Original article

Solving animal model equations through an approximate incomplete Cholesky decomposition

V Ducrocq

Institut National de la Recherche Agronomique, Station de Génétique Quantitative et Appliquée, F- 78352 Jouy-en-Josas Cedex, France

Summary − A general strategy is described for the design of more efficient algorithms to solve the large linear systems $\mathbf{Bs} = \mathbf{r}$ arising in (individual) animal model evaluations. This strategy, like Gauss–Seidel iteration, belongs to the family of "splitting methods" based on the decomposition $\mathbf{B} = \mathbf{B}^* + (\mathbf{B} - \mathbf{B}^*)$, but, in contrast to other methods, it tries to take maximum advantage of the known sparsity structure of the mixed model coefficient matrix: $\mathbf{B}^*$ is chosen to be an approximate incomplete Cholesky factor of $\mathbf{B}$. The resulting procedure requires the solution of 2 triangular systems at each iteration and 2 readings of the data and pedigree file. This approach was applied to an animal model evaluation on 15 type traits and milking ease score from the French Holstein Association with 955 288 animals and 4 fixed effects, including group effect for animals with unknown parents. Its convergence was compared with a standard iterative procedure.

genetic evaluation / animal model / computing algorithm / type trait

Résumé − Résolution des équations du modèle animal à l'aide d'une décomposition de Cholesky incomplète et approchée. *Une stratégie générale est décrite pour l'obtention d'algorithmes plus efficaces dans le but de résoudre les grands systèmes linéaires* $\mathbf{Bs} = \mathbf{r}$, *caractéristiques des évaluations de type « modèle animal». Cette stratégie, comme l'itération de Gauss–Seidel, appartient à la famille des « méthodes d'éclatement» basées sur la décomposition* $\mathbf{B} = \mathbf{B}^* + (\mathbf{B} - \mathbf{B}^*)$, *mais contrairement aux autres méthodes, elle tente de mettre à profit autant que possible la structure creuse (connue) de la matrice des coefficients des équations du modèle mixte:* $\mathbf{B}^*$ *est prise égale à une approximation de la décomposition de Cholesky incomplète de* $\mathbf{B}$. *La procédure résultante nécessite la résolution de 2 systèmes triangulaires à chaque itération et 2 lectures du fichier de données et de généalogie. Cette approche a été appliquée à une évaluation de type « modèle animal» sur 15 caractères de morphologie et une note de facilité de traite provenant de l'Unité pour la Promotion de la race Prim'Holstein, concernant 955 288 animaux et 4 effets fixes, y compris un effet groupe pour les animaux issus de parents inconnus. Sa vitesse de convergence a été comparée à une approche itérative courante.*

évaluation génétique / modèle animal / algorithme de calcul / caractère de morphologie

INTRODUCTION

In most developed countries and for all major domestic species, joint genetic evaluations of up to a few millions of animals is routinely computed using Best Linear Unbiased Prediction (BLUP). When an individual animal model is used, this supposes the solution of a linear system whose size is larger than the total number of animals to evaluate. Such a task, repeated a few times a year and for several traits of economic importance can be a real challenge, even on the most advanced computers. To reduce this computational burden, algebraic manipulations of the equations have been proposed with 2 main aims: 1), a decrease of the system size by absorption of some effects (*eg* the permanent environment effects) by use of an equivalent model (*eg* the reduced animal model of Quaas and Pollak, 1980) or by use of particular transformations (*eg* the canonical transformation which reduces multitrait evaluations to sets of single trait analyses (Thompson, 1977; Arnason, 1982, 1986); 2), an increase in the sparsity of the coefficient matrix (*eg* Quaas' transformation which makes possible the estimation of group effects for unknown parents only, with no more difficulties than if they were parents (Quaas and Pollak, 1981; Quaas, 1988) or the QP transformation which, when traits are sequentially available, makes the coefficient matrix in a multiple trait analysis much sparser, (Pollak, 1984). Unfortunately, such tools are not always applicable either because they are restricted to special data structure or because they are not always computationally advantageous. Slow convergence is not a real problem for moderate-size research applications, for which general purpose programs are available (*eg* Groeneved *et al*, 1990). But it can make routine evaluations prohibitive in the case of large-scale applications.

In any case, with or without algebraic manipulation, the linear system is virtually always too large to be solved directly and an iterative solution has to be performed. The algorithms chosen have been in most cases very simple and initially designed for the solution of general problems, without any particular attention to the type of problems animal breeders are dealing with. The most frequently used ones are the Gauss–Seidel, Successive Overrelaxation and Jacobi iterations (Golub and Van Loan, 1983). Unfortunately, for animal models, these can be extremely slow to converge (requiring up to several hundreds of iterations) especially when groups of unknown parents (Quaas, 1988) are added, or when more than one fixed effect is considered (Ducrocq *et al*, 1990).

An important breakthrough in the search for an efficient solution of animal models was the discovery of the possibility to "iterate on data", a strategy proposed by Schaeffer and Kennedy (1986) which avoids the actual computation and storage of the whole coefficient matrix. This can be considered as one of the first uses of the known nonzero structure of the mixed model equations in designing more efficient algorithms.

Indeed, a careful look at this structure gave birth to new approaches for a fast solution of some parts of these equations, in a block iterative context: Poivey (1986) showed that by considering in the inverse A^{-1} of the relationship matrix, only the diagonal and the terms relating an animal to its parent of a given sex and by correcting accordingly the right-hand side for the other terms of A^{-1} using solutions from the previous iteration, the resulting system has a very simple and sparse

Cholesky decomposition and can be solved directly. Likewise, Ducrocq *et al* (1990) proposed for the solution of the additive genetic value part of the mixed model equations the use of 2 decompositions of the type described by Poivey (1986) – considering first the relationship between an animal and its dam and second between an animal and its sire. They also proposed to absorb the equations for additive genetic values into the equations for fixed effects after correction of the right-hand side for all off-diagonal elements of $\mathbf{A}^{-1}$ ("pseudo-absorption"). Convergence was improved, but not as much as might be desired for huge routine applications. The main drawback to such an approach is the rather tedious programming.

This paper presents a more general procedure for the design of new algorithms taking maximum advantage of the known sparsity structure of the mixed model equations. An application to a large animal model evaluation is described and its performance is compared with a standard iterative procedure.

MATERIAL AND METHODS

Principles for a new algorithm

Let $\mathbf{Bs} = \mathbf{r}$ [1]

be the linear system to be solved. If $\mathbf{B}$ is very large, system [1] can be solved directly only if $\mathbf{B}^{-1}$ or $\mathbf{C}$, the Cholesky factor of $\mathbf{B}(\mathbf{C} = \mathbf{CC}', \mathbf{C}$ lower triangular) is sparse and easy to obtain. If this is not the case, consider $\mathbf{B}^*$, a matrix "close to" $\mathbf{B}$ and whose inverse or Cholesky factor is sparse and easily computed and write [1] as:

$$\left[\mathbf{B}^* + (\mathbf{B} - \mathbf{B}^*)\right]\mathbf{s} = \mathbf{r} \qquad [2]$$

Then, the following functional iterative procedure can be implemented. At iteration $(k + 1)$, solve:

$$\mathbf{B}^*\mathbf{s}^{(k+1)} = \mathbf{r} - (\mathbf{B} - \mathbf{B}^*)\mathbf{s}^{(k)} \qquad [3]$$

Expressions [3] is very general and is the base of a family of iterative algorithms known as Splitting Methods (Coleman, 1984). If $\mathbf{B}^*$ is simply a diagonal matrix whose diagonal elements are those of $\mathbf{B}$, [3] is the Jacobi iteration. If $\mathbf{B}^*$ is the lower triangular part of $\mathbf{B}$, including the diagonal, [3] is the Gauss–Seidel iteration (Golub and Van Loan, 1983).

If the starting value for $\mathbf{s}$ is taken to be $\mathbf{s}^{(0)} = \mathbf{0}$, expression [3] leads to:

$$\mathbf{B}^*\mathbf{s}^{(1)} = \mathbf{r}$$
$$\mathbf{B}^*\mathbf{s}^{(2)} = \mathbf{r} - (\mathbf{B} - \mathbf{B}^*)\mathbf{s}^{(1)}$$
$$= \mathbf{r} - \left(\mathbf{Bs}^{(1)} - \mathbf{r}\right)$$
$$= \mathbf{r} - \mathbf{\Delta}^{(1)}$$
$$\mathbf{B}^*\mathbf{s}^{(3)} = \mathbf{r} - (\mathbf{B} - \mathbf{B}^*)\mathbf{s}^{(2)}$$
$$= \mathbf{r} - \mathbf{\Delta}^{(1)} - \mathbf{\Delta}^{(2)}$$

and
$$B^*s^{(k+1)} = r - \sum_{m=1}^{k} \Delta^{(k)}$$
[4]

with
$$\Delta^{(k)} = Bs^{(k)} - r$$

ie, the right-hand side in [4] is updated at each iteration. An even more general iterative algorithm is obtained by mimicking the method of successive overrelaxation (Golub and Van Loan, 1983)

$$B^*s^{(k+1)} = \omega \left(r - \sum_{m=1}^{k} \Delta^{(k)} \right) = r^{(k)}$$
[5]

where ω is a relaxation parameter, which corresponds to the splitting of B:

$$B = \left(\frac{1}{\omega} B^* \right) + \left(B - \frac{1}{\omega} B^* \right).$$

The next paragraph will illustrate how B^* can be chosen in a particular situation. Consider the following animal model:

$$y = Xb + Z_0 a_0 + e$$
[6]

where: y is a vector of observations;

b is a vector of fixed effects;

a_0 is an n_a-vector of additive genetic values for all animals with and without records;

e is a vector of normally distributed residuals with $E(e) = 0$;

X and Z_0 are incidence matrices.

Assume $E(a_0) = Qg$ where g is an n_g-vector of group effects, defined only for the animals with unknown parents. Q is a matrix relating animals in a_0 with groups (Westell, 1984; Robinson, 1986; Quaas, 1988).

$$\text{Var} \begin{bmatrix} a_0 \\ e \end{bmatrix} = \begin{bmatrix} A_0 \sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

The mixed model equations used to compute Best Linear Unbiased Estimates of b and g and BLUP's of a_0 can be written (Quaas, 1988):

$$\begin{bmatrix} X'X & X'Z_0 & 0 \\ Z'_0X & Z'_0Z_0 + \alpha A_0^{-1} & -\alpha A_0^{-1}Q \\ 0 & -\alpha Q'A_0^{-1} & \alpha Q'A_0^{-1}Q \end{bmatrix} \begin{bmatrix} b \\ a_0 \\ g \end{bmatrix} = \begin{bmatrix} X'y \\ Z'_0y \\ 0 \end{bmatrix}$$
[7]

with $\alpha = \dfrac{\sigma_e^2}{\sigma_a^2}$

Define $\mathbf{a} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{g} \end{bmatrix}$, $\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_0^{-1} & -\mathbf{A}_0^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}_0^{-1} & \mathbf{Q}'\mathbf{A}_0^{-1}\mathbf{Q} \end{bmatrix}$

and, accordingly, $\mathbf{Z} = [\mathbf{Z}_0 \quad 0]$

Then [7] is:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \tag{8}$$

If model [6] includes only one fixed effect (which, for clarity, will be called a herd-year effect), $\mathbf{X}'\mathbf{X}$ is diagonal with hth diagonal element equal to the number of observations in herd-year h. There is at most one nonzero element per row j of $\mathbf{Z}'\mathbf{X}$ (or column of $\mathbf{X}'\mathbf{Z}$). This nonzero element is in column h and is always equal to 1 when animal j has its record in herd-year h. $\mathbf{Z}'\mathbf{Z}$ is diagonal with diagonal element equal to 1 for rows corresponding to animals with one record and 0 for animals without record and groups. Finally, if equations are ordered in such a way that progeny precede parents, and such that parents precede groups in $\mathbf{a}$, $\mathbf{A}^*$ is of the form $\mathbf{A}^* = \mathbf{L}\wedge^{-1}\mathbf{L}'$ (Quaas, 1988) where $\mathbf{L}$ is a $(n_a + n_g) \times n_a$ matrix with 3 non zero elements per column. If j_s and j_d represent the indices of the sire (or the sire's group) and the dam (or the dam's group) of animal j, column j has a 1 in row j (= the jth diagonal element of $\mathbf{L}$) and $a - 0.5$ in rows j_s and $j_d \cdot \wedge^{-1}$ is a $(n_a \times n_a)$ diagonal matrix with jth element equal to $\delta_j = 4/(m_j + 2)$ where m_j is the number of known parents of $j(m_j = 0, 1 \text{ or } 2)$.

Given this rather simple structure for $\mathbf{B}$, the coefficient matrix in [8], the following choice of $\mathbf{B}^*$ in [5] is suggested: take $\mathbf{B}^* = \mathbf{T}^*\mathbf{T}^{*'}$, where $\mathbf{T}^*$ is the incomplete Cholesky factor of $\mathbf{B}$, ie, the matrix obtained by setting t_{ij} to zero in the Cholesky factorization of $\mathbf{B}$ each time the corresponding element b_{ij} in $\mathbf{B}$ is zero (Golub and Van Loan, 1983, p 376). Equivalently, $\mathbf{B}^* = \mathbf{T}\mathbf{D}\mathbf{T}'$ where $\mathbf{T} = \{t_{ij}\}$ is lower triangular with unit diagonal elements and $\mathbf{D}$ is diagonal with positive diagonal elements d_j, and $\mathbf{T}$ and $\mathbf{D}$ are computed using the algorithm sketched in figure 1. The $\mathbf{T}\mathbf{D}\mathbf{T}'$ factorization has an important advantage over the standard Cholesky factorization: it does not require computation of square roots.

For $j = 1 \ldots n = $ dimension of $\mathbf{B}$

$$d_j = b_{jj} - \sum_{m=1}^{j-1} t_{jm}^2 d_m$$

For $i = j + 1, n$

if $b_{ij} = 0$

then $t_{ij} = 0$

else $t_{ij} = (b_{ij} - \sum_{m=1}^{j-1} t_{im}t_{jm}d_m)/d_j$

Fig 1. Computation of the incomplete factorization of a general coefficient matrix $\mathbf{B} = \{b_{ij}\}$.

A few remarks need to be made at this stage. First, it is known that in the general case, the incomplete Cholesky factorization of a positive definite matrix is

not always possible (negative numbers appear in $\mathbf{D}$ ie, the computation of diagonal elements in $\mathbf{T}^*$ requires square roots of negative numbers). It will be shown that this is never the case here.

Second, the coefficient matrix in [8] can be rewritten as:

$$
\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} (\mathbf{X}'\mathbf{X}) & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{I} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} \\ \mathbf{0} & \mathbf{Q}' \end{bmatrix} \tag{9}
$$

and it clearly appears that column h corresponding to herd-year effect h has $n_h + 1$ nonzero elements: a 1 on the diagonal, and $1/n_h$ on each of the n_h rows corresponding to animals with records in herd-year h. Hence, the incomplete $\mathbf{TDT}'$ factorization can be applied to the lower right part of the product in [9], ie, on $\mathbf{QPQ}' = (\mathbf{Z}'\mathbf{Z} + \alpha \mathbf{A}^*) - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, which is also the lower right part of [8] after absorption of the fixed effect equations.

Third, a strict application of the algorithm in figure 1 would lead to nonzero elements relating mates, as in $\mathbf{A}^*$. Given the particular structure of $\mathbf{A}^* = \mathbf{L}\Lambda^{-1}\mathbf{L}'$, we would like to have these elements being 0 too, such that the lower right part of $\mathbf{T}$ has the same nonzero structure as $\mathbf{L}$. However, $\mathbf{L}$ is a $(n_a + n_g) \times n_a$ matrix whereas the lower right part of $\mathbf{T}$ is supposed to be a $(n_a + n_g) \times (n_a + n_g)$ square matrix. We will assume (or choose) that the lower right part of $\mathbf{T}$ corresponding to groups is dense. Consequently, $\mathbf{TDT}'$ is only an approximation of the true incomplete Cholesky decomposition.

Algorithm

The computation of $\mathbf{T}$ and $\mathbf{D}$ does not require the coefficient matrix $\mathbf{B}$ to be explicitly set up. For animal j, b_{jj} in $\mathbf{B}$ is equal to:

$$
b_{jj} = x_j + \alpha \delta_j + \left[\sum_{p \text{ progeny of } j} \frac{\alpha \delta_p}{4} \right] \tag{10}
$$

with $x_j = 1 - \dfrac{1}{n_h}$ if j has its record in herd-year h and $x_j = 0$ otherwise.

Given the structure imposed on $\mathbf{T}$, the only elements that are nonzero in column j are $t_{aj}, a = j_s$ or $a = j_d$ where j_s and j_d are the indices of the sire (or sire group) and the dam (or dam group) or j. Henderson's rules (Henderson, 1976) of construction of $\mathbf{A}^*$ imply:

$$
b_{aj} = -\frac{1}{2}\alpha \delta_j \tag{11}
$$

Another consequence of the chosen structure for $\mathbf{T}$ is that the product $t_{im}t_{jm}$ in figure 1 is always 0 except when m is the herd-year effect where i and j have their records or m is a progeny of i and j. Since t_{ij} is computed only for $i = a = j_s$ and $i = a = j_d$, $t_{im}t_{jm}$ is nonzero only: 1), when j and its parent have their record in the same herd–year; and 2), when j is mated to its own sire or dam. Both events are sufficiently rare to be ignored (as $\mathbf{B}^*$ is not exact, anyway) and then:

$$
t_{j_s j} = t_{j_d j} = -\frac{1}{2}\alpha \frac{\delta_j}{d_j} \tag{12}
$$

The fact that j_s or j_d may or may not correspond to a group is irrelevant here. It is also essential to notice that t_{aj} need not be stored, as it is easily obtained from d_j.

Replacing t_{jm} by [12] in figure 1 and using [10], we get:

$$d_j = x_j + \alpha\delta_j + \left[\sum_{p \text{ progeny of } j} \left(\frac{\alpha\delta_p}{4} - \frac{\alpha^2\delta_p^2}{4d_p} \right) \right] \qquad [13]$$

For columns corresponding to groups, we have, from the structure of $\mathbf{A}^*$:

$$b_{ij} = \sum_{\substack{p \text{ progeny} \\ \text{of groups } i \text{ and } j}} \frac{\alpha\delta_p}{4} \qquad [14]$$

and $t_{im}t_{jm}$ in figure 1 is different from 0 each time m is a progeny of groups i and j.

Therefore, just before undertaking the dense factorization of $\mathbf{G}$ where $\mathbf{G}$ is the current part corresponding to groups, we have:

$$g_{jj} = \sum_{p \text{ progeny of } j} \left(\frac{\alpha\delta_p}{4} - \frac{\alpha^2\delta_p^2}{4d_p} \right) \qquad [15]$$

$$g_{ij} = \sum_{\substack{p \text{ progeny} \\ \text{of groups } i \text{ and } j}} \left(\frac{\alpha\delta_p}{4} - \frac{\alpha^2\delta_p^2}{4d_p} \right) \qquad [16]$$

(minor adaptations for terms corresponding to groups have to be made when the unknown sire and the unknown dam are in the same group).

Now, by noting that:

$$c_p = \left(\frac{\alpha\delta_p}{4} - \frac{\alpha^2\delta_p^2}{4d_p} \right) = \frac{\alpha\delta_p}{4} \left(1 - \frac{\alpha\delta_p}{d_p} \right)$$

and that $d_j = x_j + \alpha\delta_j > \alpha\delta_j$ for animals without progeny (all assumed to have a record), it follows by recurrence that $\left(1 - \dfrac{\alpha\delta_p}{d_p} \right) > 0$ for all p. Then $c_p > 0$. Therefore, from [13] and [15], $d_j > 0$ for all j and also $g_{jj} > 0$: the incomplete factorization is always possible.

Equations [12] and [13] lead to the practical algorithm given in figure 2.

1° Read the data file. Compute n_h.

2° Read the data file, with progeny preceding parents

for each animal j: $x_j = 1 - \dfrac{1}{n_h}$ if j has a record in j
$$x_j = 0 \text{ otherwise}$$

– add $x_j + \alpha\delta_j$ to d_j and store $\dfrac{1}{d_j}$

– add c_j to d_{j_s} and d_{j_d}

– if j_s and j_d are groups then

– if $j_s = j_d$ then add $2c_j$ to d_{j_d}

else add c_j to $b_{j_s j_d}$ and $b_{j_d j_s}$

Fig 2. Incomplete factorization of matrix $\mathbf{B}$ = practical case.

Iterative solution

From [5], [8] and [9], it appears that the general iterative algorithm involves at each iteration the following steps:

Solve $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Z'X(X'X)}^{-1} & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{u_b} \\ \mathbf{u_a} \end{bmatrix} = \begin{bmatrix} \mathbf{r_b} \\ \mathbf{r_a} \end{bmatrix}$ for $\mathbf{u_b}$ and $\mathbf{u_a}$ [17]

where $\begin{bmatrix} \mathbf{r_b} \\ \mathbf{r_a} \end{bmatrix}$ is the updated right-hand side.

Solve $\begin{bmatrix} \mathbf{X'X} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{v_b} \\ \mathbf{v_a} \end{bmatrix} = \begin{bmatrix} \mathbf{u_b} \\ \mathbf{u_a} \end{bmatrix}$ for $\mathbf{v_b}$ and $\mathbf{v_a}$ [18]

Solve $\begin{bmatrix} \mathbf{I} & \mathbf{(X'X)}^{-1}\mathbf{X'Z} \\ \mathbf{0} & \mathbf{T'} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{v_b} \\ \mathbf{v_a} \end{bmatrix}$ for $\mathbf{b}$ and $\mathbf{a}$ [19]

Update $\mathbf{r_b}$ and $\mathbf{r_a}$ as:

$$\begin{bmatrix} \mathbf{r_b} \\ \mathbf{r_a} \end{bmatrix} = \begin{bmatrix} \mathbf{r_b} \\ \mathbf{r_a} \end{bmatrix} - \omega \begin{bmatrix} \mathbf{X'(Xb + Za - y)} \\ \mathbf{Z'(Xb + Za - y)} + \alpha\mathbf{A}^*\mathbf{a} \end{bmatrix} \qquad [20]$$

Steps [17] to [20] can be condensed in such a way that only 2 readings of the date file are necessary at each iteration. Indeed, algebraic manipulation of these equations leads to the following requirements:

Compute $\mathbf{v_b} = \mathbf{(X'X)}^{-1}\mathbf{r_b}$ [21]

Solve $\mathbf{Tu_a} = \mathbf{r_a} - \mathbf{Z'X}\,\mathbf{v_b}$ for $\mathbf{u_a}$ [22]

Solve $\mathbf{T}'\mathbf{a} = \mathbf{D}^{-1}\mathbf{u_a}$ for $\mathbf{a}$ [23]

Compute $\mathbf{b} = \mathbf{v_b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{a}$ [24]

Update r_a and r_b.   [25]
The general sketch of the resulting algorithms is given in *Appendix 1*.

Dealing with several fixed effects

In many instances, the routine animal model evaluations involve more than one fixed effect. To adapt the algorithms to this frequent situation, one can distinguish between on one hand, a vector of fixed effects **b** with many levels (like the herd–year-(season) effects or the contemporary group effects in many applications) and on the other hand, a vector **f** of other "environmental effects" with fewer levels. Model [6] becomes:

$$\mathbf{y} = \mathbf{F}\mathbf{f} + \mathbf{X}\mathbf{b} + \mathbf{Z}_0\mathbf{a}_0 + \mathbf{e} \qquad [26]$$

where **X** and **F** are the incidence matrices corresponding to effects in **b** and **f** respectively. The resulting system of mixed model equations can be written:

$$\begin{bmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{X} & \mathbf{F}'\mathbf{Z} \\ \mathbf{X}'\mathbf{F} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{F} & \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{F}'\mathbf{y} \\ \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \qquad [27]$$

A block iterative procedure can be implemented involving at each iteration first the solution of:

$$(\mathbf{F}'\mathbf{F})\mathbf{f}^{(k+1)} = \mathbf{F}'(\mathbf{y} - \mathbf{X}\mathbf{b}^{(k)} - \mathbf{Z}\mathbf{a}^{(k)}) \qquad [28]$$

using Gauss-Seidel iteration and then, the solution of:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{b}^{(k+1)} \\ \mathbf{a}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'(\mathbf{y} - \mathbf{F}\mathbf{f}^{(k+1)}) \\ \mathbf{Z}'(\mathbf{Y} - \mathbf{F}\mathbf{f}^{(k+1)}) \end{bmatrix} \qquad [29]$$

using the algorithm described in this paper.

A LARGE-SCALE APPLICATION

Description

A BLUP evaluation based on an animal model was implemented to estimate cows' and bulls' breeding values for 15 linear type traits and milking ease score in the French Holstein breed. Records were collected by the French Holstein Association between 1986 and 1991 on 462 162 first and second lactation cows. The model considered in the analyses included an "age at calving (10 classes) × year (5) ×

region (8)" fixed effect, a "stage of lactation (15 classes) × year × region" fixed effect, a "herd × round × classifier" (36 420 classes) fixed effect and a random additive genetic value effect. Tracing back the pedigree of recorded animals, a total of 955 288 animals were evaluated. Sixty-six groups were defined according to year of birth, sex and country of origin of the animals with unknown parents. Here, **b** in [26] will refer to "herd × round × classifier" effects and **f** to other fixed effects.

The block iterative procedure described in *Dealing with several fixed effects* was implemented, using a relaxation parameter $\omega = 0.9$ in all analyses. To compare this procedure with a standard iterative algorithm, the mixed model equations were also solved using a program along the lines of Misztal and Gianola (1987), where the equations for fixed effects were solved using Gauss–Seidel iterations and the equations for additive genetic values were solved *via* second-order Jacobi iteration (see Misztal and Gianola (1987) for details). Group solutions were adjusted to average 0 at the end of each iteration, as proposed by Wiggans *et al* (1988). Indeed, this constraint had very little effect on convergence rate, as the average group effects solutions tended to a value very close to 0 anyway. Several relaxations parameters were used for the second-order Jacobi step. The same convergence criteria were computed in all cases and intermediate solutions were compared to "quasi-final" results.

RESULTS

Figures 3 and 4 illustrate for one of the traits – "rump width" ($\sigma_p = 1.4$, $h^2 = 0.25$) – the evolution of 2 convergence criteria: the norm of the change in solution between 2 consecutive iterations divided by the norm of the current solution vector (both considering elements in **a** only) and the maximum absolute change between 2 iterations. Rump width was considered as a trait representative of the average convergence rate of the procedure based on the incomplete Cholesky decomposition (hereafter referred to as ICD) for the 16 analyses: the value of the standardized norm of the change between 2 iterations obtained for rump width after 40 iterations was reached after 25 iterations for one trait, 33 to 40 iterations for 10 traits and between 41 and 46 iterations for 5 other traits. For rump width, 200 iterations with ICD and 300 iterations with the standard procedure (GS-J) were carried out and compared to intermediate solutions. The results are summarized in table I. Figure 5 shows the distribution of the changes by class of variation between 2 iterations for ICD.

Figures 3–5 and table I clearly show that convergence was much faster with ICD. Whatever the value of the relaxation parameter used, the evolution of the convergence criteria in GS-J tends to be very slow after a rather fast decline during the first iterations. This phenomenon was mentioned by Misztal *et al* (1987) and seems to worsen because of the size of the data set and the existence of 3 fixed effects in the model. The fact that ICD does not exhibit this undesirable characteristic may prove its robustness to not so well conditioned problems.

For practical purposes, 3 exact figures may be considered satisfactory for a proper ranking of all animals. Starting from 0 and with no acceleration procedure implemented (in contrast with *eg* Ducrocq *et al*, 1990), this requirement for all animals was reached for ICD after about 40 to 45 iterations. Even faster convergence may be achieved when starting from solutions of a previous evaluation or by
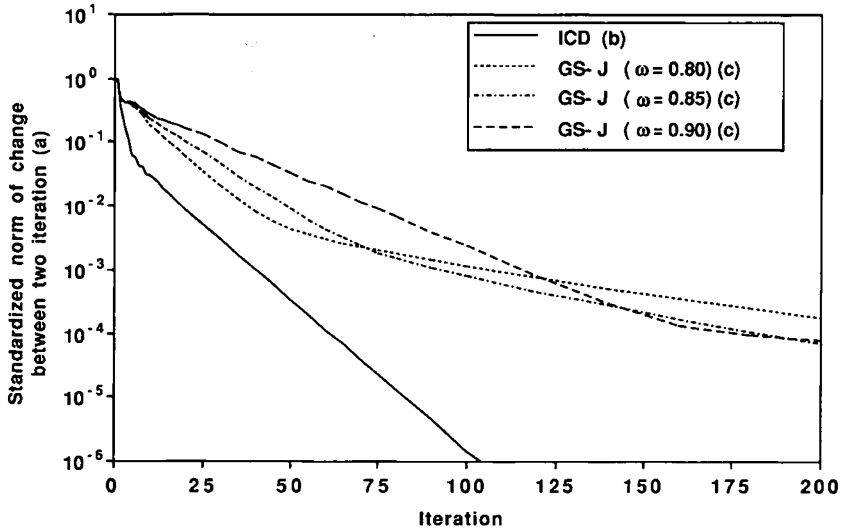
Fig 3. Standardized norm of change between 2 iterations.

$$(a) = \sqrt{\sum_i \left(a_i^{(k+1)} - a_i^{(k)}\right)^2} \Big/ \sqrt{\sum_i \left(a_i^{(k+1)}\right)^2} \; ; (b)$$ Procedure based on the incomplete Cholesky decomposition; (c) Gauss-Seidel iterative procedure with a second-order Jacobi step for genetic value effects (ω relaxation parameter).

optimizing the value of the relaxation parameter used. Figures 3 and 4 and table I show that the same convergence requirements are reached in about 150 iterations with GS-J. But the cost of each iteration is not the same: in the case of ICD, 2 copies of the data (+ pedigree) file, sorted in opposite direction, are read at each iteration, whereas GS-J requires one reading of the data file per fixed or random effect included in the model (ie, a total of 4 per iteration here) and one reading of the pedigree file. CPU time per iteration with optimized I/O operations was 12 s on an IBM 3090-17T computer for ICD and 24 s for GS-J: in both cases, the limiting factor for computation speed remains the I/O operations. Note however that a better "standard" strategy could have been designed: for example, it is possible to treat the first 2 fixed effects together in a block-iterative way as in [28] and solve for the herd-round-classifier (HRC) effect at the same time as the additive genetic value effect by reading a data file sorted by HRC, as proposed in Wiggans et al (1988). This reduces to 2 the number of copies of the data file read at each iteration. The fact that the first 2 fixed effects are not defined within HRC effects or vice-versa prevents any further reduction, as in the case described by Wiggans et al (1988). In conclusion, ICD appears to be at least 3 to 4 times faster than a standard procedure in the particular situation studied here.

For ICD, the main computing requirements in terms of core storage can be divided into 2 independent parts: 1), for the computation of d_j^{-1} (fig 2): a vector of length p, where p is the size of the coefficient matrix in [8]; 2), for the iterative
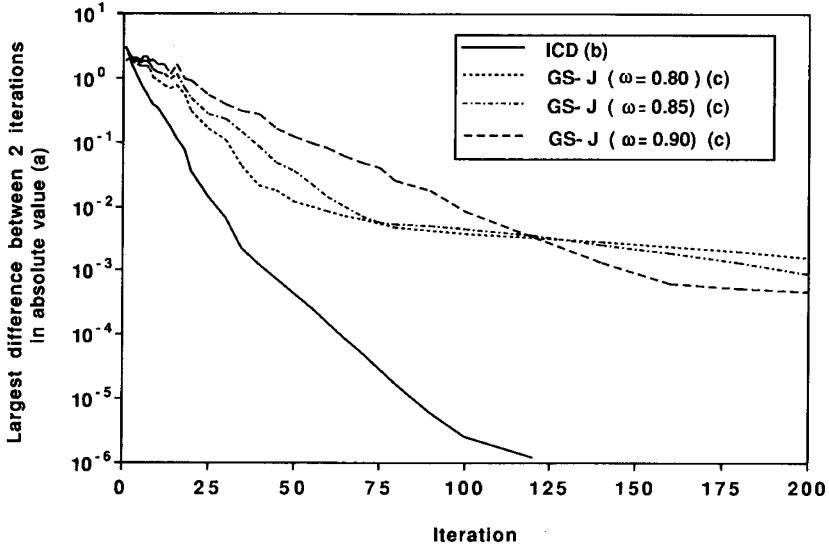
Fig 4. Maximum absolute change between 2 iterations. (a) $\sigma_p = 1.4$; (b) Procedure based on the incomplete Cholesky decomposition; (c) Gauss-Seidel iterative procedure with a second-order Jacobi step for genetic value effects (ω: relaxation parameter).
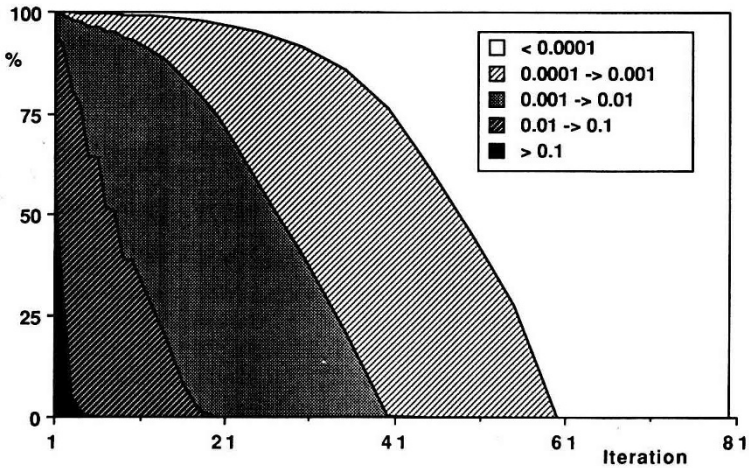

Fig 5. Distribution of absolute changes in solutions between 2 consecutive iterations by class of variation (with the procedure based on the incomplete Cholesky decomposition).

solution (*Appendix 1*): the nonzero blocks of matrices $\mathbf{F'F}$ and $\mathbf{F'X}$ (otherwise, 2 supplementary readings of the data file are necessary, in [28] and [29] and 3 vectors of length p (the right-hand side, and the current and previous vectors of solutions). This can be further reduced by noting that the previous vector of solutions for a

Table I. Comparison of solutions at intermediate iterations with "final" results.

Iteration	"Final" iteration	Correlation with final iteration	Average absolute difference	Maximum absolute difference	Standard deviation of solutions (% of final)
Based on the incomplete Cholesky decomposition					
10	200	0.986 8	0.059	0.226	92.6
20	200	0.998 6	0.020	0.074	97.3
40	200	0.999 98	0.002 4	0.008 7	99.8
60	200	1.000 000	0.000 27	0.001 0	100.0
Gauss–Seidel procedure with a second order Jacobi step (a)					
40	300	0.987 9	0.041	0.591	95.0
80	300	0.998 8	0.011	0.365	98.6
150	300	0.999 94	0.002 5	0.108	100.1

(a) $\omega = 0.9$.

is required only to compute convergence criteria and that current solutions for **a** need not to be stored for nonparent animals.

DISCUSSION AND CONCLUSIONS

It has been shown that in the context defined here the approximate incomplete Cholesky factor of the coefficient matrix always exists. However, this does not guarantee that B^* will be very close to B. In fact, when a large number of generations are considered, the terms from the true Cholesky factor of B which are ignored in the incomplete factorization may become large and one may wonder whether the proposed algorithm converges at all. Indeed, in the application presented here where pedigrees were traced up to 12 generation back in some cases, the algorithm diverges when no relaxation parameter is used: convergence seemed faster at first but about 300 estimates of genetic value (0.03%) did not stabilize and led to divergence first, of groups of unknown parents and then, of all animals' genetic values. In contrast, for a research run with a smaller data set and tracing pedigrees back for only 4 generations, the same standardized convergence criteria as in the example described here were obtained after nearly 3 times fewer iterations. This illustrates that the discrepancy between B and B^* increases with the number of generations considered. A modification of the algorithm in order to make B^* closer to B while keeping a similar sparsity structure should be investigated. For example, mate contributions to the inverse of the relationship matrix could be no longer ignored. However, this may not be necessary as it is shown in *Appendix 2* that there always exists a relaxation parameter ω such that convergence is guaranteed. Quite unfortunately, as for successive over-relaxation and second order Jacobi procedures, the optimal ω is data dependent.

The model presented in [6] and which led to the form of the incomplete Cholesky factor described in figure 2 is rather restrictive. In particular, it assumes the existence of only one fixed effect. The section *Dealing with several fixed effects* showed a way to treat models with more than one fixed effect. For more complex situations, a simple procedure would be to apply the incomplete Cholesky factorization only to the block $(\mathbf{Z'Z} + \alpha\mathbf{A}^*)$. The remaining part of the system can be solved using a more standard procedure. But then, convergence is likely to be somewhat slowed down. A trade-off is to increase the amount of information available to estimate fixed effects at each iteration by "pseudo-absorption" of genetic value equations, as described in Ducrocq *et al* (1990). Whatever the alternative chosen, it seems quite apparent that faster and more efficient algorithms from a computational viewpoint need to take larger advantage of the known sparsity structure of the mixed model equations than simpler, general-purpose algorithms. Of course, the final practical choice between competing algorithms should take into account other considerations such as programming costs and computer availability. Then, rather specialized algorithms such as the one proposed here may be attractive only for huge routine evaluations based on animal models.

APPENDIX 1

Practical algorithm sketch for the solution of the system $\mathbf{B}^*\mathbf{s} = \mathbf{r}$

In the following section, indices related to fixed and random effects (*eg*, $\mathbf{b}$ and $\mathbf{a}$ in $\mathbf{r_b}$ and $\mathbf{r_a}$) will be dropped for clarity. The indices j, j_s, j_d and h will refer to animal j, its sire (or sire's group), its dam (or dam's group) and the herd-year in which it was recorded. $\mathbf{n} = \{n_h\}$ will be the vector of number of records in each herd-year. Define $\mathbf{s} = \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix}$. Index g will refer to all groups of unknown parents.

Then, the whole algorithm for the solution de $\mathbf{B}^*\mathbf{s} = \mathbf{r}$, including steps [21] to [25], can be detailed as follows:

1) Initialize $\mathbf{r}, \mathbf{n}, \mathbf{s}$ to zero
2) Read the data file
   - For each j with record: y_j, add (ωy_j) to r_h and r_j
     add 1 to n_h
3) Compute $v_h = r_h/n_h$ for $h = 1, \ldots r_h = r_h - wn_h s_h$
4) Initialize v_j as $v_j = r_j$ for $j = 1, \ldots$
Read the data file with progeny preceding parents
For each j:
   - If j has a record: $v_j = (v_j - \omega s_h) - v_h$
   - $v_j = v_j/d_j$
   - $x_1 = 0.5\alpha\delta_j v_j$

     $v_{j_s} = v_{j_s} + x_1$

     $v_{j_d} = v_{j_d} + x_1$

5) Solve $\mathbf{T}_g\mathbf{D}_g\mathbf{T'}_g\mathbf{s}_g = \mathbf{v}_g$ for $\mathbf{s}_g$

6) Read the data file with parents preceding progeny
For each j:

- $x_2 = s_{j_s} + s_{j_d}$

  $s_j = v_j + (0.5\alpha\delta_j/d_j)x_2$

  $x_3 = \alpha\delta_j(s_j - 0.5x_2) = x_4$

- If j has a record $y_j, v_h = v_h - s_j/n_h, r_h = r_h - \omega(s_j - y_j)$

  $x_3 = x_3 + s_j - y_j$

- $r_j = r_j - \omega x_3$

  $r_{j_s} = r_{j_s} + 0.5\,\omega x_4$

  $r_{j_d} = r_{j_d} + 0.5\,\omega x_4$

7) $s_h = v_h$ for $h = 1,\dots$
8) Go to (3) until convergence.

Note that at each iteration, the computation of $s_h, h = 1,\dots$ is completed only after reaching the end of the second data file. Therefore, the update of $\mathbf{r}$ in [25] – which requires s_h – would make necessary a third reading of the data file. This can be avoided by transferring this part of the updating $(-\omega\mathbf{Z'Xb}$ in [25] to the beginning of the next iteration. This explains the term $(v_j - \omega s_h)$ in (4).

APPENDIX 2

Proof that the splitting method based on an incomplete Cholesky decomposition of the coefficient matrix always converges

The following section shows that there always exist a value of the relaxation parameter ω such that the general algorithm proposed in [5] converges for all starting values. For splitting methods, a necessary and sufficient condition for this is – with the notation used in [3] and [5] – that $\rho\left(\mathbf{I} - \omega\mathbf{B}^{*-1}\mathbf{B}\right) < 1$ where $\rho(\mathbf{M})$ is the spectral radius of $\mathbf{M}$, *ie*, the largest absolute value of the eigenvalues of $\mathbf{M}$ (Golub and Van Loan, 1983; Coleman, 1984). Indeed, if λ_i is an eigenvalue of $\mathbf{B}^{*-1}\mathbf{B}$, then λ_i is also an eigenvalue of $\mathbf{T}^{*-1}\mathbf{B}\mathbf{T}^{*-\mathbf{T}}$ where $\mathbf{B}^* = \mathbf{T}^*\mathbf{T}^{*'}$. Obviously, this matrix is nonnegative definite, so all its eigenvalues are nonnegative. Let λ_1 be the largest one. The eigenvalues of the matrix $\mathbf{I} - \omega\mathbf{B}^{*-1}\mathbf{B}$ are all of the form $(1 - \omega\lambda_i)$ and if one chooses $0 < \omega < \dfrac{1}{\lambda_1}$ then $\rho\left(\mathbf{I} - \omega\mathbf{B}^{*-1}\mathbf{B}\right) \leqslant 1$. Equality holds when $\mathbf{B}$ is singular, which is indeed the case when groups of unknown parents and another fixed effect are considered simultaneously. A constraint on solutions for groups (or the other fixed effects) can be added to avoid this problem. But is was found that convergence is faster when no constraint is included. As for Gauss–Seidel iteration, it seems that there is a built-in constraint in the iterative system preventing problems from occurring.

ACKNOWLEDGMENTS

REFERENCES

Arnason T (1982) Prediction of breeding values for multiple traits in small non random mating (horse) populations. *Acta Agric Scand* 32, 171–176

Arnason T (1986) Practical aspects in the estimation of breeding values for multi-trait selection. *World Rev Anim Prod* 22, 75–84

Coleman TE (1984) *Large Sparse Numerical Optimization. Lecture Notes in Computer Science, N° 165.* Springer-Verlag, New York, 105 p

Ducrocq V, Boichard D, Bonaiti B, Barbat A, Briend M (1990) A pseudo-absorption strategy for solving animal model equations for large data files. *J Dairy Sci* 73, 1945–1955

Golub GH, Van Loan CF (1983) *Matrix Computations.* Johns Hopkins Univ Press, Baltimore, MD, 476 p

Groeneveld E, Kovac M, Wang T (1990) PEST, a general purpose BLUP package for multivariate prediction and estimation. Proc 4th World Congr Genet Appl Livest Prod, Edinburgh, NE, July 23-27, 13, 488–491

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69-83

Misztal I, Gianola D (1987) Indirect solution of mixed model equations. *J Dairy Sci* 70, 716–723

Misztal I, Gianola D, Schaeffer LR (1987) Extrapolation and convergence criteria with Jacobi and Gauss–Seidel iteration in animal models. *J Dairy Sci* 70, 2577-2584

Poivey JP (1986) Méthode simplifiée de calcul des valeurs génétiques des femelles tenant compte de toutes les parentés. *Génét Sél Evol* 18, 321–331

Pollak EJ (1984) *Transformation for Systematically Missing Data to Facilitate Multiple Trait Evaluation.* Mimeograph, Cornell University

Quaas RL (1988) Additive genetic model with groups and relationships. *J Dairy Sci* 71, 1338–1345

Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 51, 1277–1287

Quaas RL, Pollak EJ (1981) Modified equations for sire models with groups. *J Dairy Sci* 64, 1868–1872

Robinson GK (1986) Group effects and computing strategies for models for estimating breeding values. *J Dairy Sci* 69, 3106–3111

Schaeffer LR, Kennedy BW (1986) Computing solutions to mixed model equations. Proc 3rd World Congr Genet Appl Livest Prod, Univ Nebraska, Lincoln, NE, July 16–22, 12, 382-393

Thompson R (1977) Estimation of quantitative genetic parameters. Proc Int Quant Genet, Ames, Iowa, 639–657

Westell RA (1984) Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. PhD diss, Cornell Univ, Ithaca, NY; Univ Microfilms Inc, Ann Arbor, MI

Wiggans GR, Misztal I, Van Vleck LD (1988) Animal model evaluation of Ayrshire milk yield with all lactations, herd–sire interaction and groups based on unknown parents. *J Dairy Sci* 71, 1319–1329