

A second-level diagonal preconditioner for single-step SNPBLUP

Vandenplas, Jeremie; Calus, Mario P.L.; Eding, Herwin; Vuik, Cornelis

DOI

[10.1186/s12711-019-0472-8](https://doi.org/10.1186/s12711-019-0472-8)

Publication date

2019

Document Version

Final published version

Published in

Genetics Selection Evolution

Citation (APA)

Vandenplas, J., Calus, M. P. L., Eding, H., & Vuik, C. (2019). A second-level diagonal preconditioner for single-step SNPBLUP. *Genetics Selection Evolution*, 51, 1-16. Article 30. <https://doi.org/10.1186/s12711-019-0472-8>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Open Access



A second-level diagonal preconditioner for single-step SNPBLUP

Jeremie Vandenplas^{1*} , Mario P. L. Calus¹, Herwin Eding² and Cornelis Vuik³

Abstract

Background: The preconditioned conjugate gradient (PCG) method is an iterative solver of linear equations systems commonly used in animal breeding. However, the PCG method has been shown to encounter convergence issues when applied to single-step single nucleotide polymorphism BLUP (ssSNPBLUP) models. Recently, we proposed a deflated PCG (DPCG) method for solving ssSNPBLUP efficiently. The DPCG method introduces a second-level preconditioner that annihilates the effect of the largest unfavourable eigenvalues of the ssSNPBLUP preconditioned coefficient matrix on the convergence of the iterative solver. While it solves the convergence issues of ssSNPBLUP, the DPCG method requires substantial additional computations, in comparison to the PCG method. Accordingly, the aim of this study was to develop a second-level preconditioner that decreases the largest eigenvalues of the ssSNPBLUP preconditioned coefficient matrix at a lower cost than the DPCG method, in addition to comparing its performance to the (D)PCG methods applied to two different ssSNPBLUP models.

Results: Based on the properties of the ssSNPBLUP preconditioned coefficient matrix, we proposed a second-level diagonal preconditioner that decreases the largest eigenvalues of the ssSNPBLUP preconditioned coefficient matrix under some conditions. This proposed second-level preconditioner is easy to implement in current software and does not result in additional computing costs as it can be combined with the commonly used (block-)diagonal preconditioner. Tested on two different datasets and with two different ssSNPBLUP models, the second-level diagonal preconditioner led to a decrease of the largest eigenvalues and the condition number of the preconditioned coefficient matrices. It resulted in an improvement of the convergence pattern of the iterative solver. For the largest dataset, the convergence of the PCG method with the proposed second-level diagonal preconditioner was slower than the DPCG method, but it performed better than the DPCG method in terms of total computing time.

Conclusions: The proposed second-level diagonal preconditioner can improve the convergence of the (D)PCG methods applied to two ssSNPBLUP models. Based on our results, the PCG method combined with the proposed second-level diagonal preconditioner seems to be more efficient than the DPCG method in solving ssSNPBLUP. However, the optimal combination of ssSNPBLUP and solver will most likely be situation-dependent.

Background

Since its introduction in the late 1990s [1], the preconditioned conjugate gradient (PCG) method has been the method of choice to solve breeding value estimation models in animal breeding. Likewise, the systems of linear equations of the different single-step single nucleotide polymorphism BLUP (ssSNPBLUP) models

are usually solved with the PCG method with a diagonal (also called Jacobi) or block-diagonal preconditioner [2–4]. Several studies [3–6] observed that the PCG method with such a preconditioner applied to ssSNPBLUP is associated with slower convergence. By investigating the reasons for these convergence issues, Vandenplas et al. [4] observed that the largest eigenvalues of the preconditioned coefficient matrix of ssSNPBLUP proposed by Mantysaari and Strandén [7], hereafter referred to as ssSNPBLUP_MS, resulted from the presence of the equations for single nucleotide polymorphism (SNP) effects. In their study, applying a deflated PCG (DPCG) method

*Correspondence: jeremie.vandenplas@wur.nl

¹ Animal Breeding and Genomics, Wageningen UR, P.O. 338, 6700 AH Wageningen, The Netherlands

Full list of author information is available at the end of the article



to ssSNPBLUP_MS solved the convergence issues [4]. In comparison to the PCG method, the DPCG method introduces a second-level preconditioner that annihilates the effect of the largest eigenvalues of the preconditioned coefficient matrix of ssSNPBLUP_MS on the convergence of the iterative solver. After deflation, the largest eigenvalues of the ssSNPBLUP_MS preconditioned deflated coefficient matrix were reduced and close to those of single-step genomic BLUP (ssGBLUP). As a result the associated convergence patterns of ssSNPBLUP were, at least, similar to those of ssGBLUP [4].

While it solves the convergence issues associated with ssSNPBLUP, the DPCG method requires the computation and storage of the so-called Galerkin matrix, which is a dense matrix that could be computationally expensive for very large evaluations and that requires some effort to be implemented in existing software. In addition, as implemented in Vandenplas et al. [4], each iteration of the DPCG method requires two multiplications of the coefficient matrix by a vector, instead of one multiplication for the PCG method. As a result, computing time per iteration with the DPCG method is roughly twice as long as with the PCG method. Accordingly, it is of interest to develop a second-level preconditioner that would reduce the largest eigenvalues of the preconditioned coefficient matrix of ssSNPBLUP at a lower cost than the DPCG method. As such, the aim of this study was to develop a second-level preconditioner that would decrease the unfavourable largest eigenvalues of the preconditioned coefficient matrix of ssSNPBLUP and to compare its performance to the DPCG method. The performance of the proposed second-level preconditioner was tested for two different ssSNPBLUP models.

Methods

Data

The two datasets used in this study, hereafter referred to as the reduced and field datasets, were provided by CRV BV (The Netherlands) and are the same as in Vandenplas et al. [4], in which these two datasets are described in detail.

Briefly, for the reduced dataset, the data file included 61,592 ovum pick-up sessions from 4109 animals and the pedigree included 37,021 animals. The 50K SNP genotypes of 6169 animals without phenotypes were available. A total of 9994 segregating SNPs with a minor allele frequency higher than or equal to 0.01 were randomly sampled from the 50K SNP genotypes. The number of SNPs was limited to 9994 to facilitate the computation

and the analysis of the left-hand side of the mixed model equations. The univariate mixed model included random effects (additive genetic, permanent environmental and residual), fixed co-variables (heterosis and recombination) and fixed cross-classified effects (herd-year, year-month, parity, age in months, technician, assistant, interval, gestation, session and protocol) [8].

For the field dataset, the data file included 3,882,772 records with a single record per animal. The pedigree included 6,130,519 animals. The genotypes, including 37,995 segregating SNPs, of 15,205 animals without phenotypes and of 75,758 animals with phenotypes were available. The four-trait mixed model included random effects (additive genetic and residual), fixed co-variables (heterosis and recombination) and fixed cross-classified effects (herd x year x season at classification, age at classification, lactation stage at classification, milk yield and month of calving) [9, 10].

Single-step SNPBLUP models

In this study, we investigated two ssSNPBLUP linear equations systems. The first system was proposed by Mantysaari and Strandén [7] (ssSNPBLUP_MS). This system was also investigated in Vandenplas et al. [4]. The standard multivariate model associated with the ssSNPBLUP_MS system of equations can be written as:

$$y = X\beta + \begin{bmatrix} W_n & 0 & 0 \\ 0 & W_g & W_g M_z \end{bmatrix} \begin{bmatrix} u_n \\ a_g \\ g \end{bmatrix} + e$$

where the subscripts *g* and *n* refer to genotyped and non-genotyped animals, respectively, *y* is the vector of records, β is the vector of fixed effects, u_n is the vector of additive genetic effects for non-genotyped animals, a_g is the vector of residual polygenic effects for genotyped animals, g is the vector of SNP effects and *e* is the vector of residuals. The matrices *X*, W_n and W_g are incidence matrices relating records to their corresponding effects. The matrix M_z is equal to $M_z = I_t \otimes Z$, with I_t being an identity matrix with size equal to the number of traits *t* and the matrix *Z* containing the SNP genotypes (coded as 0 for one homozygous genotype, 1 for the heterozygous genotype, or 2 for the alternate homozygous genotype) centred by their observed means.

The system of linear equations for multivariate ssSNPBLUP_MS can be written as follows:

$$C_{MS}x_{MS} = b_{MS}$$

where

$$C_{MS} = \begin{bmatrix} X'R^{-1}X & X'_nR_n^{-1}W_n & X'_gR_g^{-1}W_g & X'_gR_g^{-1}W_gM_z \\ W'_nR_n^{-1}X_n & W'_nR_n^{-1}W_n + \Sigma_{MS}^{11} & \Sigma_{MS}^{12} & \Sigma_{MS}^{13} \\ W'_gR_g^{-1}X_g & \Sigma_{MS}^{21} & W'_gR_g^{-1}W_g + \Sigma_{MS}^{22} & W'_gR_g^{-1}W_gM_z + \Sigma_{MS}^{23} \\ M'_zW'_gR_g^{-1}X_g & \Sigma_{MS}^{31} & M'_zW'_gR_g^{-1}W_g + \Sigma_{MS}^{32} & M'_zW'_gR_g^{-1}W_gM_z + \Sigma_{MS}^{33} \end{bmatrix}$$

is a symmetric positive (semi-)definite coefficient matrix,

$$x_{MS} = \begin{bmatrix} \hat{\beta} \\ \hat{u}_n \\ \hat{a}_g \\ \hat{g} \end{bmatrix} \text{ is the vector of solutions, and}$$

$$b_{MS} = \begin{bmatrix} X'R^{-1}y \\ W'_nR_n^{-1}y_n \\ W'_gR_g^{-1}y_g \\ M'_zW'_gR_g^{-1}y_g \end{bmatrix}$$

is the right-hand side with

$$R^{-1} = \begin{bmatrix} R_n^{-1} & 0 \\ 0 & R_g^{-1} \end{bmatrix}$$

being the inverse of the residual (co)variance structure matrix. The matrix Σ_{MS}^{-1} is the inverse of the covariance matrix associated with u_n , a_g and g , and is equal to

$$\Sigma_{MS}^{-1} = \begin{bmatrix} \Sigma_{MS}^{11} & \Sigma_{MS}^{12} & \Sigma_{MS}^{13} \\ \Sigma_{MS}^{21} & \Sigma_{MS}^{22} & \Sigma_{MS}^{23} \\ \Sigma_{MS}^{31} & \Sigma_{MS}^{32} & \Sigma_{MS}^{33} \end{bmatrix} \\ = G_0^{-1} \otimes \begin{bmatrix} A^{nn} & A^{ng} & A^{ng}Z \\ A^{gn} & \frac{1}{w}A^{gg} + \left(1 - \frac{1}{w}\right)Q & QZ \\ Z'A^{gn} & Z'Q & Z'QZ + \frac{m}{1-w}I \end{bmatrix}.$$

The matrix Q is equal to $Q = A^{gn}(A^{nn})^{-1}A^{ng}$, with

$$A^{-1} = \begin{bmatrix} A^{nn} & A^{ng} \\ A^{gn} & A^{gg} \end{bmatrix}$$

being the inverse of the pedigree relationship matrix. The parameter w is the proportion of variance (due to additive genetic effects) considered as residual polygenic effects and $m = 2 \sum p_o(1 - p_o)$ with p_o being the allele frequency of the o th SNP.

The second system of linear equations investigated in this study is the system of equations proposed by Gengler et al. [11] and Liu et al. [5], hereafter referred to as ssSNPBLUP_Liu. The system of linear equations for a multivariate ssSNPBLUP_Liu can be written as follows:

$$C_L x_L = b_L$$

where

$$C_L = \begin{bmatrix} X'R^{-1}X & X'_nR_n^{-1}W_n & X'_gR_g^{-1}W_g & 0 \\ W'_nR_n^{-1}X_n & W'_nR_n^{-1}W_n + \Sigma_L^{11} & \Sigma_L^{12} & \Sigma_L^{13} \\ W'_gR_g^{-1}X_g & \Sigma_L^{21} & W'_gR_g^{-1}W_g + \Sigma_L^{22} & \Sigma_L^{23} \\ 0 & \Sigma_L^{31} & \Sigma_L^{32} & \Sigma_L^{33} \end{bmatrix},$$

$$x_L = \begin{bmatrix} \hat{\beta} \\ \hat{u}_n \\ \hat{a}_g \\ \hat{g} \end{bmatrix},$$

and

$$b_L = \begin{bmatrix} X'R^{-1}y \\ W'_nR_n^{-1}y_n \\ W'_gR_g^{-1}y_g \\ 0 \end{bmatrix}.$$

The matrix Σ_L^{-1} is equal to

$$\Sigma_L^{-1} = \begin{bmatrix} \Sigma_L^{11} & \Sigma_L^{12} & \Sigma_L^{13} \\ \Sigma_L^{21} & \Sigma_L^{22} & \Sigma_L^{23} \\ \Sigma_L^{31} & \Sigma_L^{32} & \Sigma_L^{33} \end{bmatrix} \\ = G_0^{-1} \otimes \begin{bmatrix} A^{nn} & A^{ng} & 0 \\ A^{gn} & \frac{1}{w}A^{gg} + \left(1 - \frac{1}{w}\right)Q & -\frac{1}{w}A^{gg^{-1}}Z \\ 0 & -\frac{1}{w}Z'A_{gg}^{-1} & \frac{1}{w}Z'A_{gg}^{-1}Z + \frac{m}{1-w}I \end{bmatrix}$$

with $A_{gg}^{-1} = A^{gg} - Q$.

It is worth noting that the absorption of the equations associated with \hat{g} of ssSNPBLUP_Liu results in the mixed model equations of single-step genomic BLUP (ssGBLUP) for which the inverse of the genomic relationship matrix is calculated using the Woodbury formula [12]. Several studies (e.g., [13–15]) investigated the possibility of using specific knowledge of a priori variances to weight differently some SNPs in ssGBLUP. Such approaches are difficult to extend to multivariate ssGBLUP, while they can be easily applied in ssSNPBLUP by replacing the matrix $G_0^{-1} \otimes \frac{m}{1-w}I$ by a symmetric positive definite matrix B that contains SNP-specific (co)variances obtained by, e.g., Bayesian regression [5].

In the following, matrix C will refer to either C_{MS} or C_L (and similarly for the vectors x and b). In addition, the matrices C_{MS} and C_L have the same structure, and both can be partitioned between the equations associated with SNP effects (S) and the equations associated with the other effects (O), as follows:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{OO} & \mathbf{C}_{OS} \\ \mathbf{C}_{SO} & \mathbf{C}_{SS} \end{bmatrix}.$$

From this partition, it follows that $\mathbf{C}_{MS_{OO}} = \mathbf{C}_{L_{OO}}$ and that \mathbf{C}_{SO} , \mathbf{C}_{OS} , and \mathbf{C}_{SS} are dense matrices.

The PCG method

The PCG method is an iterative method that uses successive approximations to obtain more accurate solutions for a linear system at each iteration step [16]. The preconditioned systems of the linear equations of ssSNPBLUP_MS and of ssSNPBLUP_Liu have the form:

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

where \mathbf{M} is a (block-)diagonal preconditioner.

In this study, the (block-)diagonal preconditioner \mathbf{M} is defined as:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{ff} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{rr} \end{bmatrix} = \begin{bmatrix} \text{diag}(\mathbf{C}_{ff}) & \mathbf{0} \\ \mathbf{0} & \text{block_diag}(\mathbf{C}_{rr}) \end{bmatrix}$$

where the subscripts f and r refer to the equations associated with fixed and random effects, respectively and $\text{block_diag}(\mathbf{C}_{rr})$ is a block-diagonal matrix with blocks corresponding to equations for different traits within a level (e.g., an animal).

After k iterations of the PCG method applied to the Eq. (1), the error is bounded by [16, 17]:

$$\|\mathbf{x} - \hat{\mathbf{x}}_k\|_{\mathbf{C}} \leq 2\|\mathbf{x} - \hat{\mathbf{x}}_0\|_{\mathbf{C}} \left(\frac{\sqrt{\kappa(\mathbf{C}_M)} - 1}{\sqrt{\kappa(\mathbf{C}_M)} + 1} \right)^k$$

where $\mathbf{C}_M = \mathbf{M}^{-1}\mathbf{C}$, $\|\mathbf{x}\|_{\mathbf{C}}$ is the \mathbf{C} -norm of \mathbf{x} , defined as $\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x}}$, and $\kappa(\mathbf{C}_M)$ is the effective spectral condition number of \mathbf{C}_M , that is defined as $\frac{\lambda_{\max}(\mathbf{C}_M)}{\lambda_{\min}(\mathbf{C}_M)}$ with $\lambda_{\max}(\mathbf{C}_M)$ ($\lambda_{\min}(\mathbf{C}_M)$) being the largest (smallest) non-zero eigenvalue of \mathbf{C}_M .

The deflated PCG method

Vandenplas et al. [4] showed that the largest eigenvalues of the ssSNPBLUP_MS preconditioned coefficient matrix \mathbf{C}_M were larger than those of the ssGBLUP preconditioned coefficient matrix, while the smallest eigenvalues were similar. This resulted in larger effective condition numbers $\kappa(\mathbf{C}_M)$ and convergence issues for ssSNPBLUP_MS. As applied by Vandenplas et al. [4], the DPCG method annihilates the largest unfavourable eigenvalues of the ssSNPBLUP_MS preconditioned coefficient matrix \mathbf{C}_M , which resulted in effective condition numbers and convergence patterns of ssSNPBLUP_MS similar to those of ssGBLUP solved with the PCG method. The preconditioned deflated linear systems of ssSNPBLUP_MS and of ssSNPBLUP_Liu mixed model equations have the form:

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

where \mathbf{P} is a second-level preconditioner, called the deflation matrix, equal to $\mathbf{P} = \mathbf{I} - \mathbf{C}\mathbf{Z}_d\mathbf{E}^{-1}\mathbf{Z}'_d$, with the matrix \mathbf{Z}_d being the deflation-subspace matrix as defined in Vandenplas et al. [4] and $\mathbf{E} = \mathbf{Z}'_d\mathbf{C}\mathbf{Z}_d$ being the Galerkin matrix.

A second-level diagonal preconditioner

The DPCG method requires the computation and the storage of the Galerkin matrix \mathbf{E} , which is computationally expensive for very large evaluations [4]. Furthermore, as implemented in Vandenplas et al. [4], each iteration of the DPCG method requires two multiplications of the coefficient matrix \mathbf{C} by a vector, instead of one multiplication for the PCG method. Here, our aim is to develop another second-level preconditioner that decreases the largest eigenvalues of the preconditioned coefficient matrix \mathbf{C}_M at a lower cost than the DPCG method and results in smaller effective condition numbers and better convergence patterns.

To achieve this aim, we introduce a second-level diagonal preconditioner defined as:

$$\mathbf{D} = \begin{bmatrix} k_O\mathbf{I}_{OO} & \mathbf{0} \\ \mathbf{0} & k_S\mathbf{I}_{SS} \end{bmatrix} = k_O \begin{bmatrix} \mathbf{I}_{OO} & \mathbf{0} \\ \mathbf{0} & \frac{k_S}{k_O}\mathbf{I}_{SS} \end{bmatrix} = k_O\tilde{\mathbf{D}}$$

where \mathbf{I}_{OO} is an identity matrix of size equal to the number of equations that are not associated with SNP effects, \mathbf{I}_{SS} is an identity matrix of size equal to the number of equations that are associated with SNP effects, k_O and k_S are real positive numbers and $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{I}_{OO} & \mathbf{0} \\ \mathbf{0} & \frac{k_S}{k_O}\mathbf{I}_{SS} \end{bmatrix}$. Possible values for k_O and k_S are discussed below.

Therefore, the preconditioned system of Eq. (1) is modified as follows:

$$\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{C}\mathbf{x} = \mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{b}. \tag{2}$$

Hereafter, we show that the proposed second-level preconditioner \mathbf{D} applied to ssSNPBLUP systems of equations results in smaller effective condition numbers by decreasing the largest eigenvalues of the preconditioned coefficient matrices. For simplicity, the symmetric preconditioned coefficient matrix $\mathbf{D}^{-1/2}\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{-1/2}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ with $\tilde{\mathbf{C}} = \mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{-1/2}$ is used instead of $\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{C}$. Indeed, these two matrices have the same spectrum, i.e., the same set of eigenvalues. In addition, the effective condition number of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$, $\kappa(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2})$, is equal to the effective condition number of $\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}$, $\kappa(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2})$, because:

$$\begin{aligned} \kappa\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right) &= \frac{\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)}{\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)} \\ &= \frac{\lambda_{\max}\left(k_O^{-1}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)}{\lambda_{\min}\left(k_O^{-1}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)} \\ &= \frac{\lambda_{\max}\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)}{\lambda_{\min}\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)} \\ &= \kappa\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right) \end{aligned}$$

with $\lambda_{\min}\left(k_O^{-1}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right) = k_O^{-1}\lambda_{\min}\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)$, and $\lambda_{\max}\left(k_O^{-1}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right) = k_O^{-1}\lambda_{\max}\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{C}}\tilde{\mathbf{D}}^{-1/2}\right)$.

The result is that $\kappa\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ depends only on $\tilde{\mathbf{D}}$ and therefore only on the k_O/k_S ratio.

Regarding the largest eigenvalues of the preconditioned coefficient matrix $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ or equivalently of $\mathbf{D}^{-1}\mathbf{M}^{-1}\mathbf{C}$, the effect of the second-level preconditioner \mathbf{D} on $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ can be analysed using the Gershgorin circle theorem [18]. From this theorem, it follows that the largest eigenvalue of the preconditioned coefficient matrix $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ is bounded by, for all i th and j th equations:

$$\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right) \leq \max_i \left\{ \mathbf{D}_{ii}^{-1/2}\tilde{\mathbf{C}}_{ii}\mathbf{D}_{ii}^{-1/2} + \sum_{j \neq i} |\mathbf{D}_{ii}^{-1/2}\tilde{\mathbf{C}}_{ij}\mathbf{D}_{jj}^{-1/2}| \right\}. \quad (3)$$

Partitioned between the equations associated with SNP effects (S) and with the other effects (O), it follows from Eq. (3) that $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ has the following lower and upper bounds (see Additional file 1 for the derivation):

$$k_O^{-1}\lambda_{\max}\left(\tilde{\mathbf{C}}_{OO}\right) \leq \lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right) \leq k_O^{-1}\max\{a, b\}, \quad (4)$$

with $a = \max_k \left\{ \tilde{\mathbf{C}}_{OO_{kk}} + \sum_{j \neq k} |\tilde{\mathbf{C}}_{OO_{kj}}| + \sqrt{\frac{k_O}{k_S}} \sum_{j \neq k} |\tilde{\mathbf{C}}_{OS_{kj}}| \right\}$, $b = \max_l \left\{ \frac{k_O}{k_S} \tilde{\mathbf{C}}_{SS_{ll}} + \frac{k_O}{k_S} \sum_{j \neq l} |\tilde{\mathbf{C}}_{SS_{lj}}| + \sqrt{\frac{k_O}{k_S}} \sum_{j \neq l} |\tilde{\mathbf{C}}_{SO_{lj}}| \right\}$ and k and l referring to the equations not associated with and associated with SNP effects, respectively.

Therefore, for a fixed value of k_O , the upper bound of $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ will decrease with decreasing k_O/k_S ratios, up to the lowest upper bound $k_O^{-1}\max_k \left\{ \tilde{\mathbf{C}}_{OO_{kk}} + \sum_{j \neq k} |\tilde{\mathbf{C}}_{OO_{kj}}| \right\}$, that is the upper bound of $k_O^{-1}\lambda_{\max}\left(\tilde{\mathbf{C}}_{OO}\right)$.

Nevertheless, decreasing the largest eigenvalue does not (necessarily) mean decreasing the effective condition number $\kappa\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$, because $\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ could decrease at the same rate as, or faster than $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ leading to constant or larger $\kappa\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$. As such, it is required that $\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ decreases at a lower rate, remains constant, or even increases, when $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ decreases with decreasing k_O/k_S ratios. This would be achieved if $\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ is independent of k_S . Hereafter, we formulate a sufficient condition to ensure that $\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right) = k_O^{-1}\lambda_{\min}\left(\tilde{\mathbf{C}}\right)$ for any k_O/k_S ratio.

Let the matrix $\tilde{\mathbf{V}}$ be a matrix containing (column-wise) all the eigenvectors of $\tilde{\mathbf{C}}$ sorted according to the ascending order of their associated eigenvalues. The set of eigenvalues of $\tilde{\mathbf{C}}$ sorted according to their ascending order is hereafter called spectrum of $\tilde{\mathbf{C}}$. The matrix $\tilde{\mathbf{V}}$ can be partitioned into a matrix $\tilde{\mathbf{V}}_1$ storing eigenvectors associated with eigenvalues at the left-hand side of the spectrum (that includes the smallest eigenvalues) of $\tilde{\mathbf{C}}$ and a matrix $\tilde{\mathbf{V}}_2$ storing eigenvectors at the right-hand side of the spectrum (that includes the largest eigenvalues) of $\tilde{\mathbf{C}}$, and between equations associated with SNP effects or not, as follows:

$$\tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_1 \tilde{\mathbf{V}}_2] = \begin{bmatrix} \tilde{\mathbf{V}}_{O1} & \tilde{\mathbf{V}}_{O2} \\ \tilde{\mathbf{V}}_{S1} & \tilde{\mathbf{V}}_{S2} \end{bmatrix}.$$

A sufficient condition to ensure that $\lambda_{\min}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right) = k_O^{-1}\lambda_{\min}\left(\tilde{\mathbf{C}}\right)$ is that $\tilde{\mathbf{V}}_{S1} = \mathbf{0}$, $\tilde{\mathbf{V}}_{O2} = \mathbf{0}$ and that all eigenvalues associated with an eigenvector of $\tilde{\mathbf{V}}_2$ are equal to, or larger than $\frac{k_S}{k_O}\lambda_{\min}\left(\tilde{\mathbf{C}}\right)$ (see Additional file 2 for proof). Therefore, the effective condition numbers $\kappa\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ will decrease with decreasing k_O/k_S ratios until the largest eigenvalue $\lambda_{\max}\left(\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}\right)$ reaches its lower bound $k_O^{-1}\lambda_{\max}\left(\tilde{\mathbf{C}}_{OO}\right)$, as long as the sufficient condition is satisfied. In practice, the pattern of the matrix $\tilde{\mathbf{V}}$ will never be as required by the sufficient condition, because the submatrices $\tilde{\mathbf{C}}_{OS}$ and $\tilde{\mathbf{C}}_{SO}$ contain non-zero entries. However, this sufficient condition is helpful to formulate the expectation that convergence of the models will improve with decreasing k_O/k_S ratios up to a point that can either be identified from the analyses or by computing the eigenvalues of $\tilde{\mathbf{C}}$.

Analyses

Eigenvalues and eigenvectors of ssSNPBLUP_MS preconditioned coefficient matrices $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ with values of k_S from 1 to 10^5 (and $k_O = 1$) were computed for

the reduced dataset using the subroutine *dsyev* provided by Intel(R) Math Kernel Library (MKL) 11.3.2.

Using the matrix-free version of the software developed in Vandenplas et al. [4], the system of ssSNPB-LUP_MS and ssSNPB-LUP_Liu equations for the reduced and field datasets were solved with the PCG and DPCG methods together with the second-level preconditioner **D** for different values of k_S (with $k_O = 1$). The second-level preconditioner **D** was implemented by combining it with the preconditioner **M**, as $\tilde{\mathbf{M}} = \mathbf{DM}$. Accordingly, its implementation has no additional costs for an iteration of the PCG and DPCG methods. The DPCG method was applied with 5 SNP effects per subdomain [4]. To illustrate the effect of k_O , the system of ssSNPB-LUP_MS equations was also solved for the reduced dataset with the PCG method and different values of k_O (with $k_S = 1$). For both the PCG and DPCG methods, the iterative process stopped when the relative residual norm was smaller than 10^{-6} . For all systems, the smallest and largest eigenvalues that influence the convergence of the iterative methods were estimated using the Lanczos method based on information obtained from the (D)PCG method [16, 19, 20]. Effective condition numbers were computed from these estimates [17].

All real vectors and matrices were stored using double precision real numbers, except for the preconditioner, which was stored using single precision real numbers. All computations were performed on a computer with 528 GB and running RedHat 7.4 (x86_64) with an Intel Xeon E5-2667 (3.20 GHz) processor with 16 cores. The number of OpenMP threads was limited to 5 for both datasets. Time requirements are reported for the field dataset. All reported times are indicative, because they may have been influenced by other jobs running simultaneously on the computer.

Results

Reduced dataset

The spectra of the ssSNPB-LUP_MS preconditioned coefficient matrices $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ solved with the PCG method and with k_S values from 1 to 10^5 (and $k_O = 1$) are depicted in Fig. 1. It can be observed that the largest eigenvalues decreased with decreasing k_O/k_S ratios, up to $k_O/k_S = 10^{-2}$ (Fig. 1; Table 1). On the other side of the spectrum, a set of approximately 10,000 small eigenvalues that decrease with decreasing k_O/k_S ratios can be observed.

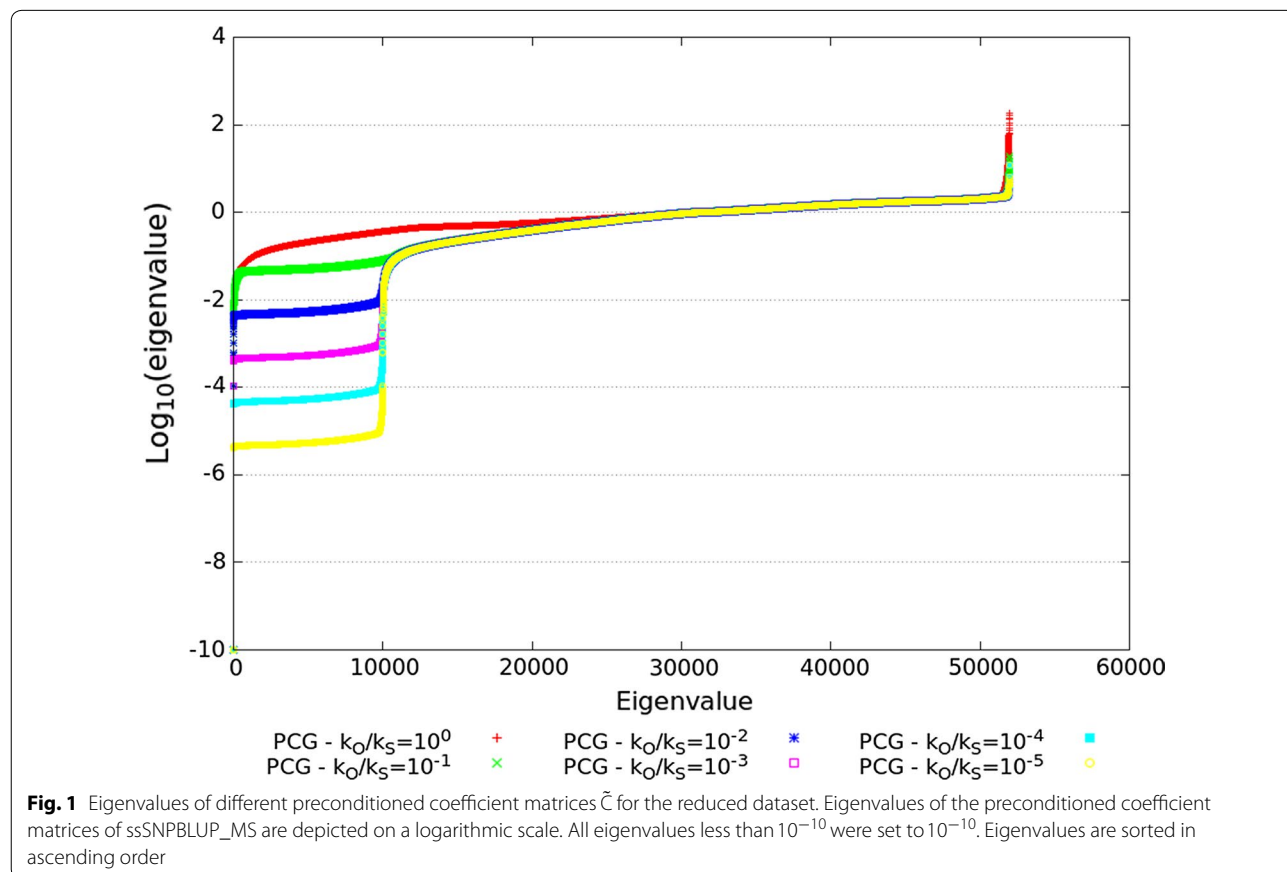


Table 1 Characteristics of preconditioned (deflated) coefficient matrices, and of PCG and DPCG methods for solving ssSNPBLUP applied to the reduced dataset

Model ^a	Method ^b	k_O^c	k_S^c	k_O/k_S	λ_{min}^d	λ_{max}^d	κ^e	N^f
MS	PCG	1	1	1	1.07×10^{-04}	1.81×10^2	1.70×10^6	1499
MS	PCG	1	2	0.5	1.07×10^{-04}	9.11×10^1	8.55×10^5	1103
MS	PCG	1	3.3	0.3	1.07×10^{-04}	5.51×10^1	5.17×10^5	862
MS	PCG	1	10^1	10^{-1}	1.07×10^{-04}	1.91×10^1	1.79×10^5	560
MS	PCG	1	10^2	10^{-2}	1.07×10^{-04}	1.19×10^1	1.12×10^5	417
MS	PCG	1	10^3	10^{-3}	1.06×10^{-04}	1.19×10^1	1.12×10^5	608
MS	PCG	1	10^4	10^{-4}	4.86×10^{-05}	1.19×10^1	2.45×10^5	1254
MS	PCG	1	10^5	10^{-5}	4.87×10^{-06}	1.19×10^1	2.45×10^6	2350
MS	PCG	10^{-1}	1	10^{-1}	1.07×10^{-03}	1.91×10^2	1.79×10^5	557
MS	PCG	10^{-2}	1	10^{-2}	1.07×10^{-02}	1.19×10^3	1.12×10^5	416
MS	PCG	10^{-3}	1	10^{-3}	1.06×10^{-01}	1.19×10^4	1.12×10^5	606
MS	PCG	10^{-4}	1	10^{-4}	4.86×10^{-01}	1.19×10^5	2.45×10^5	1254
MS	PCG	10^{-5}	1	10^{-5}	4.86×10^{-01}	1.19×10^6	2.45×10^6	2367
MS	DPCG (1)	1	1	1	1.09×10^{-04}	6.44	5.93×10^4	294
MS	DPCG (1)	1	10^5	10^{-5}	1.09×10^{-04}	6.44	5.92×10^4	293
MS	DPCG (5)	1	1	1	1.07×10^{-04}	6.44	6.03×10^4	342
MS	DPCG (5)	1	10^1	10^{-1}	1.07×10^{-04}	6.44	6.03×10^4	331
MS	DPCG (5)	1	10^2	10^{-2}	1.07×10^{-04}	6.44	6.04×10^4	385
MS	DPCG (5)	1	10^3	10^{-3}	1.06×10^{-04}	6.44	6.05×10^4	544
MS	DPCG (5)	1	10^4	10^{-4}	4.96×10^{-05}	6.44	1.30×10^5	961
MS	DPCG (5)	1	10^5	10^{-5}	4.95×10^{-06}	6.44	1.30×10^6	1456
Liu	PCG	1	1	1	1.06×10^{-04}	6.98×10^1	6.56×10^5	1401
Liu	PCG	1	10^1	10^{-1}	1.06×10^{-04}	1.19×10^1	1.12×10^5	561
Liu	PCG	1	10^2	10^{-2}	1.06×10^{-04}	1.19×10^1	1.12×10^5	563
Liu	PCG	1	10^3	10^{-3}	5.91×10^{-05}	1.19×10^1	2.02×10^5	1154
Liu	DPCG (5)	1	1	1	1.07×10^{-04}	6.44	6.05×10^4	419
Liu	DPCG (5)	1	10^1	10^{-1}	1.07×10^{-04}	6.44	6.05×10^4	399
Liu	DPCG (5)	1	10^2	10^{-2}	1.06×10^{-04}	6.44	6.05×10^4	520
Liu	DPCG (5)	1	10^3	10^{-3}	6.02×10^{-05}	6.44	1.07×10^5	1046

^aMS = ssSNPBLUP model proposed by Mantysaari and Strandén [7]; Liu = ssSNPBLUP model proposed by Liu et al. [5]

^bNumber of SNP effects per subdomain is within brackets

^cParameters used for the second-level preconditioner **D**

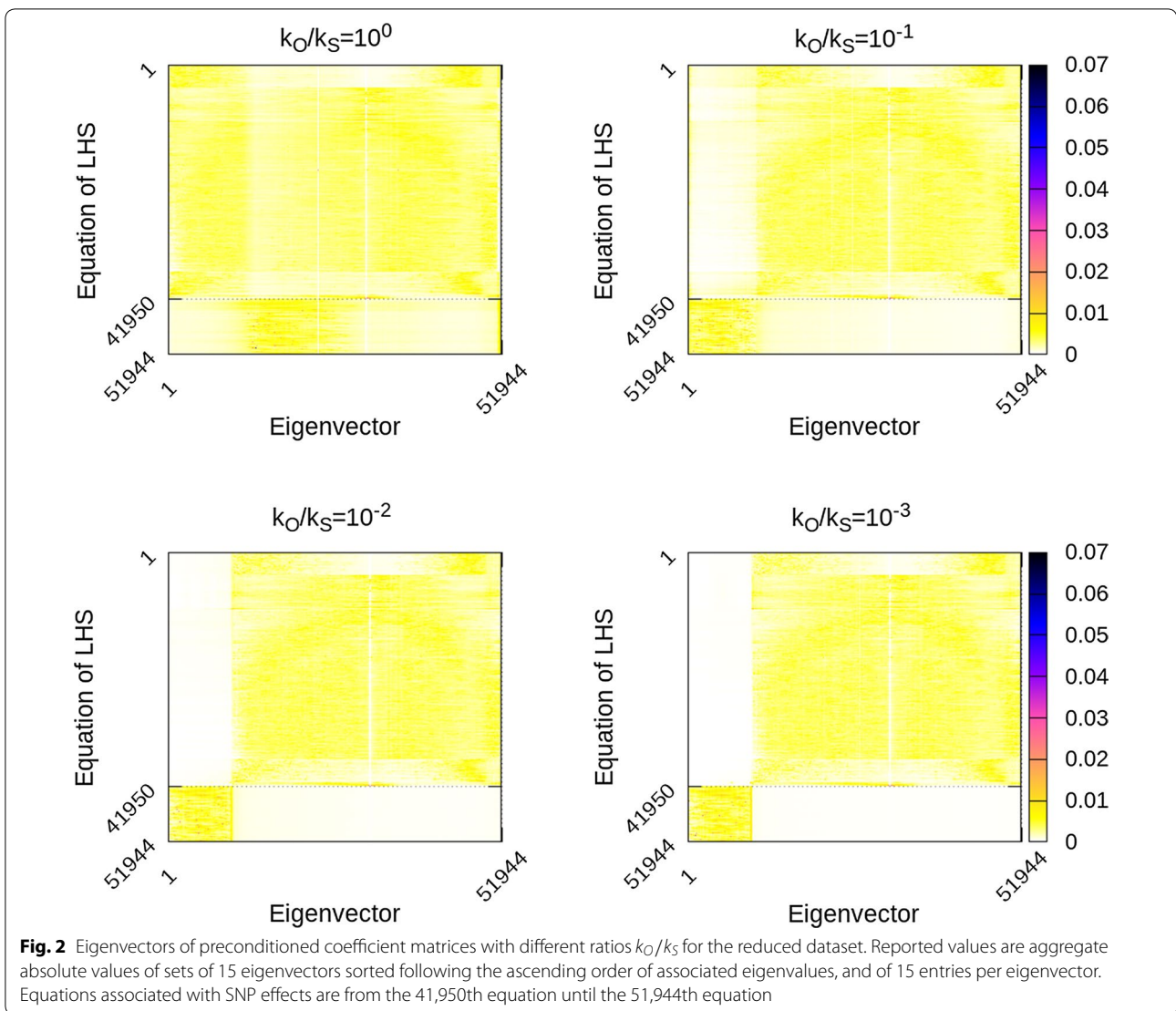
^dSmallest and largest eigenvalues of the preconditioned (deflated) coefficient matrix

^eCondition number of the preconditioned (deflated) coefficient matrix

^fNumber of iterations. A number of iterations equal to 10,000 means that the method failed to converge within 10,000 iterations

Figures 2, 3 and 4 depict all the eigenvectors of the ssSNPBLUP_MS preconditioned coefficient matrices $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ with different values of k_S (and $k_O = 1$). Non-zero eigenvector entries indicate an association of the eigenvalue (associated with this eigenvector) and the corresponding equations, while (almost) zero entries indicate no such (or a very weak) association. When $k_O/k_S = 1$, it can be observed that the smallest

eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ are mainly associated with the equations that are not associated with SNP effects. On the other side, with $k_O/k_S = 1$, the largest eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ are mainly associated with the equations that are associated with SNP effects (Figs. 2 and 3). Decreasing k_O/k_S ratios resulted in modifying the associations of the extremal eigenvalues (i.e. the smallest and largest eigenvalues) with the equations. Indeed,



decreasing k_O/k_S ratios resulted in the smallest eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ mainly associated with the equations that are associated with SNP effects, and in the largest eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ mainly associated with the equations that are not associated with SNP effects.

The extremal eigenvalues of the ssSNPBLUP_MS and ssSNPBLUP_Liu preconditioned (deflated) coefficient matrices, with various values for k_O and k_S , are in Table 1. For both ssSNPBLUP_MS and ssSNPBLUP_Liu solved with the PCG method, the largest eigenvalues of the preconditioned coefficient matrix decreased with decreasing k_O/k_S ratios to a lower value of 11.9 that was reached when $k_O/k_S = 10^{-2}$. In addition, for both models, the smallest eigenvalues remained constant with decreasing k_O/k_S ratios, until $k_O/k_S = 10^{-3}$ for ssSNPBLUP_MS and $k_O/k_S = 10^{-2}$ for ssSNPBLUP_Liu. Due to these

results, the effective condition numbers and the number of iterations to reach convergence were the smallest for $k_O/k_S = 10^{-2}$ for ssSNPBLUP_MS and for $k_O/k_S = 10^{-1}$ for ssSNPBLUP_Liu (Table 1; Figs. 5 and 6). In comparison to the PCG method without the second-level preconditioner (i.e., with $k_O = k_S = 1$), the number of iterations to reach convergence decreased by a factor of more than 3.5 for ssSNPBLUP_MS and by a factor of more than 2.4 for ssSNPBLUP_Liu. The minimum number of iterations to reach convergence with the PCG method was 417 for ssSNPBLUP_MS and 561 for ssSNPBLUP_Liu (Table 1; Figs. 5 and 6).

For the same k_O/k_S ratio, the extremal eigenvalues (i.e. the smallest and largest eigenvalues) of the different preconditioned coefficient matrices were proportional by a factor of k_O^{-1} (Table 1). Therefore, for the same k_O/k_S

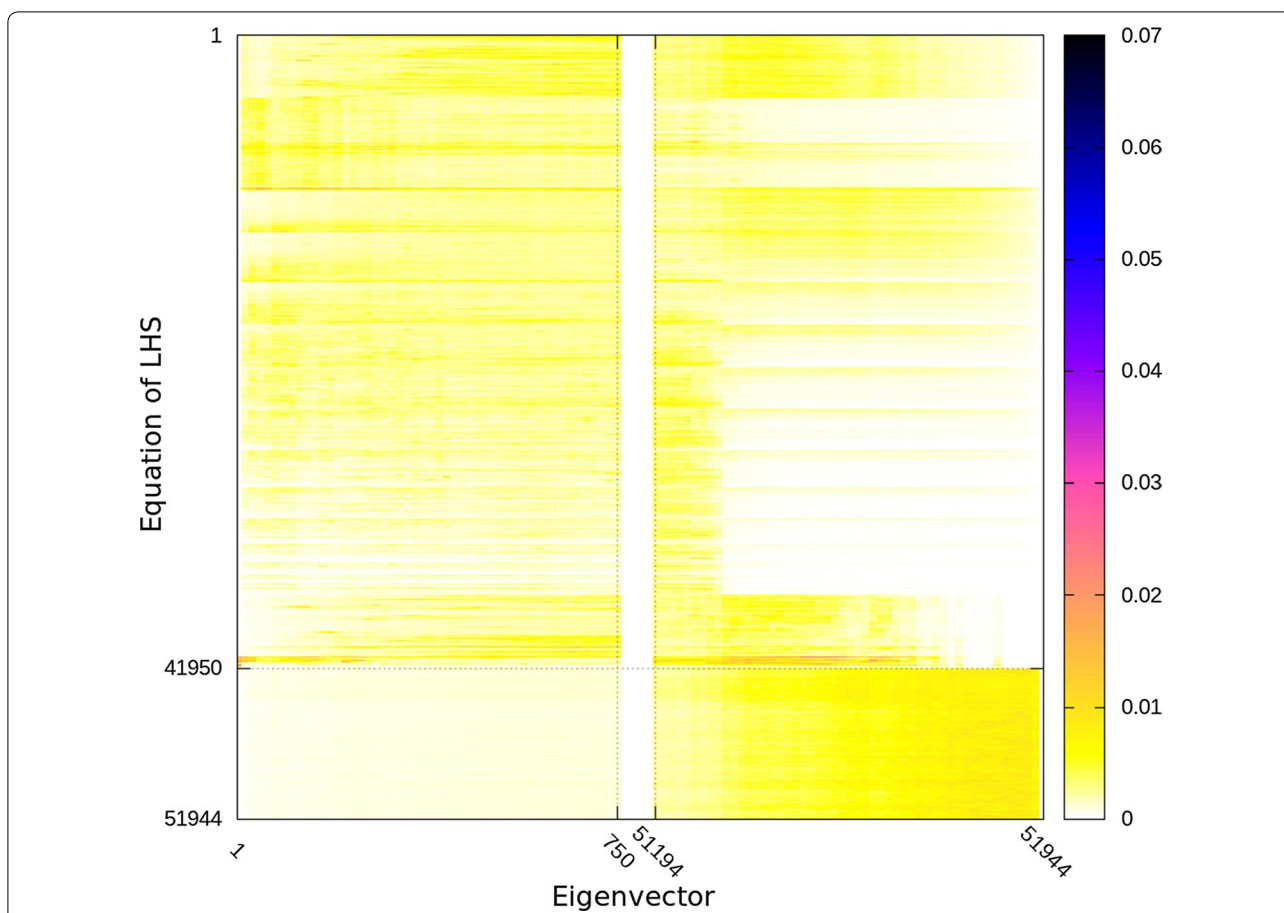


Fig. 3 Eigenvectors associated with the 750 smallest and largest eigenvalues of the preconditioned coefficient matrix with the ratio $k_O/k_S = 10^0$ for the reduced dataset. Reported values are aggregate absolute values of sets of 15 eigenvectors sorted following the ascending order of associated eigenvalues, and of 15 entries per eigenvector. Darker colors correspond to higher values. Equations associated with SNP effects are from the 41,950th equation until the 51,944th equation

ratio the effective condition numbers of the different preconditioned coefficient matrices and the associated numbers of equations to reach convergence were the same (Table 1). It is also worth noting that, for a fixed value of k_O , the largest eigenvalues decreased almost proportionally by a factor of k_S^{-1} with decreasing k_O/k_S ratios until they reached their lower bound (Table 1).

For both ssSNPBLUP_MS and ssSNPBLUP_Liu solved with the DPCG method and 5 SNPs per subdomain, the largest eigenvalues of the preconditioned deflated coefficient matrices remained constant (around 6.44) for all k_O/k_S ratios (Table 1). However, for both models, the smallest eigenvalues started to decrease for k_O/k_S ratios smaller than 10^{-3} (10^{-2}) for ssSNPBLUP_MS (ssSNPBLUP_Liu). These unfavourable decreases of the smaller eigenvalues with decreasing k_O/k_S ratios resulted in increasing the effective condition numbers and the number of iterations to reach convergence when the

second-level preconditioner **D** was applied with the DPCG method (Table 1; Figs. 5 and 6).

Field dataset

For the field dataset, regarding the extremal eigenvalues, the application of the second-level preconditioner **D** together with the PCG method led to a decrease of the largest eigenvalues of the preconditioned coefficient matrix from 1.8×10^3 for ssSNPBLUP_MS, and from 1.4×10^2 for ssSNPBLUP_Liu, to about 5. Ratios of k_O/k_S smaller than 10^{-3} for ssSNPBLUP_MS and smaller than 10^{-2} for ssSNPBLUP_Liu did not further change the largest eigenvalues (Table 2). For the DPCG method applied to ssSNPBLUP_MS, the largest eigenvalues of the preconditioned deflated coefficient matrices remained constant for all k_O/k_S ratios (Table 2). For the DPCG method applied to ssSNPBLUP_Liu, the largest eigenvalues of the preconditioned deflated coefficient matrices slightly decreased with $k_O/k_S = 10^{-1}$ and then remained

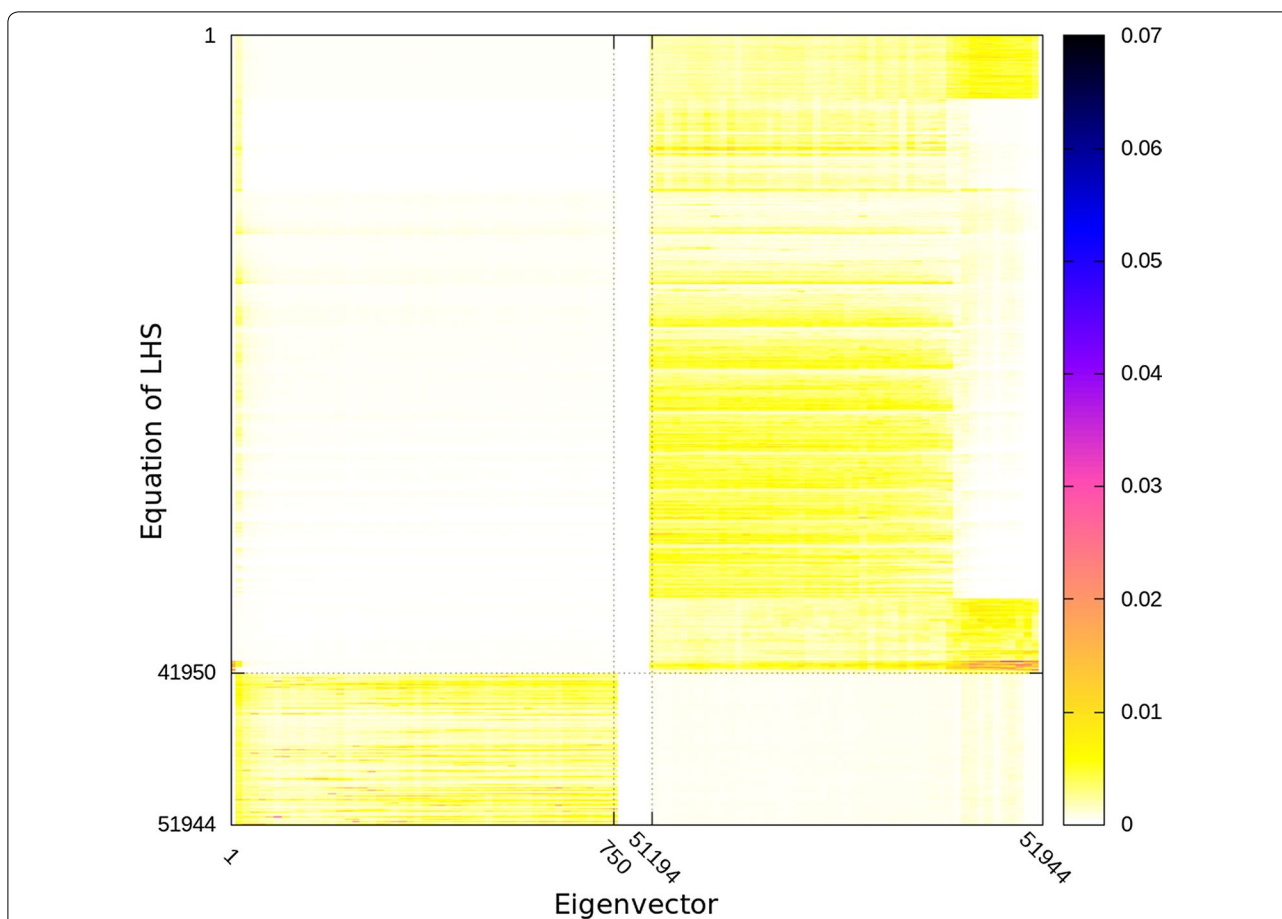


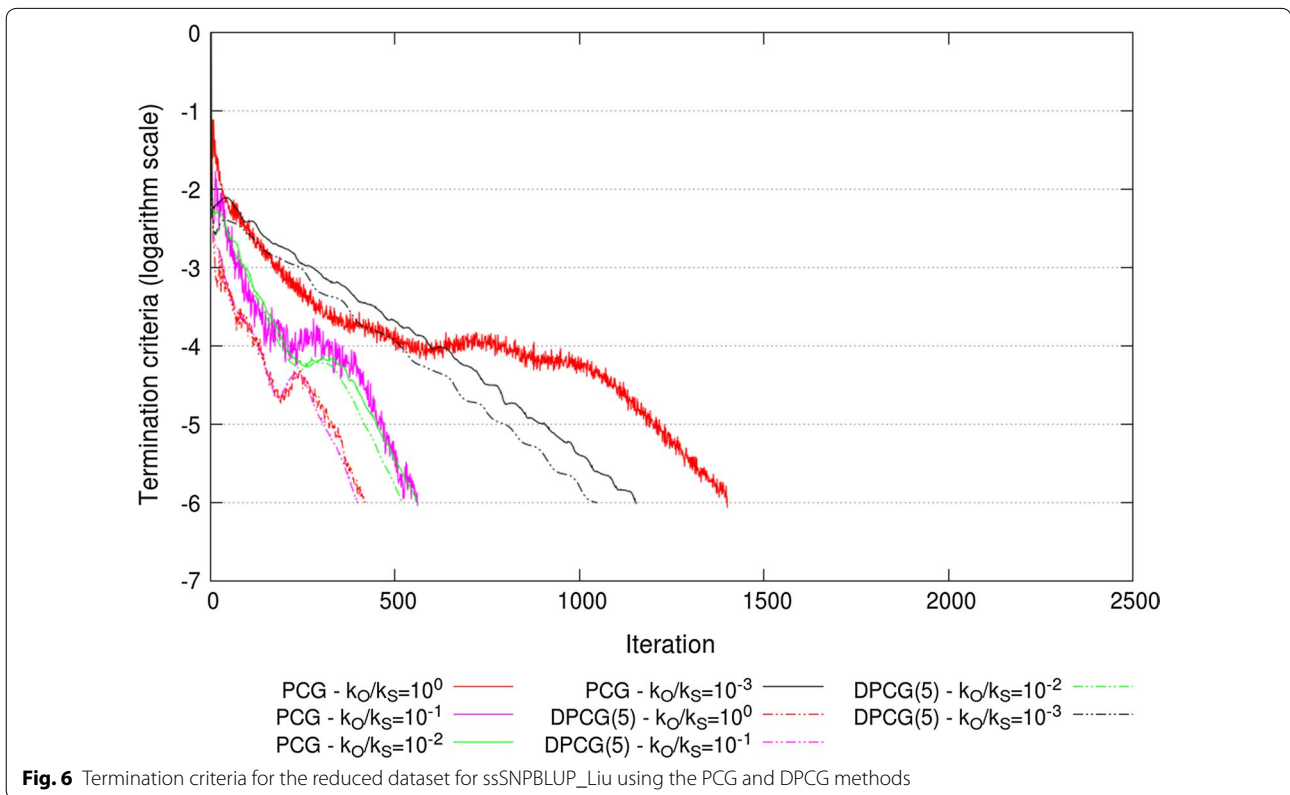
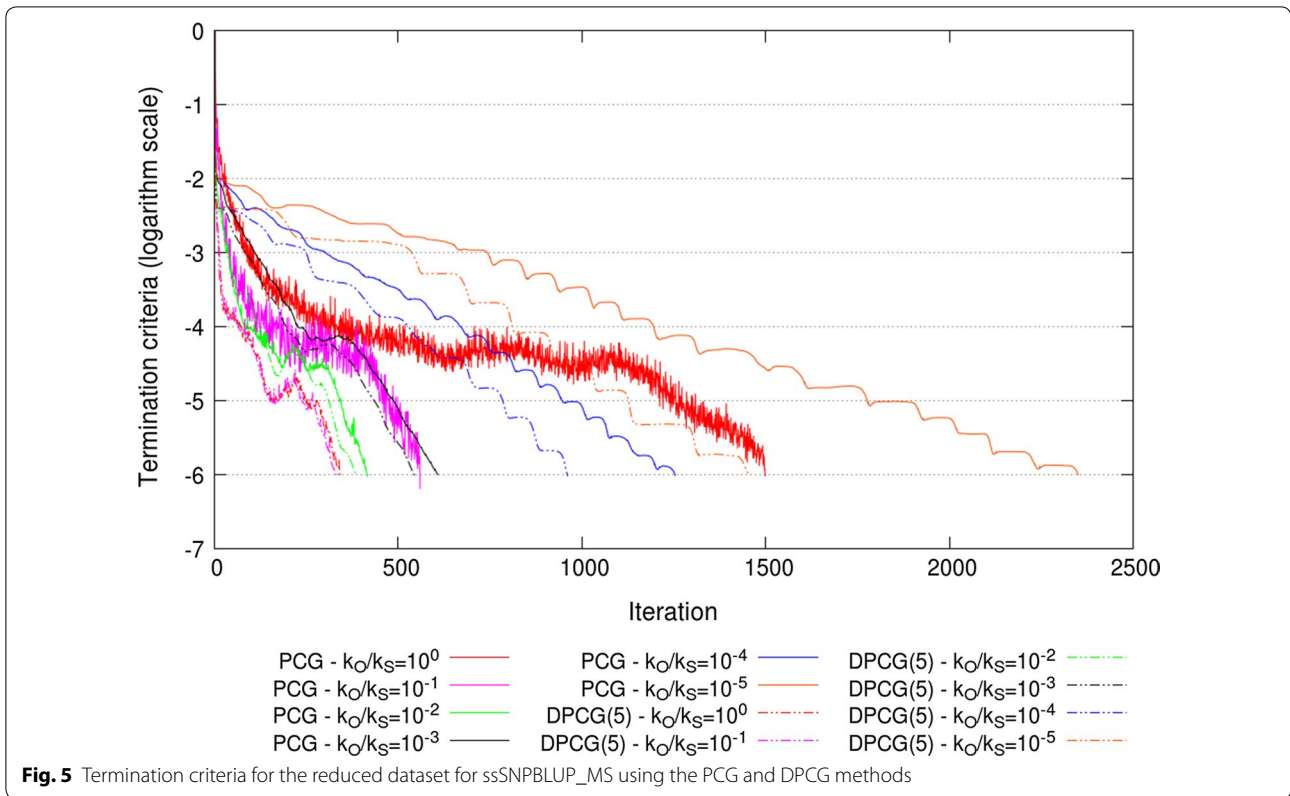
Fig. 4 Eigenvectors associated with the 750 smallest and largest eigenvalues of the preconditioned coefficient matrix with the ratio $k_O/k_S = 10^{-2}$ for the reduced dataset. Reported values are aggregate absolute values of sets of 15 eigenvectors sorted following the ascending order of associated eigenvalues, and of 15 entries per eigenvector. Darker colors correspond to higher values. Equations associated with SNP effects are from the 41,950th equation until the 51,944th equation

constant for all k_O/k_S ratios (Table 2). The application of the second-level preconditioner **D** with both the PCG and DPCG methods led to the smallest eigenvalues of the preconditioned (deflated) coefficient matrices decreasing with decreasing k_O/k_S ratios (Table 2).

These observed patterns of extremal eigenvalues resulted in an optimal $k_O/k_S = 10^{-3}$ ratio for the PCG method applied to ssSNPBLUP_MS and an optimal $k_O/k_S = 10^{-2}$ ratio for the PCG method applied to ssSNPBLUP_Liu, in terms of effective condition numbers and numbers of iterations to reach convergence (Table 2; Figs. 7 and 8). With these ratios, the PCG method converged within 3825 iterations for ssSNPBLUP_MS and within 2665 iterations for ssSNPBLUP_Liu, while the PCG method did not converge within 10,000 iterations for both models (Table 2; Figs. 7 and 8). For the DPCG method, the application of the second-level preconditioner **D** generally deteriorated the effective condition numbers and numbers of iterations to reach convergence,

for both ssSNPBLUP_MS and ssSNPBLUP_Liu. The DPCG method converged within 748 iterations for ssSNPBLUP_MS with $k_O/k_S = 1$ and within 2877 iterations for ssSNPBLUP_Liu with $k_O/k_S = 10^{-1}$ (Table 2; Figs. 7 and 8).

The total wall clock times of the iterative processes and for the complete processes (including I/O operations and computation of the preconditioners, and Galerkin matrices) for the PCG and DPCG methods are in Table 2. Across all combinations of systems of equations and solvers, the smallest wall clock time for the complete process was approximately 14,000 s for the PCG method with the second-level preconditioner **D** applied to ssSNPBLUP_Liu. Slightly greater wall clock times were needed for ssSNPBLUP_MS solved with the DPCG method (without the second-level preconditioner **D**). It is worth noting that the wall clock times needed for the computation of the inverse of the Galerkin matrix (E^{-1}) were



approximately 9700 s for ssSNPBLUP_MS and approximately 2500 s for ssSNPBLUP_Liu.

Discussion

In this study, we introduced a second-level diagonal preconditioner **D** that results in smaller effective condition numbers of the preconditioned (deflated) coefficient matrices and in improved convergence patterns for two different ssSNPBLUP mixed model equations. From the theory and based on the results, the use of the second-level preconditioner **D** results in improved effective condition numbers of the preconditioned (deflated) coefficient matrices of ssSNPBLUP by decreasing the largest eigenvalues, while the smallest eigenvalues remain constant, or decrease at a lower rate than the largest eigenvalues. In this section, we will discuss the following three points: (1) the influence of the second-level diagonal preconditioner **D** on the eigenvalues and associated eigenvectors of the preconditioned (deflated) coefficient matrices of ssSNPBLUP; (2) the application of the

second-level preconditioner in ssSNPBLUP evaluations; and (3) the possible application of the second-level preconditioner **D** to more complex ssSNPBLUP models and to models other than ssSNPBLUP.

Influence of D on the eigenvalues and associated eigenvectors

Applying the second-level preconditioner **D** with an optimal k_O/k_S ratio to the linear systems of ssSNPBLUP results in a decrease of the largest eigenvalues of the preconditioned (deflated) coefficient matrices of ssSNPBLUP. As observed by Vandenplas et al. [4] and in comparison with ssGBLUP, the largest eigenvalues that influence the convergence of the PCG method applied to ssSNPBLUP_MS were associated with SNP effects. The second-level preconditioner **D** allows a decrease of these largest eigenvalues by multiplying all entries of these SNP equations of the preconditioned coefficient matrices by a value proportional to k_O/k_S , as shown with the Gershgorin circle algorithm [18] [see Eq. (4)]. However,

Table 2 Characteristics of preconditioned (deflated) coefficient matrices, and of PCG and DPCG methods for solving ssSNPBLUP applied to the field dataset

Model ^a	Method	k_O/k_S^b	λ_{min}^c	λ_{max}^c	κ^d	N^e	Iterative time ^f	Time/iter. ^g	Total time ^h
MS	PCG	1	3.70×10^{-5}	1.75×10^3	4.74×10^7	10,000	44,808	4.5	46,081
MS	PCG	10^{-1}	1.18×10^{-5}	1.77×10^2	1.51×10^7	10,000	51,768	5.2	53,550
MS	PCG	10^{-2}	4.37×10^{-6}	1.95×10^1	4.45×10^6	6210	34,139	5.5	35,812
MS	PCG	10^{-3}	3.99×10^{-6}	5.08	1.27×10^6	3825	19,043	5.0	20,866
MS	PCG	10^{-4}	1.50×10^{-6}	5.07	3.37×10^6	7336	54,326	7.4	56,475
MS	DPCG	1	2.86×10^{-5}	4.77	1.67×10^5	748	6527	8.7	17,229
MS	DPCG	10^{-1}	1.41×10^{-5}	4.77	3.37×10^5	1211	11,864	9.8	22,947
MS	DPCG	10^{-2}	9.17×10^{-6}	4.77	5.20×10^5	1778	17,030	9.6	28,615
MS	DPCG	10^{-3}	7.50×10^{-6}	4.77	6.36×10^5	2569	23,676	9.2	35,497
Liu	PCG	1	7.38×10^{-6}	1.43×10^2	1.93×10^7	10,000	44,122	4.4	45,083
Liu	PCG	10^{-1}	3.66×10^{-6}	1.52×10^1	4.14×10^6	6049	31,085	5.1	32,018
Liu	PCG	10^{-2}	4.29×10^{-6}	5.07	1.18×10^6	2669	13,225	5.0	13,888
Liu	PCG	10^{-3}	3.51×10^{-6}	5.07	1.44×10^6	3606	20,578	5.7	21,458
Liu	PCG	10^{-4}	1.69×10^{-6}	5.07	3.00×10^6	7033	33,534	4.8	34,675
Liu	DPCG	1	5.40×10^{-6}	5.31	9.85×10^5	2877	22,791	7.9	26,521
Liu	DPCG	10^{-1}	6.91×10^{-6}	4.77	6.90×10^5	1628	14,231	8.7	18,049
Liu	DPCG	10^{-2}	5.23×10^{-6}	4.77	9.11×10^5	2234	23,244	10.4	28,057
Liu	DPCG	10^{-3}	4.31×10^{-6}	4.77	1.11×10^6	3106	34,950	11.3	39,603

^aMS = ssSNPBLUP model proposed by Mantysaari and Strandén [7]; Liu = ssSNPBLUP model proposed by Liu et al. [5];

^bParameters used for the second-level preconditioner;

^cSmallest and largest eigenvalues of the preconditioned (deflated) coefficient matrix;

^dCondition number of the preconditioned (deflated) coefficient matrix;

^eNumber of iterations. A number of iterations equal to 10,000 means that the method failed to converge within 10,000 iterations;

^fWall clock time (seconds) for the iterative process;

^gAverage wall clock time (seconds) per iteration;

^hWall clock time (seconds) for a complete process (including I/O operations)

if the k_O/k_S ratio is applied to a set of equations that are not associated with the largest eigenvalues of the preconditioned (deflated) coefficient matrices, the second-level preconditioner \mathbf{D} will not result in decreased largest eigenvalues. This behaviour was observed when the second-level preconditioner \mathbf{D} was applied to ssSNPBLUP_MS with the DPCG method for the reduced dataset (Table 1). For these scenarios, the DPCG method already annihilated all the largest unfavourable eigenvalues up to the lower bound of the largest eigenvalue that is allowed with the second-level preconditioner \mathbf{D} . Therefore, the second-level preconditioner \mathbf{D} did not further decrease the largest eigenvalues. It is worth noting that, if the DPCG method did not annihilate all the unfavourable largest eigenvalues up to the lower bound defined by Eq. (4), the application of the second-level preconditioner \mathbf{D} with the DPCG method did remove these remaining largest eigenvalues, as shown by the results for ssSNPBLUP_Liu applied to the field dataset (Table 2).

The decrease of the largest eigenvalues of the preconditioned coefficient matrices with decreasing k_O/k_S ratios (and until the lower bound is reached) can be explained by the sparsity pattern of the eigenvectors associated with the largest eigenvalues of the preconditioned coefficient matrices $\tilde{\mathbf{C}}$ of ssSNPBLUP. Indeed, Figs. 2 and 3 show that the entries that correspond to the equations that are not associated with SNP effects, are close to 0 for the eigenvectors associated with the largest eigenvalues of $\tilde{\mathbf{C}}$ of ssSNPBLUP_MS. Accordingly, if we assume that these entries are 0, i.e.,

$$\tilde{\mathbf{v}}_{max} = \begin{bmatrix} \tilde{\mathbf{v}}_{Omax} \\ \tilde{\mathbf{v}}_{Smax} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{v}}_{Smax} \end{bmatrix}$$

being an eigenvector associated with one of largest eigenvalues of $\tilde{\mathbf{C}}$, it follows that the largest eigenvalues of $\tilde{\mathbf{C}}$ multiplied by k_S^{-1} are also the eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$. These largest eigenvalues of $\tilde{\mathbf{C}}$ will therefore be equal to the largest eigenvalues of $\mathbf{D}^{-1/2}\tilde{\mathbf{C}}\mathbf{D}^{-1/2}$ until the lower bound defined by Eq. (4) is reached (see Additional file 2 for the derivation). This observation can also motivate an educated guess for an optimal k_O/k_S ratio for ssSNPBLUP with one additive genetic effect. If the largest eigenvalues $\lambda_{max}(\tilde{\mathbf{C}})$ and $\lambda_{max}(\tilde{\mathbf{C}}_{OO})$ are (approximately) known, an educated guess for the k_O/k_S ratio can be equal to $\frac{k_O}{k_S} = k_O \frac{\lambda_{max}(\tilde{\mathbf{C}}_{OO})}{\lambda_{max}(\tilde{\mathbf{C}})}$. For example, in our cases, $\lambda_{max}(\tilde{\mathbf{C}}_{OO})$ was always equal to the largest eigenvalue of the preconditioned coefficient matrix of a pedigree BLUP (results not shown). It follows that the educated guess for the field dataset is equal to 3.0×10^{-3} for ssSNPBLUP_MS and 3.5×10^{-2} for ssSNPBLUP_Liu, since $\lambda_{max}(\tilde{\mathbf{C}}_{OO}) = 5.07$. Both values are of the same

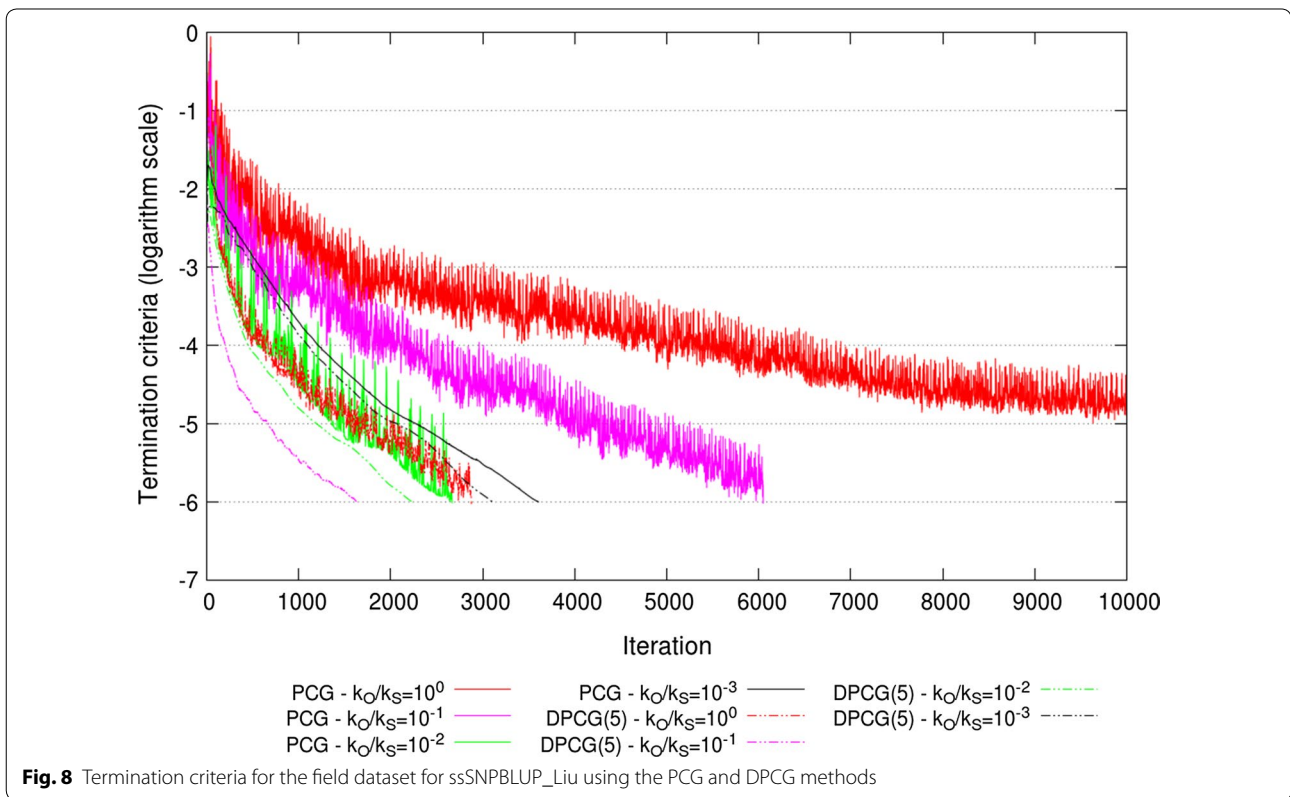
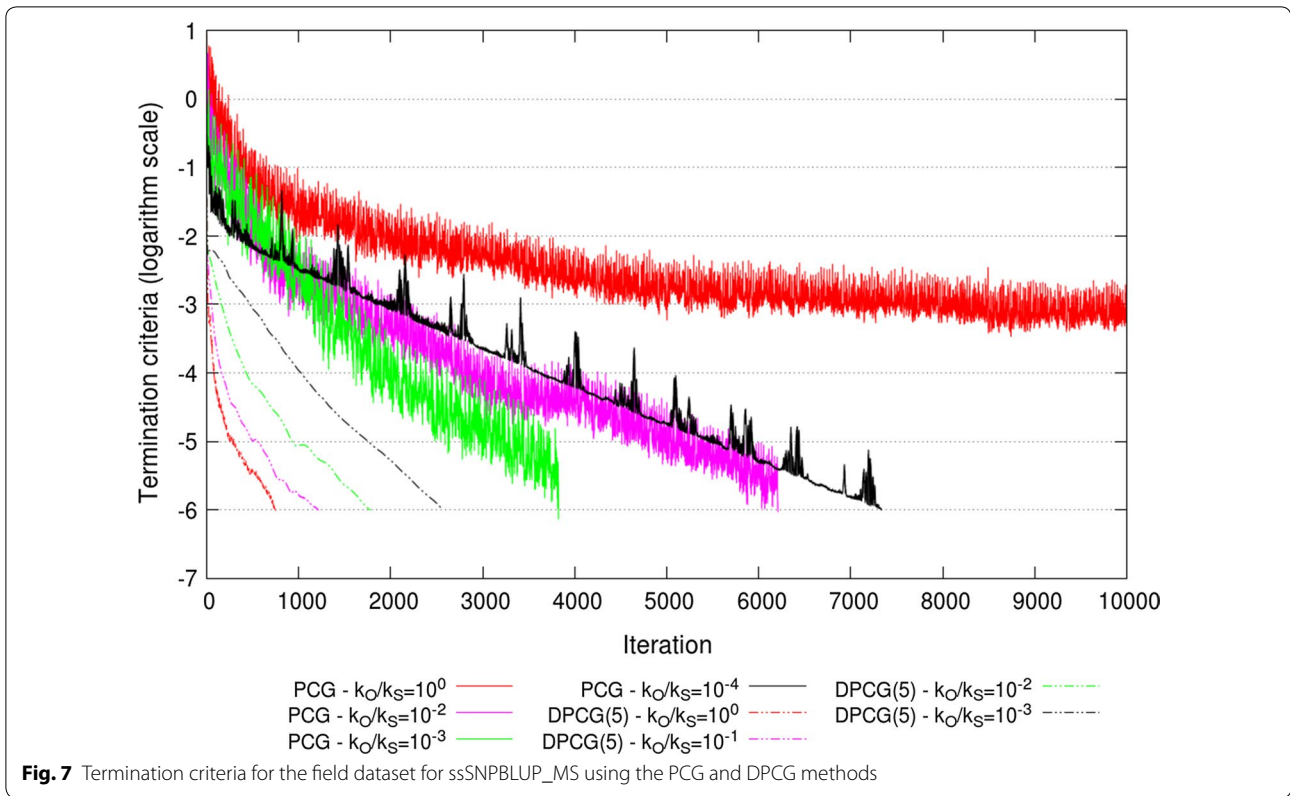
order as the corresponding optimal k_O/k_S ratios. However, the second-level preconditioner \mathbf{D} will be effective only if the smallest eigenvalues of the preconditioned coefficient matrices are not influenced, or at least less influenced than the largest eigenvalues, by the second-level preconditioner \mathbf{D} .

The decrease of the smallest eigenvalues of the preconditioned (deflated) coefficient matrices mainly depends on the sparsity pattern of the eigenvectors associated with the smallest eigenvalues. We formulated a sufficient condition such that the smallest eigenvalues remain constant when the second-level preconditioner is applied. While this sufficient condition is not fulfilled for the reduced dataset (and probably also not for the field dataset), it can help us to predict the behaviour of the smallest eigenvalues based on the sparsity pattern of the associated eigenvectors. For example, if the eigenvector associated with the smallest eigenvalue of $\tilde{\mathbf{C}}$ has mainly non-zero entries corresponding to the equations associated with SNP effects, the use of the second-level preconditioner \mathbf{D} will most likely result in a decrease of the smallest eigenvalues proportional to k_S^{-1} , which is undesirable. Other behaviours of the smallest eigenvalues of the preconditioned (deflated) coefficient matrices can lead to the conclusion that the associated eigenvectors have a different sparsity pattern, which helps understand if and how the use of the proposed second-level diagonal preconditioner will be beneficial.

Application of \mathbf{D} in ssSNPBLUP evaluations

The second-level preconditioner \mathbf{D} is easy to implement in existing software and does not influence the computational costs of a PCG iteration, since it can be merged with the preconditioner \mathbf{M} . Indeed, it is sufficient to multiply the entries of \mathbf{M}^{-1} that correspond to the equations associated with SNP effects by an optimal k_O/k_S ratio to implement the second-level preconditioner \mathbf{D} . Furthermore, the value of an optimal k_O/k_S ratio for a ssSNPBLUP evaluation can be determined by testing a range of values around the educated guess defined previously and then re-used for several subsequent ssSNPBLUP evaluations, because additional data for each new evaluation is only a fraction of the data previously used and will therefore not modify, or will modify only slightly, the properties of the preconditioned coefficient matrices $\tilde{\mathbf{C}}$.

In this study, we used the second-level diagonal preconditioner \mathbf{D} for two different ssSNPBLUP models. To our knowledge, it is the first time that ssSNPBLUP_Liu was successfully applied until convergence with real datasets [3, 5]. From our results, it seems that the preconditioned coefficient matrices of ssSNPBLUP_Liu are better conditioned than the preconditioned coefficient matrices of ssSNPBLUP_MS, leading to better convergence



patterns for ssSNPBLUP_Liu. Therefore, among all possible combinations of linear systems (i.e., ssSNPBLUP_MS and ssSNPBLUP_Liu), solvers (i.e., the PCG and DPCG methods) and the application (or not) of the second-level preconditioner **D**, it seems that ssSNPBLUP_Liu solved with the PCG method combined with the second-level preconditioner **D** is the most efficient in terms of total wall clock times and implementation. However, in our study it was tested only on two datasets and the most efficient combination of linear system and solver will most likely be situation-dependent.

Application of **D** to other scenarios

The proposed second-level preconditioner **D** can be applied and may be beneficial for ssSNPBLUP models that involve multiple additive genetic effects, or for other models that include an effect that would result in an increase to the largest eigenvalues of the preconditioned coefficient matrices. The developed theory does not require a multivariate ssSNPBLUP with only one additive genetic effect. As such, for example, if multiple additive genetic effects are fitted into the ssSNPBLUP model, such as direct and maternal genetic effects, the second-level preconditioner **D** could be used with different k_O/k_S ratios applied separately to the direct and maternal SNP effects. A similar strategy was successfully applied for ssSNPBLUP proposed by Fernando et al. [2] with French beef cattle datasets (Thierry Tribout, personal communication). Furthermore, the second-level preconditioner **D** could be used to improve the convergence pattern of models other than ssSNPBLUP. For example, with the field dataset, the addition of the genetic groups fitted explicitly as random covariables in the model for pedigree-BLUP (that is, without genomic information) led to an increase of the largest eigenvalue of the preconditioned coefficient matrix from 5.1 to 14.8. The introduction of the second-level preconditioner **D** into the preconditioned linear system of pedigree-BLUP with a $k_O/k_S = 10^{-1}$ ratio applied to the equations associated with the genetic groups reduced the largest eigenvalues to 6.0, resulting in a decrease of the effective condition number by a factor of 2.6. This decrease of the effective condition number translated to a decrease in the number of iterations to reach convergence from 843 to 660.

Conclusions

The proposed second-level preconditioner **D** is easy to implement in existing software and can improve the convergence of the PCG and DPCG methods applied to different ssSNPBLUP methods. Based on our results, the ssSNPBLUP system of equations proposed by Liu et al. [5] solved using the PCG method and the second-level preconditioner seems to be most efficient. However, the

optimal combination of ssSNPBLUP and solver will most likely be situation-dependent.

Additional files

Additional file 1. Bounds of the largest eigenvalue of the preconditioned coefficient matrix of ssSNPBLUP Derivation of the lower and upper bounds of the largest eigenvalue of the preconditioned coefficient matrix of ssSNPBLUP.

Additional file 2. Proof of the sufficient condition Proof of the sufficient condition.

Acknowledgements

The use of the high-performance cluster was made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR, Wageningen, the Netherlands).

Authors' contributions

JV conceived the study design, ran the tests, and wrote the programs and the first draft. JV and CV discussed and developed the theory. HE prepared data. CV and MPLC provided valuable insights throughout the writing process. All authors read and approved the final manuscript.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project 16022) and the Breed4Food partners Cobb Europe (Colchester, Essex, United Kingdom), CRV (Arnhem, the Netherlands), Hendrix Genetics (Boxmeer, the Netherlands), and Topigs Norsvin (Helvoirt, the Netherlands).

Ethics approval and consent to participate

The data used for this study were collected as part of routine data recording for a commercial breeding program. Samples collected for DNA extraction were only used for the breeding program. Data recording and sample collection were conducted strictly in line with the Dutch law on the protection of animals (Gezondheids- en welzijnswet voor dieren).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Animal Breeding and Genomics, Wageningen UR, P.O. 338, 6700 AH Wageningen, The Netherlands. ² CRV BV, Wassenaarweg, 20, 6843 NW Arnhem, The Netherlands. ³ DIAM, TU Delft, Van Mourik Broekmanweg, 6, 2628 XE Delft, The Netherlands.

Received: 5 March 2019 Accepted: 7 June 2019

Published online: 25 June 2019

References

- Strandén I, Lidauer M. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J Dairy Sci.* 1999;82:2779–87.
- Fernando RL, Cheng H, Garrick DJ. An efficient exact method to obtain GBLUP and single-step GBLUP when the genomic relationship matrix is singular. *Genet Sel Evol.* 2016;48:80.
- Taskinen M, Mäntysaari EA, Strandén I. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. *Genet Sel Evol.* 2017;49:36.
- Vandenplas J, Eding H, Calus MPL, Vuik C. Deflated preconditioned conjugate gradient method for solving single-step BLUP models efficiently. *Genet Sel Evol.* 2018;50:51.

5. Liu Z, Goddard M, Reinhardt F, Reents R. A single-step genomic model with direct estimation of marker effects. *J Dairy Sci.* 2014;97:5833–50.
6. Legarra A, Ducrocq V. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. *J Dairy Sci.* 2012;95:4629–45.
7. Mäntysaari EA, Strandén I. Single-step genomic evaluation with many more genotyped animals. In: Proceedings of the 67th annual meeting of the European Association for Animal Production, 29 August–2 September 2016, Belfast; 2016.
8. Cornelissen MAMC, Mullaart E, Van der Linde C, Mulder HA. Estimating variance components and breeding values for number of oocytes and number of embryos in dairy cattle using a single-step genomic evaluation. *J Dairy Sci.* 2017;100:4698–705.
9. CRV Animal Evaluation Unit. Management guides, E16: breeding value-temperament during milking; 2010. <https://www.crv4all-international.com/wp-content/uploads/2016/03/E-16-Temperament.pdf>. Accessed 15 Mar 2018.
10. CRV Animal Evaluation Unit. Statistical indicators, E-15: breeding value milking speed; 2017. https://www.crv4all-international.com/wp-content/uploads/2017/05/E_15_msn_apr-2017_EN.pdf. Accessed 15 Mar 2018.
11. Gengler N, Nieuwhof G, Konstantinov K, Goddard ME. Alternative single-step type genomic prediction equations. In: Proceedings of the 63rd annual meeting of the European Association for Animal Production, 27–31 Aug 2012; Bratislava; 2012.
12. Mäntysaari EA, Evans RD, Strandén I. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J Anim Sci.* 2017;95:4728–37.
13. Fragomeni BO, Lourenco DAL, Masuda Y, Legarra A, Misztal I. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genet Sel Evol.* 2017;49:59.
14. Raymond B, Bouwman AC, Wientjes YCJ, Schrooten C, Houwing-Duistermaat J, Veerkamp RF. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. *Genet Sel Evol.* 2018;50:49.
15. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res.* 2012;94:73–83.
16. Saad Y. Iterative methods for sparse linear systems. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics; 2003.
17. Frank J, Vuik C. On the construction of deflation-based preconditioners. *SIAM J Sci Comput.* 2001;23:442–62.
18. Varga RS. Geršgorin and his circles. Springer series in computational mathematics. Berlin: Springer; 2004.
19. Paige C, Saunders M. Solution of sparse indefinite systems of linear equations. *SIAM J Numer Anal.* 1975;12:617–29.
20. Kaasschieter EF. A practical termination criterion for the conjugate gradient method. *BIT Numer Math.* 1988;28:308–22.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

