An evaluation of ‘deflation’ to improve convergence rates for single-step genomic evaluation with the hybrid model

Karin Meyer and Andrew A. Swan
Animal Genetics and Breeding Unit
University of New England, Armidale NSW 2351

AAABG 2019
Introduction

- Single step genomic evaluation has become routine
- Extension of breeding value model most common
  - Conceptually simple adaptation of existing software
- Alternative: “Hybrid model” for additive genetic effects
  - Fit breeding values for non-genotyped
  - Fit marker effects for genotyped
  - Slow to converge when solving MME via PCG
- Suggestion: Second level pre-conditioner
  - “Deflation” of coefficient matrix

Introduction

- Single step genomic evaluation has become routine
- Extension of breeding value model most common
  - Conceptually simple adaptation of existing software
- Alternative: “Hybrid model” for additive genetic effects
  - Fit breeding values for non-genotyped
  - Fit marker effects for genotyped
  - Slow to converge when solving MME via PCG
- Suggestion: Second level pre-conditioner
  - “Deflation” of coefficient matrix
- Pilot study: apply to a practical data set

Pre-conditioned conjugate gradient

- Solve system of MME by iterations
  \[ Cx = r \]
  - Conjugate gradient scheme
  - Convergence rate depends on condition no. of \( C \)
- ‘Pre-condition’ to reduce condition number (PCG)
  \[ M^{-1}Cx = M^{-1}r \]
  - Best: \( M^{-1}C \) close to an identity matrix
  - Choice: compromise with computational burden
  - Simple but effective \( M = \text{Diag}(C) \)
Deflated PCG

- Second level pre-conditioner (DPCG)

\[ M^{-1} P C x = M^{-1} P r \quad \text{with} \quad P = I - CS (S'CS)^{-1} S' \]

- \( P \) eliminates ‘unfavourable’ eigenvalues of \( C \)
  - Projection of \( C \) on a suitable subspace defined by \( S \)
  - Reduces (‘deflates’) condition no.

- Require choice of linearly independent columns of \( S \)
  - Best: close to unfavourable eigen-vectors of \( C \)
  - Choice: compromise efficacy/extra computations

- Assign equations to non-overlapping ‘sub-domains’
  - Each row of \( S \) has one non-zero element:
    if equation \( i \in \text{sub-domain } j \) \( \rightarrow \) element \( s_{ij} = 1 \)
Deflated PCG

- Second level pre-conditioner (DPCG)

\[ M^{-1} PC x = M^{-1} Pr \quad \text{with} \quad P = I - CS (S'CS)^{-1} S' \]

- \( P \) eliminates ‘unfavourable’ eigenvalues of \( C \)
  - Projection of \( C \) on a suitable subspace defined by \( S \)
  - Reduces (‘deflates’) condition no.

- Require choice of linearly independent columns of \( S \)
  - Best: close to unfavourable eigen-vectors of \( C \)
  - Choice: compromise efficacy/extra computations

- Assign equations to non-overlapping ‘sub-domains’

  SIMPLE

  e.g. \( S = \begin{bmatrix} 1_a & 0 & 0 \\ 0 & 1_b & 0 \\ 0 & 0 & 1_c \end{bmatrix} \)
**Data & model**

- 1,206,908 records for eye muscle depth (PEMD)
- Terminal sire sheep breeds
  - Poll Dorset, Suffolk, White Suffolk and Texel
  - + 18 less numerous breeds
- 1,698,838 animals in pedigree
- 23,040 genotypes

**Model of analysis**

- 1,675,798 additive genetic effects (non-genotyped)
- 48,599 marker effects
- 93 genetic groups (random)
- 56,212 sire × flock-year (random)
- 54,094 contemporary groups (fixed)
Analyses

- Solve MME using PCG and DPCG for different $S$ and marker subsets
  - 1st level pre-conditioner: $M = \text{Diag}(C)$
- Simple GWAS: select 28,875 SNP with $p < 0.5$
- Sub-domain assignments
  - Fixed effects: single domain
  - Random effects except markers:
    - Single domain
    - As $A$ + separate domain(s) for genetic groups (1 or 93)
  - Markers: domains for subsequent chunks of equations
    - Chunk sizes: 200, 100, 50, 20, 10, 5
- Criteria: Number of iterates & wall time
Convergence rates

Sub-domains for marker effects only

<table>
<thead>
<tr>
<th>Number of subdomains</th>
<th>All SNPs (48599)</th>
<th>SNPs with p&lt;0.5 (28875)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chunk size for markers:
- None
- 200
- 100
- 50
- 20
- 10
- 5

Chunk size for genetic groups:
- None
- 93
- 1

No Deflation
Chunk size 200
Chunk size 50
Convergence rates

Subdomains for marker effects only

<table>
<thead>
<tr>
<th>Number of subdomains</th>
<th>Number of iterates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>50</td>
<td>600</td>
</tr>
<tr>
<td>10</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
</tr>
<tr>
<td>20</td>
<td>1200</td>
</tr>
<tr>
<td>100</td>
<td>1400</td>
</tr>
<tr>
<td>200</td>
<td>1600</td>
</tr>
<tr>
<td>400</td>
<td>1800</td>
</tr>
<tr>
<td>800</td>
<td>2000</td>
</tr>
<tr>
<td>1600</td>
<td>2200</td>
</tr>
<tr>
<td>3200</td>
<td>2400</td>
</tr>
<tr>
<td>4000</td>
<td>2600</td>
</tr>
</tbody>
</table>

Correlation of breeding values
- Genotyped: 0.995
- Non-genotyped: 1.000

Chunk size for markers:
- None
- 200
- 100
- 50
- 20
- 10
- 5

Chunk size for genetic groups:
- None
- 93
- 1
Convergence rates

Adding a single sub-domain for genetic groups

- All SNPs (48599)
- SNPs with p<0.5 (28875)

Number of subdomains vs. Number of iterates

- Chunk size for markers:
  - None
  - 200
  - 100
  - 50
  - 20
  - 10
  - 5

- Chunk size for genetic groups:
  - None
  - 93
  - 1
Convergence rates

Assigning a separate sub-domain for each genetic groups

- All SNPs (48599)
- SNPs with p<0.5 (28875)

Number of subdomains

Number of iterates

Chunk size for markers:
- None
- 200
- 100
- 50
- 20
- 10
- 5

Chunk size for genetic groups:
- None
- 93
- 1
Results confirm reports by Vandenplas et al. 2018

Most interesting: effectiveness of adding separate sub-domain(s) for genetic groups

Additional analyses (not shown) demonstrated

- Improved convergence when assigning small numbers of sub-domains for additional random effects
  
  e.g. additive genetic effects for non-genotyped animals in HM

- Improved convergence for breeding value model when deflating equations for genetic groups
Computing time

![Graph showing the relationship between the number of subdomains and the elapsed time for computing, with different chunk sizes for markers and genetic groups.](image)
Computing time

![Graph showing the relationship between the number of subdomains and elapsed time. The graph includes markers for different chunk sizes for markers and genetic groups.]

48599 SNPs

- **Chunk size for markers**
  - None
  - 200
  - 100
  - 50
  - 20
  - 10
  - 5

- **Chunk size for genetic groups**
  - None
  - 93
  - 1
Conclusions

- Deflation of the coefficient matrix in MME
  - Can reduce condition number
  - Can improve convergence of PCG dramatically
    - Advantageous for hybrid & breeding value models
  - Small(ish) number of separate sub-domains for multiple random effects appears most effective
  - Supplement by other means to improve condition number & reduce computational demands per iterate
    - e.g. reduced marker panel
  - Balance of convergence rates and computing times
    - depends on model, data & implementation

- Valuable addition to toolkit for genomic evaluation