

Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects

Matti Taskinen, Esa Mäntysaari, Ismo Strandén

Natural Resources Institute Finland (Luke)

31.08.2016 EAAP 2016 Belfast UK

Motivation

- Single-step genomic BLUP combines **pedigree** and **genomic marker** relationship information
- In mixed model equations (MME) this requires **inverse of combined relationship matrix** that has a **dense matrix** block for genotyped animals
- Millions of genotyped in near future \Rightarrow **numerical challenge**
- **Inversion can be avoided** with SNP-BLUP formulations having **marker effects** instead (or together with) animals effects
- Goal: numerically most efficient Single-step SNP-BLUP

Linearly equivalent MMEs by splitting variance matrix \mathbf{G}

- Fixed and random effects $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be solved from **separate equations** but usually better to use Henderson's **mixed model equations** (MME):

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Leftrightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- If variance matrix \mathbf{G} is not invertible, it could have "decomposition" $\mathbf{G} = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'$
- Part of \mathbf{G} can be "attached" to new model matrix $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{M}$ and new effects $\tilde{\mathbf{u}}$ defined:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \underbrace{\mathbf{Z}\mathbf{M}}_{\tilde{\mathbf{Z}}}\underbrace{\mathbf{G}\mathbf{M}'}_{\tilde{\mathbf{G}}}\underbrace{\mathbf{Z}'}_{\tilde{\mathbf{Z}}'} + \mathbf{R} = \tilde{\mathbf{Z}}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}' + \mathbf{R} =: \tilde{\mathbf{V}}$$

$$\tilde{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) =: \mathbf{M}\tilde{\mathbf{u}}$$

- Original effects solved using linearly equivalent MME of new effects $\tilde{\mathbf{u}}$ and variances $\tilde{\mathbf{G}}$:

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} = \tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Rightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Linearly equivalent MMEs by splitting variance matrix \mathbf{G}

- Fixed and random effects $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be solved from **separate equations** but usually better to use Henderson's **mixed model equations** (MME):

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Leftrightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- If variance matrix \mathbf{G} is not invertible, it could have "decomposition" $\mathbf{G} = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'$
- Part of \mathbf{G} can be "attached" to new model matrix $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{M}$ and new effects $\tilde{\mathbf{u}}$ defined:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \underbrace{\mathbf{Z}\mathbf{M}}_{\tilde{\mathbf{Z}}}\underbrace{\mathbf{G}\mathbf{M}'}_{\tilde{\mathbf{G}}}\underbrace{\mathbf{Z}'}_{\tilde{\mathbf{Z}}'} + \mathbf{R} = \tilde{\mathbf{Z}}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}' + \mathbf{R} =: \tilde{\mathbf{V}}$$

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) =: \mathbf{M}\tilde{\mathbf{u}}$$

- Original effects solved using linearly equivalent MME of new effects $\tilde{\mathbf{u}}$ and variances $\tilde{\mathbf{G}}$:

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} = \tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Rightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Linearly equivalent MMEs by splitting variance matrix \mathbf{G}

- Fixed and random effects $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be solved from **separate equations** but usually better to use Henderson's **mixed model equations** (MME):

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Leftrightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- If variance matrix \mathbf{G} is not invertible, it could have “decomposition” **$\mathbf{G} = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'$**

- Part of \mathbf{G} can be “attached” to new model matrix $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{M}$ and new effects $\tilde{\mathbf{u}}$ defined:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \underbrace{\mathbf{Z}\mathbf{M}}_{\tilde{\mathbf{Z}}}\underbrace{\mathbf{G}\mathbf{M}\mathbf{Z}'}_{\tilde{\mathbf{G}}}\underbrace{\mathbf{Z}'}_{\tilde{\mathbf{Z}}'} + \mathbf{R} = \tilde{\mathbf{Z}}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}' + \mathbf{R} =: \tilde{\mathbf{V}}$$

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{M}}'\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) =: \mathbf{M}\tilde{\mathbf{u}}$$

- Original effects solved using linearly equivalent MME of new effects $\tilde{\mathbf{u}}$ and variances $\tilde{\mathbf{G}}$:

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} = \tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Rightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Linearly equivalent MMEs by splitting variance matrix \mathbf{G}

- Fixed and random effects $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be solved from **separate equations** but usually better to use Henderson's **mixed model equations** (MME):

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Leftrightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- If variance matrix \mathbf{G} is not invertible, it could have “decomposition” $\mathbf{G} = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'$
- Part of \mathbf{G} can be “attached” to new model matrix $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{M}$ and new effects $\tilde{\mathbf{u}}$ defined:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \underbrace{\mathbf{Z}\mathbf{M}}_{\tilde{\mathbf{Z}}} \underbrace{\tilde{\mathbf{G}}\mathbf{M}'\mathbf{Z}'}_{\tilde{\mathbf{Z}}'} + \mathbf{R} = \tilde{\mathbf{Z}}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}' + \mathbf{R} =: \tilde{\mathbf{V}}$$

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\underbrace{\mathbf{M}'\mathbf{Z}'\mathbf{V}^{-1}}_{\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) =: \mathbf{M}\tilde{\mathbf{u}}$$

- Original effects solved using linearly equivalent MME of new effects $\tilde{\mathbf{u}}$ and variances $\tilde{\mathbf{G}}$:

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} = \tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Rightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\mathbf{X} & \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\tilde{\mathbf{Z}} + \tilde{\mathbf{G}}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Linearly equivalent MMEs by splitting variance matrix \mathbf{G}

- Fixed and random effects $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ can be solved from **separate equations** but usually better to use Henderson's **mixed model equations** (MME):

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Leftrightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- If variance matrix \mathbf{G} is not invertible, it could have “decomposition” $\mathbf{G} = \mathbf{M}\tilde{\mathbf{G}}\mathbf{M}'$
- Part of \mathbf{G} can be “attached” to new model matrix $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{M}$ and new effects $\tilde{\mathbf{u}}$ defined:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} = \underbrace{\mathbf{Z}\mathbf{M}}_{\tilde{\mathbf{Z}}} \underbrace{\tilde{\mathbf{G}}\mathbf{M}'\mathbf{Z}'}_{\tilde{\mathbf{Z}}'} + \mathbf{R} = \tilde{\mathbf{Z}}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}' + \mathbf{R} =: \tilde{\mathbf{V}}$$

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\underbrace{\mathbf{M}'\mathbf{Z}'}_{\tilde{\mathbf{Z}}'}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{M}\tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) =: \mathbf{M}\tilde{\mathbf{u}}$$

- Original effects solved using linearly equivalent MME of new effects $\tilde{\mathbf{u}}$ and variances $\tilde{\mathbf{G}}$:

$$\begin{cases} \hat{\mathbf{b}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \\ \tilde{\mathbf{u}} = \tilde{\mathbf{G}}\tilde{\mathbf{Z}}'\tilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{cases} \Rightarrow \begin{bmatrix} \hat{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\mathbf{X} & \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\tilde{\mathbf{Z}} + \tilde{\mathbf{G}}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \tilde{\mathbf{Z}}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Single-step G-BLUP relationship matrix **H**

- Inverse of **single-step G-BLUP** (ssGBLUP) relationship matrix is

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} - (\mathbf{A}_{22})^{-1} \end{bmatrix}, \text{ where } \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$$

- Genomic relationship matrix \mathbf{G}_g “adjusted” with \mathbf{A}_{22} :

$$\mathbf{G}_w = w\mathbf{A}_{22} + (1-w)\mathbf{G}_g, \text{ where } \mathbf{G}_g = \mathbf{Z}_m\mathbf{Z}_m'$$

- \mathbf{Z}_m is centered and scaled **marker matrix** and w is **polygenic proportion** between 0 and 1

- The inverse matrix \mathbf{H}^{-1} can be expressed using block LDL' decomposition

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & (\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

- Similarly, relationship matrix \mathbf{H} has decomposition

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_1 & -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ -\mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix}$$

Single-step G-BLUP relationship matrix **H**

- Inverse of **single-step G-BLUP** (ssGBLUP) relationship matrix is

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} - (\mathbf{A}_{22})^{-1} \end{bmatrix}, \text{ where } \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$$

- Genomic relationship matrix \mathbf{G}_g “adjusted” with \mathbf{A}_{22} :

$$\mathbf{G}_w = w\mathbf{A}_{22} + (1-w)\mathbf{G}_g, \text{ where } \mathbf{G}_g = \mathbf{Z}_m\mathbf{Z}_m'$$

- \mathbf{Z}_m is centered and scaled **marker matrix** and w is **polygenic proportion** between 0 and 1
- The inverse matrix \mathbf{H}^{-1} can be expressed using block LDL' decomposition

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & (\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

- Similarly, relationship matrix \mathbf{H} has decomposition

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_1 & -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ -\mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix}$$

Single-step G-BLUP relationship matrix **H**

- Inverse of **single-step G-BLUP** (ssGBLUP) relationship matrix is

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} - (\mathbf{A}_{22})^{-1} \end{bmatrix}, \text{ where } \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$$

- Genomic relationship matrix \mathbf{G}_g “adjusted” with \mathbf{A}_{22} :

$$\mathbf{G}_w = w\mathbf{A}_{22} + (1-w)\mathbf{G}_g, \text{ where } \mathbf{G}_g = \mathbf{Z}_m\mathbf{Z}'_m$$

- \mathbf{Z}_m is centered and scaled **marker matrix** and w is **polygenic proportion** between 0 and 1
- The inverse matrix \mathbf{H}^{-1} can be expressed using block LDL' decomposition

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & (\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

- Similarly, relationship matrix \mathbf{H} has decomposition

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_1 & -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ -\mathbf{A}^{21}(\mathbf{A}^{11})^{-1} & \mathbf{I}_2 \end{bmatrix}$$

Linearly equivalent Single-step G-BLUP MME

- Single-step relationship matrix \mathbf{H} can be expressed as

$$\mathbf{H} = \mathbf{M}_1 \tilde{\mathbf{G}}_1 \mathbf{M}'_1, \text{ where } \mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \text{ and } \tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \text{ or}$$

$$\mathbf{H} = \mathbf{M}_2 \tilde{\mathbf{G}}_2 \mathbf{M}'_2, \mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} & \mathbf{Z}_{imp} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{Z}_m \end{bmatrix}, \tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & w\mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (1-w)\mathbf{I}_m \end{bmatrix}$$

where $\mathbf{A}_{imp} = -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$ is **imputation operator** and $\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$ **imputed genotypes**

- New random effects $\tilde{\mathbf{u}}$ in $\tilde{\mathbf{G}}_1$: for non-genotyped $((\mathbf{A}^{11})^{-1})$ and for all animals (\mathbf{G}_w)
- $\tilde{\mathbf{G}}_2$: **residual polygenic** (RPG) effects for genotyped (\mathbf{A}_{22}), **orthogonal** marker effects (\mathbf{I}_m)
 \Rightarrow equivalent single-step SNP-BLUP (ssSNP-BLUP) ($w=0$ same as Fernando et al. (FDG))
- Original effects can be solved from linearly equivalent ssGBLUP MMEs (single trait case):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{X} & \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i + \sigma_u^{-2}\tilde{\mathbf{G}}_i^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \text{ where } \tilde{\mathbf{z}}_i = \mathbf{Z}\mathbf{M}_i$$

Linearly equivalent Single-step G-BLUP MME

- Single-step relationship matrix \mathbf{H} can be expressed as

$$\mathbf{H} = \mathbf{M}_1 \tilde{\mathbf{G}}_1 \mathbf{M}'_1, \text{ where } \mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \text{ and } \tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \text{ or}$$

$$\mathbf{H} = \mathbf{M}_2 \tilde{\mathbf{G}}_2 \mathbf{M}'_2, \mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} & \mathbf{Z}_{imp} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{Z}_m \end{bmatrix}, \tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & w\mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (1-w)\mathbf{I}_m \end{bmatrix}$$

where $\mathbf{A}_{imp} = -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$ is **imputation operator** and $\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$ **imputed genotypes**

- New random effects $\tilde{\mathbf{u}}$ in $\tilde{\mathbf{G}}_1$: for non-genotyped $((\mathbf{A}^{11})^{-1})$ and for all animals (\mathbf{G}_w)
- $\tilde{\mathbf{G}}_2$: **residual polygenic** (RPG) effects for genotyped (\mathbf{A}_{22}), **orthogonal** marker effects (\mathbf{I}_m)
 \Rightarrow equivalent single-step SNP-BLUP (ssSNP-BLUP) ($w=0$ same as Fernando et al. (FDG))
- Original effects can be solved from linearly equivalent ssGBLUP MMEs (single trait case):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{X} & \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i + \sigma_u^{-2}\tilde{\mathbf{G}}_i^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \text{ where } \tilde{\mathbf{z}}_i = \mathbf{ZM}_i$$

Linearly equivalent Single-step G-BLUP MME

- Single-step relationship matrix \mathbf{H} can be expressed as

$$\mathbf{H} = \mathbf{M}_1 \tilde{\mathbf{G}}_1 \mathbf{M}'_1, \text{ where } \mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \text{ and } \tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix} \text{ or}$$

$$\mathbf{H} = \mathbf{M}_2 \tilde{\mathbf{G}}_2 \mathbf{M}'_2, \mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} & \mathbf{Z}_{imp} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{Z}_m \end{bmatrix}, \tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & w\mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (1-w)\mathbf{I}_m \end{bmatrix}$$

where $\mathbf{A}_{imp} = -(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$ is **imputation operator** and $\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$ **imputed genotypes**

- New random effects $\tilde{\mathbf{u}}$ in $\tilde{\mathbf{G}}_1$: for non-genotyped $((\mathbf{A}^{11})^{-1})$ and for all animals (\mathbf{G}_w)
- $\tilde{\mathbf{G}}_2$: **residual polygenic** (RPG) effects for genotyped (\mathbf{A}_{22}), **orthogonal** marker effects (\mathbf{I}_m)
 \Rightarrow equivalent single-step SNP-BLUP (ssSNP-BLUP) ($w=0$ same as Fernando et al. (FDG))
- Original effects can be solved from linearly equivalent ssGBLUP MMEs (single trait case):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{X} & \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\tilde{\mathbf{z}}_i + \sigma_u^{-2}\tilde{\mathbf{G}}_i^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \tilde{\mathbf{z}}_i'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \text{ where } \tilde{\mathbf{z}}_i = \mathbf{ZM}_i$$

Efficient implementation: on-the-fly imputation

- Imputation $\mathbf{Z}_{imp} = -(\mathbf{A}^{11})^{-1} \mathbf{A}^{12} \mathbf{Z}_m$ expands genomic information to non-genotyped
- $(\mathbf{A}^{11})^{-1}$ calculated using **sparsity preserving Cholesky decomposition** of matrix \mathbf{A}^{11} :

$$\mathbf{A}^{11} = \mathbf{L}_1 \mathbf{L}_1' \Rightarrow (\mathbf{A}^{11})^{-1} = (\mathbf{L}_1')^{-1} \mathbf{L}_1^{-1}$$

where \mathbf{L}_1 is very sparse **lower triangular** matrix

- Solving with iterative methods: imputation term is always multiplied by some vector \mathbf{v} :

$$(\mathbf{A}^{11})^{-1} \mathbf{A}^{12} \mathbf{Z}_m \mathbf{v} = (\mathbf{L}_1')^{-1} \mathbf{L}_1^{-1} \mathbf{A}^{12} \mathbf{Z}_m \mathbf{v} = \mathbf{L}_1' \setminus (\mathbf{L}_1 \setminus (\mathbf{A}^{12} (\mathbf{Z}_m \mathbf{v})))$$

where \setminus indicates **forward** or **backward substitutions**

- Imputation is calculated implicitly **"on-the-fly"** during iteration without solving or storing imputed genotypes or genomic breeding values
- Only matrix-vector operations are performed

Efficient implementation: on-the-fly imputation

- Imputation $\mathbf{Z}_{imp} = -(\mathbf{A}^{11})^{-1} \mathbf{A}^{12} \mathbf{Z}_m$ expands genomic information to non-genotyped
- $(\mathbf{A}^{11})^{-1}$ calculated using **sparsity preserving Cholesky decomposition** of matrix \mathbf{A}^{11} :

$$\mathbf{A}^{11} = \mathbf{L}_1 \mathbf{L}_1' \Rightarrow (\mathbf{A}^{11})^{-1} = (\mathbf{L}_1')^{-1} \mathbf{L}_1^{-1}$$

where \mathbf{L}_1 is very sparse **lower triangular** matrix

- Solving with iterative methods: imputation term is always multiplied by some vector \mathbf{v} :

$$(\mathbf{A}^{11})^{-1} \mathbf{A}^{12} \mathbf{Z}_m \mathbf{v} = (\mathbf{L}_1')^{-1} \mathbf{L}_1^{-1} \mathbf{A}^{12} \mathbf{Z}_m \mathbf{v} = \mathbf{L}_1' \setminus (\mathbf{L}_1 \setminus (\mathbf{A}^{12} (\mathbf{Z}_m \mathbf{v})))$$

where \setminus indicates **forward** or **backward substitutions**

- Imputation is calculated implicitly **“on-the-fly”** during iteration without solving or storing imputed genotypes or genomic breeding values
- Only matrix-vector operations are performed

Equivalent Single-step formulations $\mathbf{H} = \mathbf{M}_i \tilde{\mathbf{G}}_i \mathbf{M}_i'$

$$\mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \sqrt{w}\mathbf{A}_{imp} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{I}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_3 = \begin{bmatrix} \sqrt{1-w}\mathbf{I}_1 & \sqrt{w}\mathbf{J}_1 & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_4 = \begin{bmatrix} (\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{A}_{imp}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_5 = \begin{bmatrix} \sqrt{1-w}(\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{J}_1(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_6 = \begin{bmatrix} (\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{A}_{imp}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix}$$

$$\tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_3 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_4 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_5 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_6 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{ganc} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

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

$$\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$$

Alternative RPG form

\mathbf{J}_1 picks non-genot.

\mathbf{J}_2 picks genotyped

Orthogonalized $\tilde{\mathbf{G}}_2$

$\mathbf{L} = \text{chol}(\mathbf{A}^{-1})$

$\mathbf{L}_1 = \text{chol}(\mathbf{A}^{11})$

Orthogonalized $\tilde{\mathbf{G}}_3$

Reduced \mathbf{A}_{22} of $\tilde{\mathbf{G}}_4$

$\hat{\mathbf{L}} = \text{chol}(\hat{\mathbf{A}}^{-1})$, where

$\hat{\mathbf{A}}$: genot. and ancestor

Coefficients w and $(1-w)$ moved to \mathbf{M}_i ; so that $\tilde{\mathbf{G}}_i$ can be inverted also when $w = 0$ or 1

Equivalent Single-step formulations $\mathbf{H} = \mathbf{M}_i \tilde{\mathbf{G}}_i \mathbf{M}_i'$

$$\mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \sqrt{w}\mathbf{A}_{imp} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{I}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_3 = \begin{bmatrix} \sqrt{1-w}\mathbf{I}_1 & \sqrt{w}\mathbf{J}_1 & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_4 = \begin{bmatrix} (\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{A}_{imp}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_5 = \begin{bmatrix} \sqrt{1-w}(\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{J}_1(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_6 = \begin{bmatrix} (\mathbf{L}_1')^{-1} & \sqrt{w}\mathbf{A}_{imp}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix}$$

$$\tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_3 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_4 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_5 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_6 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{ganc} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

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

$$\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$$

Alternative RPG form

\mathbf{J}_1 picks non-genot.

\mathbf{J}_2 picks genotyped

Orthogonalized $\tilde{\mathbf{G}}_2$

$$\mathbf{L} = \text{chol}(\mathbf{A}^{-1})$$

$$\mathbf{L}_1 = \text{chol}(\mathbf{A}^{11})$$

Orthogonalized $\tilde{\mathbf{G}}_3$

Reduced \mathbf{A}_{22} of $\tilde{\mathbf{G}}_4$

$$\hat{\mathbf{L}} = \text{chol}(\hat{\mathbf{A}}^{-1}), \text{ where}$$

$\hat{\mathbf{A}}$: genot. and ancestor

Coefficients w and $(1-w)$ moved to \mathbf{M}_i ; so that $\tilde{\mathbf{G}}_i$ can be inverted also when $w = 0$ or 1

Equivalent Single-step formulations $\mathbf{H}=\mathbf{M}_i\tilde{\mathbf{G}}_i\mathbf{M}'_i$

$$\mathbf{M}_1 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{A}_{imp} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_1 & \sqrt{w}\mathbf{A}_{imp} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{I}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_3 = \begin{bmatrix} \sqrt{1-w}\mathbf{I}_1 & \sqrt{w}\mathbf{J}_1 & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2 & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_4 = \begin{bmatrix} (\mathbf{L}'_1)^{-1} & \sqrt{w}\mathbf{A}_{imp}\mathbf{J}_2(\mathbf{L}'_1)^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}'_1)^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_5 = \begin{bmatrix} \sqrt{1-w}(\mathbf{L}'_1)^{-1} & \sqrt{w}\mathbf{J}_1(\mathbf{L}'_1)^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\mathbf{J}_2(\mathbf{L}'_1)^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\mathbf{M}_6 = \begin{bmatrix} (\mathbf{L}'_1)^{-1} & \sqrt{w}\mathbf{A}_{imp}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_{imp} \\ \mathbf{0} & \sqrt{w}\hat{\mathbf{J}}_2(\hat{\mathbf{L}}')^{-1} & \sqrt{1-w}\mathbf{Z}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_1 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w \end{bmatrix}$$

$$\tilde{\mathbf{G}}_2 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_3 = \begin{bmatrix} (\mathbf{A}^{11})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_4 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_5 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{all} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

$$\tilde{\mathbf{G}}_6 = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{ganc} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

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

$$\mathbf{Z}_{imp} = \mathbf{A}_{imp}\mathbf{Z}_m$$

Alternative RPG form

\mathbf{J}_1 picks non-genot.

\mathbf{J}_2 picks genotyped

Orthogonalized $\tilde{\mathbf{G}}_2$

$$\mathbf{L} = \text{chol}(\mathbf{A}^{-1})$$

$$\mathbf{L}_1 = \text{chol}(\mathbf{A}^{11})$$

Orthogonalized $\tilde{\mathbf{G}}_3$

Reduced \mathbf{A}_{22} of $\tilde{\mathbf{G}}_4$

$$\hat{\mathbf{L}} = \text{chol}(\hat{\mathbf{A}}^{-1}), \text{ where}$$

$\hat{\mathbf{A}}$: genot. and ancestors

Coefficients w and $(1-w)$ moved to \mathbf{M}_i so that $\tilde{\mathbf{G}}_i$ can be inverted also when $w=0$ or 1

Testing with small data set and simple model

Small Nordic Red dairy cattle data set with deregressed proofs of milk yield:

Animals	73579
Non-genotyped	70694
Genotyped	2885
Genotyped and their ancestors	6833
Observations	67648
Markers	37526

- Single trait, one fixed effect
- Comparing equivalent ssGBLUP formulations
- Polygenic effect proportions w : 0, 0.01, 0.2, 1
- Genomic relationship matrix \mathbf{G}_g : VanRaden 1
- One clone pair, couple of near-clones
- Implemented in Octave

Comparison statistics:

- **Efficiency**: sparsity of matrices
- **Speed**: number of iterations (PCG, without and with diagonal preconditioning)
- **Accuracy**: all equivalent ssGBLUP MMEs converged to the same results

Testing with small data set and simple model

Small Nordic Red dairy cattle data set with deregressed proofs of milk yield:

Animals	73579
Non-genotyped	70694
Genotyped	2885
Genotyped and their ancestors	6833
Observations	67648
Markers	37526

- Single trait, one fixed effect
- Comparing equivalent ssGBLUP formulations
- Polygenic effect proportions w : 0, 0.01, 0.2, 1
- Genomic relationship matrix \mathbf{G}_g : VanRaden 1
- One clone pair, couple of near-clones
- Implemented in Octave

Comparison statistics:

- **Efficiency**: sparsity of matrices
- **Speed**: number of iterations (PCG, without and with diagonal preconditioning)
- **Accuracy**: all equivalent ssGBLUP MMEs converged to the same results

Testing with small data set and simple model

Small Nordic Red dairy cattle data set with deregressed proofs of milk yield:

Animals	73579
Non-genotyped	70694
Genotyped	2885
Genotyped and their ancestors	6833
Observations	67648
Markers	37526

- Single trait, one fixed effect
- Comparing equivalent ssGBLUP formulations
- Polygenic effect proportions w : 0, 0.01, 0.2, 1
- Genomic relationship matrix \mathbf{G}_g : VanRaden 1
- One clone pair, couple of near-clones
- Implemented in Octave

Comparison statistics:

- **Efficiency**: sparsity of matrices
- **Speed**: number of iterations (PCG, without and with diagonal preconditioning)
- **Accuracy**: all equivalent ssGBLUP MMEs converged to the same results

Efficiency: Sparse matrix and decomposition sizes

Sparse matrix and decompositions sizes and **mean number of non-zeros** per row/column in Octave (cholmod):

	rows	cols	mean non-zeros row/col
A^{-1}	73579	73579	3.20*
A^{11}	70694	70694	2.39*
A^{12}	70694	2885	0.88/21.66
A^{22}	2885	2885	1.51*
L	73579	73579	2.60
L_1	70694	70694	2.57
\hat{A}^{-1}	6833	6833	3.31*
\hat{L}	6833	6833	2.66

* = Non-zeros counted from lower/upper triangular region

- All sparse matrices are very sparse in the test case
- Largest sparse matrix: 4.16 megabytes
- About 3 non-zeros on each row
⇒ growth is linear, not quadratic

Speed: Number of PCG iterations with tolerance 10^{-12}

Heritability $h^2 = 0.5$

MME	size	polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
orig.	73580	-	362	358	357	-	93	59	57
1	73580	-	536	536	536	-	152	158	171
2	111106	646	1203	987	536	847	1176	1007	171
3	181800	646	4023	3307	355	847	4058	3195	57
4	181800	193	196	191	182	442	443	853	283
5	181800	193	197	192	181	442	455	868	124
6	115054	193	196	191	182	442	443	853	283

$w=0$: same as FDG (MME 2 and 3) $w=1$: animal model

Fully orthogonalized equivalent ssSNP-BLUP MMEs (4-6):

- Speed unaffected by w and no gain from preconditioning
- Non-preconditioned seem to have lower iteration number than original single-step when heritability is low

Heritability dependency:

$$\tilde{\mathbf{z}}_i' \tilde{\mathbf{z}}_i + \lambda \tilde{\mathbf{G}}_i^{-1}$$

where $\lambda = \frac{1-h^2}{h^2}$

h^2 small \Rightarrow λ large

Speed: Number of PCG iterations with tolerance 10^{-12}

Heritability $h^2 = 0.5$

MME	size	polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
orig.	73580	-	362	358	357	-	93	59	57
1	73580	-	536	536	536	-	152	158	171
2	111106	646	1203	987	536	847	1176	1007	171
3	181800	646	4023	3307	355	847	4058	3195	57
4	181800	193	196	191	182	442	443	853	283
5	181800	193	197	192	181	442	455	868	124
6	115054	193	196	191	182	442	443	853	283

$w=0$: same as FDG (MME 2 and 3) $w=1$: animal model

Fully orthogonalized equivalent ssSNP-BLUP MMEs (4-6):

- Speed unaffected by w and no gain from preconditioning
- Non-preconditioned seem to have lower iteration number than original single-step when heritability is low

Heritability dependency:

$$\tilde{\mathbf{z}}_i' \tilde{\mathbf{z}}_i + \lambda \tilde{\mathbf{G}}_i^{-1}$$

where $\lambda = \frac{1-h^2}{h^2}$

h^2 small \Rightarrow λ large

Speed: Number of PCG iterations with tolerance 10^{-12}

Heritability $h^2 = 0.5$

Heritability $h^2 = 0.1$

MME	size	polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
orig.	73580	-	362	358	357	-	93	59	57
1	73580	-	536	536	536	-	152	158	171
2	111106	646	1203	987	536	847	1176	1007	171
3	181800	646	4023	3307	355	847	4058	3195	57
4	181800	193	196	191	182	442	443	853	283
5	181800	193	197	192	181	442	455	868	124
6	115054	193	196	191	182	442	443	853	283

		polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
		-	619	624	623	-	178	126	119
		-	758	701	744	-	191	148	139
		736	1005	907	748	438	491	429	139
		736	3422	3154	623	438	1732	1416	119
		74	74	72	71	107	105	151	82
		74	73	70	70	107	107	151	49
		74	74	72	71	107	105	151	82

$w=0$: same as FDG $w=1$: animal model
(MME 2 and 3)

Fully orthogonalized equivalent ssSNP-BLUP MMEs (4-6):

- Speed unaffected by w and no gain from preconditioning
- Non-preconditioned seem to have lower iteration number than original single-step when heritability is low

Heritability dependency:

$$\tilde{\mathbf{z}}_i' \tilde{\mathbf{z}}_i + \lambda \tilde{\mathbf{G}}_i^{-1}$$

where $\lambda = \frac{1-h^2}{h^2}$

h^2 small $\Rightarrow \lambda$ large

Speed: Number of PCG iterations with tolerance 10^{-12}

Heritability $h^2 = 0.5$

Heritability $h^2 = 0.1$

MME	size	polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
orig.	73580	-	362	358	357	-	93	59	57
1	73580	-	536	536	536	-	152	158	171
2	111106	646	1203	987	536	847	1176	1007	171
3	181800	646	4023	3307	355	847	4058	3195	57
4	181800	193	196	191	182	442	443	853	283
5	181800	193	197	192	181	442	455	868	124
6	115054	193	196	191	182	442	443	853	283

		polygenic proportion w							
		no preconditioning				with preconditioning			
		0.00	0.01	0.20	1.00	0.00	0.01	0.20	1.00
		-	619	624	623	-	178	126	119
		-	758	701	744	-	191	148	139
		736	1005	907	748	438	491	429	139
		736	3422	3154	623	438	1732	1416	119
		74	74	72	71	107	105	151	82
		74	73	70	70	107	107	151	49
		74	74	72	71	107	105	151	82

$w=0$: same as FDG $w=1$: animal model
(MME 2 and 3)

Fully orthogonalized equivalent ssSNP-BLUP MMEs (4-6):

- Speed unaffected by w and no gain from preconditioning
- Non-preconditioned seem to have lower iteration number than original single-step when heritability is low

Heritability dependency:

$$\tilde{\mathbf{z}}_i' \tilde{\mathbf{z}}_i + \lambda \tilde{\mathbf{G}}_i^{-1}$$

where $\lambda = \frac{1-h^2}{h^2}$

h^2 small \Rightarrow λ large

Conclusions

- Very simple procedure to derive **linearly equivalent** ssGBLUP MMEs
- Several linearly equivalent Single-step G-BLUP MMEs:
 - ▶ Build-in **residual polygenic effects**
 - ▶ **Avoiding inversion** of genomic relationship matrix: equivalent Single-step SNP-BLUP MMEs
- Efficient implementation:
 - ▶ **On-the-fly imputation** of genomic information
 - ▶ Utilizing **sparse matrices** and **sparse decompositions** \Rightarrow **Millions of genotyped possible**
- Preliminary results:
 - ▶ Stable and low iteration counts for **fully orthogonalized** ssSNP-BLUP MMEs for all polygenic proportions w without preconditioning
 - ▶ Non-preconditioned ssSNP-BLUP seems to be comparable to diagonally preconditioned original Single-step depending on **heritability**



Sparse matrix and decomposition sizes (larger)

Pedigree of **9,225,136 animals, 1,539,182 “simulated” genotyped** (90% of bulls, 50% of cows born last 6 years, and 10% of other cows):

	rows	cols	non-zeros	mean nnz	memory usage
A^{-1}	9225136	9225136	34393022*	3.73*	595.18 megabytes*
A^{11}	7695153	7695153	14785215*	1.92*	284.31 megabytes*
A^{12}	7695153	1529983	15615397	2.03/10.21	249.94 megabytes
A^{22}	1529983	1529983	3992410*	2.61*	72.59 megabytes*
L	9225136	9225136	26452169	2.87	474.01 megabytes
L_1	7695153	7695153	15381760	2.00	297.47 megabytes
\hat{A}^{-1}	4292835	4292835	15865044*	3.70*	274.83 megabytes*
\hat{L}	4292835	4292835	12211956	2.84	219.09 megabytes

*Non-zeros and memory usage of symmetric matrices counted from the lower/upper triangular region only.

Number of `pcg()` iterations using tolerance 10^{-12}

Heritability 0.5

MME	size	polygenic proportion w											
		no preconditioning						with preconditioning					
		0.00	0.01	0.10	0.20	0.30	1.00	0.00	0.01	0.10	0.20	0.30	1.00
orig.	73580	-	362	358	358	355	357	-	93	62	59	57	57
1	73580	-	536	512	536	523	536	-	152	153	158	158	171
2	111106	646	1203	1085	987	886	536	847	1176	1099	1007	950	171
3	181800	646	4023	3660	3307	3040	355	847	4058	3658	3195	2873	57
4	181800	193	196	193	191	190	182	442	443	702	853	920	283
5	181800	193	197	194	192	190	181	442	455	720	868	917	124
6	115054	193	196	193	191	190	182	442	443	702	853	920	283

$w = 1$: animal model

$w = 0$: same as FDG
(MME 2 and 3)

Heritability 0.1

orig.	73580	-	619	622	624	624	623	-	178	131	126	122	119
1	73580	-	758	733	701	728	744	-	191	156	148	145	139
2	111106	736	1005	970	907	836	748	438	491	459	429	398	139
3	181800	736	3422	3315	3154	2934	623	438	1732	1592	1416	1247	119
4	181800	74	74	72	72	71	71	107	105	130	151	160	82
5	181800	74	73	72	70	71	70	107	107	132	151	162	49
6	115054	74	74	72	72	71	71	107	105	130	151	160	82