Deflated preconditioned conjugate gradient method for solving single-step single nucleotide polymorphism BLUP

J. Vandenplas, H. Eding, M.P.L. Calus, C. Vuik

February 12, 2018

WAGENINGEN UNIVERSITY & RESEARCH

100 years 1918 — 2018

CRV

TU Delft

Breed4Food is dedicated to be the leading research consortium in animal breeding, genetics and genomics enabling the Breed4Food partners to breed better products to benefit society's needs.

Single-step models

- Prediction of genomic breeding values
  - Genotyped and non-genotyped animals
- Single-step GBLUP
  - Animal-based model
  - Limited?
- Single-step SNPBLUP
  - SNP-based model
  - Several equivalent formulations
  - No limitation?!

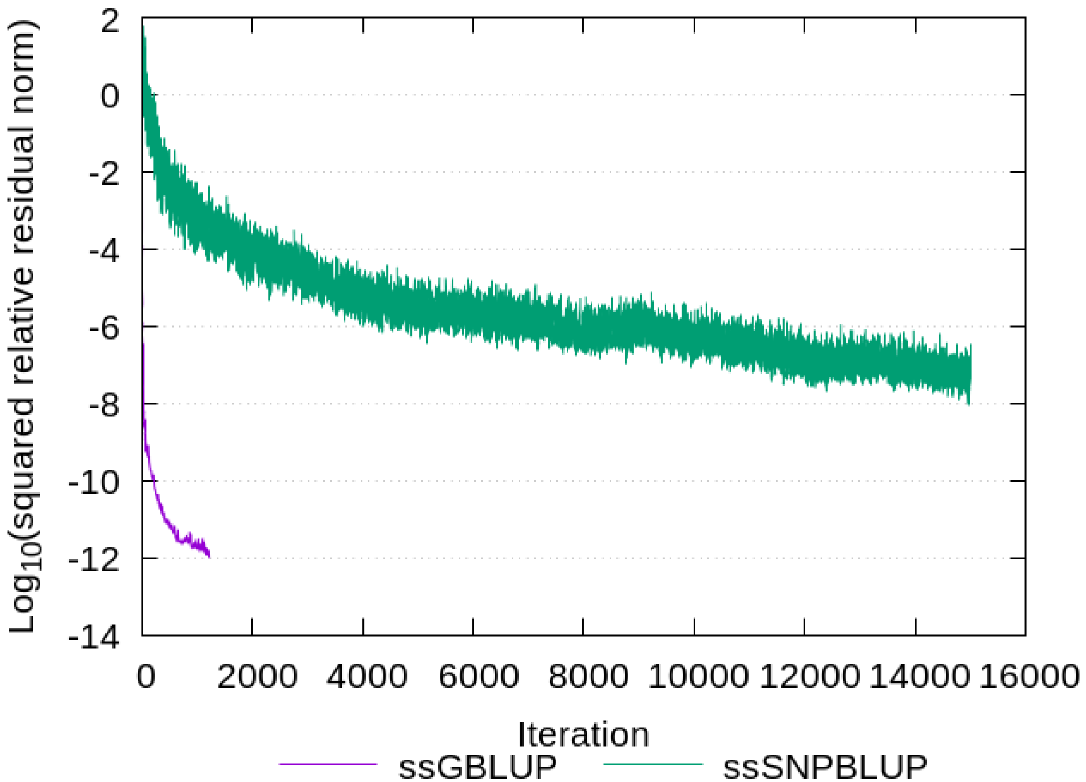
Fernando et al., 2016; Legarra et al., 2014; Liu et al., 2014; Taskinen et al., 2017

Single-step SNPBLUP

System of equations has the form **Cx = b**

➔Iterative solver: Preconditioned Conjugate Gradient

➔Convergence issues often reported...



WAGENINGEN UNIVERSITY & RESEARCH

100years 1918 — 2018

Liu et al., 2014; Manzanilla Pech et al. 2017; Taskinen et al., 2017

Aim

1. Comparison of properties of coefficient matrices of ssGBLUP and ssSNPBLUP

→ Comprehension of convergence patterns of PCG

2. Implementation of a Deflated PCG method for solving efficiently ssSNPBLUP

Equivalent single-step models

ssGBLUP:

$$\mathbf{y} = \mathbf{Xb} + \begin{bmatrix} \mathbf{W}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_g \end{bmatrix} \begin{bmatrix} \mathbf{u}_n \\ \mathbf{u}_g \end{bmatrix} + \mathbf{e}$$

ssSNPBLUP:

$$\mathbf{y} = \mathbf{Xb} + \begin{bmatrix} \mathbf{W}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_g & \mathbf{W}_g\mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{u}_n \\ \mathbf{a}_g \\ \mathbf{g} \end{bmatrix} + \mathbf{e}$$

$$\mathbf{u}_g = \mathbf{a}_g + \mathbf{Zg}$$

Addition of SNP effects

$\mathbf{b}$: fixed effects

$\mathbf{u}_n$, $\mathbf{u}_g$: aggregate GEBVs for (non-)genotyped animals

$\mathbf{a}_g$: residual polygenic effects for genotyped animals

Fernando et al., 2016; Mantysaari and Stranden, 2016

Conjugate Gradient (CG)

- Successive approximations to obtain a more accurate solution of **x** by solving

$$\mathbf{Cx=b}$$

- Convergence
  - Function of the effective condition number of **C**

$$\kappa(\mathbf{C}) = \frac{largest\ eigenvalue\ of\ \mathbf{C}}{(non-zero)\ smallest\ eigenvalue\ of\ \mathbf{C}}$$

  - Smaller condition number → faster convergence

Preconditioned CG (PCG)

- Improvement of the condition number from $\kappa(\mathbf{C})$ to $\kappa(\mathbf{M}^{-1}\mathbf{C})$ by introducing a preconditioner $\mathbf{M}$

$$\mathbf{M}^{-1}\mathbf{C}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b}$$

- In animal breeding
  - PCG often implemented
  - Usually: $\mathbf{M} = diag(\mathbf{C})$ (or a variant)

Deflated PCG (DPCG)

- Improvement of the condition number from $\kappa(\mathbf{M}^{-1}\mathbf{C})$ to $\kappa(\mathbf{M}^{-1}\mathbf{P}\mathbf{C})$ by introducing a second-level preconditioner **P**

$$\mathbf{M}^{-1}\mathbf{P}\mathbf{C}\mathbf{x} = \mathbf{M}^{-1}\mathbf{P}\mathbf{b}$$

P chosen such that unfavourable eigenvalues are set to 0 (deflated)

Deflated PCG (DPCG)

- **P** = deflation matrix

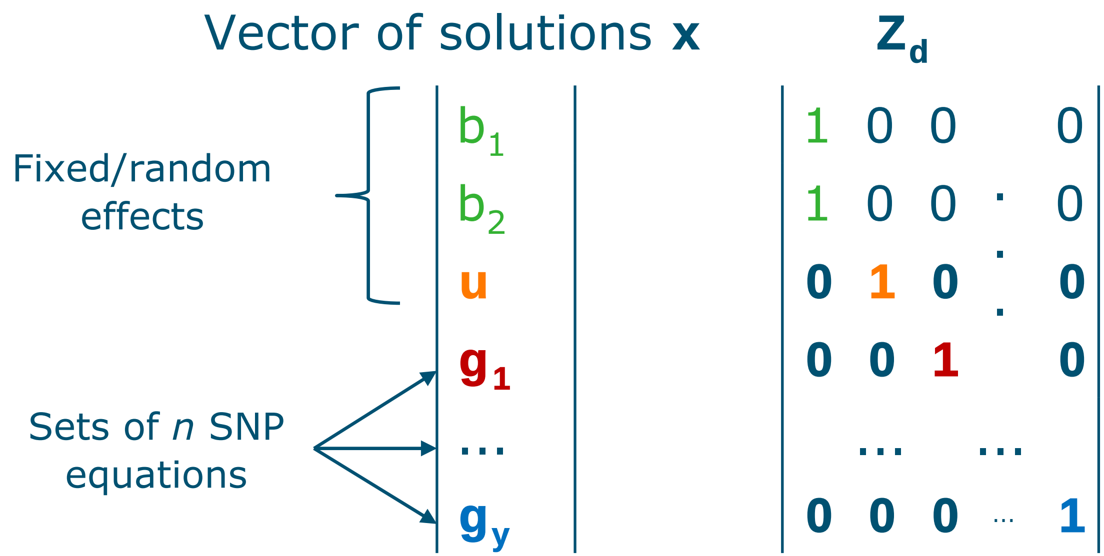
  $= \mathbf{I} - \mathbf{C}\mathbf{Z_d}(\mathbf{Z_d}'\mathbf{C}\mathbf{Z_d})^{-1}\mathbf{Z_d}'$

- $\mathbf{Z_d}$ contains the deflation vectors
  - Approximation of the same space of the span of unfavourable eigenvectors

  → Set of deflation vectors ≈ Set of unfavourable eigenvectors

Deflated PCG (DPCG)

- Setting-up of $\mathbf{Z_d}$ following a subdomain deflation approach
  - 1 subdomain per fixed/random effect
  - 1 subdomain per set of *n* (1, 5, 50, or 200) SNP equations



Sparse $\mathbf{Z_d}$ → efficient implementation → small extra-costs

Data

- 61,592 Ovum pick-up sessions
- 37,021 animals
  - 4,109 phenotyped animals
  - 6,169 genotyped animals (without phenotype)
- 9,994 segregating SNPs

- Heritability = 0.35
- Residual polygenic variance = 5%

Cornelissen et al. 2017

Results – Spectra and condition numbers

Model	Method	Smallest eigenvalue	Largest eigenvalue	κ	#iter.
ssG	PCG	$1.1*10^{-4}$	11.9	$1.1*10^{5}$	273
ssSNP	PCG	$1.1*10^{-4}$	**181.0**	**$1.7*10^{6}$**	**1497**

ssSNPBLUP vs ssGBLUP

- Unchanged smallest eigenvalues
- Increased largest eigenvalues

→ Larger condition number

→ Increased number of iterations

Results – Spectra and condition numbers

Model	Method	Smallest eigenvalue	Largest eigenvalue	κ	#iter.
ssG	PCG	$1.1*10^{-4}$	11.9	$1.1*10^{5}$	273
ssSNP	PCG	$1.1*10^{-4}$	**181.0**	**$1.7*10^{6}$**	**1497**
	DPCG (200)	$1.1*10^{-4}$	99.4	$9.3*10^{5}$	1195

200 SNPs equations per subdomain

- ➔Unchanged smallest value
- ➔Decreased largest eigenvalue
- ➔Better condition number after deflation

Results – Spectra and condition numbers

Model	Method	Smallest eigenvalue	Largest eigenvalue	κ	#iter.
ssG	PCG	$1.1*10^{-4}$	11.9	$1.1*10^{5}$	273
ssSNP	PCG	$1.1*10^{-4}$	**181.0**	**$1.7*10^{6}$**	**1497**
	DPCG (200)	$1.1*10^{-4}$	99.4	$9.3*10^{5}$	1195
	DPCG (50)	$1.1*10^{-4}$	40.5	$3.8*10^{5}$	880
	DPCG (5)	$1.1*10^{-4}$	**6.0**	**$5.6*10^{4}$**	**338**
	DPCG (1)	$1.1*10^{-4}$	**6.0**	**$5.4*10^{4}$**	**240**

1 and 5 SNPs per subdomain

- Similar (decreased) condition numbers
- #iterations similar as ssGBLUP
- Reduction of #iter. by up to a factor 6!

Conclusions

- ssSNPBLUP - PCG: larger eigenvalues
  - ➔ Larger condition number
- ssSNPBLUP - Deflated PCG
  - Treats the largest unfavourable eigenvalues
  - ➔ Smaller condition number
  - ➔ Faster convergence (similar to ssGBLUP)
- Similar pattern on large and multivariate ssSNPBLUP

Thank you!