

Genetic evaluation using unsymmetric single step genomic methodology with large number of genotypes

I. Aguilar¹, A. Legarra², S. Tsuruta³, I. Misztal³

¹INIA Las Brujas

²INRA Toulouse

³University of Georgia

Single-Step evaluation

- Unified approach with pedigree, phenotypic and genomic markers information considered simultaneously
- Pedigree-based relationships augmented by genomic relationship matrix (Miształ et al. 2009)

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

$$\mathbf{H} = \mathbf{A} + \mathbf{A}_{\Lambda}$$

\mathbf{A} - conventional numerator relationship matrix

\mathbf{A}_{Λ} - matrix modified to account for genomic relationships

Single step GBLUP - H Inverse

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22}) \\ (\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix}$$

Legarra et al., 2009

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

Aguilar et al., 2010
Christensen & Lund, 2010

Inverse of the regular pedigree relationship matrix

...and avoiding « double counting »

Correcting for genomic relationships...

Computations in single-step

- Add extra relationship matrix to current software
- Cubic cost for inversion relationship matrices



ORIGINAL ARTICLE

Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation

I. Aguilar^{1,2}, I. Misztal¹, A. Legarra³ & S. Tsuruta¹

1 Animal and Dairy Science Department, University of Georgia, Athens, GA, USA

2 Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, Uruguay

3 INRA, UR631 SAGA, BP 52627, 32326 Castanet-Tolosan, France

Computing time for different number of genotyped individuals

Blade Server INIALB 24 CPUs 750 GB RAM

Intel(R) Xeon(R) CPU E7- 4870 @ 2.40GHz

Number of genotypes	Genomic Relationship Matrix	
	Creation	Inversion
10k	0.6 m	0.1 m
30k	5.4 m	3 m
50k	15 m	14 m
70k	30 m	36.4 m
100k	60 m	122 m
120k	140 m	208 m
150k	215 m	406 m

Alternatives to Single-Step

- Legarra and Ducrocq 2012
- Strategies to implement single-step GBLUP with no inversion relationship matrices
 - Iterative single-step
 - Unsymmetric single-step
- Test with simulated data

Unsymmetric and non inverse ssGBLUP

(Legarra & Ducroq 2012)

Regular equations

$$\begin{array}{ccc}
 \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W}'_1 & \mathbf{X}'_2\mathbf{W}'_2 \\
 \mathbf{W}'_1\mathbf{X}_1 & \mathbf{W}'_1\mathbf{W}'_1 + \alpha_u\mathbf{A}^{11} & \alpha_u\mathbf{A}^{12} \\
 \mathbf{W}'_1\mathbf{X}_2 & \alpha_u\mathbf{A}^{12} & \mathbf{W}'_2\mathbf{W}'_2 + \alpha_u\mathbf{A}^{22} \\
 0 & 0 & \alpha_u\mathbf{I} \\
 0 & 0 & \alpha_u\mathbf{I}
 \end{array}$$

Unsymmetric

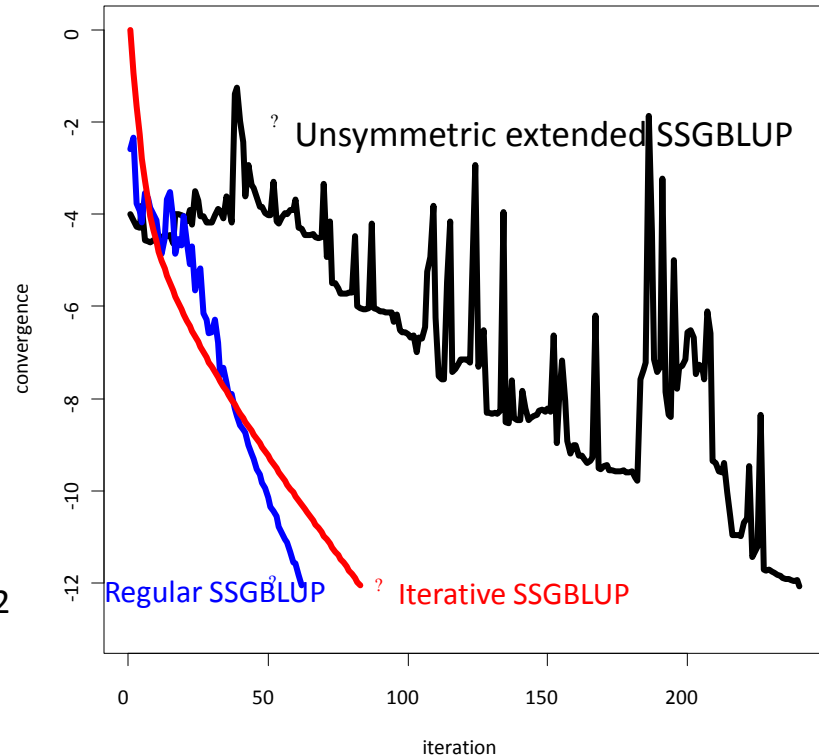
$$\begin{array}{cc|c}
 0 & 0 & \hat{\mathbf{b}} \\
 0 & 0 & \hat{\mathbf{u}}_1 \\
 \alpha_u\mathbf{I} & -\alpha_u\mathbf{I} & \hat{\mathbf{u}}_2 \\
 \alpha_u\mathbf{A}_{22} & 0 & -\hat{\phi} \\
 0 & \alpha_u\mathbf{G} & -\hat{\gamma}
 \end{array}
 =
 \begin{array}{c}
 \mathbf{X}'\mathbf{y} \\
 \mathbf{W}'_1\mathbf{y}_1 \\
 \mathbf{W}'_2\mathbf{y}_2 \\
 0 \\
 0
 \end{array}$$

- No need for inverse of genomic relationship matrix and pedigree relationship matrix for genotyped
- Unsymmetric system of equations

Results simulation

(Legarra & Ducrocq 2012)

- Using simulated data all strategies arrived to the same solution
 - 28000 pedigrees
 - 1800 genotypes
- Reasonable computing time:
 - 2 s « regular » (with \mathbf{G} and \mathbf{A}_{22} already inverted)
 - 47 s « unsymmetric »
 - 286 s « iterative »



Implementation on iteration on data program unsymmetric BiCGSTAB

- Iteration on data program (BLUP90IOD)
Tsuruta et al 2001
- Solvers:
 - BiCGSTAB (van de Vorst 1992)
 - BiCGSTAB(l) (Sleijpen & Fokkema 1993)
- Different preconditioner
 - Simple & Unsymmetric preconditioner

Solving $Ax=b$ by Bi-Conjugate Gradient Stabilized algorithm (Bi-CGSTAB)

Compute $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$ for some initial guess $\mathbf{x}^{(0)}$

Choose $\tilde{\mathbf{r}}$ (for example, $\tilde{\mathbf{r}} = \mathbf{r}^{(0)}$)

for $i = 1, 2, \dots$

$$\rho_{i-1} = \tilde{\mathbf{r}}' \mathbf{r}^{(i-1)}$$

if $\rho_{i-1} = 0$ method fails

if $i = 1$

$$\mathbf{p}^{(i)} = \mathbf{r}^{(i-1)}$$

else

$$\beta_{i-1} = \frac{\rho_{i-1}}{\rho_{i-2}} \frac{\alpha_{i-1}}{\omega_{i-1}}$$

$$\mathbf{p}^{(i)} = \mathbf{r}^{(i-1)} + \beta_{i-1} (\mathbf{p}^{(i-1)} - \omega_{i-1} \mathbf{v}^{(i-1)})$$

endif

$$\text{solve } M^{-1} \hat{\mathbf{p}} = \mathbf{p}^{(i)}$$

$$\mathbf{v}^{(i)} = A \hat{\mathbf{p}}$$

$$\alpha_i = \rho_{i-1} / \tilde{\mathbf{r}}' \mathbf{v}^{(i)}$$

$$\mathbf{g} = \mathbf{r}^{(i-1)} - \alpha_i \mathbf{v}^{(i)}$$

check norm of \mathbf{g} ; if small enough: set $\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} + \alpha_i \hat{\mathbf{p}}$ and stop

$$\text{solve } M \hat{\mathbf{g}} = \mathbf{g}$$

$$\mathbf{t} = A \hat{\mathbf{g}}$$

$$\omega_i = \mathbf{t}' \mathbf{g} / \mathbf{t}' \mathbf{t}$$

$$\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)} + \alpha_i \hat{\mathbf{p}} + \omega_i \hat{\mathbf{g}}$$

$$\mathbf{r}^{(i)} = \mathbf{g} - \omega_i \mathbf{t}$$

check convergence; continue if necessary

for continuation it is necessary that $\omega_i \neq 0$

end

Uses 2 LHS by vector multiplications

Matrix-vector operations for unsymmetric ssGBLUP

$$LHS^* p = \begin{bmatrix} X'X & X'_1 W_1 & X'_2 W_2 & 0 & 0 \\ W'_1 X_1 & W'_1 W_1 + \alpha A^{11} & \alpha A^{12} & 0 & 0 \\ W'_1 X_2 & \alpha A^{21} & W'_2 W_2 + \alpha A^{22} & \alpha I & -\alpha I \\ 0 & 0 & \alpha I & \alpha A_{22} & 0 \\ 0 & 0 & \alpha I & 0 & \alpha G \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}$$

Sequential operations

$$LHS^* p = \begin{bmatrix} X' X p_1 + X'_1 W_1 p_2 + X'_2 W_2 p_3 \\ W'_1 X_1 p_1 + W'_1 W_1 p_2 \\ W'_1 X_2 p_1 + W'_2 W_2 p_3 \\ 0 \\ 0 \end{bmatrix}$$

→ Contributions due to records

$$+ \begin{bmatrix} 0 \\ \alpha A^{11} p_2 + \alpha A^{12} p_3 \\ \alpha A^{21} p_2 + \alpha A^{22} p_3 \\ 0 \\ 0 \end{bmatrix}$$

→ Contributions due to relationships

$$+ \begin{bmatrix} 0 \\ 0 \\ \alpha l p_4 - \alpha l p_5 \\ \alpha l p_3 + \alpha A_{22} p_4 \\ \alpha l p_3 + \alpha G p_5 \end{bmatrix}$$

→ Contributions due to genomics

Final Score US Holstein data set

(Aguilar et al., 2010)

- US Holstein Final Score – May 2009
 - 10,466,066 records (1955-2009)
 - 6,232,548 cows
 - 9,100,106 pedigrees
- Genotypes from AIPL
 - Illumina BovineSNP50 chip
 - 6,508 bulls
 - 38,416 SNPs

Final Score US Holstein data set

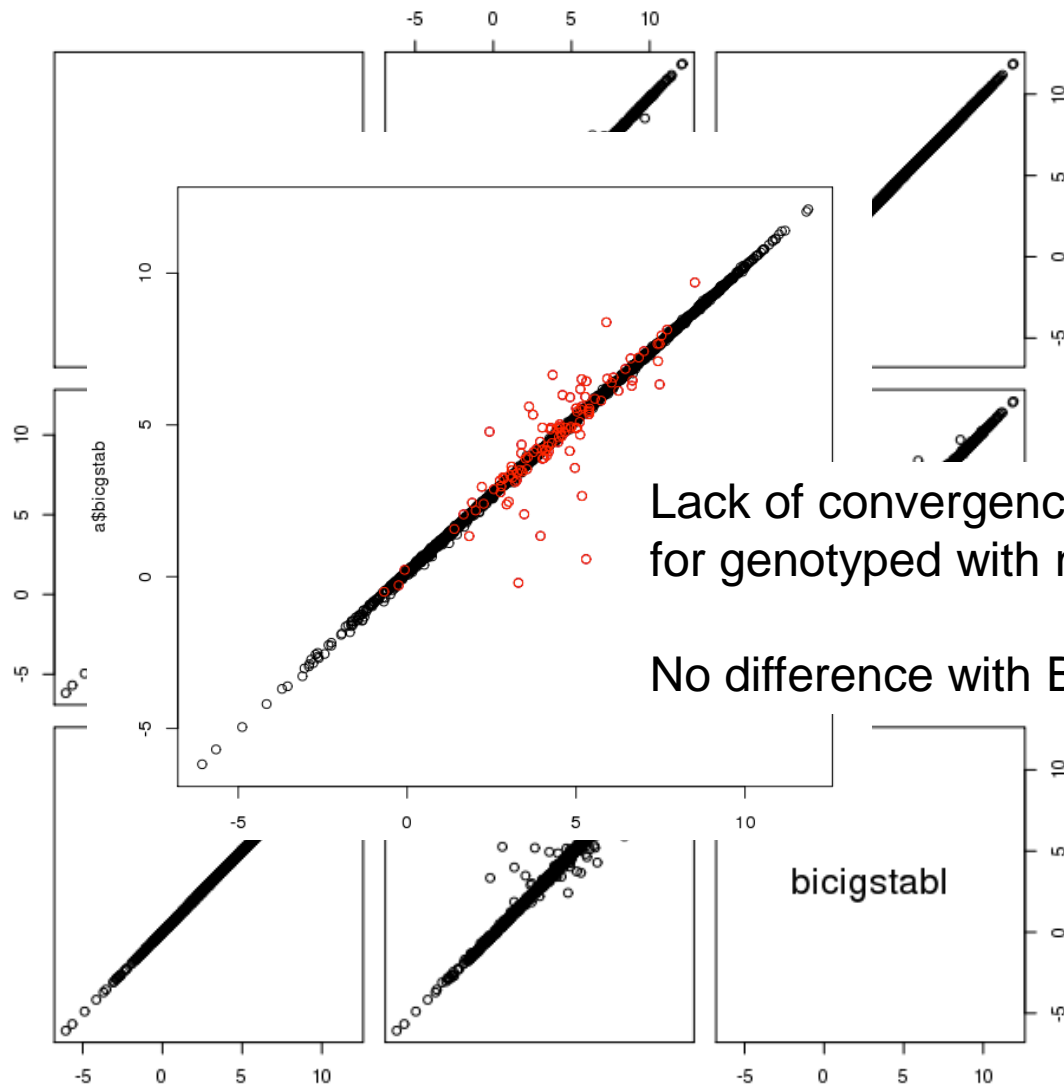
Correlations of EBV

	hinv	bicgstab	bicigstabl
hinv	1.0000000	0.9985626	0.9999728
bicgstab	0.9985626	1.0000000	0.9985127
bicigstabl	0.9999728	0.9985127	1.0000000

Statistics EBV

hinv	bicgstab	bicigstabl
Min. : -6.069	Min. : -6.181	Min. : -6.075
1st Qu.: 2.995	1st Qu.: 3.178	1st Qu.: 3.000
Median : 4.484	Median : 4.689	Median : 4.505
Mean : 4.540	Mean : 4.740	Mean : 4.556
3rd Qu.: 6.170	3rd Qu.: 6.394	3rd Qu.: 6.192
Max. : 11.873	Max. : 12.105	Max. : 11.905

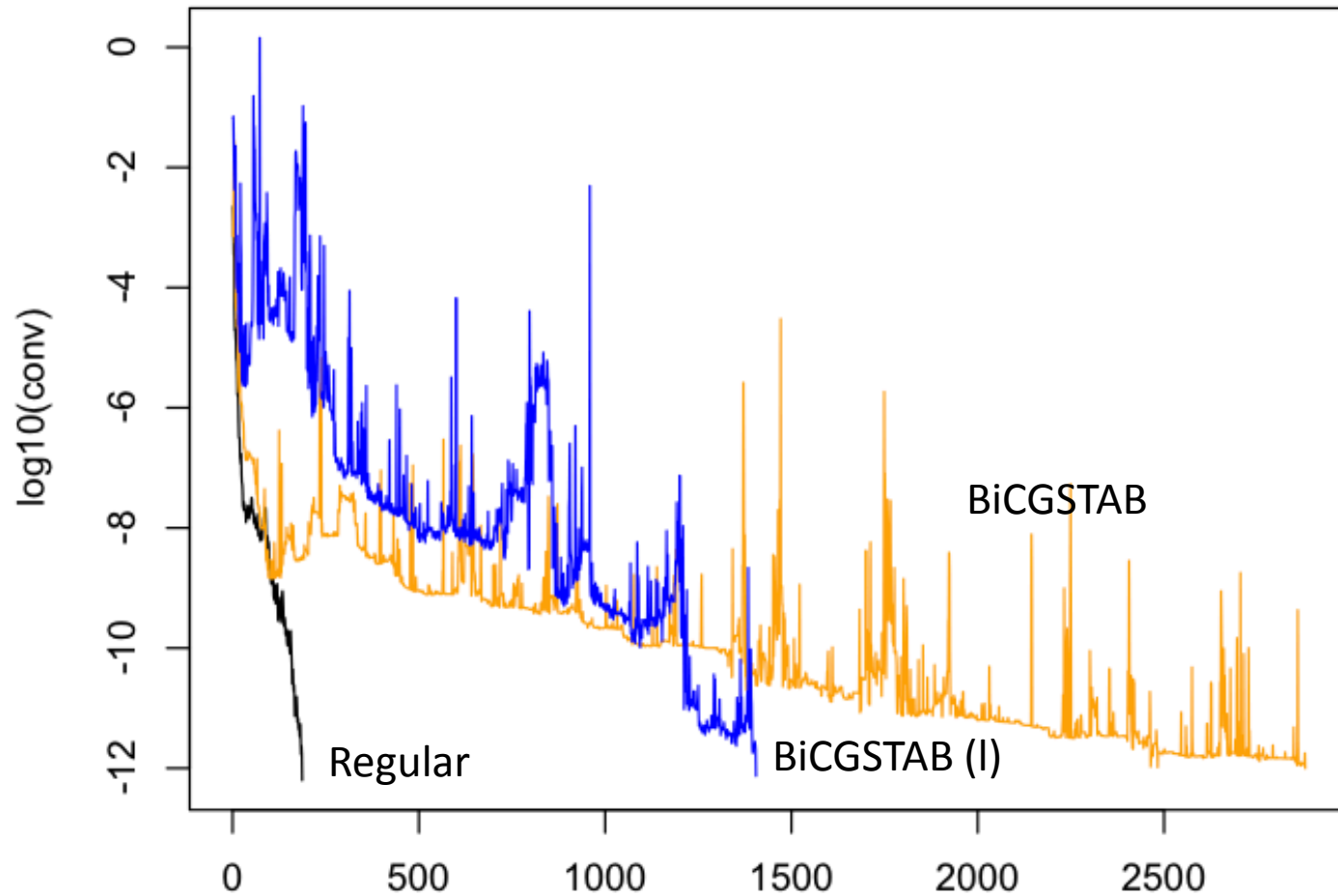
Final Score US Holstein data set



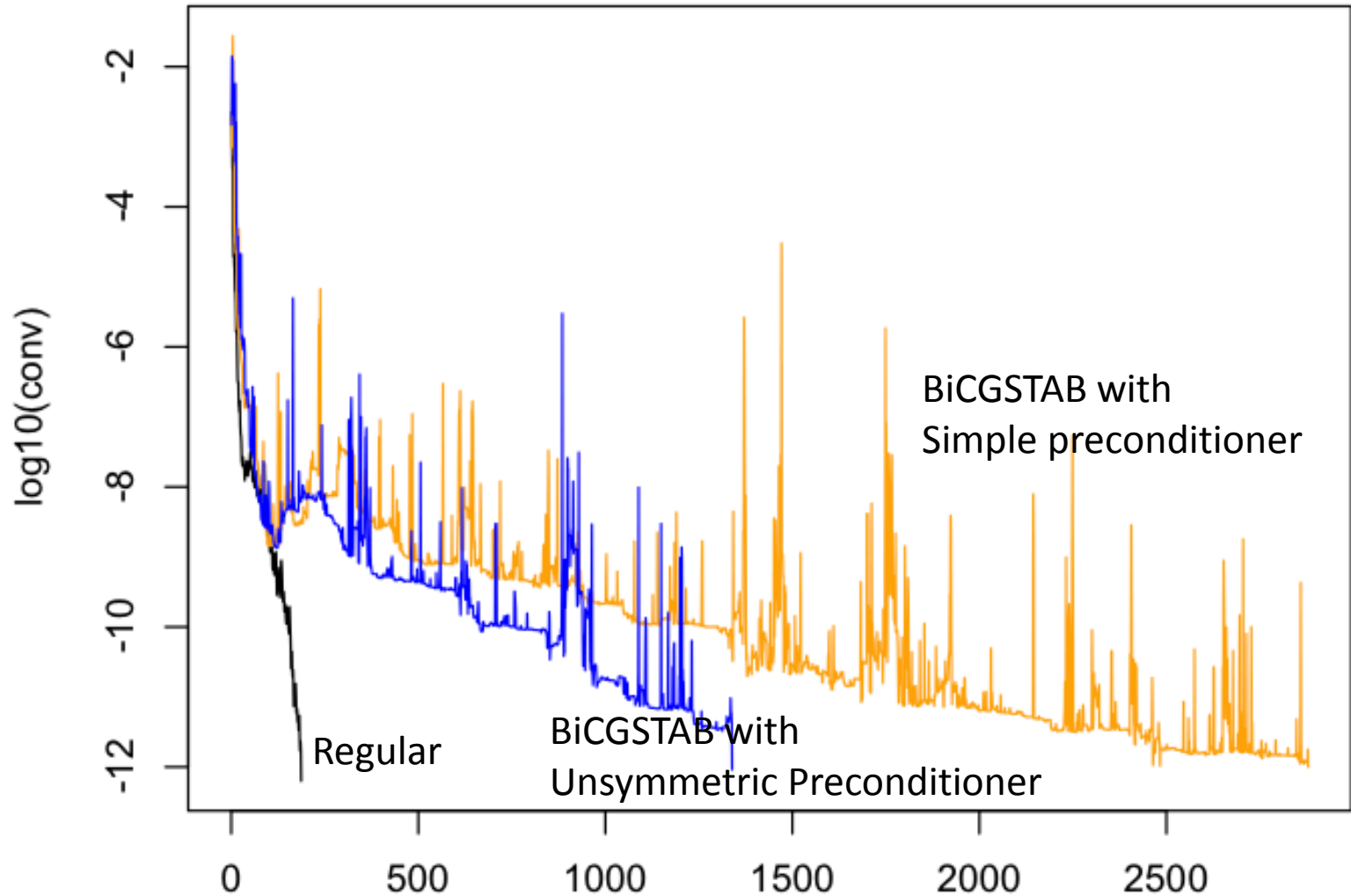
Lack of convergence with BiGSTAB
for genotyped with no pedigree information

No difference with BiCGSTAB (I)

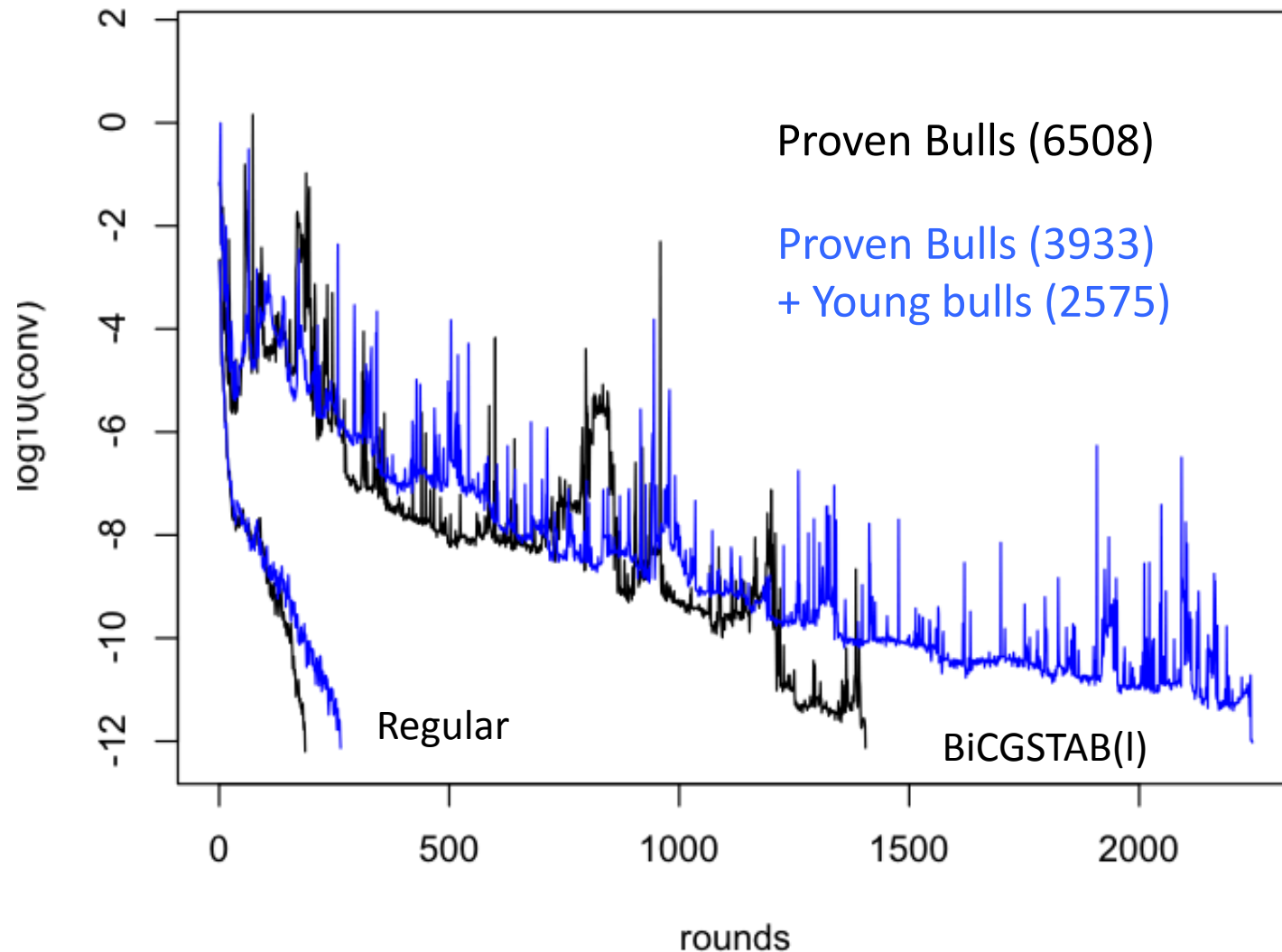
Effect of Solver method on convergence



Effect of preconditioner on convergence



Effect of data structure on convergence



Summary

- Feasibility of unsymmetric ssGBLUP for large scale genetic evaluation
- Convergence affected
 - Different preconditioners
 - Data structure of genotyped individuals
- Slow of convergence with BiCGSTAB solver
=> lower correlation of EBV for particular cases
- Need to consider strict convergence criteria
- More sophisticated preconditioner for unsymmetric ssGBLUP

Acknowledgements

Holstein Association

AIPL