



USEFUL ALGORITHMS FOR THE GENETIC EVALUATION
OF LIVESTOCK

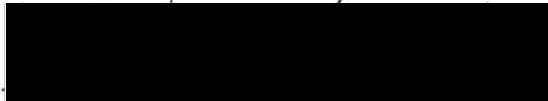
By
Bruce Tier

A thesis submitted for the degree of
Doctor of Philosophy
of The University of New England
December 1997

DECLARATION

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree.

I certify that to the best of my knowledge, any help received in preparing this thesis, and all sources used, have been acknowledged in this thesis.

.....


ACKNOWLEDGEMENTS

This thesis was undertaken and completed while I had the good fortune to be employed at the Animal Genetics and Breeding Unit under a series of research grants provided by the Meat Research Corporation. Much of the work in this thesis was developed as part of these projects.

While this work stems from my own ideas, I must acknowledge the significant contributions that others have made - either in stimulating ideas, posing problems that required solving and/or working together to solve them. They are Dr. Hans-Ulrich Graser and Dr. Johann Sölkner. An invitation from Hans to work in Germany provided the initial stimulation to start the whole study and helped realise Chapters 6 and 8. Johann's visit to AGBU and our exchange of ideas led to the formalisation of some of my ideas described in Chapters 3 and 4.

I had considerable encouragement from many colleagues at and visitors to AGBU and UNE to begin and complete this work. These include Steve Smith, Asko Mäki-Tanila, Karin Meyer, Steve Barwick, Tom Long, Ina Höschele, David Johnston, Johann Sölkner and Johan van Arendonk. I am indebted to my supervisors, Stuart Barker, Keith Hammond and Hans-Ulrich Graser.

Thanks especially to Shirley, Celia and Nadine for their encouragement.

ABSTRACT

The development of Best Linear Unbiased Prediction (BLUP) by Henderson (1973) equipped animal breeders with a tool to evaluate animals for their genetic potential much more efficiently than had been previously possible. However, BLUP requires solving a large system of simultaneous equations (the Mixed Model Equations: MME) which is computationally demanding. It also requires knowledge of the covariances among all random effects. Covariances among additive genetic effects are described by a function of the Numerator Relationship Matrix (\mathbf{A}). A simple method to compute its inverse (\mathbf{A}^{-1}) was necessary before BLUP could be adopted widely. Such a method was found by Henderson (1975, 1976) whereby \mathbf{A}^{-1} could be written directly from a list of animals, their pedigrees and the inbreeding coefficients of all parents. Now, BLUP is the method of choice for genetic evaluation of livestock and its use is widespread across species and countries. This thesis presents some computer algorithms which reduce the time and memory required to evaluate livestock using BLUP. In six chapters, questions relating to the building of the numerator relationship matrix and its inverse, computing inbreeding coefficients and storing and solving the MME are addressed.

Firstly, an alternative approach to building and inverting the numerator relationship matrix using partitioned matrix theory is described. This method is used to demonstrate how gametes can be included in an animal model. This is achieved by treating the gametes as completely inbred animals. Modifications to the usual rules for building \mathbf{A} and \mathbf{A}^{-1} are presented for these ‘animals’. The equivalence of this approach to a combined animal and gametic approach is demonstrated.

An efficient method for computing inbreeding coefficients is described. It identifies the subset of elements of \mathbf{A} which are required to compute its diagonal elements, from which inbreeding coefficients can be easily calculated. The sparse nature of the subset is exploited by using sparse storage techniques. This method minimises the amount of computation that is done, but uses a relatively large amount of memory when compared with some other methods.

Analysis of the coefficient matrix of the MME for multiple trait animal models revealed considerable repetition of some small (trait x trait) blocks. A method to exploit this repetition by storing only one copy of each unique block in the MME is described. This method allows the coefficient matrix to be stored implicitly, requiring a minimum amount of memory, while minimising the amount of input required while solving the equations. In addition, the method of storage and iteration allowed a partial factoring out of the inverse of the genetic covariance matrix among traits (\mathbf{G}^{-1}). This reduced significantly the amount of computation, and hence time, required during each iteration. The algorithm was subsequently tested on two populations of dairy cattle and found to be highly efficient, and it has been implemented already in a number of national evaluation systems.

Methods of factoring out more of the common terms in the coefficient matrix of a multiple trait model were described. Firstly, a further factoring out of \mathbf{G}^{-1} was found to be possible for multiple trait animal models. Additional factoring out resulting from the repeated occurrences of the same values in rows of \mathbf{A}^{-1} was also illustrated. This resulted in a substantial reduction of the amount of computation required per iteration. Secondly, methods of combining the implicit representation of the MME for reduced animal models and factoring out common terms were explored. These exploited the structure of \mathbf{A} and common terms occurring between random and fixed effects. Possible improvements in efficiency were found to be greater for animal models than for reduced animal models. Finally, a transformation of the MME was presented which completely factored out \mathbf{G}^{-1} for multiple trait animal models. This reduced the amount of time required for solving the MME to almost one seventh of that taken by an explicit representation.

The question as to which of these methods to use was left open. Implicit representation of the MME minimises the storage and input requirements. Some population structures are better analysed with a reduced animal model and others with an animal model. Thus population structure was found to have implications for the choice of transformation or factoring technique.

Contents

ABSTRACT	iv
1 Introduction	1
2 Literature Review	3
2.1 Introduction	3
2.2 Best Linear Unbiased Prediction	4
2.3 The Numerator Relationship Matrix \mathbf{A} and its inverse	6
2.4 Solving the Mixed Model Equations	9
2.5 Multiple Trait Models	10
2.6 Transformations of the Mixed Model Equations	11
2.7 Conclusion	12
3 Inverting \mathbf{A} using partitioned matrix theory	13
3.1 Introduction	13
3.2 Rules for \mathbf{A} in matrix form	13
3.3 Inverting the numerator relationship matrix: the partitioned matrix approach	14

4	Analysing gametic variation with an animal model	16
4.1	Introduction	16
4.2	Construction of \mathbf{A} and \mathbf{A}^{-1} : the usual rules	17
4.3	Treating haploids or gametes as conceptual diploids	18
4.4	Equivalence of the model including ‘conceptual diploids’ to the Gametic Model: an example	19
4.4.1	The genetic model	19
4.4.2	The additive + gametic model	20
4.4.3	The conceptual diploid model	21
4.4.4	Equivalence of models	22
4.4.5	Numerical example	23
4.5	Discussion	26
5	An efficient method for computing inbreeding coefficients	27
5.1	Introduction	27
5.2	Computing Inbreeding Coefficients Using the Tabular Method	28
5.3	The Recursive Pedigree Method	29
5.4	Computational Considerations	32
5.5	Simulation Study	33
5.6	Discussion	34
6	Predicting breeding values using an implicit representation of the mixed model equations for a multiple trait animal model	43
6.1	Introduction	44
6.2	Analysis of the mixed model equations	45

6.3	Accommodating Inbreeding	49
6.4	The Implicit Algorithm	50
6.5	Solving the Equations	51
6.6	Example from the Bavarian Dairy Industry	52
6.7	Discussion	54
7	Factoring out common terms in the mixed model equations	56
7.1	Introduction	56
7.2	Analysis	57
7.3	Non-zero elements in \mathbf{A}^{-1}	61
7.4	Example	63
7.5	Factoring out terms between fixed and random effects	64
7.6	Algorithm	64
7.7	Factoring out terms in the reduced animal model	65
7.8	Discussion	68
8	A Computationally Efficient Transformation of the Mixed Model Equations for Multiple Trait Models	70
8.1	Introduction	70
8.2	Models	71
8.3	Transformation	72
8.4	Results and Discussion	73
9	General Discussion	75

List of Tables

4.1	Example data for maternal imprinting analysis (from Schaeffer et al., 1989)	23
4.2	Relationship matrix for pedigrees in Table 4.1 including maternal gametes treated as being diploid	24
4.3	Inverse of the relationship matrix in Table 4.2	24
5.1	Pedigree of a sample population	29
5.2	Numerator Relationship Matrix (\mathbf{A}) for the sample population	30
5.3	The subset of \mathbf{A} required to compute inbreeding coefficients	31
5.4	Elements of \mathbf{A} required for the flagged element to be computed.	32
5.5	Space and time (CPU seconds on a Gould NP1) required to compute inbreeding coefficients by the recursive pedigree method.	35
5.6	State of the vectors at various stages of the algorithm when determining the diagonal of \mathbf{A} for the sample population in Table 5.1.	37
6.1	Contributions to \mathbf{A}^{-1} for animals with different numbers of parents identified	47
6.2	Contributions to \mathbf{A}^{-1} augmented with groups (Westell et al., 1988) for animals with different numbers of parents identified.	53
7.1	Example population	61
7.2	\mathbf{A}^{-1} for the sample population in Table 7.1 ignoring inbreeding	62

7.3	Numbers of different types of off-diagonal elements that occur in the rows of a lower or upper triangular representation of \mathbf{A}^{-1}	63
7.4	Contributions to the MME by the i th animal in the animal model	66
7.5	Contributions to the MME by the i th animal in the gametic model	66