

# 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  $\hat{\mathbf{b}}$  and random effects  $\hat{\mathbf{u}}$ :

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- ▶ Involves inverses of relationship and residual variance matrices  $\mathbf{H}^{-1}$  and  $\mathbf{R}^{-1}$
- ▶ Some individuals have **genomic** relationship information  $\mathbf{G}_g = \mathbf{Z}_m\mathbf{Z}'_m$ , all are related by **pedigree** relationship  $\mathbf{A}$ :

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} - (\mathbf{A}_{22})^{-1} \end{bmatrix}$$

- ▶ Adjusted genomic relationship matrix:

$$\mathbf{G}_w = w\mathbf{A}_{22} + (1 - w)\mathbf{G}_g$$

Here,  $\mathbf{A}^{-1}$  is **sparse** whereas  $\mathbf{G}_g$  is usually a **full** matrix

## Karush-Kuhn-Tucker (KKT) Matrix Equation

- ▶ Alternative **Karush-Kuhn-Tucker** (KKT) matrix equation:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}\mathbf{Z}' \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{e}} \end{bmatrix}, \text{ where } \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{X}' \\ \mathbf{X} & \mathbf{V} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}$$

and  $\mathbf{V} = \mathbf{Z}\mathbf{H}\mathbf{Z}' + \mathbf{R}$  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  $\mathbf{H}$  expressed as:

$$\mathbf{H} = (\mathbf{L}')^{-1}\tilde{\mathbf{G}}\mathbf{L}^{-1}, \text{ where } \mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \mathbf{L}\mathbf{L}'$$

- ▶ Genomic information separated to:

$$\tilde{\mathbf{G}} = \mathbf{I} + (1 - w)\mathbf{P}_{\mathbf{L}_1}^\perp (\mathbf{L}'_2\mathbf{Z}_m\mathbf{Z}'_m\mathbf{L}_2 - \mathbf{I})\mathbf{P}_{\mathbf{L}_1}^\perp$$

where  $\mathbf{I}$  is identity matrix, and  $\mathbf{P}_{\mathbf{L}_1}^\perp$  **orthogonal projection**

- ▶ **Extended Single-step KKT MME** with **Mendelian sampling**  $\hat{\boldsymbol{\phi}}$  and original random effects  $\hat{\mathbf{u}}$ :

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\boldsymbol{\phi}} \\ \hat{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}' \\ \mathbf{0} & \mathbf{0} & \mathbf{L} & \mathbf{Z}' \\ \mathbf{0} & \mathbf{L}' & \tilde{\mathbf{G}} & \mathbf{0} \\ \mathbf{X} & \mathbf{Z} & \mathbf{0} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{y} \end{bmatrix}$$

- ▶ Sparse  $\mathbf{X}$ ,  $\mathbf{Z}$ , and  $\mathbf{L}$  matrices are separated from the **full** genomic information parts in  $\tilde{\mathbf{G}}$

## Block Anti-Triangular Factorization

- ▶ KKT matrix equations solved using **block anti-triangular (BAT) factorization**

- ▶ **Sparsity preserving QR** decomposition of sparse parts:

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{L}' \\ \mathbf{X} & \mathbf{Z} \end{bmatrix} = \mathbf{Q}\mathbf{R}_0 = [\mathbf{Q}_2 \ \mathbf{Q}_1] \begin{bmatrix} \mathbf{0} \\ \mathbf{R}_1 \end{bmatrix}$$

where  $\mathbf{Q}$  is **orthogonal** and  $\mathbf{R}_1$  sparse upper triangular

- ▶ Extended KKT matrix written as:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}' \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{K}' \\ \mathbf{K} & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{R}'_1 \\ \mathbf{0} & \mathbf{Q}'_2\mathbf{T}\mathbf{Q}_2 & \mathbf{Q}'_2\mathbf{T}\mathbf{Q}_1 \\ \mathbf{R}_1 & \mathbf{Q}'_1\mathbf{T}\mathbf{Q}_2 & \mathbf{Q}'_1\mathbf{T}\mathbf{Q}_1 \end{bmatrix}, \text{ where } \mathbf{T} = \begin{bmatrix} \tilde{\mathbf{G}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

- ▶ Reparametrized anti-triangular matrix equation **solved similarly to forward and backward substitutions**, starting from upper right corner:

$$\mathbf{R}'_1\mathbf{f}_1 = \mathbf{0} \Leftrightarrow \mathbf{f}_1 = \mathbf{0}, \text{ where } \mathbf{f} = \begin{bmatrix} \mathbf{f}_2 \\ \mathbf{f}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}'_2 \\ \mathbf{Q}'_1 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\mathbf{e}} \end{bmatrix} = \mathbf{Q}' \begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\mathbf{e}} \end{bmatrix}$$

- ▶ Next,  $\mathbf{f}_2$  solved using **iterative solution method** from:

$$(\mathbf{Q}'_2\mathbf{T}\mathbf{Q}_2)\mathbf{f}_2 = \mathbf{Q}'_2 \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}$$

where no inversion of full genomic parts is needed

- ▶ Finally, original effects solved from:

$$\mathbf{R}_1 \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \mathbf{Q}'_1 \begin{bmatrix} -\tilde{\mathbf{G}}\hat{\boldsymbol{\phi}} \\ \mathbf{y} - \mathbf{R}\hat{\mathbf{e}} \end{bmatrix}, \text{ where } \begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\mathbf{e}} \end{bmatrix} = \mathbf{Q}\mathbf{f} = \mathbf{Q}_2\mathbf{f}_2$$

## Results

- ▶ Nordic Red dairy cattle data set
- ▶ Single trait: deregressed proofs of milk yield
- ▶ One fixed effect
- ▶ In numbers:

Non-genotyped	70694	Memory usage:	
Genotyped	2885	Sparse QR:	
Observations	67648	$\mathbf{R}_1$	18 MB
Markers	37526	$\mathbf{Q}$	560 MB

- ▶ **Number of iterations** (conjugate gradient, tol  $10^{-6}$ ):

	polygenic proportion $w$			
heritability $h^2$	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

## 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