Solving Single-step GBLUP MME by Extended KKT Equations and Block Anti-Triangular Factorization

M. Taskinen, E.A. Mäntysaari, I. Strandén
Natural Resources Institute Finland (Luke), Myllytie 1, FI-31600 Jokioinen, Finland

Introduction

- Single-step GBLUP mixed model equations
- Expressed as alternative extended KKT matrix equation
- Separates very sparse “animal model” and full genomic information parts
- Sparse and full parts solved with different numerical approaches

Single-step GBLUP Mixed Model Equations

- Henderson’s mixed model equations (MME) of fixed \( \hat{b} \) and random effects \( \hat{u} \):
  \[
  \begin{pmatrix}
  \hat{b} \\
  \hat{u}
  \end{pmatrix} = \begin{pmatrix}
  X'X & X'Z \\
  Z'X & Z'Z + H^{-1}
  \end{pmatrix}^{-1}
  \begin{pmatrix}
  X'y \\
  Z'y
  \end{pmatrix}
  \]
- Involves inverses of relationship and residual variance matrices \( H^{-1} \) and \( R^{-1} \)
- Some individuals have genomic relationship information \( G_g = Z_g Z_{gg} \), all are related by pedigree relationship \( A \):
  \[
  H^{-1} = A^{-1} + \begin{pmatrix}
  0 & 0 \\
  0 & G_g^{-1} - (A_{gg})^{-1}
  \end{pmatrix}
  \]
- Adjusted genomic relationship matrix:
  \[
  G_w = wA_{gg} + (1 - w)G_g
  \]
  Here, \( A^{-1} \) is sparse whereas \( G_g \) is usually a full matrix

Karush-Kuhn-Tucker (KKT) Equation

- Alternative Karush-Kuhn-Tucker (KKT) matrix equation:
  \[
  \begin{pmatrix}
  \hat{b} \\
  \hat{u}
  \end{pmatrix} = \begin{pmatrix}
  I & 0 \\
  0 & HZ + R
  \end{pmatrix}
  \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix}, \text{ where } \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix} = \begin{pmatrix}
  X'X^{-1} & 0 \\
  X'V & y
  \end{pmatrix}
  \]
  and \( V = ZH + R \) needs to be invertible
- Symmetric anti-triangular, i.e. naturally indefinite with both positive and negative eigenvalues
- Not ideal for iterative solution methods and not widely used

Extended Single-step KKT MME

- Relationship matrix \( H \) expressed as:
  \[
  H = (L')^{-1}GL^{-1}, \text{ where } L = \begin{pmatrix}
  L_1 \\
  L_2
  \end{pmatrix} \text{ and } A^{-1} = LL'
  \]
- Genomic information separated to:
  \[
  G = I + (1 - w)P_{ee}^T(Z_g Z_g' + L_1 L_2 - I)P_{ee}^T
  \]
  where \( I \) is identity matrix, and \( P_{ee} \) is orthogonal projection
- Extended Single-step KKT MME with Mendelian sampling \( \phi \) and original random effects \( \hat{u} \):
  \[
  \begin{pmatrix}
  \hat{b} \\
  \hat{u} \end{pmatrix} = \begin{pmatrix}
  0 & 0 & X'X^{-1} & 0 \\
  0 & L' & Z' & 0
  \end{pmatrix}
  \begin{pmatrix}
  \hat{b} \\
  \hat{u} \\
  \hat{e} \\
  \hat{e}
  \end{pmatrix}
  \]
- Sparse \( X, Z, \) and \( L \) matrices are separated from the full genomic information parts in \( G \)

Block Anti-Triangular Factorization

- KKT matrix equations solved using block anti-triangular (BAT) factorization
- Sparsity preserving QR decomposition of sparse parts:
  \[
  K = \begin{pmatrix}
  0 & L' \\
  X & Z
  \end{pmatrix} = QR = [Q_2 Q_1] \begin{pmatrix}
  0 \\
  R_1
  \end{pmatrix}
  \]
  where \( Q \) is orthogonal and \( R_1 \) sparse upper triangular
- Extended KKT matrix written as:
  \[
  \begin{pmatrix}
  0 & 0 & K'T & 0 \\
  0 & Q & 0 & TQ
  \end{pmatrix}
  \begin{pmatrix}
  R' \\
  0 & Q_1 & 0
  \end{pmatrix}
  \]
  where \( T = \begin{pmatrix}
  G & 0 \\
  0 & R
  \end{pmatrix} \)
- Reparametrized anti-triangular matrix equation solved similarly to forward and backward substitutions, starting from upper right corner:
  \[
  R_1 b + \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix} = Q_2 y, \text{ where } \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix} = \begin{pmatrix}
  Q_1 \phi \\
  Q_2 \phi
  \end{pmatrix}
  \]
- Next, \( f_2 \) solved using iterative solution method from:
  \[
  (Q_2 TQ_2)^{-1} f_2 = Q_2 y
  \]
  where no inversion of full genomic parts is needed
- Finally, original effects solved from:
  \[
  R_1 b + \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix} = Q_1 G y - \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix}, \text{ where } \begin{pmatrix}
  \hat{b} \\
  \hat{e}
  \end{pmatrix} = Qf = Q_2 f_2
  \]

Results

- Nordic Red dairy cattle data set
- Single trait: deregressed proofs of milk yield
- One fixed effect
- In numbers:
  \[
  \begin{array}{lccccc}
  \text{Non-genotyped} & 70694 & \text{Genotyped} & 2885 & \text{Observations} & 67648 \\
  \text{Markers} & 37526 & \text{Memory usage:} & \begin{array}{cccc}
  \text{Sparse QR} & 18 MB \\
  Q & 560 MB
  \end{array}
  \end{array}
  \]
  Number of iterations (conjugate gradient, tol 10^-6):
  \[
  \begin{array}{cccccccc}
  \text{polygenic proportion } w & 0.01 & 0.10 & 0.20 & 0.30 \\
  0.5 & 24 & 22 & 20 & 17 \\
  0.2 & 24 & 22 & 19 & 18 \\
  0.1 & 28 & 27 & 26 & 25
  \end{array}
  \]

Conclusions

- Single-step GBLUP was expressed as extended KKT matrix equation
- Sparse “animal model” and full genomic parts separated
- Solving: block anti-triangular factorization, sparse QR decomposition, and iterative method for full part
- Low iteration numbers
- Applicability with large data sets depends on performance of sparse QR decomposition