Handling Non Positive Definite Relationship Matrices In Mixed Models

A.R. Gilmour[1]

Introduction

It is not uncommon to form relationship matrices for use in genetic evaluation from marker data. Such relationship matrices may pertain to a particular locus or represent an average relationship over the whole genome. The mixed model equations used for genetic evaluation typically require the inverse relationship matrix. The paper presents a reformulation, introduced by Thompson (2009), of the mixed model equations which allows for the relationship matrix to be singular. It outlines this procedure as implemented in ASReml 3 and discusses consequences of using a non-positive definite relationship matrix.

Material and methods

Standard Mixed model. The mixed model is typically written as

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}u + e$$

where y is a vector of observations, $\boldsymbol{\beta}$ is a vector of fixed effects with design matrix $\mathbf{X}$, u is a vector of random effects with design matrix $\mathbf{Z}$, and e is a vector of residual errors. Without loss of generality but for clarity of presentation, we consider the simple model where $u \sim N(0,\sigma_a^2\mathbf{A})$ and where $e \sim N(0, \sigma^2\mathbf{I})$. The mixed model equations typically used to solve for $\boldsymbol{\beta}$, u, and e given $\mathbf{A}$, $\mathbf{X}$, $\mathbf{y}$ and $\mathbf{Z}$ are

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\gamma\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} \mathbf{X}'y \\ \mathbf{Z}'y \end{bmatrix}$$

This formulation requires that $\mathbf{A}$, the relationship matrix in our context, be invert-able.

Alternative form of Mixed model equations. An equivalent form is

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \cdot \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{I} \\ \cdot & \mathbf{I} & -1/\gamma\,\mathbf{A} \end{bmatrix} \begin{bmatrix} \beta \\ u \\ w \end{bmatrix} = \begin{bmatrix} \mathbf{X}'y \\ \mathbf{Z}'y \\ 0 \end{bmatrix}$$

---

[1] Cargo Vale, Cargo, NSW, 2800, Australia

where w is a set of effects of no particular interest. It is easily shown that absorbing the additional set of equations into the portion involving the design matrices yields the original formulation. The advantage of the second form though is that the $\mathbf{A}$ equations need not be absorbed first. In particular, any singularity in $\mathbf{A}$ itself can be overcome by absorbing the corresponding row of $\mathbf{Z'Z}$ first.

Implementation in ASReml. Generally, $\mathbf{A}$ will only have a few singularities and it is convenient when the mixed model equations must be solved repeatedly, to form a partially absorbed 'A inverse' by absorbing the non-singular rows. Let 'A inverse' first be written as

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{-A} \end{bmatrix}$$ and then the rows and columns of $\mathbf{A}$ reordered with permutation matrix $\mathbf{P}$ so

that the first s rows of $\mathbf{PAP'}$ will be singular when adjusted for the remaining rows. Rewrite

as $$\begin{bmatrix} \mathbf{0} & \mathbf{P'} \\ \mathbf{P} & \mathbf{-PAP'} \end{bmatrix}$$. Absorbing the non-singular rows of $\mathbf{PAP'}$ leaves an 'A inverse'

matrix of the form

$$\begin{bmatrix} \mathbf{P_2}'(\mathbf{P_2AP_2}')^{-1}\mathbf{P_2} & \mathbf{P_1}'+\mathbf{P_2}'(\mathbf{P_2AP_2}')^{-1}\mathbf{P_2AP_1}' \\ \mathbf{P_1}+\mathbf{P_1AP_2}'(\mathbf{P_2AP_2}')^{-1}\mathbf{P_2} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{P_1}$ selects the singular subset of $\mathbf{A}$ and $\mathbf{P_2}$ selects the non-singular subset.

This can be directly used in the mixed model equations after inserting the s extra equations for the singular rows of $\mathbf{A}$. These extra rows amount to a Lagrange constraint on the solution. The solution algorithm for the mixed model equations also needs to delay processing of these extra rows until something appears on the diagonal which happens when its corresponding data equation has been processed.

Results and discussion

When a relationship matrix is formed from marker data, the algorithms used sometimes create singular or negative definite equations. In the past several pragmatic approaches have been taken to overcome the resultant problems when these matrices are inverted and incorporated into the mixed model equations.

One approach if there are no singularities is to ignore the fact that there are negative roots. There is a real probability that the solution of the mixed models will fail for certain parameter (γ) values, especially in the context of REML algorithms where the Information matrix may become negative definite. It is also possible that the solutions will be

unacceptable. While this approach is possible it is not recommended. In any case, care should be taken to check the plausibility of resulting predictions.

Another approach is to add a small constant to some or all of the diagonal elements to make the matrix positive definite. It is not evident how big this constant should be, and it is hardly necessary to add it to all diagonal elements when the problem involves only some of them.

Mishchenko et al. (2008), in proposing a method for singular or negative semi-definite matrices, proposed using a modified relationship matrix based on the positive (or dominant) eigenvalues. Their method was limited to a very simple design and so is not a general solution. However, the strategy of setting negative eigenvalues to zero and reconstituting the matrix is one possibility which can be used in conjunction with the solution strategy outlined above.

Another approach would be 'bending', that is reconstructing the matrix after shrinking the eigen values toward their mean until the smallest was positive.

Oikawa and Yasuda (2009) present a method of calculating an additive relationship matrix when genetically identical animals appear in the pedigree (say through cloning). This is a special case which could also be handled by the method proposed here.

There are a few issues that arise. When the mixed model equations are being solved as part of a REML algorithm, it is necessary to calculate the determinant of the **A** matrix when calculating the REML Log likelihood. Since the order of permutation to identify a singular subset is not unique, the determinant of the non-singular portion is not unique, but will be order dependent. Further, an adjustment of the number of effects will be needed in the calculation of the likelihood value.

The procedure outlined here has been implemented in ASReml 3 (Gilmour 2009). When the user supplies a relationship matrix, ASReml 3 checks whether it is positive definite and reports the number of negative pivots and/or singularities to the user. Extension of the underlying concept to handle REML estimation of singular variance matrices, say with a Cholesky factorization, is in progress.

Conclusions

Non positive definite relationship matrices can arise as a direct result of cloning, or in the process of calculating relationships from marker data, either for a particular locus or averaged across the genome. The method outlined here is one approach to handling these matrices in a mixed model setting.

References

Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R., 2009. ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, HPI 1ES, UK

Mishchenko, K., Holmgren, S., and Ronnegard, L. (2008) *J. Comp. Meth. Sci. Eng.* **8**: 53-67

Takuro Oikawa and Kazuhiro Yasuda *Genet. Sel. Evol.* 2009, 41:25 doi:10.1186/1297-9686-41-25

Thompson, R., (2009) In *Proc. Assoc Adv Ani Breeding and Genet*. 18