Eliminating the Approximation Bias in NCE Accuracy Computations with High Performance Gibbs Sampling

B. L. Golden, D. J Garrick, R. Fernando and S. N. Newman

BIF Accuracy

$$ACC_{BIF} = 1 - \sqrt{\sigma^2_{PEV} / \sigma^2_g}$$

Objective:

Determine if a Gibbs Sampler is better then current approximations for calculating the prediction error variance used in (BIF) accuracy computations.

Better =
- High correlation to inverse diagonal MME elements
- Sufficiently fast on production size problems.
- Provides more information

MT Gibbs Sampler Code

Ax=b

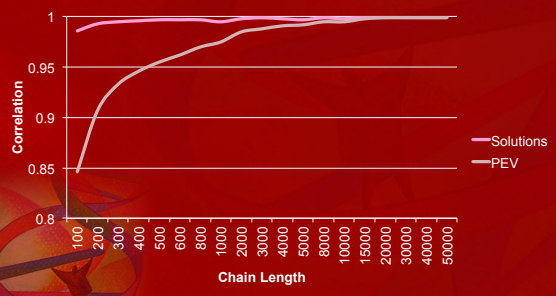
```
...
for( sample=0; sample<nSamples; sample++ )
{
    for( j=0; j<x.num_elem; j++ )
    {
        ax = rowMult( &A, &x, j ) - x.v[j] * diagA.v[j];
        xHat = ( b.v[j] - ax ) / diagA.v[j];
        x.v[j] = xHat + nrdGpu() * sqrtInvDiagA.v[j];
    }
}
...
```

Does Gibbs Sampler give the same PEV as the inverse elements?

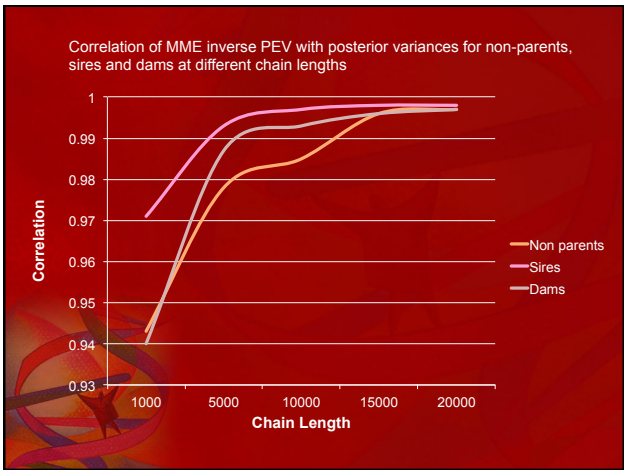
$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

$$\lambda = \frac{\sigma^2_e}{\sigma^2_g}$$

Correlation of posterior means and variances with PCG Solutions and MME inverse elements at different chain lengths



Power Genetics post weaning gain – 23,784 animals, 27,632 equations.
Started with Posterior Mean



Do parallel chains give the same answer as 1 long chain?

- Multi-trait
  - BW, WW, MILK, PWG, SC
  - 550,303 equations
  - 12,712,825 NNZ

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$Var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} A \otimes G_0 & 0 \\ 0 & R \end{bmatrix}$$

Correlations of Posterior Means for Various Sampling Strategies

	sol	1 chain of 20k	1 chain of 100k	5 chains of 4k	10 chains of 2k	10 chains of 3k with 1k burn	5 chains of 20k
sol		.996	1.000	.999	1.000	.996	.999
1 chain of 20k			.998	.994	.996	.987	.993
1 chain of 100k				.998	.999	.993	.997
5 chains of 4k					.999	.998	.999
10 chains of 2k						.996	.999
10 chains of 3k with 1k burn							.999
5 chains of 20k							

Correlations of Posterior Variances for Various Sampling Strategies

	1 chain of 20k	1 chain of 100k	5 chains of 4k	10 chains of 2k	10 chains of 3k with 1k burn	5 chains of 20k
1 chain of 20k		1.000	.999	.999	.999	.999
1 chain of 100k			.999	.999	.999	1.000
5 chains of 4k				.999	.999	.999
10 chains of 2k					1.000	.999
10 chains of 3k with 1k burn						.999

Can Gibbs Sampler be sufficiently fast on production size problems?

Different parallel chain lengths to produce 20,000 samples. Started with PCG solutions

# of Chains	Length of Each Chain	Time (1.29V)	Time (1.5V)	
1	20,000	9m 3s	6m 4s	
5	4,000	2m 27s	1m 58s	
10	2,000	2m 44s	2m 10s	

550,303 Equations
12,712,825 NNZ
5 Genetic Components

Real World Problem

ASA BW, WW, MILK

- 32,173,703 Equations
- 647,105,967 NNZ
- Time to PCG solution 57.706s
- Time to 10,000 samples from 10 chains at 1k samples each (0 burnin):

196m 14s

Example previous work using MCMC Sampler for PEV

- Garcia-Cortes, et al., 1992
- Van Tassell and Van Vleck, 1996
- Hickey, et al., 2009

Conclusions

- Gibbs sampler resulted in high quality PEV estimates that converged to the inverse of the MME;
- Multi-core, overclocking and heterogeneous computing helped make implementing Gibbs Sampling for PEV tractable;
- Parallel chains started with PCG solutions provided a performance improvement and the same answers.

Future Work

Hickey et al., should be performance programmed and tested with longer parallel chains.