ALTERNATIVE IMPLEMENTATIONS OF PRECONDITIONED CONJUGATE GRADIENT ALGORITHMS FOR SOLVING MIXED MODEL EQUATIONS

D.J. Garrick[1,2], B.L. Golden[2] and D.P. Garrick[1,2]

[1]AL Rae Centre of Genetics and Breeding, Massey University, Hamilton, 3240 New Zealand
[2]Theta Solutions LLC, Atascadero, CA, 93422 USA

SUMMARY

Mixed model equations encountered in pedigree and genomic analyses are typically solved using an iterative preconditioned conjugate gradient algorithm. That algorithm requires a preconditioning matrix chosen to improve the condition number of the problem. Convergence is very fast when an appropriate preconditioning matrix is used, but some equations fail to converge unless an effective preconditioner can be found, and that is not always straightforward, especially in genomic analyses. Some preconditioning and absorption options are compared in the context of a national cattle evaluation for growth traits using a multi-trait single-step marker effects model. It is demonstrated that computing time is largely determined by the number of iterations required to obtain convergence, rather than the complexity of the equations or preconditioning. Further, a reliable convergence statistic for general applications remains problematic.

INTRODUCTION

Mixed linear models that include fixed effects other than the mean, and random effects other than the residuals, are fundamental to theoretical and applied aspects of animal breeding. Most genetic improvement programs rely on routine multiple-trait prediction that involves finding the solution to sets of simultaneous equations we refer to as the mixed model equations (e.g. Henderson, 1975) that are typically large, sparse, symmetric and positive semi definite. Early applications of national evaluation programs explicitly formed every contribution to the left- and right-hand sides of the equations, frequently after absorbing fixed effects such as herd-year-season, and then solved the resultant sparse set of equations using Gauss-Seidel (GS) iteration (e.g. Van Vleck and Dwyer 1985). Later applications avoided the accumulation of every element of the left- and right-hand sides and instead used iteration on data (Schaeffer and Kennedy 1986) to recreate matrix and vector elements as required. Gauss-Seidel iteration was replaced by the sometimes problematic but typically much faster approach of preconditioned conjugate gradient (PCG) (Berger *et al*. 1989; Stranden *et al*. 1999; Tsuruta *et al*. 2001).

A nice property of GS iteration is that every iteration results in a solution that when multiplied by the left-hand side coefficient matrix will more closely agree with the right hand side vector. However, GS can be slow to converge, and convergence tends to slow down with every subsequent iteration. In contrast, PCG tends to converge quite quickly, but in finite arithmetic the system is prone to rounding errors and to loss of conjugacy that can result in successive iterations being poorer fits than previous iterations. Performance is sensitive to the condition number of the equations, which can be dramatically influenced by the choice of preconditioner matrix. Calculating the perfect preconditioner matrix for a given problem is more effort than solving the equations. Finally, it is hard to know exactly when to stop iterating and accept the current iteration as a practical solution to the mixed model equations.

The adoption of single-step models for national evaluation that include both genotyped and non-genotyped animals in the same evaluation has created some additional challenges in obtaining PCG solutions. First, some submatrices of the mixed model equations are no longer sparse, and second, the equations are more likely to lose conjugacy than mixed model equations based on pedigree relationship matrices, at least when historically used diagonal or block preconditioners are applied. The objective of this study was to compare the performance of some alternate PCG implementations

in the context of a multiple-trait single-step national cattle evaluation.

MATERIALS AND METHODS

The American Hereford Association runs its genetic evaluation on a weekly basis that includes performance and genotypic data along with pedigree records comprising about 2.5 million US and Canadian Hereford cattle. The complete evaluation comprises nine multiple-trait single-step marker effects models (Fernando *et al*. 2016) to publish 16 different EPD (Golden *et al*. 2018). Mixed model equations are solved using PCG, then the PCG solutions are used to seed parallel Markov chain Monte Carlo analyses using single-site Gibbs sampling to estimate prediction error variances (PEV) to calculate reliabilities, and PEV for contrasts between groups of one or more animals (Garrick *et al*. 2018). This paper reports the PCG solving performance for the multiple trait growth model. The model equations for each correlated trait in that analysis are

$$\begin{aligned} y_B &= J_B j_B + X_B b_B + P_B p_B + Z_B a_B + M_B m_B + Z_B^n u_B^n + Z_B^g S_B \alpha_B + e_B \\ y_W &= J_W j_W + X_W b_W + P_W p_W + Z_W a_W + M_W m_W + Z_W^n u_W^n + Z_W^g S_W \alpha_W + e_W \\ y_G &= J_G j_G + X_G b_G + Z_G a_G + Z_G^n u_G^n + Z_G^g S_G \alpha_G + e_G \end{aligned}$$

where y_i is a vector of phenotypic observations on *B*=birth weight, *W*=weaning weight, or *G*=post weaning gain, j_i is a fixed covariate accounting for the difference in expected value between genotyped and non-genotyped founders for each trait, b_i are all the other fixed effects, p_i are the random permanent environmental effects of the dam for birth or weaning weight, a_i are the random additional polygenic effects of each trait, m_i are the random maternal genetic effects of birth or weaning weight, u_i^n are the direct breeding values for non-genotyped animals for each trait, α_i are the random marker or SNP effects for each trait, and e_i are the random residual effects for each trait. The J_i matrices are formed from a vector of 1's corresponding to genotyped individuals and an imputed value for non-genotyped animals, X_i, P_i, Z_i, M_i, are incidence matrices for fixed effects, maternal permanent environmental effects, direct genetic effects, and direct maternal effects, respectively, Z_i^n, and Z_i^g are direct effect incidence matrices for non-genotyped and genotyped individuals with phenotypes, and S_i are marker matrices for centred SNP covariates for genotyped animals. The variance-covariance matrices and their inverses for this single-step marker effects model and its mixed model equations are in Fernando *et al*. (2016) and Garrick *et al*. (2018).

Two approaches to characterise convergence during PCG iteration are the two-norm of the preconditioned residual divided by the number of effects (which we denote the iteration *residual*), and the two-norm of the raw residual, divided by the two-norm of the right-hand side, which we denote as *cr* (following Lidauer *et al*. 2015). That is, for solving equations denoted by coefficient matrix, solution and right-hand side as $Cs = r$, based on the preconditioned equations $P^{-1}Cs = P^{-1}r$, the vector of raw residuals at iteration k is $\varepsilon^k = r - C\hat{s}^k$, which is used every round of iteration to compute the $residual = \varepsilon^{k\prime}P^{-1}\varepsilon^k/\text{length}(\varepsilon^k)$, and $cr = \sqrt{\varepsilon^{k\prime}\varepsilon^k}/\text{r}'\text{r}$, for all effects, or separately for each effect in the mixed model equations (i.e., j_i, b_i, p_i, a_i, m_i, u_i, and α_i).

Two options were compared for the preconditioning matrix, the simplest representing the inverse of the diagonal elements of the mixed model equations (i.e. diagonal preconditioning), and the other replacing the preconditioner elements for the fixed effects by the actual inverse of the submatrix of the mixed model equations for fixed effects, namely $(X'R^{-1}X)^{-1}$, either separately for each trait, or with one block for all three traits.

Two options for forming the mixed model equations were compared, one which explicitly fitted all the effects other than the random residual effects shown in the model equation above, and a reduced order set of equations in which fixed effects, b_i, for all three traits had been absorbed. The absorbed

equations can be represented by striking out the rows and columns of the mixed model equations corresponding to the fixed effects to be absorbed, then subtracting some terms from the coefficient matrix and right-hand side to eliminate the absorbed equations. For the simplest mixed model equations represented by the model equation $y = Xb + Zu + e$ with $var(u) = G$ and $var(e) = R$, the complete mixed model equations would have order defined by the number of fixed effects plus the number of random effects and be given by

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix},$$ whereas the absorbed equations would have order defined by the number of random effects as in the equations

$$[Z'R^{-1}Z + G^{-1} - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}Z][u] = [Z'R^{-1}y - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}y].$$

RESULTS AND DISCUSSION

The number of iterations and computing times per iteration for BOLT PCG software on a 256Gb RAM Ubuntu server using one 12 Gb Titan V graphics processing unit are shown in Table 1 for the complete and absorbed sets of mixed model equations for various stopping criteria. Correlations between solutions for each factor from different approaches all exceeded 0.99 if not 0.999.

Table 1. Numbers of PCG iterations to achieve alternative stopping criteria in the North American Hereford multiple-trait single-step growth analysis using block or diagonal preconditioning of full or absorbed equations

Mixed Model Equations	Preconditioner	Stopping Criteria Change in residual 1e-10	1e-11	1e-12	1e-13	cr 1e-5	Time per iter
Complete[1]	Diagonal	1,768	1,826	4,373	5,435	2,617	0.17s
Complete[2]	Diagonal	1,575	3,227	4,627	6,182	2,624	0.15s
Complete[2]	Block	2,483	2,858	2,858	2,858	2,502	0.15s
Absorbed	Diagonal	2,123	3,289	3,386	6,965	8,641	0.21s

[1]Separate submatrix blocks for *J* factor and *X* factor for each trait

[2]Single submatrix block for *J* factors for B,W,G and another for *X* factors for B,W,G

The total computing time for PCG solution of the multi-trait single-step marker effects model varied from 4 minutes to 24 minutes but was influenced to a much greater extent by the number of iterations (1,575 to 8,641) required for convergence than by the computing time per iteration (0.15 to 0.21 s). The absorbed equations if formed explicitly are much less sparse than the complete set of mixed model equations, but the computing effort was little affected by the absorption of effects as the matrix multiplications were done in parts. This is not surprising as easily shown by denoting the coefficient matrix for the full equations to solve as $\begin{bmatrix} S^{-1} & T \\ T' & Q \end{bmatrix}$, where S^{-1} represents the fixed effects block diagonal partition to be absorbed, Q represents the block diagonal partition for all the other effects, and T represents the block off-diagonal partition between the effects being absorbed and the remaining effects, then the left-hand-side of the absorbed equations can be represented as $[Q - T'ST]$. Each iteration of PCG involves multiplying the coefficient matrix by a work vector, denoted w, as in $w' = [w_b' \quad w_u]$, which for the complete equations requires computing $S^{-1}w_b$, Tw_u, $T'w_b$, and Qw_u, whereas for the absorbed equations it would requires computing Qw_u and $T'STw_u$. The latter term can be computed in parts as $T'(S(Tw_u))$, first involving the matrix-vector product Tw_u, then pre-multiplying this vector by S then pre-multiplying that product by T'. The only difference in effort between applying the PCG algorithm to the full or the absorbed equations is the computation of the matrix

product involving S^{-1} rather than the matrix product involving S. In many mixed model equations, the sparsity and complexity of S^{-1} is similar to that of S, for example for $(X'R^{-1}X)^{-1}$ and $X'R^{-1}X$. Computation of the matrix-vector products in the full equations can be done in parallel, whereas the part equations requires the multiplications to be undertaken serially, involving the product of the first matrix-vector as the vector used in the second matrix-vector multiplication.

Changes in the number of iterations required to meet a given stopping criterion occur due to rounding errors and loss of conjugacy even when there is no change to the elements of the mixed model equations, or to the method of preconditioning, as shown by comparing rows 1 and 2 of Table 1 when the complete mixed model equations were partitioned into submatrices by factor and trait compared to when the factors for J were pooled across traits into one submatrix, and the factors for X were pooled across trait into another submatrix.

Using a block diagonal structure rather than a diagonal matrix for preconditioning fixed effects was initially slower but reached convergence much faster for higher convergence thresholds.

Changes in the number of iterations by absorbing the fixed effects did not have a consistent effect on the number of iterations. This is partly because the process of absorption reduces the two-norm in the denominator of the *cr* criterion, making the same tolerance (i.e. *cr* <1e-5) much more strict than in the complete mixed model equations.

Changes in the number of iterations by changing stopping criteria (from *residual* to *cr*) or the tolerance of the stopping criteria, resulted in reranking of the performance of the algorithms. The *residual* statistic is not a good stopping criterion because it tends to bounce around from iteration to iteration, but can occasionally achieve very small changes between iterations that result in apparent convergence that is not supported by the *cr* statistic. However, the *cr* statistic is sensitive to parameterisation of the mixed model equations, as shown by the effect of absorption, which also makes that criteria problematic for routine use.

CONCLUSIONS

The results demonstrate that uniformly appropriate convergence criteria for PCG systems are challenging to identify. Minor changes to the manner in which the mixed model equations are parameterised can have considerable influence on performance and run time, most notably by influencing the number of iterations required to achieve a given definition of convergence. Alternative blocking structures, preconditioning matrices, and parameterisation of models can notably influence results.

REFERENCES

Berger P.J., Luecke G.R. and Hoekstra J.A. (1989) J. Dairy Sci. 72:514.
Fernando R.L., Cheng H., Golden B.L., and Garrick D.J. (2016) Genet. Sel. Evol. 48:96.
Garrick D.J., Cheng H., Golden B.L. and Fernando R.L. (2017) Proc. Assoc. Advmt. Anim. Breed. Genet. 22:269.
Garrick D.P., Golden B.L. and Garrick D.J. (2018) WCGALP 2018:974.
Golden B.L., Spangler M.L., Snelling W.M. and Garrick D.J. (2018) Proc. 11th Beef Impr. Fed. Genet. Pred. Workshop 11:14.
Henderson C.R. (1973) J. Anim. Sci. Issue Symp. 70:23.
Lidauer M., Matilainen K., Mantysaari E., Pitkanen T., Taskinen M. and Stranden I. (2015) http://www.luke.fi/mix99
Schaeffer L.R. and Kennedy B.W. (1986) J. Dairy Sci. 69:575.
Stranden I. and Lidaeur M. (1999) J. Dairy Sci. 82:2779.
Tsuruta S., Misztal I. and Stranden I. (2001) J. Anim. Sci. 79:1166.
Van Vleck L.D. and Dwyer D.J. (1985) J. Dairy Sci. 68:1006.