An inversion-free method to compute genomic predictions using an animal model approach

Jørgen Ødegård¹, Per Madsen² & Theo Meuwissen³

¹AquaGen AS
²Aarhus University
³Norwegian University of Life Sciences
One-step genomic evaluation

- Combines the genomic and pedigree-based relationship matrices into a single unified matrix
  - Used in an equivalent animal model
- Phenotypic data from genotyped and ungenotyped animals can be utilized jointly
One-step genomic animal model
(Aguillar et al. 2010, Christensen & Lund 2010)

- Standard Henderson’s mixed model equations (HMME-1)
  - \[
  \begin{bmatrix}
  X'X & X'Z \\
  Z'X & Z'Z + H^{-1}\lambda
  \end{bmatrix}
  \begin{bmatrix}
  \hat{\beta} \\
  \hat{g}
  \end{bmatrix}
  =
  \begin{bmatrix}
  X'y \\
  Z'y
  \end{bmatrix}
  \]
  \[
  \lambda = \frac{\sigma^2_e}{\sigma^2_g}
  \]
  - \[
  H = \begin{bmatrix}
  H_{11} & H_{12} \\
  H_{21} & H_{22}
  \end{bmatrix}
  = \begin{bmatrix}
  G_\omega & G_\omega A_{11}^{-1} A_{12} \\
  A_{21} A_{11}^{-1} G_\omega & A_{21} A_{11}^{-1} G_\omega A_{11}^{-1} A_{12} + A_{22} - A_{21} A_{11}^{-1} A_{12}
  \end{bmatrix}
  \]
  - \(G_\omega\) = the genomic relationship matrix (or some function of it)
  - The inverse of \(H\) has a simple structure:
    - \[
    H^{-1} = \begin{bmatrix}
    H_{11}^{11} & H_{12}^{12} \\
    H_{21}^{11} & H_{22}^{12}
    \end{bmatrix}
    = \begin{bmatrix}
    G_\omega^{-1} - A_{11}^{-1} & 0 \\
    0 & 0
    \end{bmatrix}
    + A^{-1} = \begin{bmatrix}
    G_\omega^{-1} + A^0 & A^{12} \\
    A^{21} & A^{22}
    \end{bmatrix}
    \]
    - \(A^0 = A^{11} - A_{11}^{-1}\)
    - Equations of ungenotyped animals are identical to the pedigree-based equations
Limitations of HMME-1

- Requires $G_\omega^{-1}$
  - Cannot be set up directly
  - May not be of full rank
  - Inversion of big-sized matrices are computationally demanding
    May eventually be impossible if dimension >100,000

- Requires $A^0 = A^{11} - A_{11}^{-1}$
  - $A_{11}^{-1}$ is hard to compute unless all ancestors are included among the genotyped animals
  - $A^0$ can be set up directly if ancestors are included among the genotyped animals
Weighted least squares (WLS) method

- The WLS equations:
  - \( \hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \)
  - \( \hat{g} = \sigma_g^2 HZ'V^{-1}(y - X\hat{\beta}) = \sigma_g^2 HZ'\hat{s} \)
    - \( V = HZH'\sigma_g^2 + I\sigma_e^2 \)
    - \( \hat{s} = V^{-1}(y - X\hat{\beta}) \), i.e. \( X\hat{\beta} + Vs = y \)

\[
\begin{bmatrix}
X'X & X'ZH'\sigma_g^2 \\
X & ZHZ'\sigma_g^2 + I\sigma_e^2
\end{bmatrix}
\begin{bmatrix}
\\
\hat{\beta} \\
\hat{s}
\end{bmatrix} =
\begin{bmatrix}
X'y \\
y
\end{bmatrix}
\]
Alternative Henderson’s mixed model equations (HMME-2) (Henderson, 1984)

\[
\begin{bmatrix}
X'X & X'Z \\
HZ'X & HZ'Z + I \lambda
\end{bmatrix}
\begin{bmatrix}
\hat{\beta} \\
\hat{g}
\end{bmatrix}
= 
\begin{bmatrix}
X'y \\
HZ'y
\end{bmatrix}
\]

• The method does not require the inverse relationship matrix
• The entire $H$ must be set up
• Method implemented in the DMU software using both direct and iterative solvers
Hybrid MME (HMME-3)

- The genetic effects are split into genotyped and ungenotyped animals \( \hat{\mathbf{g}} = \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} \)
  - Equations for genotyped animals are similar to HMME-2
  - Equations for ungenotyped animals are as in HMME-1

\[
\begin{bmatrix}
X'X & X'Z_1 \\
G_\omega X_1'X & G_\omega Z_1'Z_1 + (G_\omega A^0 + I)\lambda \\
Z_2'X & A^{21}\lambda \\
\end{bmatrix}
\begin{bmatrix}
X'Z_2 \\
G_\omega A_{12}\lambda \\
Z_2Z_2 + A^{22}\lambda \\
\end{bmatrix}
\begin{bmatrix}
\hat{\beta} \\
\hat{\mathbf{g}}_1 \\
\hat{\mathbf{g}}_2 \\
\end{bmatrix}
= \begin{bmatrix}
X'y \\
G_\omega Z_1'y_1 \\
Z_2'y_2 \\
\end{bmatrix}
\]

- The entire \( \mathbf{H} \) is not needed
Limitations of the one-step approach

- All approaches need $A_{11}^{-1}$ (or $A^0$)
- Can be set up directly if $H_{11}$ is expanded with relationships of ungenotyped ancestors

$$H_{11} = \begin{bmatrix} H_{00} & H_{01} \\ H_{10} & G_\omega \end{bmatrix}$$
How to expand $H_{11}$?

- Imputation of missing genotypes
  - Not always possible

$$H_{11} = \begin{bmatrix} G_{\omega(00)} & G_{\omega(01)} \\ G_{\omega(10)} & G_{\omega(11)} \end{bmatrix}$$
How to expand $H_{11}$?

- Set up $H_{11}$ through linkage analysis
  - Pedigree information will be used to incorporate ungenotyped ancestors of genotyped animals
  - Genomic IBD matrix

$$H_{11} = \begin{bmatrix} G_{LA(00)} & G_{LA(01)} \\ G_{LA(10)} & G_{LA(11)} \end{bmatrix}$$
How to expand $H_{11}$?

- Fill in ungenotyped ancestors using pedigree relationships
  - Simple, but not necessarily the best

$$H_{11} = \begin{bmatrix} A_{00} & A_{01} \\ A_{10} & G_\omega \end{bmatrix}$$
Is the one-step approach appropriate?

- Parents genotyped and offspring ungenotyped
  - Infers the most likely relationship structure among ungenotyped offspring
  - Appropriate
- Parents ungenotyped and offspring genotyped
  - The one-step will modify founder relationships such that they explain the genomic relationships
- Co-segregation of loci among relatives will cause deviations from pedigree relationships
  - Occur spontaneously
- The one-step approach explains effects of Mendelian sampling and linkage by relationships between distant founders
  - Unjustified and will likely give bias
Example: Inappropriate one-step analysis

- Assume the following pedigree of two paternal half-sibs
Example: Inappropriate one-step analysis

A =
\[
\begin{bmatrix}
1 & 0 & 0 & 0.5 & 0.5 \\
0 & 1 & 0 & 0.5 & 0 \\
0 & 0 & 1 & 0 & 0.5 \\
0.5 & 0.5 & 0 & 1 & 0.25 \\
0.5 & 0 & 0.5 & 0.25 & 1
\end{bmatrix}
\]

The paternal half-sibs (4&5) are genotyped:

\[ G_\omega = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix} \]

The slightly increased relationship can be attributed to co-segregation of linked loci. However, the one-step relationship matrix modifies the ancestral relationships (matrix sorted from oldest to youngest)

H =
\[
\begin{bmatrix}
1.016 & 0.008 & 0.008 & 0.520 & 0.520 \\
0.008 & 0.993 & 0.015 & 0.493 & 0.027 \\
0.008 & 0.015 & 0.993 & 0.027 & 0.493 \\
0.520 & 0.493 & 0.027 & 1.000 & 0.300 \\
0.520 & 0.027 & 0.493 & 0.300 & 1.000
\end{bmatrix}
\]
Implications

- Solving large-scale MME without inverting $G_\omega$ is possible
- Expanding $G_\omega$ with ungenotyped ancestors may be needed
  - For solving large-scale MME (using HMME-3)
  - For obtaining proper predictions of genetic effects
- Linkage analysis may be used to construct $G_\omega$ including ungenotyped ancestors