THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., VOL.39 No.1, 1995

Diagonalization of variance-covariance matrices using F-G algorithm

Ali. A.K.A[1] and C.Y. Lin[2]

INTRODUCTION

A major limitation of applying multitrait mixed model analysis is the increased number of equations to be solved. The technique of canonical transformation has been used to reduce computing costs when all traits are recorded for each animal with equal design matrices and one random classification in the model. Basically canonical transformation involves transformation of all correlated traits into uncorrelated canonical traits, such that single trait analysis can be performed on these canonical traits. The use of canonical transformation as presented by Thompson (1977), Meyer (1984), and Jensen and Mao (1988) is applicable only to the model with one random factor.

Lin and Smith (1990) applied F-G algorithm to transform a multitrait into a unitrait mixed model that equal design matrices for all traits and contains more than one random effect. The class of models was restricted to those in which the covariance matrices for all random effects, including residual, can be diagonalized simultaneously. Another application of F-G algorithm was used by Ali (1993) to restrict maximum likelihood estimation of variance components, where both algorithms, F-G of Flury and Gautschi (1984) and modified F-G of Clarkson (1987) were applied to three sets of mixed model coefficient matrices in animal breeding. Close estimate to exact REML solutions were obtained for traits with low heritability. The objective of this study is to use the maximum likelihood and least square versions of F-G algorithm to simultaneously diagonalize variance-covariance matrices which are generated using different multiple trait parameters.

The diagonalization process runs as follows:

1) In one step diagonalization (diagonal case): diagonal herd and non-diagonal sire and error variance-covariance matrices were diagonalized simultaneously.
2) In one step diagonalization (nondiagonal case): nondiagonal herd, sire and error variance-covariance matrices were diagonalized simultaneously.

3) In two diagonalization: diagonal herd and nondiagonal sire and error variance-covariance matrices diagonalized simultaneously.

F-G algorithm:

F-G algorithm was developed by Flury (1984), Flury (1988), and Flury and Gautschi (1984). The algorithm diagonalizes $A_1 \ldots\ldots A_K$ p.d. matrices by finding an orthogonal matrix $B_{P\times p}$ such that:

$$\phi(B) = \prod_{i=1}^{k} \{Det[diag(B' A_i B)]/Det(A_i)\}^{n_i} \text{ is minimum} \qquad (1)$$

F-G algorithm minimizes ϕ by iteration on two levels. The F level. F algorithm consists of rotation of all $p(p-1)/2$ pairs of vectors of B, so F algorithm yields a converging sequence of orthogonal matrices $B_o, B_1, B_2 \ldots$ such that $\phi(B^{f+1}) \leq \phi(B^f)$. However, G level, G algorithm finds an orthogonal matrix $Q_{2\times 2}$ by solving the equation.

$$q_1' \left(\sum_{i=1}^{k} n_i \frac{d_{i1} - d_{i2}}{d_{i1} d_{i2}} T_i \right) q_2 \qquad (2)$$

$$\text{where} \qquad T_i\,(2x2) = \begin{bmatrix} b_l' A_i b_l & b_l' A_i b_j \\ b_j' A_i b_l & b_j' A_i b_j \end{bmatrix} \qquad \begin{matrix} i = 1,2,\ldots k \\ 1 \leq l < j \leq p \end{matrix}$$

$$dij = q_j' T_i q_j \qquad (i = 1,\ldots\ldots\ldots,k; j = 1,2)$$

The b_l and bj are the l and j columns of B and q_1 and q_2 are the columns of Q.

The iteration of G algorithm yields the sequence of orthogonal matrix $Q_0, Q_1, \ldots\ldots$ converging to solution of (2). The matrix Q is an orthogonal matrix which rotates each pair of B by an angle θ.

Flury and Constantine (1985) applied F-G algorithm on two sets of p. d. matrices and obtained for each set an orthogonal matrix B which simultaneously diagonalize the two matrices to a near diagonal form. Clarkson (1987) modified the F-level of F-G algorithm and improved its performance by reducing the number of operation required for computing each pair of orthogonal vectors $B_p=(b_l, b_j)$ in B. An orthogonal matrix P is found such that $P = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ where s and c are sine and cosine of the rotation angle ($c^2 + s^2 = 1$). The updated versions of vectors b_j, b_l are computed as $B_n = B_p P$. That is, $b_j^n = Cb_j + Sb_l$ and $b_l^n = -Sb_j + Cb_l$.

Clarkson (1988) used a least squares algorithm for diagonalizing symmetric and not necessarily positive definite matrices. The function to be minimized by least squares is: $\phi = \sum_{i=1}^{k} \sum_{j=1}^{p} \sum_{l=1}^{P} n_i \, d_{ijl}$ where d_{ijl} is j lth element of the matrix $\Delta = B' A_i B$ and n_i is a matrix of weights.

MATERIALS AND METHODS

A) Consider a linear model for t variates:

Y = X b + W h + Z s + e where Y = data vector of t variates (or traits). b = vector of fixed effects (year), h = vector of random herd effect, s = vector of random sire effect. e = vector of random residual effects. and X. W. Z. = known incidence matrices associated with vectors b. h. and s. respectively.

The expectations and variance-covariance matrices of random vectors are:

$$E\begin{bmatrix} h \\ s \\ e \end{bmatrix} = \begin{bmatrix} o \\ o \\ o \end{bmatrix} \text{ and } V\begin{bmatrix} h \\ s \\ e \end{bmatrix} = \begin{bmatrix} I_p * H & O & o \\ o & A_q * S & o \\ o & O & I_N * R \end{bmatrix}$$

where H. S and R are herd. sire and residual variance-covariance matrices. respectively. N, q and p are the number of records. sires and herds. respectively. A is the numerator relationship matrix among sires and * indicates direct product between two matrices. It is assumed that h. s and e are mutually uncorrelated and R is positive definite (p.d.).

B) Given H. S and R matrices, the effects of herd, sire and residual can be simulated from a multivariate normal distribution. It should be noted that variances and covariances of S and R can be easily computed based on the assumed values of h^2, genetic and phenotypic correlations (r_g and r_p) and sire variances (σ_s^2). Different seeds were used as starting values in simulation.

C) Ten sires were mated randomly with 250 dams and each dam had one daughter. The daughters were distributed randomly over herds. Herd and year variances were estimated as follows:

Head variance (σ_h^2) = herd proportion* ($\sigma_s^2 + \sigma_e^2$) / (1 - herd proportion).

Year variance (σ_y^2) = year proportion* ($\sigma_h^2 + \sigma_s^2 + \sigma_e^2$) / (1 - year proportion).

D) Year effect was generated as a random variable but was treated as fixed. First generation was computed such as each dam was mated randomly with sire from the base population. Each daughter received the same herd effect but different year effect. Therefore: Daughter record = year effect + herd effect + ½ (sire effect + dam effect) + Mendelian sampling effect.

E) Maximum likelihood (ML) and least squares (LS) versions of F-G algorithm (one-step diagonalization) were used to:

   1. estimate orthogonal matrices B_M and B_L respectively to diagonalize simultaneously herd, sire and error variance-covariance matrices where herd variance-covariance matrix is diagonal (diagonal case) and sire and error variance-covariance matrices are non-diagonal.
   2. estimate two orthogonal matrices B_M and B_L to diagonalize simultaneously nondiagonal herd, sire and error variance-covariance matrices (non-diagonal case).

F) ML and LS versions of F-G algorithm (two-step diagonalization were used to estimate two orthogonal matrices B_M and B_L to diagonalize simultaneously $L^{-1\prime}SL^{-1}$ and $L^{-1\prime}RL^{-1}$. Cholesky decomposition can be used to factorize herd variance-covariance matrix such that $H=L'L$, where L is an upper triangular matrix. In two-step diagonalization as described by Lin and Smith (1990), since $L^{-1\prime}\,SL^{-1}$ and $L^{-1\prime}RL^{-1}$ are positive definite and symmetric matrices, an orthogonal matrix B can be found to diagonalize (or nearly diagonalize) simultaneously these two p.d. matrices and if $U=B'L$, then the matrix U would simultaneously diagonlize H.S and R such that:

$U'RU = B'L^{-1\prime}\,R\,L^{-1}B = D$; $U'SU = B'L^{-1\prime}\,SL^{-1}B = C$; $U'HU = B'L^{-1}\,(L'L)\,L^{-1}\,B = I$.

G) Testing for Closeness to Diagonality:

   1. Computing the function: $$Q(A_i) = \frac{Det\,[diag\,(B'A_iB)]}{Det\,(A_i)}$$

$MQ = \prod_{i=1}^{k} Q(A_i)$ if $B=B_M$ where B_M is based on ML version of F-G

$LQ = \prod_{i=1}^{k} Q(A_i)$ if $B=B_L$ where B_L is based on LS version of F-G

Note that if A_i is diagonal, $Q(A_i) = 1$. If all A_i matrices are diagonal, MQ = LQ = 1.

2. Computing the values of SS_U, SS_{MT} and SS_{LT} where

   SS_U = ratio of sum of squares (SS) of the off-diagonal elements in A_i to SS of the diagonal elements in A_i.

   SS_{MT} = ratio of SS of the off-diagonal elements in $B'A_iB$ to SS of the diagonal elements in $B'A_iB$, where the matrix B was estimated by ML version of F-G algorithm.

   SS_{LT} = ratio of SS of the off-diagonal elements in $B'A_iB$ to SS of the diagonal elements in $B'A_iB$, where the B matrix was estimated by LS version of F-G algorithm.

3. Computer the function E_p such that $Ep = 100\,[e_{ii} - d_{ii}]\,/d_{ii}$ where d_{ii} = the values of the diagonal elements of $B'A_iB$, and e_{ii} = the corresponding eigenvalues of the diagonal elements of $B'A_iB$, where the matrix $B = B_M$ or $B = B_L$.

RESULTS AND DISCUSSION

Table 1 shows different herd proportions and different combinations of h^2, r_g and r_p used to compute herd, sire and error variance-covariance matrices. Ten combinations were chosen to represent: 1) Low, medium and high herd proportions, and 2) Low, medium and high genetic and phenotypic parameters (h^2, r_g, r_p). Combination 2 represents milk traits (milk, fat and protein yields). Combination 4 is an example of fitness or reproduction traits. High heritability estimates ($h^2 > .4$) were given in combination 3 and 5 for traits like fat, protein and solids percentages as given by Schmidt and Van Vleck (1974). Combinations 1 and 6 represent parameters for some type traits, Hypothetical situations, like high herd proportion (.9), were given to clarify the behaviour of MQ and LQ. Combinations (1,6), (2,7), (4,11) and (3,5,12) have the same parameters, but combinations within each set differ in herd proportion.

Table 2 shows the values of MQ and LQ for different genetic combinations. In the diagonal case, combinations 1,4,6,8 and 11 had values of MQ 1.101, 1.041, 1.100, 1.044 and 1.041, respectively. The correpsonding values of LQ were 1.141, 1.114, 1.102, 1.049 and 1.049, respectively. The values of MQ and LQ were close to one. These combinations were characterized by low h^2 and low phenotypic and genetic able 1. Combinations of h^2, r_g and r_p with different herd proportions

Combination	Herd Proportions	h^2			r_g			r_p		
		1	2	3	1	2	3	1	2	3
1	.1	.050	.070	.030	.200	.250	.150	.220	.270	.200
2	.1	.250	.220	.200	.800	.900	.920	.900	.920	.950
3	.1	.400	.420	.450	.800	.900	.920	.900	.920	.950
4	.3	.001	.008	.005	.100	.080	.050	.120	.100	.080
5	.3	.400	.420	.450	.800	.900	.920	.900	.920	.950
6	.5	.050	.070	.030	.200	.250	.150	.220	.270	.200
7	.5	.250	.220	.200	.800	.900	.920	.900	.920	.950
8	.7	.050	.080	.100	.100	-.200	-.150	.150	-.250	.200
9	.7	.400	.420	.450	.100	.080	.050	.120	.100	.080
10	.7	.400	.420	.450	.800	-.900	-.920	.900	-.920	-.950
11	.9	.001	.008	.005	.100	.080	.050	.120	.100	.080
12	.9	.400	.420	.450	.800	.900	.920	.900	.920	.950

correlations. As a consequence, herd, sire and residual variance-covariance matrices have large diagonal elements relative to the off-diagonal elements and both versions of F-G algorithm produced nearly diagonal matrices.

Combination 9 with high h^2 and low r_g and r_p gave values of MQ and LQ close to one (1.015 and 1.036), so genetic and phenotypic correlations are determining factors in this combination. On the other hand, the rest of the combinations have values of MQ and LQ larger than one. These combinations were characterized by high values of h^2, r_g and r_p. Consequently, herd, sire and residual variance-covariance matrices are far from being simultaneously diagonlized by F-G algorithm.

In non-diagonal herd variance-covariance matrices (i.e. some correlations exist between traits which may be due to common environmental conditions), the values of LQ were higher than those of MQ for all combinations. As expected, the values of both MQ and LQ in non-diagonal herd variance-covariance matrix

Table 2. Values of MQ[1] and LQ[2] for one and two step diagonalizations by different parameter combinations

Combinations	MQ			LQ		
	One step		Two step	One step		Two step
	Diag H[3]	NDiag H[4]		Diag H	NDiag H	
1	1.101	1.246	1.062	1.141	3.683	1.161
2	1.430	1.837	1.059	1.531	9.538	1.223
3	1.279	1.658	1.094	1.446	8.906	1.228
4	1.041	1.228	1.029	1.114	2.529	1.029
5	1.276	1.650	1.033	1.434	8.898	1.037
6	1.100	1.263	1.062	1.102	4.094	1.161
7	1.423	1.838	1.056	1.627	130.417	115.016
8	1.044	1.222	1.045	1.049	1.332	1.131
9	1.015	1.166	1.001	1.036	1.579	1.002
10	1.276	1.629	170.881	2.244	711.674	19.036
11	1.041	1.228	1.566	1.049	3.957	2.432
12	1.276	1.651	1.094	1.369	935.209	1.230

[1] $MQ = \prod_{i=1}^{3} |diag(B'A_iB)| / |A_i|$ where $B=B_M$.

[2] $LQ = \prod_{i=1}^{3} |diag(B'A_iB)| / |A_i|$ where $B=B_L$.

[3] Diag H = diagonal herd variance-covariance matrix.

[4] Ndiag H = non-diagonal herd variance-covariance matrix.

were higher than the corresponding values computed from diagonal herd variance-covariance matrix. Valuesranged from 1.332 to 4.094. High parameter combinations showed some large values. As defined previously, the function of MQ or LQ is the product of ratios. and each ratio was computed as the determinant of the diagonal transformed matrix $B'A_i B$ to the determinant of the untransformed matrix A_i. for MQ for low parameter combinations (1.4,6,8 and 11) were 1.246, 1.2228. 1.263. 1.222 and 1.228, respectively. However, high parameter combinations (2.3.5.7.10 and 12) gave values of MQ 1.837, 1.658.

Table 3. Values of SS_U, SS_{MT}, and SS_{LT} for diagonal herd, sire, and error variance-covariance matrices by parameter combinations.[1]

Combination	SS_U		SS_{MT}			SS_{LT}		
	(S)	(R)	(H)	(S)	(R)	(H)	(S)	(R)
1	17.491	.020	.020	.019	.006	.045	.014	4.814
2	.797	.154	.154	.001	.001	.109	.002	1.793E-6
3	.797	.107	.108	.001	.001	.068	.004	9.073E-7
4	008	015	.001	.005	.002	.010	.039	.027E-2
5	.797	1.450	.107	001	.001	.068	.025	.027e-2
6	17.392	19.523	.021	001	.001	013	.024	.011
7	1.255	1.256	.154	001	.002	.107	.006	.001
8	.039	.083	.007	.004	.017	.006	.001	.024
9	129.158	63.367	.005	.001	.003	.001	.005	.012
10	800	689	.107	001	.001	.053	.013	.020
11	008	.015	.001	005	.014	.000	.008	.014
12	.800	.690	.107	.001	.001	.000	1.161	1.355

[1] SS_U = Ratio of SS of the off-diagonal to diagonal elements of untransformed matrix

SS_{MT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by ML version of F-G algorithm.

SS_{LT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by LS version of F-G algorithm.

SS_U(S), SS_U(R) = SS_U of sire and error variance-covariance matrices, respectively

SS_{MT}(H), SS_{MT}(S), SS_{MT}(R) = SS_{MT} of herd, sire and error variance-covariance matrices, respectively.

SS_{LT}(H), SS_{LT}(S), SS_{LT}(R) = SS_{LT} of herd, sire and error variance-covariance matrices, respectively.

1.650, 1.838, 1.629 and 1.651, respectively, on the other hand, values of LQ for low parameter combinations ranged from 1.332 to 4.094. High parameter combinations showed some large values. As defined previously, the function of MQ or LQ is the product of ratios, and each ratio was computed as the determinant of the diagonal transformed matrix B'A, B to the determinant of the untransformed matrix A_i.

Therefore, one can expect large values of MQ or LQ under the following situations: 1) one or more of the untransformed matrix has a small determinant (close to singularity); 2) the ratio of the largest to the smallest eigenvalue is large for any of the untransformed matrix; for example, this ratio is 54:1 and 71:1 for sire and error variance-covariance matrices in combination 7 (has large value of LQ); or 3) large determinants of the diagonal matrix $B'A_iB$ (e.g., the large values of LQ in combinations 10 and 12 in non-diagonal case is due mainly to large determinants of the transformed herd, sire and error variance-covariance matrices. Large values of LQ were observed because LQ is for positive and non-positive definite matrices. So, large values of LQ were noticed for combination with near singularity matrices; for example $L^{-1}{}'SL^{-1}$ and $L^{-1}{}'RL^{-1}$ of combination 7 (LQ = 155.016) have determinants $4.935*10^{-6}$ and .011; so two step diagonalization works better in transforming variance-covariance matrices to near diagonal form. The difference between MQ and LQ values for combinations within each set (1,5), (2,7), (4,11), and (3,5,12) is due to the difference in herd proportion.

Table 3 shows the values of SS_U, SS_{MT} and SS_{LT} for herd, sire and error variance-covariance matrices for the diagonal case. SS_{MT} was comupted using the transformation matrix B_M, and SS_{LT} was computed using B_L. The values of SS_U for diagonal herd variance-covariance matrix are zero for all combinations. However, SS_{MT} and SS_{LT} get larger than zero after transformation. On the other hand, SS_{MT} and SS_{LT} for sire and error variance-covariance matrices have smaller values than SS_U for the same effects. This reduction in sums of squares is due mainly to the effect of orthogonal transformation of B matrix on reducing the sum of squares of off-diagonal elements (Ali, 1993).

Table 4 shows the values of SS_U, SS_{MT} and SS_{LT} for herd, sire and error variance-covariance matrices in non-diagonal case. A considerable reduction in sum of squares of the off-diagonal elements of herd, sire and error variance-covariance matrices was observed. Greater reduction in sum of squares was observed for error variance-covariance matrices.

Table 4. Values of $SS_{U}{}^{1}$, SS_{MT}, and SS_{LT} for non-diagonal herd, sire and error variance-covariance matrices by parameter combinations.

Combination	SS_{U}			SS_{LT}			SS_{LT}		
	(H)	(S)	(R)	(H)	(S)	(R)	(H)	(S)	(R)
1	32.610	11.564	22.934	.012	.051	.013	.052	.035	.000
2	22.134	.807	.773	.316	.003	.001	.214	.006	.000
3	16.995	.807	.677	.216	.004	.002	.159	.001	.001
4	13.532	41.699	81.751	.055	.025	.028	.076	.048	.003
5	16.679	.807	.676	.214	.004	.002	.157	.011	.000
6	15.388	73.521	22.934	.0186	.576	.013	.017	.087	.016
7	22.132	.807	.773	.317	.003	.000	.206	.011	.001
8	11.494	98.597	10.358	.638	.404	.410	1.347	.111	.042
9	16.670	41.891	82.113	.028	.014	.021	.007	.035	.064
10	16.670	.774	.696	.231	.001	.001	.126	.208	.199
11	13.532	41.699	81.751	.055	.025	.028	.000	.112	.086
12	16.662	.807	.677	.214	.004	.002	.000	1.137	1.684

SS_{U} = Ratio of sums of squares(SS) of the off-diagonal to diagonal elements of untransformed matrix

SS_{MT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by ML version of F-G algorithm.

SS_{LT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by LS version of F-G algorithm.

SS_{U}(S), SS_{U}(R) = SS_{U} of sire and error variance-covariance matrices, respectively.

SS_{MT}(H), SS_{MT}(S), SS_{MT}(R) = SS_{MT} of herd, sire and error variance-covariance matrices, respectively.

SS_{LT}(H), SS_{LT}(S), SS_{LT}(R) = SS_{LT} of herd, sire and error variance-covariance matrices, respectively.

Testing near diagonality by sums of squares is based on the criteria of minimizing the sum of squares of off-dingonal elements of weighted function of $B'A_iB$ (Clarkson, 1988). An alternative approach is to include the diagonal elements in computing sum of squares of each matrix of $B'A_iB$ or one might give different weights to the off-diagonal elements. Reduction in sum of squares gets larger with two step diagonalization.

Table 5. Values of SS_U[1], SS_{MT}[2] and SS_{LT}[3] for two-step diagonalization by different parameter combinations

Combination	SS_U[4] $L^{-1'}SL^{-1}$	SS_U[5] $L^{-1'}RL^{-1}$	SS_{MT}[6] $Q'L^{-1'}SL^{-1}Q$	SS_{MT}[7] $Q'L^{-1'}RL^{-1}Q$	SS_{LT}[8] $Q'L^{-1'}SL^{-1}Q$	SS_{LT}[9] $Q'L^{-1'}RL^{-1}Q$
1	4.951	9.258	.016	.025	.007	$.373*10^{-8}$
2	.672	.579	$.150*10^{-4}$	$.422*10^{-5}$	$.374*10^{-4}$	$.113*10^{-8}$
3	.652	.579	$.108*10^{-4}$	$.715*10^{-5}$	$.369*10^{-4}$	$.938*10^{-8}$
4	$.215*10^{9}$	48.649	$.623*10^{-4}$	$.201*10^{-3}$	.488	$.300*10^{-10}$
5	.692	.553	$.177*10^{-4}$	$.798*10^{-6}$	$.232*10^{-4}$	$.302*10^{-8}$
6	14.953	9.258	$.160*10^{-3}$	$.248*10^{-3}$	$.871*10^{-3}$	$.492*10^{-10}$
7	.667	$.196*10^{12}$	$.150*10^{-4}$	.848	.379	1.158
8	22.194	11.884	$.719*10^{-4}$	$.212*10^{-3}$	$.699*10^{-3}$	$.118*10^{-9}$
9	86.527	46.399	$.638*10^{-5}$	$.271*10^{-5}$	$.131*10^{-4}$	$.214*10^{-8}$
10	.653	.579	.811	.750	.161	.117
11	204.085	48.649	.104	$.166*10^{-3}$	.337	$.726*10^{-10}$
12	.653	.579	$.110*10^{-4}$	$.715.10^{-5}$	$.372*10^{-4}$	$.938*10^{-8}$

[1] SS_U = Ratio of sums of squares (SS) of the off-diagonal to diagonal elements of untransformed variance-covariance matrix

[2] SS_{MT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by ML version of F-G algorithm.

[3] SS_{LT} = Ratio of SS of the off-diagonal to diagonal elements of transformed matrix by LS version of F-G algorithm.

[4] SS_U $L^{-1'}SL^{-1}$ = SS_U of sire variance-covariance matrix transformed with triangular matrix L^{-1}

[5] SS_U $L^{-1'}SL^{-1}$ = SS_U of error variance-covariance matrix transformed with triangular matrix L^{-1}

[6] SS_{MT} $Q'L^{-1'}SL^{-1}Q$ = SS_{MT} of sire variance-covariance matrix transformed by $L^{-1}Q$

[7] SS_{MT} $Q'L^{-1'}SL^{-1}Q$ = SS_{MT} of error variance-covariance matrix transformed by $L^{-1}Q$

[8] SS_{LT} = SS_{LT} of sire variance-covariance matrix transformed by $L^{-1}Q$

[9] SS_{LT} = SS_{LT} of error variance-covariance matrix transformed by $L^{-1}Q$

Table 6. Values of E_p for sire and error variance-covariance matrices in two-step diagonalization.

Combination	$L^{-1}SL^{-1}$	$B'L^{-1\prime}SL^{-1}B$ (ML)	$B'L^{-1\prime}SL^{-1}B$ (LS)	$L^{-1\prime}RL^{-1}$	$B'L^{-1\prime}RL^{-1}B$ (ML)	$B'L^{-1\prime}RL^{-1}B$ (LS)
1	144.471	61.132	14.968	184.067	4.201	0.001
	-78.002	-38.660	-8.991	90.671	-.731	0.020
	-94.204	-1.001	-17.843	-95.162	-8.802	0.002
3	160.016	.052	.162	182.242	.033	0.000
	-88.000	.221	6.423	-72.601	3.651	0.024
	-94.378	-3.500	1.143	-95.412	-8.804	0.064
5	154.201	.088	.118	188.488	.011	0.000
	-79.774	.571	.382	-91.003	-.109	-.004
	-89.739	-3.579	.329	-99.129	.001	.178
8	10.791	1.304	85.821	39.519	3.968	0.001
	-8.478	-1.103	-14.949	-14.269	2.271	0.001
	-8.039	-1.399	-30.400	-25.682	-8.955	0.003
12	158.872	.001	25.899	183.989	.002	.003
	-80.357	.002	.002	-90.349	3.568	.003
	-94.345	.019	91.099	-95.439	-8.902	.005

Table 5 shows the values of sums of squares SS_U for $L^{-1\prime}SL^{-1}$ and $L^{-1\prime}RL^{-1}$, and the values of SS_{MT}, and SS_{LT} after applying F-G diagonalization algorithm. Sums of squares of $B'(L^{-1\prime}SL^{-1})B$ and $B'(L^{-1\prime}RL^{-1})B$ of both maximum likelihood and least squares versions get smaller after transformation by F-G algorithm. Moreover, these sums of squares get close to zero for combinations (3, 5 and 12) characterized by high heritability and high genetic correlations.

In this study H was decomposed first such that $H=L'L$ where L is an upper triangular matrix, followed by simultaneous diagonalization of $L^{-1\prime}SL^{-1}$ and $L^{-1\prime}RL^{-1}$. Alternatively, one can find L such that $L'L = S$ (or $L'L = R$) and simultaneously diagonalize $L^{-1\prime}HL^{-1}$ and $L^{-1\prime}RL^{-1}$ (or $L^{-1\prime}SL^{-1}$). In practice, $B'L^{-1\prime}SL^{-1}B$ and

$B'L^{-1\prime}R^{-1}L^{-1}B$ are not exactly diagonal but are diagonally dominant, i.e. each off-diagonal element is very small relative to its diagonal element in that row.

The function E_p measures the deviation of each eigenvalue from the corresponding diagonal element of the transformed matrix. In complete diagonality all values of E_p are zero and all diagonal elements of the matrix are identical with the respective eigenvalues. As shown in Table 6, the two-step diagonalization reduced the values of E_p. The matrices $B'(L^{-1\prime}SL^{-1})B$ and $B'(L^{-1\prime}RL^{-1})B$ are close to diagonality because their eigenvalues are very close to their respective diagonal elements than $L^{-1\prime}SL^{-1}$ and $L^{-1\prime}RL^{-1}$. Furthermore, $B'(L^{-1\prime}RL^{-1})B$ are closer to diagonality than $B'(L^{-1\prime}SL^{-1})B$ because the function E_p for the former is close to zero in most combinations(Table 6).

Using different initial matrices for running a Fortran program of F-G algorithm to compute the transformation matrix B may result in a different orthogonal matrix B since each orthogonal matrix B represents a local minimum for minimizing both functions of MQ and LQ of F-G algorithm. This is due mainly to the structure of the function Q which is the product of k functions each involving a different untransformed matrix. On the other hand, using different initial matrix B might result in a unique transformation matrix B which is a global minimum for minimizing both functions of MQ and LQ of F-G algorithm. An identity matrix or the matrix of eigenvectors of different untransformed matrix could be used as initial matrices for computing the transformation matrix B (B_M or B_L).

CONCLUSIONS

Both ML and LS versions of F-G algorithm will not achieve complete diagonalization for all different combinations. However, simultaneous diagonalization of three matrices is closer to diagonality in diagonal case than in non-diagonal case. In one step diagonalization, F-G algorithm can achieve near diagonality with low parameter combinations. Matrices with small determinant (i.e. near singularity), give large values of Q measures of deviation from diagonality. Diagonalization based on F-G algorithm works better on matrices with large diagonals relative to the off-diagonal elements. Using equal or unequal weights showed no effect on the orthogonal matrix B computed by both versions of F-G algorithm. Simultaneous diagonalization by F-G algorithm has an important application in animal breeding, mainly in transforming a multitrait mixed model with more than one random classification into a unitrait analysis. The only class of models applicable is that in which variance-covariance matrices for all random effects can be diagonalized simultaneously. Based on the results of this study, the maximum likelihood version of F-G algorithm was found to be more effective in achieving simultaneous diagonalization than the least squares version and thus was preferable for transforming multitrait into unitrait mixed model analysis.

REFERENCES

1. Ali, A.K.A. (1993). Application of F-G diagonalization algorithm to restricted maximum likelihood estimation of variance components. J. of King Saud University (Agric. Sciences) 6:2.
2. Clarkson, D.B. (1987). A remark on algorithm As 211. The F-G diagonalization. Appl. Statist. 37:147.
3. Clarkson, D.B. (1988). A least square version of algorithm. As 211: The F-G diagonalization algorithm. Appl. Statist. 38:317.
4. Flury, B.N. (1984). Common principal components in K groups. J. Am. Stat. Assoc. 79:892.
5. Flury, B.N. (1988). Common principal components and related multivariate models. Wiley, New York, NY
6. Flury, B.N. and W. Gautschi (1984). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. SIAM, J. Scientific and Computing 7:169.
7. Flury, B.N. and G. Constantine (1985). The F-G diagonalization algorithm. Appl. Statist. 35:177.
8. Meyer, K. (1985). Maximum likelihood estimation of variance components for multitrait mixed model with equal design matrices. Biometrics. 41:153.
9. Jensen, J. and I.L. Mao (1988). Transformation algorithm in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait: a review. J. Anim. Sci. 66:2750.
10. Lin, C.Y. and S.P. Smith (1990). Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects. J. Dairy Sci. 73:2494.
11. Schmidt, G.H. and L.D. Van Vleck (1974). Principles of Dairy Science. W.H. Freeman and Co., San Francisco.