Computing genetic evaluations through application of generalized least squares to an animal model

G.F.S. HUDSON

Department of Animal Sciences, University of Maryland,
College Park, Maryland 20742, U.S.A.

Summary

The animal model for performance data is rewritten in the form of a fixed model with uncorrelated residuals. This transformation allows the use of computationally efficient methods for solving generalized least squares problems to obtain best linear unbiased predictions of breeding values. Application of a specific algorithm to the transformed model is described and compared with more traditional approach of obtaining solutions to the mixed model equations through an iterative process. The new approach may have merit for recursive prediction of breeding values from sequentially collected data.

Key words : Animal model, QR algorithm, best linear unbiased prediction.

Résumé

Calcul des valeurs génétiques par application des moindres carrés
généralisés à un modèle animal

Le modèle statistique d'interprétation des données de contrôle de performances est réécrit sous la forme d'un modèle à effets fixés, avec des résidus non corrélés. Cette transformation permet d'utiliser des méthodes de résolution des problèmes de moindres carrés généralisés, efficaces sur le plan calculatoire, pour obtenir les meilleures' prédictions linéaires sans biais des valeurs génétiques. L'application d'un algorithme spécifique au modèle transformé est décrite et comparée à l'approche plus classique d'obtention de solutions aux équations du modèle mixte par un processus itératif. L'approche proposée peut présenter un intérêt pour une prédiction récursive des valeurs génétiques à partir de données recueillies de façon séquentielle.

Mots clés : Modèle animal, algorithme QR, meilleure prédiction linéaire sans biais.

I. Introduction

Initial research into the use of mixed model equations to obtain best linear unbiased predictions of breeding values or transmitting abilities concentrated on application of separate sire and cow models to dairy cattle data, based on the pioneering work of

C.R. Henderson at Cornell University (see THOMPSON, 1979 for a review). In 1976, HENDERSON & QUAAS described an animal model for simultaneous evaluation of males and females based on available performance data from both evaluated and related animals. The animal model has advantages over the separate sire and cow models :

— for non sex-limited traits the sire's own performance record is used which is equivalent in reliability to many half-sib progeny records for traits with high heritability ;

— all relationships can be used, completely accounting for selection among dams ;

— nonrandom mating of bulls causes no bias in the evaluations ;

— evaluation of females is improved (compared with a within-herd cow model) by incorporation of across herd relationships and by direct incorporation of sires' evaluation rather than an approximation.

A major disadvantage of the animal model is the commonly large number of equations to be solved. For example, implementing the animal model for dairy cattle evaluation in the northeastern U.S. involved over 1,500,000 animal equations and nearly one quarter million fixed effect equations (WESTELL, 1984). Attempts have been made to reduce the computational effort involved in using the animal model. A major contribution was made by QUAAS & POLLAK (1980) who used a gametic model for records of animals without progeny, thus the only equations needed were those of parent and ancestor animals. A practical application of this « reduced animal model » to swine evaluation by HUDSON & KENNEDY (1985) required only 10 to 20 p. 100 of the animal equations needed by the full animal model.

Recursive prediction allows further reduction of the number of equations to be solved by evaluating only animals of interest. For example, dead animals need not be directly evaluated although information from these animals is retained (HUDSON, 1984). Recursive prediction techniques require the variance-covariance matrix of estimates and prediction errors, and therefore preclude the use of iterative methods to obtain solutions. This paper describes a procedure for representing the animal model as a fixed model. Advanced computational methods for solving generalized least squares problems can then be applied to obtain the necessary variances and covariances for recursive prediction.

II. The animal model

The animal model is demonstrated here for a simple situation :

— each animal has a single record on one trait ;

— each animal with a record has both sire and dam identified ;

— some ancestors with recorded progeny, but without performance records of their own, do not have identified sire or dam and are a random sample from a base population ;

— the only fixed effect is the mean of records observed in a particular period (years, seasons, etc.) ;

— animals may be inbred and the model allows overlapping generations.

Relaxation of these assumptions will be indicated where appropriate.

The $n \times 1$ vector of data, $\mathbf{y}_i$, observed in the ith period is described by the equation

$$\mathbf{y}_i = \mathbf{1}_i \, \mu_i + \mathbf{a}_i + \mathbf{e}_i \tag{1}$$

where

$\mathbf{1}_i$ is a vector of n_i ones,

μ_i is the mean of $\mathbf{y}_i$

$\mathbf{a}_i$ is the vector of additive genetic effects (breeding values) of animals with records in $\mathbf{y}_i$, random with mean zero and variance $\mathbf{A}_i \sigma_a^2$, and

$\mathbf{e}_i$ is the vector of residuals, random with mean zero and variance $\mathbf{I} \sigma^2$.

The breeding value of the kth animal in $\mathbf{a}_i$ can be represented as

$$a_k = .5 a_k^s + .5 a_k^d + v_k \tag{2}$$

where a_k^s and a_k^d are breeding values of the sire and dam of the kth animal, and v_k represents Mendelian sampling and is defined as the deviation of a_k from $.5 \, (a_k^s + a_k^d)$, the mid-parent value. Equation (2) is the gametic model of QUAAS & POLLAK (1980). If all the parents of animals represented in $\mathbf{a}_i$ are in $\mathbf{a}_{i-1}$ then the vector representation of (2) is

$$\mathbf{a}_i = \mathbf{T}_{i, \, i-1} \, \mathbf{a}_{i-1} + \mathbf{v}_i \tag{3}$$

where $\mathbf{T}_{i, \, i-1}$ is an $n_i \times n_{i-1}$ matrix relating offspring in $\mathbf{a}_i$ to parents in $\mathbf{a}_{i-1}$. The kth row of $\mathbf{T}_{i, \, i-1}$ has $.5$ in columns corresponding to the parents of the kth animal in $\mathbf{a}_i$, and zeros elsewhere. For overlapping generations, equation (3) is written as

$$\mathbf{a}_i = \sum_{j=0}^{i-1} \mathbf{T}_{i, \, j} \, \mathbf{a}_j + \mathbf{v}_i. \tag{4}$$

For $i > 0$, var $\mathbf{v}_i = \mathbf{D}_i \, \sigma_a^2 : \mathbf{D}_i$ is a diagonal matrix with the kth diagonal equal to $.5 - .25 \, (F_s + F_d)$, F_s and F_d are Wright's coefficients of inbreeding for the sire and dam of the kth animal, and σ_a^2 is the additive genetic variance in the base population. For $i = 0$, $\mathbf{a}_0$ is the vector of breeding values of ancestors from the base population. If these animals are a random sample, define $\mathbf{a}_0 \equiv \mathbf{v}_0$ with var $\mathbf{v}_0 = \mathbf{I} \sigma_a^2$. The recursive equation (4) can be succinctly written for all animals as

$$\mathbf{a} = \mathbf{Ta} + \mathbf{v} \tag{5}$$

where $\mathbf{a} = (\mathbf{a}_i)$, $\mathbf{v} = (\mathbf{v}_i)$ for $i = 0, 1, ..., N$, and $\mathbf{T}$ is a block matrix with subdiagonal blocks equal to $\mathbf{T}_{i, \, j}$ of (4), and null blocks on and above the diagonal. For example, for 3 periods of data (5) is

$$
\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}
=
\begin{bmatrix} \varnothing & & & \\ \mathbf{T}_{1, \, 0} & \varnothing & & \\ \mathbf{T}_{2, \, 0} & \mathbf{T}_{2, \, 1} & \varnothing & \\ \mathbf{T}_{3, \, 0} & \mathbf{T}_{3, \, 1} & \mathbf{T}_{3, \, 2} & \varnothing \end{bmatrix}
\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}
+
\begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} .
$$

Rearrangement of (5) leads to

$$\mathbf{a} = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{v} \tag{6}$$

Thus (HENDERSON, 1976 ; QUAAS, 1976 ; THOMPSON, 1977)

$$\text{var } \mathbf{a} = (\mathbf{I} - \mathbf{T})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{T})^{-t}, \tag{7}$$

and (HENDERSON, 1976 ; QUAAS, 1976)

$$\text{var}^{-1}\mathbf{a} = (\mathbf{I} - \mathbf{T})^{t} \mathbf{D}^{-1} (\mathbf{I} - \mathbf{T}). \tag{8}$$

In (7) and (8), $\mathbf{D}$ is block diagonal matrix of $\mathbf{D}_i$, superscript t indicates matrix transpose, and superscript $-$ t indicates transposition of the inverse matrix.

The model for all data from N periods is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{a}_* + \mathbf{e} \tag{9}$$

where $\mathbf{y} = (\mathbf{y}_i)$, $\mathbf{X} = \Sigma^+ \mathbf{1}_i$, $\boldsymbol{\mu} = (\mu_i)$, $\mathbf{a}_* = (\mathbf{a}_i)$, and $\mathbf{e} = (\mathbf{e}_i)$ with $i = 1, 2, ..., N$ and Σ^+ indicating direct matrix summation.

The complete data equation (9) can be combined with the animal equation by writing (5) as

$$\mathbf{0} = (\mathbf{I} - \mathbf{T}) \mathbf{a} - \mathbf{v} \tag{10}$$

then (9) and (10) together are

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & (\varnothing \quad \mathbf{I}) \\ \varnothing & (\mathbf{I} - \mathbf{T}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{a} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ -\mathbf{v} \end{bmatrix} \tag{11}$$

with $\mathbf{0}$ and $\varnothing$ being null vectors and matrices of appropriate order. DUNCAN & HORN (1972) describe (11) as a linear dynamic recursive model which forms the basis of recursive prediction (HUDSON, 1984). Equations similar to (11) have been described for a general (i.e., with $\mathbf{T}$ null) mixed model by DEMPSTER et al. (1981, 1984) and for a sire model by FRIES (1984). Similar equations also have been presented for fixed models in terms of ridge regression (MARQUARDT, 1970) and for variable selection in multiple regression problems (ALLEN, 1974).

For data from three periods equations (11) have seven « rows »

$$\begin{matrix} 1. \\ 2. \\ 3. \\ 4. \\ 5. \\ 6. \\ 7. \end{matrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_1 & & & \mathbf{I}_1 & & \\ & \mathbf{1}_2 & & & \mathbf{I}_2 & \\ & & \mathbf{1}_3 & & & \mathbf{I}_3 \\ & & & \mathbf{I}_0 & & \\ & & & -\mathbf{T}_{1,0} & \mathbf{I}_1 & \\ & & & -\mathbf{T}_{2,0} & -\mathbf{T}_{2,1} & \mathbf{I}_2 \\ & & & -\mathbf{T}_{3,0} & -\mathbf{T}_{3,1} & -\mathbf{T}_{3,2} & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ -\mathbf{v}_0 \\ -\mathbf{v}_1 \\ -\mathbf{v}_2 \\ -\mathbf{v}_3 \end{bmatrix} .$$

Other models familiar to animal breeders can be derived from (11). Animals represented in $\mathbf{a}_3$ are nonparents, i.e., they have their own performance records but no progeny data. If row 7 of the example is subtracted from row 3, the latter becomes

$$\mathbf{y}_3 = \mathbf{1}_3 \mu_3 + \sum_{j=0}^{2} \mathbf{T}_{3,\,j}\, \mathbf{a}_j + (\mathbf{e}_3 + \mathbf{v}_3)$$

(thus eliminating $\mathbf{a}_3$) and the model becomes the reduced animal model of QUAAS and POLLAK (1980). If $\mathbf{y}_i$ contains only daughter records and $\mathbf{a}_i$ contains only male breeding values, then suitable redefinition of $\mathbf{e}_i$, $\mathbf{T}_{i,\,j}$ and $\mathbf{D}_i$ generates either the sire model or the maternal grandsire model of QUAAS et al. (1979). Fixed effects other than the period mean can be incorporated by replacing μ_i with a vector $\boldsymbol{\beta}_i$ and replacing $\mathbf{1}_i$ with appropriately defined incidence (or regressor) matrices. Fixed effects may fall into two categories (HUDSON, 1984) : those common to all data (e.g., sex, age, $\boldsymbol{\beta}$) and those specific to data collected in the ith period (e.g. year-season means, $\boldsymbol{\beta}_i$). To avoid rank deficiency problems in $\mathbf{X}$, fixed effects should be defined so that $\boldsymbol{\beta}$ and all $\boldsymbol{\beta}_i$ are jointly estimable. If no pedigree information is missing, the $\mathbf{T}_{i,\,j}$ completely account for selection and no genetic groups are required in the model.

The utility of (11) is demonstrated by treating $\mathbf{a}$ as fixed and setting up the generalized least squares normal equations

$$\left[\begin{matrix} \begin{matrix} \mathbf{X}^t\mathbf{X} & (\varnothing & \mathbf{X}^t) \\ \varnothing & \begin{bmatrix} \varnothing & \varnothing \\ \varnothing & \mathbf{I} \end{bmatrix} \\ \mathbf{X} & \end{matrix} + (\mathbf{I} - \mathbf{T})^t\, \mathbf{D}^{-1}\, (\mathbf{I} - \mathbf{T})\, \alpha \end{matrix} \right] \left[\begin{matrix} \hat{\mu} \\ \hat{\mathbf{a}} \end{matrix} \right] = \left[\begin{matrix} \mathbf{X}^t\mathbf{y} \\ \mathbf{0} \\ \mathbf{y} \end{matrix} \right] \qquad (12)$$

with $\alpha = \sigma^2/\sigma_a^2$ and $\hat{\ }$ indicating solution, not parameter. Equations (12) are identical to the mixed model equations for (9), therefore $\hat{\mu}$ and $\hat{\mathbf{a}}$ are the best linear unbiased estimator and predictor of μ and $\mathbf{a}$. See DEMPSTER et al. (1981) for a Bayesian derivation of (11) and (12) from a general mixed model. Thus equations (11) represent a method by which an animal model for multiperiod data can be written in terms of a fixed model with heterogeneous variances. There are numerous computing algorithms for least squares problems applied to fixed linear models, and some may provide means by which animal breeders can easily process periodic data through the use of (11).

III. The QR algorithm for generalized least squares

A. General principles

The QR algorithm (STEWART, 1973 ; LAWSON & HANSON, 1974 ; VAN LOAN, 1976) for solving generalized least squares problems is described here for the fixed linear model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \; \mathbf{e} \sim (\mathbf{0}, \mathbf{E}) \qquad (13)$$

with $\mathbf{X}$ having full column rank, q, and $\mathbf{E} = \mathbf{C}\mathbf{C}^t$. Define $\mathbf{C}$ as the lower triangular decomposition of $\mathbf{E}$. If $\mathbf{E}$ is diagonal (as is the variance-covariance matrix of the

residuals in (11)), then **C** is also diagonal. Define $\tilde{X}$ and $\bar{y}$ as solutions to $C [\tilde{X} \ \bar{y}] = [X \ y]$ and write the model for $\bar{y}$ as

$$\bar{y} = \tilde{X}b + \tilde{e}, \ \tilde{e} \sim (0, I). \tag{14}$$

Thus a standardization of variables has transformed the model to one with uncorrelated residuals with unit variances. Now define **Q** such that

1) **Q** is orthogonal (i.e., $Q^t \ Q = I$),

$$2) \ Q\tilde{X} = \begin{bmatrix} \tilde{R} \\ \varnothing \end{bmatrix},$$

$$3) \ Q\bar{y} = \begin{bmatrix} \bar{y}_{\bullet} \\ \bar{y}_{\bullet\bullet} \end{bmatrix}, \text{ and}$$

4) $\tilde{R}$ is an upper triangular matrix with order and rank equal to the rank of **X**, q.

The generalized least squares solution to (13) is the solution to

$$Q\tilde{X}\hat{b} = Q\bar{y} \tag{15}$$

or, equivalently,

$$\begin{bmatrix} \tilde{R} \\ \varnothing \end{bmatrix} \hat{b} = \begin{bmatrix} \bar{y}_{\bullet} \\ \bar{y}_{\bullet\bullet} \end{bmatrix} \tag{16}$$

or, equivalently,

$$\tilde{R}\hat{b} = \bar{y}_{\bullet}. \tag{17}$$

That $\hat{b}$ is the generalized least squares solution to (13) is demonstrated by premultiplying (15) by $X^t Q^t$ which yields the familiar normal equations. Solving equations (17) is trivial because $\tilde{R}$ is, by definition, upper triangular. Less easy is determining **Q**, which is the product of a series of q orthogonal matrices $Q_q, Q_{q-1}, ... Q_1$ with Q_i defined so that, for $\tilde{x}_i$ being the ith column of $Q_{i-1}Q_{i-2} ... Q_1\tilde{X}$, $Q_i\tilde{x}_i$ is a vector with zeros in all elements below the ith element, thus satisfying the second and fourth conditions above. Each Q_i is a Householder reflection matrix described next.

B. Householder matrices

For any vector **w** the corresponding Householder matrix is $H = I - 2uu^t/u^tu$ where $u = w$ with the first element replaced by $w_1 + \delta (w^tw)^{.5}$; $\delta = 1$ if $w_1 \geq 0$ and $\delta = -1$ if $w_1 < 0$. Thus $Hw = [\delta (w^tw)^{.5}, 0, 0, ... 0]^t$, i.e., the Householder matrix has « zeroed-out » all but the first element of **Hw**. Straightforward multiplication shows that **H** is orthogonal. To zero-out elements below the diagonal in $\tilde{x}_i$, define w_i as $(\tilde{x}_{i,i} \ \tilde{x}_{i+1,i} ... \tilde{x}_{n,i})^t$ and then

$$Q_i = \begin{bmatrix} I_{i-1} & \varnothing \\ \varnothing & H_i \end{bmatrix}.$$

Note that to apply the QR algorithm the matrix **Q** in (15) need not be explicitly created or stored. Householder reflections are applied directly to $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$. For example, to apply the Householder reflection $\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^t / \mathbf{u}^t\mathbf{u}$ to a vector $\mathbf{z}$ (either a column of $\tilde{\mathbf{X}}$ or $\tilde{\mathbf{y}}$) simply requires subtracting from $\mathbf{z}$ a scalar multiple of $\mathbf{u}$ (GOULT *et al.*, 1974) : $\mathbf{H}\mathbf{z} = \mathbf{z} - (2\mathbf{u}^t\mathbf{z}/\mathbf{u}^t\mathbf{u})\mathbf{u}$. In situations that requires retaining **Q** for future use, then nonzero elements of **u** can occupy the zeroed-out elements of **w** (LAWSON & HANSON, 1974).

The coefficient matrix of (11) is both sparse and highly structured ; the design of the QR algorithm can utilize both properties. For example, the Householder matrix can be constructed to operate on only nonzero elements. Appropriate reordering of rows of $[\tilde{\mathbf{X}} \ \tilde{\mathbf{y}}]$ can exploit the structure of the equations.

C. Updating the QR with new data

Data are often collected sequentially over an extended period of time. New data are combined with old data to provide updated solutions ; with the QR algorithm this updating procedure is relatively easy. Suppose the QR algorithm has been applied to the data and incidence matrix of the transformed model (14) so that $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{y}}_*$ of (17) exist. New data, $\mathbf{y}_2$, are collected which fit the model

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{b} + \mathbf{e}_2, \ \mathbf{e}_2 \sim (\mathbf{0}, \mathbf{E}_2) \tag{18}$$

with cov $(\mathbf{e}, \mathbf{e}_2) = \varnothing$. Model (18) is transformed to

$$\tilde{\mathbf{y}}_2 = \tilde{\mathbf{X}}_2\mathbf{b} + \tilde{\mathbf{e}}_2, \ \tilde{\mathbf{e}}_2 \sim (\mathbf{0}, \mathbf{I}) \tag{19}$$

and the QR algorithm is applied to

$$\begin{bmatrix} \tilde{\mathbf{R}} \\ \tilde{\mathbf{X}}_2 \end{bmatrix} \hat{\mathbf{b}} = \begin{bmatrix} \tilde{\mathbf{y}}_* \\ \tilde{\mathbf{y}}_2 \end{bmatrix} \tag{20}$$

to obtain

$$\tilde{\mathbf{R}}_2\hat{\mathbf{b}} = \tilde{\mathbf{y}}_{*2}. \tag{21}$$

Householder reflections are applied to (20) so that $\tilde{\mathbf{X}}_2$ is zeroed-out. Note that **Q** of (15) is not required for the updating procedure.

The updating described here should not be confused with recursive prediction by which new solutions are obtained from new data along with previous solutions and their variances (HUDSON, 1984). This updating procedure simply combines the new data with the old equations and applies the QR algorithm.

D. Variance-covariance matrix of solutions

The variance-covariance matrix of $\hat{\mathbf{b}}$ is required to generate confidence intervals and to test hypotheses about $\hat{\mathbf{b}}$, and is also required in recursive estimation and prediction (HUDSON, 1984). Obtaining $\hat{\mathbf{b}}$ by solving (17) does not requires $\tilde{\mathbf{R}}^{-1}$ because $\tilde{\mathbf{R}}$ is triangular. However var $\hat{\mathbf{b}} = \tilde{\mathbf{R}}^{-1}\tilde{\mathbf{R}}^{-t}\sigma^2$. Thus the inverse may be required in certain applications. Inverting a triangular matrix is straightforward, but, in certain cases, var $\hat{\mathbf{b}}$

itself is not needed. For example, to test the hypothesis $\mathbf{K'b = m}$ requires var $\mathbf{K'\hat{b} = K'}$ var $\hat{\mathbf{b}}$ $\mathbf{K = K'\bar{R}^{-1}\bar{R}^{-1}K}$. LAWSON & HANSON (1974) suggest computing this as $\mathbf{AA'}$ where $\mathbf{A}$ is solution to $\mathbf{AR = K'}$. Diagonal elements of $\bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}^{-1}$ are needed for confidence intervals and are calculated by the sum of squares of elements in each row of $\bar{\mathbf{R}}^{-1}$.

IV. Comparing the QR algorithm with normal equations

Criteria by which computer algorithms are often compared include central processing unit time required for various parts of the process, amount of storage needed and accuracy of final solutions. If $\mathbf{X}$ is square, then the QR algorithm requires two-thirds the storage locations of the normal equations method (LAWSON & HANSON, 1974). As the number of rows of $\mathbf{X}$ increases relative to the number of columns the advantage of the QR over normal equations decreases. For example, if there are 5 times as many rows as columns in $\mathbf{X}$, then the QR requires as much as 90 p. 100 of the storage locations needed by the normal equations. If the ratio of number of rows to number of columns exceeds 50, then storage requirements of each method are essentially equal.

LAWSON & HANSON (1974) discuss in detail the accuracy of the QR algorithm and compare various methods. In general, to obtain solutions of comparable accuracy, the normal equations must be computed with higher precision arithmetic than the QR algorithm. This is only of concern to animal breeders if the linear model contains covariates, and even then careful avoidance of collinearity and scaling of variables can lessen problems due to loss of accuracy. If $\mathbf{X}$ is an incidence matrix with no regressors each method produce solutions of equal accuracy. Ranking animals for the purpose of making selection decisions certainly does not require solutions to machine accuracy.

The major criterion for comparing the QR algorithm with normal equations is computer time. If $\mathbf{X}$ is square and has no special exploitable structure then the QR method requires the same number of computer operations as setting up and solving the normal equations. If, as is usual, the number of rows in $\mathbf{X}$ exceeds the number of columns then the normal equations method is approximately twice as fast as the QR method. If $\mathbf{X}$ is an incidence matrix with no regressor variables then further time savings are available because generating the normal equations requires only summation operations and no multiplications. Applying the QR algorithm to incidence matrices also reduces the computational work compared with the general case. For example, triangularizing $\mathbf{X}$ in (11) is rapid due to the simple structure of $\mathbf{X}$. Additional time can be saved by applying normal equations to (11) because $(\mathbf{I} - \mathbf{T})'\mathbf{D}^{-1} (\mathbf{I} - \mathbf{T})$ of (12) can be generated directly by the methods of HENDERSON (1976) and QUAAS (1976), which requires less work than zeroing-out $(\mathbf{I} - \mathbf{T})$ in (11). The actual time requirements of applying QR to (11) have yet to be determined.

Animal breeders do not require solutions of great accuracy and often use iterative methods such as successive overrelaxation (GOULT et al., 1974) to solve (12). This approach has been investigated by numerous authors, but BLAIR & POLLAK (1984) conclude : « Sufficiently accurate ranking of animals for selection purposes is achieved long before random effect solutions converge, especially if the [reduced animal model of QUAAS & POLLAK (1980)] is used ». Even their criterion for convergence, the mixed model equivalent of $c = (\hat{e}'\mathbf{X'X}\hat{e}/\mathbf{y'X'Xy})^{.5} < .0001$, ($\hat{e}$ is the estimated residual vector) did not involve solutions accurate to machine precision. VAN VLECK & EDLIN (1984) evaluated 484 Holstein bulls for calving difficulty of their calves : 4 iterations produced

solutions deemed sufficiently accurate with $c < .0005$; an extra 16 iterations yielded $c < .0001$. GOULT *et al.* (1974) suggested that, for any system of non-symmetric non-sparse equations : « ... an iterative method may have the advantage... provided the number of iterations needed to give the accuracy desired is less than about [one-third the number of equations] ». This obviously held in the case of VAN VLECK & EDLIN (1984). The advantage of iterative methods is even greater if the equations are sparse and symmetric, as (12) often are. BLAIR & POLLAK (1984) did note that more iterations were required to obtain an accurate indication of genetic trend, than to rank animals for selection.

V. Discussion

Generation and iterative solving of mixed model equations is, for a large class of linear models common in animal breeding, rapid and straightforward. Sufficiently accurate solutions may be obtained after only a few iterations, especially if the ranking of animals is the only concern. If estimates of genetic trend are required, many more iterations are needed. Under what conditions, then, may the QR algorithm be superior (in terms of computer usage) than the more traditional methods ? The answer relies on exploiting the triangularity of $\tilde{R}$.

First, equations (17) are easy to solve and the solution to any particular equation, the kth say, requires only solutions from $k + 1$ to N. In order to solve the kth equation, solutions from 1 to $k - 1$ are not needed. Thus fixed effect solutions are not required to compute estimated breeding values. If equations are ordered as shown in the example, then ancestor equations are not required to obtain solutions for younger animals. These unneeded equations may in fact be discarded after they have been triangularized.

Second, updating $\tilde{R}$ with new data is quite straightforward. The only equations that change are those of parents and common fixed effects ; new equations are needed for new fixed effects and new animals. Once the update to $\tilde{R}_2$ has been accomplished, solving (21) is trivial. Although updating mixed model equations is also straightforward, those updated equations then need to be re-iterated.

Third, inverting a triangular matrix is easy, thus obtaining variances of fixed solutions and of errors of prediction is also easy. This may facilitate the use of recursive prediction of breeding values which is not possible if solutions are obtained by iterating mixed model equations.

References

ALLEN D.M., 1974. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics,* **16**, 125-127.

BLAIR H.T., POLLAK E.J., 1984. Comparison of an animal model and an equivalent reduced animal model for computational efficiency using mixed model methodology. *J. Anim. Sci.,* **58**, 1090-1096.

DEMPSTER A.P., RUBIN D.B., TSUTAKAWA R.K., 1981. Estimation in covariance components models. *J. Am. Stat. Assoc.,* **76**, 341-353.

DEMPSTER A.P., SELWYN M.R., PATEL C.M., ROTH A.J., 1984. Statistical and computational aspects of mixed model analysis. *Appl. Stat.,* **33**, 203-214.

DUNCAN D.B., HORN S.B., 1972. Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Stat. Assoc.,* **67**, 815-821.

FRIES L.A., 1984. *A study of weaning weights in Hereford cattle in the state of Rio Grande do Sul, Brazil,* Ph.D. Thesis, Iowa State Univ., Ames, Iowa.

GOULT R.J., HOSKINS R.F., MILNER J.A., PRATT M.J., 1974. *Computational methods in linear algebra,* 204 pp., Wiley, New York.

HENDERSON C.R., 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics,* **32**, 69-83.

HENDERSON C.R., QUAAS R.L., 1976. Multiple trait evaluation using relatives' records. *J. Anim. Sci.,* **43**, 1188-1197.

HUDSON G.F.S., 1984. Extension of a reduced animal model to recursive prediction of breeding values. *J. Anim. Sci.,* **59**, 1164-1175.

HUDSON G.F.S., KENNEDY B.W., 1985. Genetic evaluation of swine for growth rate and backfat thickness. *J. Anim. Sci.,* **61**, 83-91.

LAWSON C.L., HANSON R.J., 1974. *Solving least squares problems,* 340 pp., Prentice-Hall, Englewood, New Jersey.

MARQUARDT D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics,* **12**, 519-562.

QUAAS R.L., 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics,* **32**, 949-953.

QUAAS R.L., EVERETT R.W., McCLINTOCK A.C., 1979. Maternal grandsire model for dairy sire evaluation. *J. Dairy Sci.,* **62**, 1648-1654.

QUAAS R.L., POLLAK E.J., 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.,* **51**, 1277-1287.

STEWART G.W., 1973. *Introduction to matrix computations,* 441 pp., Academic Press, New York.

THOMPSON R., 1977. The estimation of heritability with unbiased data. II. Data available on more than two generations. *Biometrics,* **33**, 497-504.

THOMPSON R., 1979. Sire Evaluation. *Biometrics,* **35**, 339-353.

VAN LOAN C.F., 1976. *Lectures in least squares.* Technical Report TR 76-279, Dept. of Comp. Sci., Cornell Univ., Ithaca, New York.

VAN VLECK L.D., EDLIN K.M., 1984. Multiple trait evaluation of bulls for calving ease. *J. Dairy Sci.,* **67**, 3025-3033.

WESTELL R.A., 1984. *Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle.* Ph.D. Thesis, Cornell Univ., Ithaca, New York.