

DATA DEPENDENT FACTORS AFFECTING CONVERGENCE IN ANIMAL MODELS

H. H. SWALVE, R. VOGEL-LACKENBERG, R. REENTS, and E. BRUNS
Institute of Animal Breeding and Genetics
D-3400 Göttingen, FRG

SUMMARY

For dairy cattle evaluation systems that use an animal model, algorithms are needed that aim for a maximum reduction in computing time. Different data sets may vary in convergence rate for solving equations. In West Germany, herds generally are small, which allows block inversion of equations for a system structured by herds. Using this strategy, a rapid rate of convergence was observed for data sets of 67,953 and 190,842 cows. If starting values were used, 35 rounds of iteration were sufficient. Use of starting values was more important for larger data sets. Partial exclusion of pedigree information had little impact on convergence rate.

INTRODUCTION

With large-scale animal models (e.g., joint dairy sire and cow evaluation), large systems of equations have to be solved iteratively. Iteration often is done using Gauss-Seidel techniques. Convergence of predictors for large systems of equations is a crucial point as emphasized by several researchers (Westell, 1984; Wiggans *et al.*, 1988). The aim of the present study was to analyze convergence behavior of such a system under various conditions given by specific data sets.

MATERIAL AND METHODS

The applied algorithm followed the general strategy of Westell (1984) to take advantage of the structure of the data. Dairy data naturally can be grouped by herds, which leads to block structures in mixed model equations. In addition, the algorithm was tailored according to the model used. Under the general model

$$y = Xb + Zu + e$$

where y is a vector of observations, X and Z are design matrices pertaining to fixed and random effects, respectively, and b , u , and e are unknown vectors of fixed effects, additive genetic effects of animals (random), and residuals. Under a single trait model, variances (Var) are computed as $\text{Var}(u) = A\sigma_u^2$ and $\text{Var}(e) = I\sigma_e^2$, where A is the numerator relationship matrix and I is an identity matrix.

For a population with small herds such as in West Germany, a herd class (HC) model by which herds are grouped according to average production still seems justified. The system of equations for a single-trait model was organized in such a way that block structures appeared on the diagonal of the coefficient matrix. The first block included year, HC-year, and season-year effects; the following blocks were animals' additive genetic effects, and animals were grouped by herd. The two last blocks comprised dams without records and sires. A relatively dense off-diagonal area in the coefficient matrix was the sire-by-cow block; less dense blocks were dams without records by cows, cows in herds by cows in herds, sires by dams, and cows by fixed effects. Solutions were obtained by applying inversion and iteration techniques.

Iteration started by obtaining quasi-solutions for fixed effects by inversion of the respective block of the coefficient matrix. Results for fixed effects were used to correct right-hand sides (RHS) for animals ordered by herds, for which then quasi-solutions also were computed using inversion. Correction for relationships across herds was done as soon as solutions from previous herds were available. Finally, solutions for dams without records and sires were obtained by iteration after correction for contributions by cows ordered in herds. Solutions for cows then were used for correction of RHS of fixed effects. A relaxation factor of .35 was used from round 5, altered to .45 in round 15, and then reset to .35 at round 35.

Two data sets from two regions of northwestern Germany were used. Data from region 1 included 67,953 cows in 2578 herds and 2859 sires; data from region 2 comprised 190,842 cows in 4404 herds and 11,276 sires. Both data sets were taken from the routine set used for evaluation of Friesian cows and bulls in northern Germany. Data were provided by Rechenzentrum zur Förderung der Landwirtschaft in Niedersachsen (Agricultural Computing Center, Verden).

RESULTS AND DISCUSSION

Rate of convergence was examined by iterating for 100 rounds under continuous computation of suitable criteria. In Table 1, some parameters of convergence are given for bulls and cows using data from region 2. Correlations between estimates of a given round and round 100 rapidly approached 1. Differences between estimates from a specific round and those of round 100 were relatively small after round 35, and average absolute differences approached 1 kg milk for cows and bulls. Relative difference between consecutive solutions (C_d) as proposed by Misztal *et al.* (1987) confirmed that convergence occurred relatively fast.

Table 1. Convergence parameters for milk yield evaluations (\hat{u}) for region 2.

Gender	Iteration round (i)	Correlation of \hat{u}_i with \hat{u}_{100}	$\hat{u}_i - \hat{u}_{100}$			Mean $ \hat{u}_i - \hat{u}_{100} $	$-\log_{10} C_d^4$
			Min ¹	Max ²	SD ³		
Bulls	12	.9983	-120.6	71.3	18.4	24.6	2.27
	24	.9997	-41.7	2.9	7.4	10.6	2.52
	36	.9999	-14.2	14.9	1.0	1.2	3.08
	48	1.0000	-.1	6.8	1.2	1.7	3.83
	60	1.0000	-.1	4.0	.7	1.0	3.92
	72	1.0000	.0	2.1	.4	.5	4.11
	84	1.0000	.0	1.0	.2	.2	4.32
	96	1.0000	.0	.2	.0	.1	4.52
	100	4.59
Cows	12	.9993	-84.6	34.1	11.8	29.8	2.50
	24	.9999	-33.9	1.5	4.8	13.4	2.56
	36	1.0000	-7.1	8.1	.6	1.2	3.04
	48	1.0000	-.3	4.3	.7	2.1	3.88
	60	1.0000	-.2	2.6	.4	1.2	3.95
	72	1.0000	-.1	1.4	.2	.7	4.15
	84	1.0000	.0	.6	.1	.3	4.35
	96	1.0000	.0	.1	.0	.1	4.56
	100	4.60

¹Min = minimum; ²Max = maximum; ³SD = standard deviation; ⁴ C_d = relative difference between consecutive solutions.

Convergence of solutions for region 1 after excluding part of pedigree data at random is in Table 2. The complete data set included records from 61,582 cows. Excluding dam information for about 50% (30,377) of these cows affected convergence; however, most of the effect was after 50 rounds. Excluding data from 8450 of 12,586 dams with records or from about 50% (30,377) of sires did not affect convergence rate substantially.

Table 2. Convergence¹ of solutions for region 1 data with various pedigree information excluded.

Gender	Iteration rounds	All data ²	Pedigree information excluded		
			Dam ³	Dam with record ⁴	Sire ⁵
Bulls	10	1.62	1.67	1.63	1.59
	20	1.76	1.83	1.79	2.15
	30	1.98	2.07	1.99	2.42
	40	2.63	2.55	2.58	2.85
	50	3.68	2.88	3.32	3.54
Cows	10	1.70	1.77	1.71	1.71
	20	1.80	1.89	1.83	2.36
	30	1.91	2.00	1.93	2.60
	40	2.62	2.51	2.56	2.95
	50	3.58	2.84	3.24	3.64

¹ $-\log_{10}C_d$ where C_d = relative difference between consecutive solutions; ²61,582 cows with records; ³dam information excluded for 30,377 of 61,582 cows; ⁴information from dams with records excluded for 8450 of 12,586 cows that had dams with records; ⁵sire information excluded for 30,377 of 61,582 cows.

In routine applications, estimates from previous evaluations are used as starting values to speed convergence. Use of starting values was simulated, and results are in Table 3 for both regions. The initial data sets included records with first calvings from 1979 to 1986. Then 4-month periods of data from 1987 were added, and results from the initial run were used as starting values. Convergence rates also were calculated for solutions to the complete data sets (first calvings from 1979 to 1987) without using starting values and were similar to those for the initial data set. Convergence improved if solutions from the initial data set were used as starting values for later, more complete data sets. The more new data that were added, the greater was the effect on convergence rate. However, these effects were not extreme. Benefit from using starting values is more obvious for region 2. In any case, convergence indicated by values for $\log_{10}C_d$ between -3.0 and -4.0 is reached after 50 rounds of iteration.

CONCLUSIONS

Under the algorithm applied, rate of convergence was affected only slightly by missing pedigree information or the use of starting values. Starting values would be more important for larger data sets. For the data analyzed, 35 rounds of iteration were sufficient if starting values were used. A reason for the quick rate of convergence probably is the direct quasi-solution for all animals within herds, which is feasible under West German conditions.

Table 3. Convergence¹ of solutions with and without use of starting values² by date of first calving and region.

Region	Gender	Iteration rounds	Date of first calving				
			1/79- 12/86 (NSV)	1/79- 4/87 (SV)	1/79- 8(10)/87 ³ (SV)	1/79- 12/87 (SV)	1/79- 12/87 (NSV)
Region 1	Bulls	6	.98	1.80	1.77	1.72	.98
		12	1.74	2.33	2.33	2.32	1.75
		18	1.84	2.62	2.63	2.62	1.86
		24	1.75	2.62	2.59	2.56	1.76
		30	1.98	2.73	2.71	2.70	1.98
		36	2.29	3.04	3.03	3.01	2.29
		42	2.90	3.69	3.66	3.63	2.85
		48	3.57	4.29	4.23	4.11	3.50
		50	3.68	4.50	4.44	4.37	3.68
		(No.)		2679	2788	2809	2859
	Cows	6	1.14	2.01	1.95	1.91	1.14
		12	1.80	2.52	2.52	2.51	1.82
		18	1.79	2.63	2.62	2.60	1.80
		24	1.77	2.61	2.59	2.57	1.78
		30	1.91	2.69	2.69	2.67	1.91
		36	2.22	2.98	2.97	2.96	2.21
		42	2.88	3.65	3.63	3.58	2.82
		48	3.55	4.31	4.23	4.11	3.41
		50	3.69	4.52	4.45	4.33	3.58
		(No.)		59,679	63,092	64,039	67,953
Region 2	Bulls	6	1.02	1.81	1.76	1.68	1.02
		12	1.57	2.35	2.32	2.27	1.56
		18	1.58	2.50	2.50	2.48	1.60
		24	1.60	2.54	2.54	2.52	1.60
		30	1.78	2.65	2.65	2.64	1.77
		36	2.28	3.10	3.10	3.08	2.25
		42	2.91	3.96	3.93	3.84	2.87
		48	2.91	3.78	3.81	3.83	2.97
		50	2.92	3.76	3.79	3.82	2.98
		(No.)		10,624	11,060	11,180	11,276
	Cows	6	1.13	2.05	1.99	1.92	1.14
		12	1.64	2.54	2.53	2.50	1.63
		18	1.61	2.57	2.57	2.56	1.61
		24	1.64	2.56	2.57	2.56	1.64
		30	1.83	2.69	2.70	2.69	1.81
		36	2.23	3.06	3.06	3.04	2.20
		42	3.10	4.09	4.01	3.89	3.05
		48	2.98	3.81	3.85	3.88	3.05
		50	2.97	3.80	3.83	3.86	3.04
		(No.)		168,954	175,933	182,375	190,842

¹ $-\log_{10}C_d$ where C_d = relative difference between consecutive solutions; ²NSV = starting values not used, SV = solutions from 1/79-12/86 used as starting values; ³region 1 (8/87), region 2 (10/87).

REFERENCES

- MISZTAL, I., GIANOLA, D. and SCHAEFFER, L.R. 1987. J. Dairy Sci. 70: 2577-2584.
 WESTELL, R.A. 1984. Ph.D thesis, Cornell University, Ithaca, NY.
 WIGGANS, G.R., MISZTAL, I. and VAN VLECK, L.D. 1988. J. Dairy Sci. 71(Suppl. 2): 54-69.