{2}(n-p^*)} \prod_{i=1}^r (1 + \gamma_1 \Delta_i)^{-\frac{1}{2}} \\
 &\quad \cdot \exp\left\{-\left(\frac{1}{2}\right) \frac{1}{\sigma_2^2} \left[ S_2 + \sum_{i=1}^r \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i} \right]\right\}.
 \end{aligned}$$

Thus, the log-likelihood function associated with any set of  $n - p^*$  linearly independent error contrasts differs by no more than an additive constant from the function

$$\begin{aligned}
 L_1(\underline{\gamma}; \underline{y}) &= k_3 - \frac{1}{2} \{ (n-p^*) \log(\sigma_2^2) \\
 &\quad + \sum_{i=1}^r \log(1 + \gamma_1 \Delta_i) \}
 \end{aligned}$$

$$+ \frac{1}{\sigma_2^2} \left[ S_2 + \sum_{i=1}^r \frac{\tilde{\epsilon}_i^2}{1 + \gamma_1 \Delta_i} \right] \}, \quad \underline{\gamma} \in \Omega_3^*, \quad (9.10)$$

where  $k_3$  is a scalar that does not depend on  $\underline{\gamma}$ .

Subsequently, we use the notation  $\sum_i$  to represent  $\sum_{i=1}^r$ .

Differentiating expression (9.10), we find that

$$\frac{\partial L_1}{\partial \gamma_1} = - \left( \frac{1}{2} \right) \left[ \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} - \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right], \quad (9.11)$$

$$\frac{\partial L_1}{\partial \sigma_2^2} = - \left( \frac{1}{2} \right) (n-p^*) \frac{1}{\sigma_2^4} \left\{ \sigma_2^2 - \left( \frac{1}{n-p^*} \right) \left( S_2 + \sum_i \frac{\tilde{\epsilon}_i^2}{1 + \gamma_1 \Delta_i} \right) \right\} \quad (9.12)$$

and that

$$\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} = \left( \frac{1}{2} \right) \left[ \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} - 2 \left( \frac{1}{\sigma_2^2} \right) \sum_i \frac{\Delta_i^2 \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^3} \right], \quad (9.13)$$

$$\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2} = - \left( \frac{1}{2} \right) \frac{1}{\sigma_2^4} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \quad (9.14)$$

$$\frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2} = \left( \frac{1}{2} \right) (n-p^*) \frac{1}{\sigma_2^6} \left[ \sigma_2^2 - 2 \left( \frac{1}{n-p^*} \right) \left( S_2 + \sum_i \frac{\tilde{\epsilon}_i^2}{1 + \gamma_1 \Delta_i} \right) \right], \quad (9.15)$$

which establishes results (3.82) - (3.86).

Since  $\mathbb{E}(\tilde{\epsilon}_i^2) = \sigma_2^2 (1 + \gamma_1 \Delta_i)$  ( $i = 1, \dots, r$ )

and since  $\mathbb{E}(S_2) = r_2 \sigma_2^2 = \sigma_2^2 (n - p^* - r)$ , we have that

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) = - \left(\frac{1}{2}\right) \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} ,$$

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right) = - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} ,$$

$$\mathbb{E}\left(\frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2}\right) = \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} [(n-p^*) \sigma_2^2 - 2\mathbb{E}(S_2)]$$

$$- 2 \sum_i \frac{\mathbb{E}(\tilde{\epsilon}_i^2)}{1 + \gamma_1 \Delta_i} ]$$

$$= \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} [-(n - p^*) \sigma_2^2 + 2r\sigma_2^2 - 2r\sigma_2^2 ]$$

$$= - \left(\frac{1}{2}\right)(n-p^*) \frac{1}{\sigma_2^4} ,$$

which establishes results (3.87) - (3.89).

Equating expressions (3.37) and (9.12), we find that

$$\hat{\sigma}_2^2(\gamma_1) = \frac{1}{n-p^*} (S_2 + \sum_i \frac{\tilde{\epsilon}_i^2}{1 + \gamma_1 \Delta_i}) \quad (9.16)$$

which establishes result (3.81).

It follows from (9.10) that the concentrated log-likelihood function associated with any set of  $n-p^*$  linearly independent error contrasts differs by no more than an additive constant from the function

$$L_c(\gamma_1; \underline{y}) = k_4 - \frac{1}{2} \{ (n - p^*) \log[\hat{\sigma}_2^2(\gamma_1)] \\ + \sum_i \log(1 + \gamma_1 \Delta_i) \} ,$$

where  $k_4$  is a scalar that does not depend on  $\gamma_1$ . Differentiating, we find that

$$\frac{\partial L_c}{\partial \gamma_1} = - \left(\frac{1}{2}\right) \left[ \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} - \frac{1}{\hat{\sigma}_2^2(\gamma_1)} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right]$$

and that

$$\frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1} = \left(\frac{1}{2}\right) \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} - \frac{1}{\hat{\sigma}_2^2(\gamma_1)} \sum_i \frac{\Delta_i^2 \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^3} \\ + \left(\frac{1}{2}\right) \left(\frac{1}{n-p^*}\right) \left[ \frac{1}{\hat{\sigma}_2^2(\gamma_1)} \right]^2 \left[ \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right]^2 ,$$

which establishes results (3.90) and (3.91)

Since  $\lambda_{11} = \frac{1}{r} \text{tr}(C)$  [Section III.B.2], we have that  $\lambda_{11} = \frac{1}{r} \sum_{i=1}^r$

$\Delta_i = \bar{\Delta}$ . Substituting this expression for  $\lambda_{11}$  into the expressions for  $k^*(\gamma_1)$  and  $K^*(\gamma_1)$  given in Section III.B.4.b, we find that

$$k^*(\gamma_1) = (1 + \gamma_1 \bar{\Delta}) \left(1 + \gamma_1 \bar{\Delta} + \frac{S_1}{S_2}\right) \frac{\partial L_c}{\partial \gamma_1},$$

$$K^*(\gamma_1) = [2 \bar{\Delta}(1 + \gamma_1 \bar{\Delta}) + \bar{\Delta} \frac{S_1}{S_2}] \frac{\partial L_c}{\partial \gamma_1} \\ + (1 + \gamma_1 \bar{\Delta}) \left(1 + \gamma_1 \bar{\Delta} + \frac{S_1}{S_2}\right) \frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1},$$

which establishes results (3.92) and (3.93).

Using  $\lambda_{11} = \bar{\Delta}$ , together with the expressions for  $\frac{\partial L_R}{\partial \psi_1}$  [result (3.53)] and  $\frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1}$  given in Section III.B.3.b, we find that

$$\frac{\partial L_R}{\partial \psi_1} = -\frac{1}{\bar{\Delta}} \left(\frac{1}{\psi_1^2}\right) \frac{\partial L_c}{\partial \gamma_1} \quad [\text{since } w_{11} = \frac{1}{\lambda_{11}} = \frac{1}{\bar{\Delta}}],$$

$$\frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1} = \left(\frac{1}{\bar{\Delta}}\right)^2 \frac{1}{\psi_1^4} \frac{\partial^2 L_c}{\partial \gamma_1 \partial \gamma_1} + 2 \left(\frac{1}{\bar{\Delta}}\right) \frac{1}{\psi_1^3} \frac{\partial L_c}{\partial \gamma_1},$$

which establishes results (3.94) and (3.95).

Using results (9.1) - (9.5) and (9.7), together with results (9.11) - (9.15) and (9.16), we find that

$$\frac{\partial L_v}{\partial \sigma_1^2} = \frac{1}{\sigma_2^2} \frac{\partial L_1}{\partial \gamma_1},$$

$$\frac{\partial L_v}{\partial \sigma_2^2} = \frac{\partial L_1}{\partial \sigma_2^2} - \frac{1}{\sigma_2^4} (\sigma_1^2) \frac{\partial L_1}{\partial \gamma_1}$$

$$= - \left(\frac{1}{2}\right) (n-p^*) \frac{1}{\sigma_2^4} [\sigma_2^2 - \hat{\sigma}_2^2(\gamma_1)]$$

$$+ \left(\frac{1}{2}\right) \left(\frac{1}{\sigma_2^4}\right) \sigma_1^2 \left[ \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} - \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right]$$

$$= - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} [\sigma_2^2 (n - p^* - \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i})$$

$$- (n-p^*) \hat{\sigma}_2^2(\gamma_1) + \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}] , \quad (9.17)$$

$$\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} = \frac{1}{\sigma_2^4} \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1},$$

$$\begin{aligned}
\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} &= -\frac{1}{\sigma_2^4} \frac{\partial L_1}{\partial \gamma_1} + \frac{1}{\sigma_2^2} \frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2} - \frac{1}{\sigma_2^6} (\sigma_1^2) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \\
&= \left(\frac{1}{2}\right) \frac{1}{\sigma_2^4} \sum_i \frac{\Delta_i}{1 + \gamma \Delta_i} - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^6} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
&\quad - \left(\frac{1}{2}\right) \frac{1}{\sigma_2^6} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
&\quad - \frac{1}{\sigma_2^6} (\sigma_1^2) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \\
&= -\frac{1}{\sigma_2^4} \left[ \gamma_1 \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} - \left(\frac{1}{2}\right) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \right. \\
&\quad \left. + \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2} &= \frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2} - 2 \left(\frac{1}{\sigma_2^4}\right) (\sigma_1^2) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2} \\
&\quad + 2 \left(\frac{1}{\sigma_2^6}\right) (\sigma_1^2) \frac{\partial L_1}{\partial \gamma_1} \\
&\quad + \frac{1}{\sigma_2^8} (\sigma_1^4) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{2}\right) (n-p^*) \frac{1}{\sigma_2^6} [\sigma_2^2 - 2 \hat{\sigma}_2^2(\gamma_1)] \\
&\quad + \sigma_1^2 \left(\frac{1}{\sigma_2^8}\right) \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
&\quad + \left(\frac{1}{2}\right) \sigma_1^4 \left(\frac{1}{\sigma_2^8}\right) \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
&\quad - \sigma_1^4 \left(\frac{1}{\sigma_2^{10}}\right) \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^3} \\
&\quad - \sigma_1^2 \left(\frac{1}{\sigma_2^2}\right) \left[ \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} - \frac{1}{\sigma^2} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right] \\
&= \left(\frac{1}{2}\right) \frac{1}{\sigma_2^6} \{ \sigma_2^2 [n - p^* - 2 \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} + \gamma_1^2 \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2}] \\
&\quad - 2 [(n-p^*) \hat{\sigma}_2^2(\gamma_1) - 2 \gamma_1 \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
&\quad + \gamma_1^2 \sum_i \frac{\Delta_i^2 \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^3}] \},
\end{aligned}$$

which establishes results (3.96) - (3.100).

It follows from results (9.4), (9.6), and (9.8) that

$$\mathbb{E}\left(\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2}\right) = \frac{1}{\sigma_2^4} \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) ,$$

$$\mathbb{E}\left(\frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2}\right) = \frac{1}{\sigma_2^2} \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right)$$

$$- \frac{1}{\sigma_2^4} (\gamma_1) \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) ,$$

and [since  $\mathbb{E}\left(\frac{\partial L_1}{\partial \gamma_1}\right) = 0$ ]

$$\mathbb{E}\left(\frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2}\right) = \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \sigma_2^2 \partial \sigma_2^2}\right)$$

$$- 2 \gamma_1 \left(\frac{1}{\sigma_2^2}\right) \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \sigma_2^2}\right)$$

$$+ \gamma_1^2 \left(\frac{1}{\sigma_2^4}\right) \mathbb{E}\left(\frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}\right) ,$$

which establishes results (3.101) - (3.103).

Note that, in the special case  $c=1$ ,  $w_{11} = \frac{1}{\lambda_{11}}$ ,  $w_{21} = -\frac{1}{\lambda_{11}}$ , and  $w_{22} = 1$ . Consequently, it follows from (3.35) and (3.36) that

$$\frac{\partial L_1^*}{\partial m_1} = \frac{1}{m_2 \lambda_{11}} \frac{\partial L_1}{\partial \gamma_1} = \frac{1}{\lambda_{11}} \frac{\partial L_v}{\partial \sigma_1^2}, \quad (9.18)$$

$$\frac{\partial L_1^*}{\partial m_2} = -\frac{1}{\lambda_{11}} \frac{\partial L_v}{\partial \sigma_1^2} + \frac{\partial L_v}{\partial \sigma_2^2}, \quad (9.19)$$

$$\frac{\partial^2 L_1^*}{\partial m_1 \partial m_1} = \left[ \frac{1}{\lambda_{11}} \right]^2 \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2}, \quad (9.20)$$

$$\frac{\partial^2 L_1^*}{\partial m_1 \partial m_2} = \frac{1}{\lambda_{11}} \left( -\frac{1}{\lambda_{11}} \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} + \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} \right), \quad (9.21)$$

$$\begin{aligned} \frac{\partial^2 L_1^*}{\partial m_2 \partial m_2} = & -\frac{1}{\lambda_{11}} \left( -\frac{1}{\lambda_{11}} \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} + \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} \right) \\ & - \frac{1}{\lambda_{11}} \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} + \frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2}. \end{aligned} \quad (9.22)$$

Using results (9.18) - (9.22), together with  $\lambda_{11} = \bar{\Delta}$ , we find that

$$q_1^*(\underline{m}) = m_1^2 \frac{\partial L_1^*}{\partial m_1}$$

$$= m_1^2 \left( \frac{1}{\bar{\Delta}} \right) \frac{\partial L_V}{\partial \sigma_1^2},$$

$$q_2^*(\underline{m}) = m_2^2 \frac{\partial L_1^*}{\partial m_2}$$

$$= - \sigma_2^2 \left( \frac{1}{\bar{\Delta}} \right) \frac{\partial L_1}{\partial \gamma_1} + \sigma_2^4 \frac{\partial L_V}{\partial \sigma_2^2}$$

$$= - \left( \frac{1}{2} \right) \left[ - \sigma_2^2 \left( \frac{1}{\bar{\Delta}} \right) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} + \left( \frac{1}{\bar{\Delta}} \right) \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right]$$

$$- \left( \frac{1}{2} \right) \left[ \sigma_2^2 (n - p^* - \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}) \right]$$

$$- (n - p^*) \hat{\sigma}_2^2(\gamma_1) + \gamma_1 \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} ]$$

$$= - \left( \frac{1}{2} \right) \left\{ \sigma_2^2 [n - p^* - \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}] \right.$$

$$\left. - [(n - p^*) \hat{\sigma}_2^2(\gamma_1) - \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2}] \right\}$$

$$\frac{\partial q_1^*}{\partial m_1} = 2 m_1 \frac{\partial L_1^*}{\partial m_1} + m_1^2 \frac{\partial^2 L_1^*}{\partial m_1 \partial m_1}$$

$$= \frac{m_1}{\bar{\Delta}} \left[ 2 \frac{\partial L_v}{\partial \sigma_1^2} + \frac{m_1}{\bar{\Delta}} \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} \right],$$

$$\frac{\partial q_1^*}{\partial m_2} = m_1^2 \frac{\partial^2 L_1^*}{\partial m_1 \partial m_2}$$

$$= - m_1^2 \left( \frac{1}{\bar{\Delta}} \right)^2 \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} + m_1^2 \left( \frac{1}{\bar{\Delta}} \right) \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2}$$

$$= - (\sigma_2^2 + \bar{\Delta} \sigma_1^2)^2 \left( \frac{1}{\bar{\Delta}} \right)^2 \frac{1}{\sigma_2^4} \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}$$

$$+ (\sigma_2^2 + \bar{\Delta} \sigma_1^2)^2 \left( \frac{1}{\bar{\Delta}} \right) \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} \quad [\text{since } m_1 = \sigma_2^2 + \bar{\Delta} \sigma_1^2]$$

$$= - (1 + \gamma_1 \bar{\Delta})^2 \left( \frac{1}{\bar{\Delta}} \right) \left( \frac{1}{\bar{\Delta}} \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \right)$$

$$- (1 + \gamma_1 \bar{\Delta})^2 \left( \frac{1}{\bar{\Delta}} \right) [\delta_1 \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1}]$$

$$- \left(\frac{1}{2}\right) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}$$

$$+ \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2}]$$

$$= - (1 + \gamma_1 \bar{\Delta})^2 \left[ \frac{1}{\bar{\Delta}} \left[ \frac{1}{2} (1 + \gamma_1 \bar{\Delta}) \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \right. \right.$$

$$\left. - \left(\frac{1}{2}\right) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} + \frac{1}{\sigma_2^2} \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right],$$

$$\frac{\partial q_2^*}{\partial m_1} = \frac{\partial^2 L_1}{m_2^2 \partial m_1 \partial m_2}$$

$$= m_2^2 \left(\frac{1}{m_1^2}\right) \frac{\partial q_1^*}{\partial m_2}$$

$$= \left(\frac{1}{1 + \gamma_1 \bar{\Delta}}\right)^2 \frac{\partial q_1^*}{\partial m_2},$$

$$\frac{\partial q_2^*}{\partial m_2} = 2 \frac{\partial L_1^*}{m_2 \partial m_2} + m_2^2 \frac{\partial^2 L_1^*}{\partial m_2 \partial m_2}$$

$$= - 2 \sigma_2^2 \left[ \frac{1}{\bar{\Delta}} \right] \frac{\partial L_v}{\partial \sigma_1^2} + 2 \sigma_2^2 \frac{\partial L_v}{\partial \sigma_2^2}.$$

$$\begin{aligned}
& + \sigma_2^4 \left[ \frac{1}{\Delta} \right]^2 \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_1^2} \\
& - 2 \sigma_2^4 \left[ \frac{1}{\Delta} \right] \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} \\
& + \sigma_2^4 \frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2} \\
& = - 2 \left[ \frac{1}{\Delta} \right] \frac{\partial L_1}{\partial \gamma_1} + 2 \sigma_2^2 \frac{\partial L_v}{\partial \sigma_2^2} \\
& + \left[ \frac{1}{\Delta} \right]^2 \frac{\partial^2 L_1}{\partial \gamma_1 \partial \gamma_1} \\
& - 2 (\sigma_2^4) \left[ \frac{1}{\Delta} \right] \frac{\partial^2 L_v}{\partial \sigma_1^2 \partial \sigma_2^2} \\
& + (\sigma_2^4) \frac{\partial^2 L_v}{\partial \sigma_2^2 \partial \sigma_2^2} \\
& = - \left( \frac{1}{2} \right) \frac{1}{\sigma_2^2} \left\{ [-2 \sigma_2^2 \left[ \frac{1}{\Delta} \right] \sum_i \frac{\Delta_i}{1 + \gamma_{1i} \Delta_i} \right. \\
& \quad \left. + 2 \left[ \frac{1}{\Delta} \right] \sum_i \frac{\Delta_i \xi_i^2}{(1 + \gamma_{1i} \Delta_i)^2} \right\} .
\end{aligned}$$

$$\begin{aligned}
& + [ 2 \sigma_2^2 (n-p^*) - 2 \sigma_2^2 \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \\
& - 2 (n-p^*) \hat{\sigma}_2^2(\gamma_1) + 2 \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} ] \\
& + [- \sigma_2^2 \left( \frac{1}{\bar{\Delta}} \right)^2 \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
& + 2 \left( \frac{1}{\bar{\Delta}} \right)^2 \sum_i \frac{\Delta_i^2 \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^3} ] \\
& + [- 2 (\sigma_2^2) \left( \frac{1}{\bar{\Delta}} \right) \gamma_1 \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
& + 4 \left( \frac{1}{\bar{\Delta}} \right) \gamma_1 \sum_i \frac{\Delta_i^2 \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^3} \\
& + 2 (\sigma_2^2) \left( \frac{1}{\bar{\Delta}} \right) \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}
\end{aligned}$$

$$\begin{aligned}
& - 4 \left\{ \frac{1}{\bar{\Delta}} \right\} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \Big] \\
& + [- \sigma_2^2 (n-p^*) + 2 \sigma_2^2 \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \\
& - \sigma_2^2 \gamma_1^2 \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
& + 2 (n-p^*) \hat{\sigma}_2^2 (\gamma_1) \\
& - 4 \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \\
& + 2 (\gamma_1^2) \sum_i \frac{\Delta_i^2 \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^3} \Big] \Big\} \\
& = - \left( \frac{1}{2} \right) \frac{1}{\sigma_2^2} \left\{ \sigma_2^2 [n - p^* - \left( \frac{1}{\bar{\Delta}} \right)^2 (1 + \gamma_1 \bar{\Delta})^2 \sum_i \frac{\Delta_i^2}{(1 + \gamma_1 \Delta_i)^2} \right] \\
& - 2 \left( \frac{1}{\bar{\Delta}} \right) (1 + \gamma_1 \bar{\Delta}) \left[ \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right. \\
& \left. - \frac{1}{\bar{\Delta}} (1 + \gamma_1 \bar{\Delta}) \sum_i \frac{\Delta_i^2 \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^3} \right] \Big\},
\end{aligned}$$

which establishes results (3.104) - (3.109).

By definition,

$$m_1^*(\psi_1) = \psi_1 \left( \psi_1 \frac{s_1}{s_2} + 1 \right) \frac{\partial L_R}{\partial \psi_1}$$

and, according to the expression for  $\frac{\partial m_1^*}{\partial \psi_1}$  given in Section III.B.4.a,

$$\begin{aligned} \frac{\partial m_1^*}{\partial \psi_1} &= m_1^*(\psi_1) \frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1} + \left[ \left( \psi_1 \frac{s_1}{s_2} + 1 \right) + \psi_1 \frac{s_1}{s_2} \right] \frac{\partial L_R}{\partial \psi_1} \\ &= \psi_1 \left( 1 + \psi_1 \frac{s_1}{s_2} \right) \frac{\partial^2 L_R}{\partial \psi_1 \partial \psi_1} + \left( 1 + 2 \psi_1 \frac{s_1}{s_2} \right) \frac{\partial L_R}{\partial \psi_1}, \end{aligned}$$

which establishes results (3.110) - (3.111).

Part (iv) of Lemma 2.7, together with result (9.14) implies that

$$\tilde{v}_1' \tilde{v}_1 = \frac{1}{\sigma_2^4} \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}$$

and, hence, that

$$\tilde{b}_1' \tilde{b}_1 = \gamma_1^2 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}. \quad (9.23)$$

Similarly, part (ii) of Lemma 2.7, together with result (9.11), implies,

for  $\gamma_1 > 0$ , that

$$\frac{1}{\gamma_1} [q_1 - \text{tr}(T_{11})] = \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}$$

and, hence, that

$$q_1 - \text{tr}(T_{11}) = \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \quad (9.24)$$

or, equivalently,

$$\text{tr}(T_{11}) = q_1 - \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} . \quad (9.25)$$

Also,

$$\begin{aligned} \hat{\sigma}_2^2(\gamma_1) &= \frac{1}{n-p^*} (\sigma_2^2) (\underline{y} - \underline{X}\tilde{\underline{\alpha}})' \underline{V}^{-1} (\underline{y} - \underline{X}\tilde{\underline{\alpha}}) \quad [\text{using (3.39)}] \\ &= \left( \frac{1}{n-p^*} \right) \underline{y}' (\underline{y} - \underline{X}\tilde{\underline{\alpha}} - \underline{Z}\tilde{\underline{b}}) \quad [\text{using (2.14)}] . \end{aligned} \quad (9.26)$$

Substituting from expressions (9.24) and (9.16) in expression (9.17), we find that

$$\frac{\partial L_v}{\partial \sigma_2^2} = - \left( \frac{1}{2} \right) \sigma_2^4 \left\{ \sigma_2^2 [n - p^* - q_1 + \text{tr}(T_{11})] \right.$$

$$- \left[ S_2 + \sum_i \frac{\tilde{t}_i^2}{1 + \gamma_1 \gamma_i} - \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right] \Bigg\}.$$

Comparing this expression with expression (2.34), we obtain the identity

$$\begin{aligned} (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}}) \\ = S_2 + \sum_i \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i} - \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}. \end{aligned} \quad (9.27)$$

Now,

$$\begin{aligned} h_1(\underline{\sigma}) &= \frac{\tilde{\underline{b}}_1' \tilde{\underline{b}}_1}{q_1 - \text{tr}(T_{11})} \\ &= \frac{\gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}}{\sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}} \quad [\text{using (9.23) and (9.24)}], \end{aligned}$$

$$\begin{aligned} h_2(\underline{\sigma}) &= \frac{(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})'(\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})}{n - p^* - [q - \text{tr}(T)]} \\ &= \frac{S_2 + \sum_i \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i} - \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2}}{n - p^* - \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}} \\ &\quad [\text{using (9.24) and (9.27)}], \end{aligned}$$

$$h_2^*(\underline{\sigma}) = \frac{\underline{y}'(\underline{y} - \underline{x}\tilde{\alpha} - \underline{z}\tilde{b})}{n-p^*}$$

$$= \hat{\sigma}_2^2(\gamma_1) \quad [\text{using (9.26)}]$$

$$= \left(\frac{1}{n-p^*}\right) \left(S_2 + \sum_i \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i}\right) \quad [\text{using (9.16)}],$$

$$\epsilon_1(\underline{\sigma}) = \frac{\tilde{b}_1' \tilde{b}_1 + \sigma_1^2 \text{tr}(T_{11})}{q_1}$$

$$= \frac{1}{q_1} \left\{ \gamma_1^2 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} + \sigma_1^2 \left[ q_1 - \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \right] \right\}$$

[using (9.23) and (9.25)] ,

$$\epsilon_2^*(\underline{\sigma}) = \left(\frac{1}{n-p^*}\right) \{(\underline{y} - \underline{x}\tilde{\alpha} - \underline{z}\tilde{b})'(\underline{y} - \underline{x}\tilde{\alpha} - \underline{z}\tilde{b})$$

$$+ \sigma_2^2 [q - \text{tr}(T)]\}$$

$$= \left(\frac{1}{n-p^*}\right) \left[ S_2 + \sum_i \frac{\tilde{t}_i^2}{1 + \gamma_1 \Delta_i} - \gamma_1 \sum_i \frac{\Delta_i \tilde{t}_i^2}{(1 + \gamma_1 \Delta_i)^2} \right.$$

$$\left. + \sigma_1^2 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i} \right] \quad [\text{using (9.24) and (9.27)}] ,$$

$$\epsilon_2(\underline{\sigma}) = \frac{1}{n} \{ (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})' (\underline{y} - \underline{x}\tilde{\underline{\alpha}} - \underline{z}\tilde{\underline{b}})$$

$$+ \sigma_2^2 [p^* + q - \text{tr}(T)] \}$$

$$= \frac{1}{n} \{ S_2 + \sum_i \frac{\tilde{\epsilon}_i^2}{1 + \gamma_1 \Delta_i} - \gamma_1 \sum_i \frac{\Delta_i \tilde{\epsilon}_i^2}{(1 + \gamma_1 \Delta_i)^2}$$

$$+ \sigma_2^2 [p^* + \gamma_1 \sum_i \frac{\Delta_i}{1 + \gamma_1 \Delta_i}] \}$$

[using (9.24) and (9.27)] ,

which establishes results (3.112) - (3.117).