MiX99 features for large models

Martin Lidauer

MiX99 workshop 2014, Tuusula, Finland
Which models are large?

Models for which, given a certain computing environment, solving is reaching limits

- Solving time is long or even unacceptable long
- Memory requirements are reaching limitations
- Hard disk space is limited

Large? Maybe if:

- Number of traits $>> 10$
- Number of records $>> 1$ million
- Number of animals $> 1$ million
- Number of equations $> 50$ million
- Non-linear models
MiX99’s large-model features aim for

- Reducing computing time
  - Avoiding reading of redundant information
  - Optimizing iteration-on-data calculations by providing information about the structure of the data and model
  - Improving convergence
  - Parallel computing

- Reducing memory requirements
  - Options for fitting models with reduced dimension
  - Reduced memory use option for iteration-on-data
Features for large models

MiX99 is offering several opportunities to speed up solving or reduce memory requirements

- Data sorting
- Input data design
  - Grouping of traits
  - Table values for covariables
- Parallel processing
- Pre-conditioning
- Detection of convergence
- Options for fitting models with reduced dimensions

The suitability of a feature depends on the model
Data sorting

Beneficial for data with repeated observations and essential for parallel computing

- 3 sorting levels:
  - by data blocks
    herd, cohort, country, ...
  - by relationship code within block
    animal ID, sire ID, progeny ID, ...
  - CAUTION: For models with a maternal or paternal effect and a sire-maternal-grand-sire relationship matrix the sire ID must not be used as sorting variable!
  MiX99 stores relationship information only once for all records with same relationship code!
- by trait group within relationship code
Data sorting

- CLIM syntax
  ```
  DATASORT BLOCK=Herd PEDIGREECODE=Animal
  ```

- Sorting by BLOCK or PEDIGREECODE is optional
- Sorting by trait group code is mandatory
- If BLOCK is specified, it has to be given in the pedigree and data file

Both, pedigree and data file have to be sorted by the BLOCK variable

Only data file has to be sorted by the PEDIGREECODE variable
Input data design

For certain multiple-trait models it is possible to structure data in a way that significantly speeds up iteration-on-data.

For such models MiX99 allows grouping of traits, given:

- traits are measured at different time or environment;
- and there exists no residual correlations between traits of different trait groups;
- or the residual correlation between traits of different trait groups is modelled by a random (e.g. permanent environment) effect.
Input data design

Example A: RRM with 4 traits: milk and protein in 1\textsuperscript{st} and 2\textsuperscript{nd} lactation

Data file
without grouping of traits

<table>
<thead>
<tr>
<th>Classification variables</th>
<th>Covariables</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hrd  Ani  Ag1  Ag2  Se1  Se2  C11  C12  C21  C22  M1  P1  M2  P2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.  .  7  0  5  0  .  .  -16.  -16.  12.1  3.40  -16.  -16.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  7  0  6  0  .  .  -16.  -16.  8.7  3.52  -16.  -16.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  0  17  0  10  -16.  -16.  .661  .035  -16.  -16.  28.2  3.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  0  17  0  10  -16.  -16.  .430  .087  -16.  -16.  32.7  -16.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

data file
with grouping of traits

<table>
<thead>
<tr>
<th>Classification variables</th>
<th>Covariables</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hrd  Ani  Trg  Age  Sea  Cvl  Cv2  Mlk  Prt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.  .  1  7  5  .967  .042  12.1  3.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  1  7  6  .562  .084  8.7  3.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  2  17  10  .661  .035  28.2  3.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34  10  2  17  10  .430  .087  32.7  -16.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CLIM syntax

**DATASORT BLOCK**=Hrd **PEDIGREECODE**=Ani

**TRAITGROUP**  Trg

**MODEL**

Mlk(1) = Hrd LCurve(1 Cvl Cv2| Sea) Age Ani G(1 Cvl Cv2| Ani)  
Prt(1) = Hrd LCurve(1 Cvl Cv2| Sea) Age Ani G(1 Cvl Cv2| Ani)  
Mlk(2) = Hrd LCurve(1 Cvl Cv2| Sea) Age Ani G(1 Cvl Cv2| Ani)  
Prt(2) = Hrd LCurve(1 Cvl Cv2| Sea) Age Ani G(1 Cvl Cv2| Ani)
Input data design

For some models with regression functions it is possible to store covariables in a table

Example A: There are only 305 different sets of covariables

Table showing covariables:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hrd</td>
<td>Ani</td>
<td>Trg</td>
<td>Age</td>
<td>Sea</td>
<td>DIM</td>
<td>M</td>
<td>P</td>
</tr>
<tr>
<td>34</td>
<td>10</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>268</td>
<td>12.1</td>
<td>3.40</td>
</tr>
<tr>
<td>34</td>
<td>10</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>301</td>
<td>8.7</td>
<td>3.52</td>
</tr>
<tr>
<td>34</td>
<td>10</td>
<td>2</td>
<td>17</td>
<td>10</td>
<td>15</td>
<td>28.2</td>
<td>3.37</td>
</tr>
<tr>
<td>34</td>
<td>10</td>
<td>2</td>
<td>17</td>
<td>10</td>
<td>48</td>
<td>32.7</td>
<td>-16.</td>
</tr>
</tbody>
</table>

CLIM syntax:

```
TABLEFILE covar.tab
TABLEINDEX DIM
DATASORT BLOCK=Hrd PEDIGREECODE=Ani
.
TRAITGROUP Trg
.
MODEL
  Mlk(1) = Hrd LCurve(1 t1 t2| Sea) Age Ani G(1 t1 t2| Ani)
  Prt(1) = Hrd LCurve(1 t1 t2| Sea) Age Ani G(1 t1 t2| Ani)
  Mlk(2) = Hrd LCurve(1 t1 t2| Sea) Age Ani G(1 t1 t2| Ani)
  Prt(2) = Hrd LCurve(1 t1 t2| Sea) Age Ani G(1 t1 t2| Ani)
  .
```
Input data design

Example B: Udder Health Evaluation Model for Nordic RDC

- MT random regression model
  - 3 TD-SCS traits: 1., 2. & 3. lactation
  - 4 clinical mastitis traits: 1. (2 traits), 2. & 3. lactation
  - 2 udder type traits
- Residual correlations are modelled by VCV matrix for PE effects
- 85 million records, 5.8 million animals, 157 million equations in MME

CLIM syntax

...
Parallel processing

- MiX99 has solver programs that can use several CPUs /cores at the same time
- Best speedup by means of minimizing communication
  - maximum data locality within process
  - ordering the equations in the MME to get a nearly doubly-bordered block diagonal form for the coefficient matrix
- MiX99 provides two features to meet these requirements
  - Sorting of the data by a suitable variable to get data locality (DATASORT BLOCK=<sorting variable>)
  - Arranging model effects within or across blocks (WITHINBLOCKORDER <effect names>)
Parallel processing

Example C:
milk yield is modelled by a RR model including effects for:
herd-test-day, age, lactation curve × year-season, PE, and animal effect

CLIM syntax

```
DATASORT BLOCK=HERD PEDIGREECODE ANI
WITHINBLOCKORDER HTD PE G
MODEL
 milk = HTD AGE LC(t1 t2 t3 t4 t5| YS) &
      PE(t1 t2 t3| ANI) G(t1 t2 t3| ANI)
PARALLEL 4 1
```

Non-zeros of coefficient matrix form doubly-bordered block diagonal matrix

Only common block equations and equations of herd-changers need to be communicated between cores

Common block equations << 10%
Parallel processing

Example D:
Nordic test-day yield evaluation for Holstein
- 252 million observations
- 380 million unknowns
- 2500 PCG iterations
- MiX99 parallel solver using different number of cores
Preconditioning

• PCG would converge in one step if all eigenvalues of $CM^{-1}$ would be the same, i.e. the preconditioner matrix $M$ would be equal to the coefficient matrix $C$

• Hence, try to find a preconditioner matrix $M$ that approximates $C$

• However better preconditioner matrices often increase computations

• MiX99 allows to specify for each effect the preconditioner type
  • Diagonal
  • Block diagonal (block size equal to number of traits, or size of VCV-matrix)
  • Full block (only for across block fixed effects)

• A good choice to start with:
  • **Fixed effects**: Block diagonals
  • **Random effects**: Diagonals
Preconditioning

Example E:

- Multiple trait RR test-day model with the following effects:
  - fixed effects: herd-year, age, year-season, lactation curve × year-season
  - random effects: herd-TD, RR functions for herd-curve, pe and animal

- Preconditioner alternatives
  - A: **Diagonal** for all effects
  - B: **Block diagonal** for all effects (block sizes: fixed 9; random 9, 27, 36, 36)
  - C: **Full block** for all fixed effects, **Block diagonal** for all random effects

<table>
<thead>
<tr>
<th>Preconditioner alternative</th>
<th>Number of Iterations</th>
<th>Solving Time (min)</th>
<th>Size of Pre-conditioner (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Diagonals</td>
<td>3725</td>
<td>56.3</td>
<td>8</td>
</tr>
<tr>
<td>B: Block diagonal</td>
<td>584</td>
<td>13.6</td>
<td>140</td>
</tr>
<tr>
<td>C: Block diagonal + Full block</td>
<td>598</td>
<td>24.0</td>
<td>250</td>
</tr>
</tbody>
</table>
Convergence

• By definition, given $C$ is positive definite, each conjugate gradient step will yield estimates which are closer to the true solutions
• However, convergence characteristics is affected by many factors
• Overall, larger and more complex models will require more iterations to reach convergence
• Very poor convergence, or even divergence, indicates that the model is ill-conditioned and requires improvements
  • Variance components: are matrices almost singular?
  • Size of pedigree versus phenotype information?
  • Quality of pedigree?
  • Sparseness of observations in multiple trait models?
  • Confounding of environmental and genetic effects?
  • Error in the model input instructions?
When to stop iterations?

MiX99 reports 3 convergence indicators (norms)

- **CA:** Relative difference between left-hand and right-hand side of the additive genetic effect equations
- **CR:** Relative difference between left-hand and right-hand side of the MME
- **CD:** Relative differences between solutions of consecutive iteration rounds

- \( cd(k) < 10^{-5} \) indicates convergence, often \( cd(k) < 10^{-4} \) is enough

For routine evaluations, optimal stopping point depends on publishing precision of EBVs

\[
c_a(k) = \sqrt{\frac{(r - C\hat{a}(k))^T (r - C\hat{a}(k))}{(r^T r)}}
\]

\[
c_r(k) = \sqrt{\frac{(r - C\hat{s}(k))^T (r - C\hat{s}(k))}{(r)^T (r)}}
\]

\[
c_d(k) = \sqrt{\frac{(\hat{s}(k) - \hat{s}(k-1))^T (\hat{s}(k) - \hat{s}(k-1))}{(\hat{s}(k))^T (\hat{s}(k))}}
\]

where \( r, s, a \) = vector of right-hand side, solutions, add. gen. effects; \( C \) = coefficient matrix of MME; and \( k \) = iteration round
Convergence indicators

Example E: Convergence when applying different preconditioning
Reducing model dimension

Complex multiple-trait or random regression models are often over parameterized

• Investigating the eigenvalues of applied variance components often reveals possibilities for reducing the model dimensions

• Advantage:
  • Significant reduction in memory requirements
  • Improved convergence characteristics
  • Reduced solving time

The next presentation will deal with this topic

THANK YOU