An Iterative Implementation of the Single Step Approach for Genomic Evaluation which Preserves Existing Genetic Evaluation Models and Software

V. Ducrocq[1] and A. Legarra[2]

[1]*INRA, UMR1313 Animal Genetics and Integrative Biology (GABI), 78352 Jouy en Josas, France;*
[2]*INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), F-31326 Castanet-Tolosan, France*
email : vincent.ducrocq@jouy.inra.fr

Abstract

The "single step approach" of Misztal et al (2009) and Christensen et al (2010) combines phenotypic, genomic and pedigree information into a single BLUP analysis. To adapt its implementation to most current genetic and genomic evaluation models, we propose an iterative solution procedure. Its main benefits are the fact that only moderate modifications of existing software are required and that it offers a framework for extension to more complex genetic or genomic models or their approximation.

Keywords: genomic evaluation, single step approach, iterative solution

Introduction

The current genomic evaluation models implemented worldwide usually involve multiple steps: a) classical genetic evaluations from which b) corrected phenotypes such as daughter yield deviations (DYD) or deregressed proofs (DP) are obtained for a genotyped reference subpopulation; c) computation of direct genomic values (DGV); d) blending of genomic and phenotypic information. These successive steps may lead to biases, for example if genetic and genomic evaluations are not expressed on consistent scales or even more importantly, when only partial information is used. Patry and Ducrocq (2011a) showed that genomic pre-selection of young bulls leads to biased genetic breeding values when candidates culled based on their DGV are not included, which in turn leads to biased DYD for genomic evaluations. Vitezica *et al.* (2011) showed that in presence of family-based selection, genotyped animals were not a random sample of the population, leading to bias as well. Another drawback of multiple step procedures is that non genotyped animals do not benefit directly from genomic information on relatives, except if another extra step is added (Patry and Ducrocq, 2011b).

To circumvent these limitations, Misztal *et al.* (2009; see also Legarra *et al.,* 2009; Aguilar *et al.,* 2010) and Christensen and Lund (2010) proposed a single step approach in which a joint evaluation of phenotypic and genomic information is performed simultaneously for all animals. The approach appears as an extension of the usual mixed model equations (and therefore as an extension of GBLUP). The implementation of the single step approach is not complicated in the simplest cases but may become much more challenging for advanced models (test day, multiple trait, threshold models…) requiring demanding software adaptation.

Our objective here is to develop an iterative procedure for the solution of the single step mixed model equations which requires only moderate software modifications and which can be adapted to a broad variety of genetic and genomic evaluation models.

Methods

The single step GBLUP model

Consider a single trait analyzed with the classical mixed model for genetic evaluations:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wu} + \mathbf{e} \qquad (1)$$

with the usual notations.

Let the subscripts 1 and 2 correspond to non genotyped and genotyped animals respectively,

$\alpha_u = \sigma_e^2 / \sigma_u^2$ and let $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ be the numerator relationship matrix. Assume $\text{var}(\mathbf{u}_2) = \mathbf{G}\,\sigma_u^2$, Note that $\mathbf{G}$ may be a genomic relationship matrix (VanRaden, 2009) but can have other forms too, if another genomic evaluation method is preferred; for instance, Zhang *et al.* (2010) and Legarra *et al.* (2011) suggested to (pre-)estimate $\mathbf{G}$ through, respectively, BayesB or the Bayesian Lasso, giving more weight to SNPs of large effect.

The mixed model equations for the single step approach are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'W'} \\ \mathbf{W'X} & \mathbf{W'W} + \alpha_u \mathbf{H}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{W'y} \end{pmatrix} \quad (2)$$

where $\mathbf{H} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}$

Another parameterization

In (1), one can decompose the additive genetic value $\mathbf{u}$ into a "strictly polygenic" part $\mathbf{u}^*$ and an independent deviation $\mathbf{d}$ due to genomic information, with $\mathbf{d}=\mathbf{u}-\mathbf{u}^*$, so:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{W}(\mathbf{u}^*+\mathbf{d}) + \mathbf{e} \quad (3)$$

The deviation $\mathbf{d}_1$ for non genotyped information is obtained by regression on genomic contribution $\mathbf{d}_2$ from genotyped individuals:

$$\mathbf{d}_1 = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{d}_2 \quad (4)$$

Therefore, model (1) can be written as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}\mathbf{b} + \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 \end{bmatrix}\begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{W}_2 \end{bmatrix}\mathbf{d}_2 + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$
$$(5)$$

If $\text{var}(\mathbf{u}_2) = \mathbf{G}\,\sigma_u^2$ and assuming that $\text{cov}(\mathbf{u}_2^*, \mathbf{d}_2) = \mathbf{0}$, we have:

$$\text{var}(\mathbf{u}_2) = \text{var}(\mathbf{u}_2^* + \mathbf{d}_2) = \mathbf{A}_{22}\sigma_u^2 + \text{var}(\mathbf{d}_2)$$

It follows that $\text{var}(\mathbf{d}_2) = (\mathbf{G} - \mathbf{A}_{22})\sigma_u^2$

The corresponding mixed model equations are of the form $\mathbf{M}^* \mathbf{x}^* = \mathbf{z}^*$, i.e.:

$$\mathbf{M}^* \begin{bmatrix} \mathbf{b} \\ \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 + \mathbf{X}_2'\mathbf{y}_2 \\ \mathbf{W}_1'\mathbf{y}_1 \\ \mathbf{W}_2'\mathbf{y}_2 \\ \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{W}_1'\mathbf{y}_1 + \mathbf{W}_2'\mathbf{y}_2 \end{bmatrix} \quad (6)$$

with $\mathbf{M}^*$ equal to:

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2 & \mathbf{X}_1'\mathbf{W}_1 & \mathbf{X}_2'\mathbf{W}_2 & \mathbf{X}_1'\mathbf{W}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1} + \mathbf{X}_2'\mathbf{W}_2 \\ \mathbf{W}_1'\mathbf{X}_1 & \mathbf{W}_1'\mathbf{W}_1 + \alpha_u\mathbf{A}^{11} & \alpha_u\mathbf{A}^{12} & \mathbf{W}_1'\mathbf{W}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{W}_2'\mathbf{X}_2 & \alpha_u\mathbf{A}^{21} & \mathbf{W}_2'\mathbf{W}_2 + \alpha_u\mathbf{A}^{22} & \mathbf{W}_2'\mathbf{W}_2 \\ \{\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{W}_1'\mathbf{X}_1 + \mathbf{W}_2'\mathbf{X}_2\} & \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{W}_1'\mathbf{W}_1 & \mathbf{W}_2'\mathbf{W}_2 & \{\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{W}_1'\mathbf{W}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1} + \mathbf{W}_2'\mathbf{W}_2 + \alpha_u(\mathbf{G} - \mathbf{A}_{22})^{-1}\} \end{bmatrix}$$
$$(7)$$

Define $\mathbf{S} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$. $\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$

System (6) is equivalent to:

$$\mathbf{S}^{-T}\mathbf{M}^*(\mathbf{S}^{-1}\mathbf{S})\,\mathbf{x}^* = \mathbf{S}^{-T}\mathbf{z}^* \quad (8)$$

The development of expression (8) leads to a system $\mathbf{M}\,\mathbf{x} = \mathbf{z}$ where:

$$\mathbf{x} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{d}_2 \end{bmatrix} \quad (9)$$

and:

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2 & \mathbf{W}_1'\mathbf{W}_1 & \mathbf{X}_2'\mathbf{W}_2 & \mathbf{0} \\ \mathbf{W}_1'\mathbf{X}_1 & \mathbf{W}_1'\mathbf{W}_1 + \alpha_u\mathbf{A}^{11} & \alpha_u\mathbf{A}^{12} & \mathbf{M}_{24} \\ \mathbf{W}_2'\mathbf{X}_2 & \alpha_u\mathbf{A}^{21} & \mathbf{W}_2'\mathbf{W}_2 + \alpha_u\mathbf{A}^{22} & \mathbf{M}_{34} \\ \mathbf{0} & \mathbf{M}_{42} & \mathbf{M}_{43} & \mathbf{M}_{44} \end{bmatrix}$$

with:

$$\mathbf{M}_{24} = \mathbf{M}_{42}^{'} = -\alpha_u (\mathbf{A}^{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} + \mathbf{A}^{12})$$

$$\mathbf{M}_{34} = \mathbf{M}_{43}^{'} = -\alpha_u (\mathbf{A}^{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} + \mathbf{A}^{22})$$

$$\mathbf{M}_{44} = \alpha_u (\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}^{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} + \mathbf{A}^{12})$$
$$+ \alpha_u (\mathbf{A}^{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} + \mathbf{A}^{22}) + \alpha_u (\mathbf{G} - \mathbf{A}_{22})^{-1}$$

Using the rules for the inverse of a partitioned matrix (Searle, 1982), we have:

$$\mathbf{M}_{24} = \mathbf{M}_{42}^{'} = \mathbf{0}, \ \mathbf{M}_{34} = \mathbf{M}_{43}^{'} = -\alpha_u \mathbf{A}_{22}^{-1} \text{ and}$$

$$\mathbf{M}_{44} = \alpha_u \left\{ \mathbf{A}_{22}^{-1} + (\mathbf{G} - \mathbf{A}_{22})^{-1} \right\}$$

Therefore, system (8) simplifies to:

$$\begin{bmatrix} \mathbf{X}_1^{'}\mathbf{X}_1 + \mathbf{X}_2^{'}\mathbf{X}_2 & \mathbf{X}_1^{'}\mathbf{W}_1 & \mathbf{X}_2^{'}\mathbf{W}_2 & \mathbf{0} \\ \mathbf{W}_1^{'}\mathbf{X}_1 & \mathbf{W}_1^{'}\mathbf{W}_1 + \alpha_u \mathbf{A}^{11} & \alpha_u \mathbf{A}^{12} & \mathbf{0} \\ \mathbf{W}_2^{'}\mathbf{X}_2 & \alpha_u \mathbf{A}^{21} & \mathbf{W}_2^{'}\mathbf{W}_2 + \alpha_u \mathbf{A}^{22} & -\alpha_u \mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} & -\alpha_u \mathbf{A}_{22}^{-1} & \alpha_u \left\{ \mathbf{A}_{22}^{-1} + (\mathbf{G} - \mathbf{A}_{22})^{-1} \right\} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{b} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^{'}\mathbf{y}_1 + \mathbf{X}_2^{'}\mathbf{y}_2 \\ \mathbf{W}_1^{'}\mathbf{y}_1 \\ \mathbf{W}_2^{'}\mathbf{y}_2 \\ \mathbf{0} \end{bmatrix} \quad (10)$$

Applying again the rules for the inverse of a partitioned matrix (Searle, 1982), it can be shown that the absorption of the last equation of system (10) into the first ones leads to system (2): models (1) and (3) are equivalent.

An iterative solution of the single step mixed model equations

Isolating the first three blocks of equations in (10) from the last one suggests the following iterative solution procedure:

Algorithm A:

A0) Assume $\hat{\mathbf{d}}_2 = \mathbf{0}$

A1) Solve for **b**, $\mathbf{u}_1$ and $\mathbf{u}_2$:

$$\begin{bmatrix} \mathbf{X}_1^{'}\mathbf{X}_1 + \mathbf{X}_2^{'}\mathbf{X}_2 & \mathbf{X}_1^{'}\mathbf{W}_1 & \mathbf{X}_2^{'}\mathbf{W}_2 \\ \mathbf{W}_1^{'}\mathbf{X}_1 & \mathbf{W}_1^{'}\mathbf{W}_1 + \alpha_u \mathbf{A}^{11} & \alpha_u \mathbf{A}^{12} \\ \mathbf{W}_2^{'}\mathbf{X}_2 & \alpha_u \mathbf{A}^{21} & \mathbf{W}_2^{'}\mathbf{W}_2 + \alpha_u \mathbf{A}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X}_1^{'}\mathbf{y}_1 + \mathbf{X}_2^{'}\mathbf{y}_2 \\ \mathbf{W}_1^{'}\mathbf{y}_1 \\ \mathbf{W}_2^{'}\mathbf{y}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \alpha_u \mathbf{A}_{22}^{-1} \hat{\mathbf{d}}_2 \end{bmatrix} \quad (11)$$

A2) Solve for $\mathbf{d}_2$:

$$\left[\mathbf{A}_{22}^{-1} + (\mathbf{G} - \mathbf{A}_{22})^{-1} \right] \mathbf{d}_2 = \mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2 \quad (12)$$

A3) Iterate A1) and A2) until convergence

At convergence, equation (12) is:

$$\left[\mathbf{A}_{22}^{-1} + (\mathbf{G} - \mathbf{A}_{22})^{-1} \right] \hat{\mathbf{d}}_2 = \mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2 \text{ or:}$$
$$(\mathbf{G} - \mathbf{A}_{22})^{-1} \hat{\mathbf{d}}_2 = \mathbf{A}_{22}^{-1} (\hat{\mathbf{u}}_2 - \hat{\mathbf{d}}_2) = \mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2^{*}$$

After manipulations, it follows that:

$$\hat{\mathbf{d}}_2 = \hat{\mathbf{u}}_2 - \mathbf{A}_{22} \mathbf{G}^{-1} \hat{\mathbf{u}}_2 \quad (13)$$

and therefore $\hat{\mathbf{u}}_2^{*} = \mathbf{A}_{22} \mathbf{G}^{-1} \hat{\mathbf{u}}_2$. $\quad (14)$

Now, if we plug expression (13) into the second term of the right hand side of (11), we have:

$$\alpha_u \mathbf{A}_{22}^{-1} \hat{\mathbf{d}}_2 = \alpha_u (\mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2 - \mathbf{G}^{-1} \hat{\mathbf{u}}_2) = \alpha_u (\hat{\mathbf{t}}_A - \hat{\mathbf{t}}_G)$$

The solution algorithm becomes (<u>Algorithm B</u>):

B0) Assume $\hat{\mathbf{d}}_2 = \mathbf{0}$

B1) Solve (11) for **b**, $\mathbf{u}_1$ and $\mathbf{u}_2$

B2) Compute $\mathbf{t}_A = \mathbf{A}_{22}^{-1} \hat{\mathbf{u}}_2$

or solve $\mathbf{A}_{22} \mathbf{t}_A = \hat{\mathbf{u}}_2$ for $\mathbf{t}_A$

B3) Compute $\mathbf{t}_G = \mathbf{G}^{-1} \hat{\mathbf{u}}_2$

or solve $\mathbf{G} \mathbf{t}_G = \hat{\mathbf{u}}_2$ for $\mathbf{t}_G$

B4) Iterate B1) and B2) until convergence

Note that in the iterative algorithm, the genomic information is included only at one place, for the computation of $\mathbf{t}_G$ in step B3. One can use here a **G** matrix corresponding to any genomic evaluation method, e.g., Bayes A to Z, Lasso, PLS, Elastic Net … or a BLUP on QTL as in the French genomic evaluation (Boichard *et al.,* 2010). All computations remain unchanged as long as $\mathbf{t}_A$ and $\mathbf{t}_G$ are consistent, i.e., as long as **G** is really equal to var(**u₂**).

Adaptation when the genomic evaluation is not GBLUP

So far, we assumed that $\hat{\mathbf{u}}_2$ in step B3 came from (11) but it can also be obtained using existing genomic evaluation software. After all, the solutions must be the same at convergence, if exactly the same information is used.

There are several reasons to prefer such strategy: first, we are not only interested in the GEBV ($\hat{\mathbf{u}}_2$) but also in the estimated (SNP or haplotype) markers effects. Second, there may be some phenotypes from genotyped animals that we would like to *exclude* from the genomic evaluation part: for example, in most European countries and Canada, own records from bull dams are excluded for the estimation of marker effects, because of fear of preferential treatment. Third and perhaps even more importantly, there may also be some extra phenotypes from genotyped reference animals that we would like to *include* in the genomic evaluation part only. A typical example is phenotypes (usually deregressed international EBV) from foreign bulls in multinational reference populations. Finally, genomic evaluations software are already available and our goal is to minimize changes compared to the existing situations.

The iterative nature of Algorithms A or B suggests to use (11) to compute at each iteration corrected phenotypes (DYD or DP) to be used as input data in the genomic evaluation software. For example, suppose that we are interested in genotyped males. For each one of them, one can correct their daughters' records for fixed effects and half their mates' EBV and absorb the corresponding equations as it is currently done. This leads to a system:

$$\left(\mathbf{W}_2'\Psi_2\mathbf{W}_2 + \alpha_u\mathbf{A}^{22}\right)\hat{\mathbf{u}}_2^\circ = \mathbf{W}_2'\Psi_2\,(\mathbf{D}\hat{\mathbf{Y}}\mathbf{D}_2) - \alpha_u\mathbf{A}_{22}^{-1}\hat{\mathbf{d}}_2$$

$$(15)$$

for which it is hoped that $\hat{\mathbf{u}}_2^\circ \approx \hat{\mathbf{u}}_2$, where Ψ_2 is a diagonal matrix of EDC (Equivalent daughter contribution) and $\mathbf{D}\hat{\mathbf{Y}}\mathbf{D}_2$ is a vector of DYD updated at each iteration for the current values of $\hat{\beta}$ and $\hat{\mathbf{u}}_1$.

Take as an example a BLUP evaluation on QTL haplotypes (Boichard et al, 2010), which can be written as:

$$y_i^* = \mu + a_i + \sum_j \left(h_{ij1} + h_{ij2}\right) + \varepsilon_i \qquad (16)$$

where y_i^* is a corrected phenotype (DYD or DP), h_{ij1} and h_{ij2} are haplotype effects 1 and 2 at QTL j for animal i and u_i is a residual polygenic effect. In matrix notation,

$$\mathbf{y}^* = \mu\mathbf{1} + \mathbf{a} + \mathbf{Nh} + \varepsilon$$

where $\mathbf{N}$ is the incidence matrix relating phenotypes to haplotypes. Let $\mathbf{F}$ be proportional to the (co)variance matrix of haplotype effects and θ the proportion of the total genetic variance attributed to QTL. Then the genomic evaluation software currently used can be run at each iteration of Algorithm B to get estimates of $\mathbf{a}$, $\mathbf{h}$ and the relevant elements of $\mathbf{u}_2$ as well as $\mathbf{t}_G$, using:

$$\mathbf{G} = \mathrm{Var}(\mathbf{a} + \mathbf{Nh}) = (1-\theta)\mathrm{Var}(\mathbf{a}) + \theta\mathrm{Var}(\mathbf{Nh})$$

$$= (1-\theta)\mathbf{A}_{22} + \theta\,\mathbf{NFN}' \qquad (17)$$

Discussion

The interest of an iterative approach is conceptual as well as computational: Splitting the evaluation into two separates the difficulty (and wealth) of national evaluations with their (trait dependent) complex models (huge data sets, multiple traits, heterogeneity of variances, unknown parent groups, threshold models, etc) from that of the genomic evaluation, with other problems (modeling, imputation of missing genotypes, computations…). It allows us to focus on one problem at a time.

Modifying a national evaluation software to include a correction of the right hand side should be relatively easy. Further, because these are two separate procedures, changing one does not imply to change the other. Moving, for instance, from GBLUP to BayesB or any other approach would imply changing just a part (B3) of the system. The fact that the method applies to any genomic evaluation system, not only to GBLUP, as far as $\mathbf{G}$ describes covariances among breeding values (as emphasized by Legarra *et al.,* 2009) must be underlined.

The computations as described are rather simple. Matrices $\mathbf{G}$ and $\mathbf{A}_{22}$ (and their inverses) can be stored in core for small numbers of genotyped animals. For large numbers, B2 and B3 can be solved by iteration on data, repeatedly computing the products $\mathbf{Gt}_G$ and $\mathbf{A}_{22}\mathbf{t}_A$. For example for GBLUP, the first product can be computed as $\mathbf{Z}\,(\,\mathbf{\Lambda}\,(\,\mathbf{Z'}\,\mathbf{t}_G\,)\,)$, where $\mathbf{Z}$ describes the genotypes of $\mathbf{u}_2$ and $\mathbf{\Lambda}$ is a diagonal matrix at a cost of $O(mp)$ operations if m is the number of genotyped animals and p the number of markers; the second product can be calculated using Colleau's algorithm, at a cost of $O(n)$ operations where n is the number of animals related to the genotyped ones. This algorithm computes the product $\mathbf{w} = \mathbf{At}$ as the solution to the system $\mathbf{A}^{-1}\mathbf{w} = \mathbf{T}^{-1}\mathbf{D}^{-1}\mathbf{T}^{-T}\mathbf{w} = \mathbf{t}$, i.e., solving twice an extremely sparse triangular system of equations by reading the pedigree file twice. In order to compute $\mathbf{A}_{22}\mathbf{t}_A$, some appropriate elements of $\mathbf{t}$ have to be set equal to 0. Misztal *et al.* (2009) and Aguilar *et al.* (2011) describe the algorithm and provide a Fortran code.

An unsolved problem is the estimation of reliabilities because for large applications, the inverses of $\mathbf{G}$ and $\mathbf{A}_{22}$ will not be available. But the decomposition (5) isolating the contribution of the genomic information independent from the rest suggests that an extension of the approach of Harris and Johnson (1998) could perhaps be conceived.

A pertinent question is: why should one consider a single step evaluation if current multi-step evaluations work? One reason is the problem of bias above mentioned. Bias will plague national evaluations if early selection based on genomic proofs becomes a rule. But the other reason is the elegance and power of a single, unified framework, such as BLUP was in relation to contemporary comparison or selection indexes.

Acknowledgments

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S. & Lawlor, T.J. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci. 93,* 743-752.

Aguilar, I., Misztal, I., Legarra, A. & Tsuruta, S. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet. 128,* 422-428.

Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M.-N., Boscher, M.-Y., Druet, T., Genestout, L., Eggen, A., Journaux, L., Ducrocq, V. & Fritz, S. 2010. Genomic Selection in French dairy Cattle. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production.* Leipzig, Germany, August 1-6, 2010, Communication 716

Christensen, O.F. & Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol. 42,* 2.

Colleau, J.J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol. 34,* 409–421.

Legarra, A., Aguilar, I. & Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci. 92,* 4656–4663.

Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F. & Fritz, S. 2011. Improved Lasso for genomic selection. *Genet. Res. 93,* 77–87.

Misztal, I., Legarra, A. & Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci. 92,* 4648-4655.

Patry, C. & Ducrocq, V. 2011a. Evidence of biases in genetic evaluations due to genomic pre-selection in dairy cattle. *J. Dairy Sci. 94,* 1011-1020.

Patry, C. & Ducrocq, V. 2011b. Accounting for genomic pre-selection in national genetic dairy cattle evaluations. *Genet. Sel. Evol. 43,* 30.

Searle, S.R. 1982. *Matrix algebra useful for statistics.* John Wiley & Sons., 438p

Van Raden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. 2009. Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92,* 16–24.

Vitezica, Z.G., Aguilar, I., Misztal, I. & Legarra, A. 2011. Bias in genomic predictions for populations under selection. *Genet. Res. 93,* 357-366.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D-J. & Zhang, Q. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *Plos ONE 5(9): e12648.*