Use of high-performance computing in animal breeding

J. Vandenplas*† & N. Gengler*

* University of Liège, Gembloux Agro-Bio Tech – Gembloux, Belgium
† National Fund for Scientific Research – Brussels, Belgium
Contact: jvandenplas@ulg.ac.be

Background

- Most programs for variance components estimations do not allow parallel computing
  - Limited problem sizes
- High-performance computing facilities more and more available
  - Increased performance and problem sizes

Hardware - software

- Intel® Core™ i7 - 3.20 GHz - RAM: 64GB
- REMLF90 and associated libraries (Misztal et al., 2014)
- Solvers + Inversions
  - FSPAK90 (Misztal et al., 2014)
  - PARDISO 5.0.0 (Schenk et al., 2007, 2008; Luisier et al., 2013)
  - METIS 4.0.3 (Karypis and Kumar, 1999)
- OpenMP (http://openmp.org)

Solver implementations

- Reference: original FSPAK90 with multiple minimum degree (MMD) algorithm (ORI)
- FSPAK90 + optimization of « do » loops + OpenMP directives (FMD)
- FSPAK90 + METIS 4.0.3 + optimization of « do » loops + OpenMP directives (FMT)
- PARDISO 5.0.0 + MMD option (PMD)
- PARDISO 5.0.0 + METIS 4.0.1 option (PMT)

Objective: Effective modifications for REMLF90 to reduce real time

Conclusion: Speedups per iteration until 4.5 times for genomic models and until 74.6 times for other models with 6 threads

Data | Context | Model | # traits | # animals | # genotypes | # equations | # VCE |
--- | --- | --- | --- | --- | --- | --- | --- |
D1 | Dairy cattle | TDM | 1 | 102,065 | 0 | 175,887 | 4 |
D2 | | RR-TDM | 1 | 102,065 | 0 | 462,656 | 19 |
D3 | | RR-TDM | 2 | 102,065 | 0 | 925,318 | 65 |
J1 | Jumping horses | RAM | 1 | 55,594 | 0 | 120,463 | 3 |
J2 | | RAM | 2 | 101,262 | 0 | 433,662 | 9 |
P1 | Pigs | SM | 1 | 85,807 | 0 | 426,347 | 13 |
P2 | | SM | 2 | 85,807 | 0 | 1,132,960 | 72 |
S1 | Simulated | AM | 1 | 203,600 | 0 | 203,702 | 2 |
S2 | | SSGM | 1 | 203,600 | 9592 | 203,702 | 2 |
S3 | | SSGM | 1 | 203,600 | 19,155 | 203,702 | 2 |

VCE: variance components estimates different from 0; (RR-)TDM: (random regressions-)test day model; (R)AM: (repeatability) animal model; SM: splines model; SSGM: single-step genomic model

Comparisons

- Real times (s) averaged on 5 REMLF90 iterations (without reordering time)
- Speedup: ratio of average real time of ORI over average real time of PMD or PMT with 1, 2, 4 or 6 threads

Results

**Speedups for the implementations (achieving the highest speedup with 6 threads) for the 10 datasets with 1, 2, 4 and 6 threads**

- Maximum speedup (74.6 times) for P2 with PMT and 6 threads (123s per iteration on average)
- Lowest speedups (between 1.8 and 4.5 times with PMD and 6 threads) for S1, S2 and S3

**Real times averaged on 5 iterations with 1 thread for the 10 datasets**

- PARDISO 5.0.0 (PMD or PMT)
  - Most efficient
- Performance
  - Function of the reordering algorithm
- METIS (FMT or PMT)
  - Mostly most efficient
  - Not appropriate for SSGM

**Contact:** jvandenplas@ulg.ac.be

The authors acknowledge I. Aguilar, D. Colignon, D. François, K. Gaertner, G. Karypis and S. Tsuruta for their advices, and C. Bastin, M. Dufrasne and S. Janssens for their permissions to use their datasets.

J. Vandenplas acknowledges the National Fund for Scientific Research for his Research Fellow position and the ULg for a travel grant.