OPTIMIZING GENOMIC PREDICTION:
STRATEGIES TO OBTAIN INVERSE OF LARGE RELATIONSHIP MATRICES

P. Faux 1, I. Misztal 2 and N. Gengler 1, 3
1Animal Science Unit, Gembloux Agro-Bio Tech, University of Liège, B-5030 Gembloux, Belgium
2Department of Animal and Dairy Science, University of Georgia, Athens 30602, GA, USA
3National Fund for Scientific Research, B-1000 Brussels, Belgium

CONTEXT: Genomic prediction of breeding values

- Genomic prediction (GBLUP) aims to link the phenotypic variability to the genome-wide variability, which is detected by thousands of Single Nucleotide Polymorphisms (SNP)
- An innovative and simple procedure (single-step GBLUP, Misztal et al., 2009) performs genomic prediction by using all 3 sources of information at the same time (phenotypes, genotypes and pedigrees)
- Key point of ssGBLUP: use of an enhanced relationship matrix (H. Legarra et al., 2009) as covariance matrix between random genetic effects, instead of pedigree-based relationship matrix (A) in mixed model equations (MME)

\[ \begin{bmatrix} X X & X Z \\ Z X & Z Z + c I \end{bmatrix} H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{A}^{-1} \end{bmatrix} \]

Left-hand side of MME

Structure of inverse of H

- In order to get H^{-1}, 3 matrices have to be inverted:
  - A \rightarrow Sparse inverse easily computed using particular rules (see below)
  - G \rightarrow Genomic relationship matrix between genotyped animals; inverse is dense
  - A_{A} \rightarrow Pedigree-based relationship matrix between genotyped animals; inverse may be sparse in some cases (e.g. weakly-related animals)

METHODS: Creation of a sparse triangular decomposition of the inverse

- Direct creation of inverse of A (Henderson, 1976):
  \[ A^{-1} = (T^{-1})^T D T^{-1} \]
- T^{-1} is an identity matrix (I(n)) that is filled in its lower part with max. (2n-3) elements equal to -0.5
- D^{-1} is a diagonal matrix with max. 3 different values: 2, 4/3 or 1
- Based on this a priori knowledge of values, any element of A^{-1} can directly be computed without computation of the whole matrix
- Rules are modified to fit the case of G by the creation of a sparse approximation of T^{-1} (T^{−1}):
  \[ G^{-1} = (T^{-1})^T D G (T^{-1}) \]
- For a given animal, elements of T^{-1} are determined by linear regression of relationships coefficients of closely-related animals (i.e. having a relationship with this animal lower than p, an arbitrary threshold) on the relationships coefficients they have with this animal
- Computing the reverse equation, D = T^{−1}G(T^{−1})^T, we obtain a “close-to-diagonal” matrix D
- To approximate inverse of D, it is either submitted to a new round (with a new p) into the same process of approximation, or computed as diagonal matrix made of inverse of its diagonal elements
- Because T^{−1} is sparse, creation of G^{−1} involves less computations

OBJECTIVE: Approximation of inverse of large relationship matrices

- G and A_{A} inverted so far using dense matrix inversion algorithms (e.g. Gaussian elimination)
- Increase of number of genotyped animals = cubic increase of computation cost of these inverses
- Therefore, an algorithm for a direct approximation of the inverses of G and A_{A} is required

RESULTS: Different tests on 2 sets

- 1,716 dairy bulls: Several rounds of approximation of inverse of G, assessments of closeness of approximation and sparsity, correlation between MME’s solutions for genotyped animals computed either approximate or real inverse (TABLE 1)
- 4,536 broiler chickens: Several rounds of approximation of inverse of A_{A}, same criterions as for dairy bulls and measure of elapsed CPU time for inversion (TABLE 2)

<table>
<thead>
<tr>
<th>Round</th>
<th>p</th>
<th>MSD1</th>
<th>% of 02</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.21</td>
<td>66.53 * 10^{-4}</td>
<td>92.35</td>
<td>0.79</td>
</tr>
<tr>
<td>2</td>
<td>0.017</td>
<td>34.87 * 10^{-4}</td>
<td>10.78</td>
<td>0.91</td>
</tr>
<tr>
<td>3</td>
<td>0.009</td>
<td>16.81 * 10^{-4}</td>
<td>2.28</td>
<td>0.97</td>
</tr>
<tr>
<td>4</td>
<td>0.005</td>
<td>5.97 * 10^{-4}</td>
<td>0.86</td>
<td>0.99</td>
</tr>
<tr>
<td>5</td>
<td>0.003</td>
<td>1.82 * 10^{-4}</td>
<td>0.49</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round</th>
<th>p</th>
<th>MSD1</th>
<th>% of 02</th>
<th>r</th>
<th>CPU time1 (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.13</td>
<td>1.63 * 10^{-6}</td>
<td>98.52</td>
<td>0.77</td>
<td>170.27</td>
</tr>
<tr>
<td>2</td>
<td>0.005</td>
<td>2.92 * 10^{-6}</td>
<td>96.19</td>
<td>&gt;0.99</td>
<td>371.47</td>
</tr>
</tbody>
</table>

1. Statistics for Mean Square Differences, i.e. averaged square difference between elements of real and approximated inverses
2. CPU time have to be compared with elapsed CPU time (1805 s) for real inversion of A_{A} using a non-optimized gaussian elimination algorithm

CONCLUSIONS: A new insight

- For both A_{A} and G, the algorithm tends to provide real inverse: decreasing MSD and increasing r show that approximated inverse comes closer to real inverse throughout rounds
- As the triangular decomposition stays highly sparse for A_{A}, the algorithm is suitable for inversion of this matrix
- For the case of G, this algorithm brings a new insight by computing a covariance matrix of MME for genetic random effects that blends both genomic and pedigreed informations

REFERENCES

ACKNOWLEDGEMENTS
This research is part of the ARPI-FPS grant PhD-113-118 (Fonds National de la Recherche, Luxembourg) project NextGenGUES done under the Private-Public-Partnership agreement between CONVIS (Etterb, Luxembourg) and UG (UGA) (formerly PSUGIS). Nicolas Gengler (Prestigious Senior Research Achievable) acknowledges the National Fund for Scientific Research (FNRS) for the support. The authors gratefully acknowledge support by the Federation Association USA Inc. (Stratford, VT), the Cooperative Dairy DNA Repository (Beltsville, MD) and Colibi-Ventures (Silver Spring, MD).

Contact: pierre.faux@ulg.ac.be