A recipe for multiple trait deregression

I. Strandén and E.A. Mäntysaari

MTT Agrifood Research Finland, Biometrical Genetics, 31600 Jokioinen
Ismo.Stranden@MTT.fi

Abstract

Standard method to deregress estimated breeding values use block iteration where one block is solved by a direct method, and the other block by a Gauss-Seidel type update. We used an iterative method for the first block, and root finding methods to accelerate solving the other block. The iterative method allowed use of existing software for breeding value evaluation, and gave a flexible program. The acceleration methods improved convergence of the deregression method considerably. The best acceleration method gave convergence in less than 10 iterations but more than 100 iterations were needed when no acceleration was used.

Introduction

Deregression of estimated breeding values (EBV) was introduced by Goddard (1985), and described by Jairath et al. (1998) for single trait, and by Schaeffer (2001) for multiple trait models. Solving deregressed EBVs or proofs requires setting up mixed model equations with a general mean as an unknown fixed effect. But, unlike in regular mixed model equations, the EBVs of some animals are known but their observations are unknown.

Deregression is a non-linear problem. Solving of the non-linear deregression problem is computationally difficult due to slow convergence. In addition, the presented algorithms require inversion of a large matrix or solving a linear system explicitly. We used different strategies in these steps. We used an iterative procedure based on preconditioned conjugate gradient (PCG) for solving the linear model. We accelerated the non-linear solving problem by root finding methods. We implemented deregression into existing mixed model software used for EBV calculations.

In this study, we compared three acceleration methods with the original solving method. The acceleration methods were bisection, secant, and Broyden's method. We show performance of these methods in the small data sets presented in Schaffer (2001).

Methods

Multivariate deregression

Consider a T trait deregression problem. There are q_b bulls having EBVs, and q_a ancestors without EBVs. For ease of presentation, assume that the bulls have EBVs in all T traits. The base equation system in deregression is (Schaeffer, 2001)

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{R^{-1}X} & \mathbf{R^{-1}} + \mathbf{A}^{bb} \otimes \mathbf{G}_0^{-1} & \mathbf{A}^{ba} \otimes \mathbf{G}_0^{-1} & \mathbf{A}^{bg} \otimes \mathbf{G}_0^{-1} \\ \mathbf{0} & \mathbf{A}^{ab} \otimes \mathbf{G}_0^{-1} & \mathbf{A}^{aa} \otimes \mathbf{G}_0^{-1} & \mathbf{A}^{ag} \otimes \mathbf{G}_0^{-1} \\ \mathbf{0} & \mathbf{A}^{gb} \otimes \mathbf{G}_0^{-1} & \mathbf{A}^{ga} \otimes \mathbf{G}_0^{-1} & \left(\mathbf{A}^{gg} + \mathbf{I}\right) \otimes \mathbf{G}_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\mathbf{t}}_b \\ \hat{\mathbf{t}}_a \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_\mu \\ \mathbf{r}_b \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad [1]$$

where $\hat{\boldsymbol{\mu}}$ is a $T \times 1$ vector of fixed unknown trait means, $\mathbf{X}$ is incidence matrix relating the bull EBVs to the appropriate trait means, $\hat{\mathbf{t}}_b$ is $Tq_b \times 1$ vector of known EBVs adjusted by the general mean ($\hat{\mathbf{t}}_b = \mathbf{a}_b - \mathbf{X}\hat{\boldsymbol{\mu}}$), $\hat{\mathbf{t}}_a$ is $Tq_a \times 1$ vector of unknown ancestor genetic effects, $\hat{\mathbf{g}}$ is vector of random unknown phantom parent group effects,

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{bb} & \mathbf{A}^{ba} & \mathbf{A}^{bg} \\ \mathbf{A}^{ab} & \mathbf{A}^{aa} & \mathbf{A}^{ag} \\ \mathbf{A}^{gb} & \mathbf{A}^{ga} & \mathbf{A}^{gg} \end{bmatrix}$$ is inverse of the

numerator relationship matrix augmented with phantom parent groups, $\mathbf{G}_0$ is known $T \times T$ genetic (co)variance matrix for the T traits, and $\mathbf{R}$ is known residual (co)variance matrix that may have weights. The right hand side is unknown but has the following relationships: $\mathbf{r}_b = \mathbf{R}^{-1}\mathbf{y}$,

$\mathbf{r}_\mu = \mathbf{X}'\mathbf{r}_b$ where $\mathbf{y}$ has the unknown deregressed proofs.

The right hand side ($\mathbf{r}_\mu$, $\mathbf{r}_b$), the general mean ($\hat{\boldsymbol{\mu}}$), the ancestral EBVs ($\hat{\mathbf{t}}_a$), and the phantom parent groups ($\hat{\mathbf{g}}$) are solved by block iteration. Iteration k in the algorithm is ($k=1,2,\ldots$) :

1. Solve $\hat{\mathbf{t}}_a^{[k+1]}$ and $\hat{\mathbf{g}}^{[k+1]}$ in [1] given current solutions for the other unknowns.

2. Calculate $\mathbf{r}_b^{[k+1]}$ and $\mathbf{r}_\mu^{[k+1]}$ given the other unknowns by making the matrix times vector product in [1].

3. Update the general mean:

   $\hat{\boldsymbol{\mu}}^{[k+1]} = \hat{\boldsymbol{\mu}}^{[k]} - \boldsymbol{\Delta}^{[k]}$

   where $\boldsymbol{\Delta}^{[k]} = \left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{r}_\mu^{[k+1]} - \mathbf{X}'\mathbf{r}_b^{[k+1]}\right)$.

Note that step 2 gives zero $\boldsymbol{\Delta}^{[k]}$ only at convergence. Thus, iteration of the 3 steps is continued until convergence. Our convergence criteria was

$$\left| \Delta_i^{[k]} / \hat{\mu}_i^{[k]} \right| < 10^{-6}$$

for all traits i. At convergence the deregressed EBVs are $\mathbf{y} = \mathbf{R}\mathbf{r}_b$.

Acceleration methods

The solving step 3 for the general mean calculates change in iteration ($\boldsymbol{\Delta}^{[k]}$) for the original update. Instead of using that to update the general mean solution, the change can be considered as value for a function at $\hat{\boldsymbol{\mu}}^{[k]}$. Our problem is to find $\hat{\boldsymbol{\mu}}$ having zero $\boldsymbol{\Delta}$, or a root finding problem.

We considered the following methods (e.g. Press et al., 1992) to find general mean
- None. Original update method
- Bisection
- Secant
- Broyden

Starting values for the trait means were means of EBVs by trait. However, for the bisection method minimum and maximum values of EBVs and their mean were used.

The acceleration methods change step 3 of the algorithm. Consider the secant method. It is a one dimensional root finding method. Thus, let $\Delta_i^{[k]}$ be the function value at $\hat{\boldsymbol{\mu}}^{[k]}$ for trait i, and $\hat{\mu}_i^{[k]}$ be the value for general mean for trait i at iterate k. The secant method update for trait i is

$$\mu_i^{[k+1]} = \mu_i^{[k]} - \frac{\mu_i^{[k]} - \mu_i^{[k-1]}}{\Delta_i^{[k]} - \Delta_i^{[k-1]}}$$

Each trait mean is updated independently of the others. The current and previous iterate values are needed in the update. Thus, the secant method update was used from iteration two onwards.

Broyden's method is a quasi-Newton root finding method for multivariate non-linear problems. The update formula is

$$\boldsymbol{\mu}^{[k+1]} = \boldsymbol{\mu}^{[k]} - \mathbf{J}^{-1[k]}\boldsymbol{\Delta}^{[k]}$$

where $\mathbf{J}^{-1[k]}$ estimates inverse of the Jacobian matrix at iteration k. Update k of the inverse Jacobian matrix estimate by the Broyden's method is

$$\mathbf{J}^{-1[k]} = \mathbf{J}^{-1[k-1]} + \frac{\boldsymbol{\delta}^{[k]} - \mathbf{J}^{-1[k-1]}\boldsymbol{\Delta}^{[k]}}{\boldsymbol{\delta}^{[k]\prime}\mathbf{J}^{-1[k-1]}\boldsymbol{\Delta}^{[k]}}\boldsymbol{\delta}^{[k]\prime}\mathbf{J}^{-1[k-1]}$$

where $\boldsymbol{\delta}^{[k]} = \boldsymbol{\mu}^{[k]} - \boldsymbol{\mu}^{[k-1]}$. Thus, no matrix inversion is needed. Like for the secant method, the Broyden's method was used from iteration two onwards. Initial value for the inverse Jacobian matrix was identity matrix.

Additional calculations needed by the acceleration methods were small in comparison to calculations in steps 1 and 2 of the algorithm. Thus, comparison of the methods will be based on number of iterations needed by the methods.

PCG iteration

In step 1 of the algorithm, unknowns $\hat{\mathbf{t}}_a$ and $\hat{\mathbf{g}}$ are solved from mixed model equations given other parameters. We implemented deregression in a standard iteration on data solver of mixed model equations, in our case MiX99 (Lidauer and Strandén, 1999). The solver uses preconditioned conjugate gradient (PCG) iteration.

The standard PCG was modified so that solutions of some effects (here $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{t}}_b$) in the PCG algorithm remained unchanged. This was done by not updating these solutions, and other vectors associated with these effects in the PCG algorithm.

Material

Schaeffer (2001) presented small data sets from two countries denoted A and B, and calculated deregressed proofs. We used the same data and variance components to compare the different acceleration methods, and performance of our deregression implementation. For country A, there were EBVs for first, second, and third lactation 305-d protein yield. Four analyses were done: first

lactation only (1), first and second lactation (MT_{1+2}), all lactations with original variance components (MT_{1+2+3}) and with all covariances set to zero (ST). Country B EBVs were 305-d protein yield and somatic cell scores (SCS). Four analyses were done: protein only, SCS only, both together with original variance components (MT_{p+s}) and with all covariances set to zero (ST).

Results and discussion

The function value (Δ) given the general mean (μ) is in Figure 1 for country B, protein. The function is linear but flat. Thus, slow convergence can be expected by the non-accelerated method. In addition, starting value for the non-accelerating method is very important. Similar figures can be drawn for all the traits: linear and quite flat.

The acceleration methods showed their strength (Table 1). The number of PCG calls was lowest with the secant and Broyden's methods. The number was always more than 100 when no acceleration was used, but stayed under 10 using the Broyden's method. The secant method was as good for single trait models and even better when multiple independent traits were analyzed. The PCG calls could have been calls for a direct solving method instead of the iterative solver. Thus, the results apply when direct method is used.

Total number of PCG iterations showed the same as the number of PCG calls, the secant and Broyden's methods were best. The calculated deregressed proofs were the same as those presented in Schaffer (2001) for the full multivariate models.

We have tested the deregression implementation with real data sets and found the acceleration methods perform well. Convergence of the general mean effects was affected by definition of genetic groups: the more groups the faster convergence. However, changes in genetic group definition had only a small effect on actual deregressed proof

estimates when measured by correlation. Effect of genetic correlations on the deregression results is still to be studied.

Conclusion

Deregression was easily implemented within existing BLUP solving software. Now, the same statistical models (sire and animal relationships) are available, and the same pedigree can be used for EBV and deregression calculations.

Results clearly showed that the secant and Broyden's methods were the best for calculating deregressed proofs for a single trait model. In the multiple trait analyses, Broyden's method needed least number of iterations for convergence. However, when all covariances between the traits were zero, the secant method was best. Thus, none of the implemented acceleration methods were universally best.

References

Goddard, M. 1985. A method of comparing sires evaluated in different countries. *Livest. Prod. Sci.* 13, 321-331.

Jairath, L., Dekkers, J.C., Schaeffer, L.R., Liu, Z., Burnside, E.B. & Kolstad B. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81, 550-562.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992. Numerical recipes in FORTRAN – The art of scientific computing. Cambridge University Press.

Lidauer, M. & Strandén, I. 1999. Fast and flexible program for genetic evaluation in dairy cattle. *Interbull bulletin* 20, 19-24.

Schaeffer, L.R. 2001. Multiple trait international bull comparisons. *Livest. Prod. Sci.* 69, 145-153.

Table 1. Number of PCG calls (total number of PCG iterations in parenthesis) by solving method in the different data sets.

Country	Data	None	Bisection	Secant	Broyden
A	Lactation 1	138 (1656)	16 (191)	4 (48)	3 (36)
	MT_{1+2}	154 (1849)	250 (3093)	8 (97)	6 (73)
	MT_{1+2+3}	169 (2197)	269 (3497)	39 (506)	7 (91)
	All, ST^1	138 (1794)	18 (233)	4 (51)	8 (104)
B	protein	713 (8556)	16 (192)	4 (49)	3 (36)
	SCS	149 (1506)	12 (128)	4 (47)	3 (34)
	$MT_{protein + SCS}$	748 (8976)	444 (5329)	6 (73)	6 (72)
	All ST^1	713 (8556)	16 (193)	4 (49)	6 (72)

[1] Multiple trait analysis with all covariances set to zero.
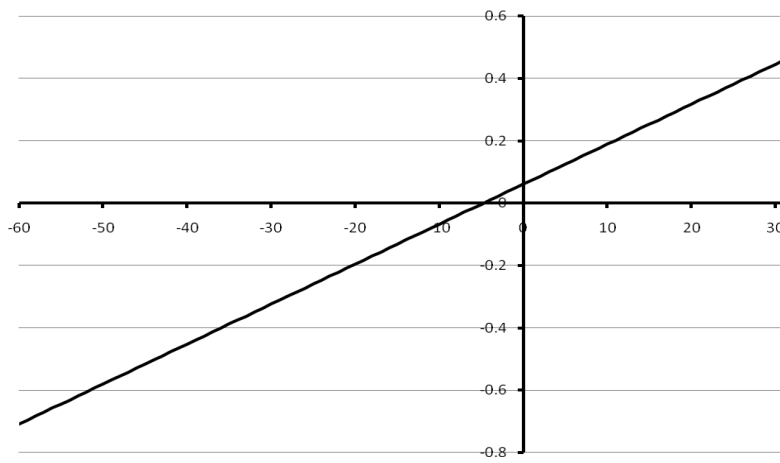

Figure 1. Function value ($\Delta^{[k]}$) with different values of general mean $\hat{\mu}^{[k]}$. The convergence is achieved in $\hat{\mu}^{[k]} = -4.75$ Example is for 305-d protein yield in country B.