Relationship matrices
and
Iterative construction of their inverses

Pierre Faux & Nicolas Gengler
University of Liège, Gembloux Agro-Bio Tech, Belgium
AFR Beneficiary (FNR)
Summary

• Theoretical measurement of similarity between related

• Construction of the matrix $A$

• Direct creation of the inverse of $A$

• Iterative construction of inverses using the inverse factorization

• Application to the case of the inverse of $G$
Theoretical measurement of similarity between related individuals
(“Génétique Quantitative”, Gengler, 2012)
Relationships and Similarity

- When observing phenotypes:
  - A son has a certain similarity with his father
- Relationships might help while predicting unknown animals
- How to scale this similarity? Theoretical view:
  - What is similarity at a molecular level?
  - How to express this similarity from molecular to individual level?
Relationship Measurement

• Let us have a biallelic gene (either yellow or red)

• 2 possible configurations
  – Homozygous (identical)

  – Heterozygous (non-identical)
• Homozygosity is imputed to the same common ancestor
• The « allele’s pathway » is known
IBD at a molecular level …

… is expressed as inbreeding at the individual level

– Inbreeding coefficient \( (F) = P(\text{IBD}) \), is function of \( d \)
– Inbreeding coefficient = ½ of relationship coefficient \( (a_{X,Y}) \) between parents,
– Relationship coefficient is only **additive part** of relationship
How to assess P(IBD) for animal Z?

Different gene configurations:

Thus, \( P(\text{IBD}) = \frac{2}{16} = 0.125 = F_Z \)

And \( a_{X,Y} = 2 \times 0.125 = 0.25 \)
Easier method to assess $P(\text{IBD})$

- Count $n$, lowest number of steps to join X to Y through each of $x$ common ancestors
- Relationship coefficient between X and Y:
  \[ a_{X,Y} = (2^{-n})_1 + (2^{-n})_2 + ... + (2^{-n})_x \]

- Halfsibs: $2^{-2} = 0.25$
- Sibs: $2^{-2} + 2^{-2} = 0.5$
- Child to parent: $2^{-1} = 0.5$
- Grandchild to grandparent: $2^{-2} = 0.25$
• May ease more complicated cases
  – Example: various common ancestors at various levels

- $a_{K,J} = 2^{-5} + 2^{-3} + 2^{-3} = 0.28125$
  
- $= \frac{1}{32} + \frac{1}{8} + \frac{1}{8} = \frac{9}{32}$
Construction of the additive genetic relationship matrix
Tabular Method: Views

• The previous method needs to be streamlined
  – May be executed iteratively
  – Use of a symmetric matrix \( n\times n \) (noted \( A \)) that will content relationships between all individual

• Requires to re-organize pedigree data
  – A given animal always needs to appear after both parent

• Proceeds by filling successive squares
  – From square 1*1 to square \( n\times n \)
  – For animal \( i \), firstly row \( i \) from 1 to \( i \), including element \( (i,i) \), secondly column \( i \) from 1 to \( i-1 \) (using symmetry)
  – At each row, sum of weighted contributions from sire, dam, and animal himself
Tabular Method: Rules

1. For a given population of \( n \) individuals (sorted by generation order), define a square matrix of size \( n \)

2. At line \( j \), for all \( i \) younger than \( j \), the relationship between \( i \) and \( j \) is equal to the sum of the half of the relationship of known parents of \( j \) with \( i \)

3. At line \( j \), diagonal element equal 1, plus half of the relationship between parents of \( j \)

4. At line \( j \), paste line \( j \) (from 1 to \( i \)) in column \( j \)

5. Go back to point 2.
Tabular Method: Iterative process

- For animal $i$, firstly row $i$ from 1 to $i$, including element $\left(i, i\right)$, secondly column $i$ from 1 to $i-1$.
Tabular Method: Example

• Numbered pedigree:

<table>
<thead>
<tr>
<th>Animal</th>
<th>Père</th>
<th>Mère</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>E</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>G</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>H</td>
<td>D</td>
<td>F</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td>J</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td>K</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>L</td>
<td>J</td>
<td>K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Animal</th>
<th>Père</th>
<th>Mère</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>
# Tabular Method: Example

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0,5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0,5</td>
<td>0,5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Tabular Method: Example

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5*</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>+ 0.5*</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>+ 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1+0*1/2</td>
</tr>
</tbody>
</table>

= 7 | 0 | 0.5 | 0.5 | 0 | 0 | 0.25 | 1 |
### Tabular Method: Example

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/2</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>1/2</td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td></td>
<td>1/2</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
<td>3/16</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
<td>1/4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
<td></td>
<td>1</td>
<td>1/4</td>
<td>1/2</td>
<td>1/8</td>
<td>1/8</td>
<td>5/16</td>
<td>7/32</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
<td>1/4</td>
<td>1</td>
<td>1/8</td>
<td>1/2</td>
<td>1/2</td>
<td>5/16</td>
<td>13/32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1/4</td>
<td>1/4</td>
<td>1/2</td>
<td></td>
<td>1/2</td>
<td>1/8</td>
<td>1</td>
<td>1/16</td>
<td>1/16</td>
<td>17/32</td>
<td>19/64</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1/4</td>
<td>1/4</td>
<td>1/2</td>
<td>1/8</td>
<td>1/2</td>
<td>1/16</td>
<td>1</td>
<td>1/2</td>
<td></td>
<td>17/32</td>
<td>33/64</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1/4</td>
<td>1/4</td>
<td></td>
<td>1/8</td>
<td>1/2</td>
<td>1/16</td>
<td>1/2</td>
<td>1</td>
<td>9/32</td>
<td>41/64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1/8</td>
<td>1/4</td>
<td>1/8</td>
<td>1/4</td>
<td>1/4</td>
<td>5/16</td>
<td>5/16</td>
<td>17/32</td>
<td>17/32</td>
<td>33/32</td>
<td>42/64</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1/16</td>
<td>1/4</td>
<td>3/16</td>
<td>1/8</td>
<td>3/8</td>
<td>7/32</td>
<td>13/32</td>
<td></td>
<td></td>
<td>41/64</td>
<td>73/64</td>
<td></td>
</tr>
</tbody>
</table>
Direct creation of the inverse of the matrix A
• Additive genetic relationship matrix $A$ will structure the covariance of breeding values of animals in a related population.

• Used in the LHS of the MME for an animal model:

\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + \lambda \cdot A^{-1}
\end{bmatrix}
\]
Factorization of A

• Root-free Cholesky factorization of matrix $A$
  \[ A = TDT' \]

• Structures of matrices $T$ and $D$:
  - $T$: for any animal $i$, with sire $s$ and dam $d$, the row is filled in with the same rules as for $A$:
    \[ T_i = 0.5T_s + 0.5T_d ; \quad T_{i,i} = 1 \]
  - $D$: for any animal $i$, only one element, equal to
    \[ D_{i,i} = \begin{cases} 
    1 & | \quad 0.75 - 0.25(F_p) \quad | \quad 0.5 - 0.25(F_s + F_d) 
    \end{cases} \]
Inverse of the factorization of $A$

- Factorization of the inverse of $A$
  \[ A^{-1} = (T^{-1})' D^{-1} T^{-1} \]

- Structures of matrices $T^{-1}$ and $D^{-1}$:
  - $T^{-1}$: for any animal $i$, with sire $s$ and dam $d$, only 3 elements are filled in
    \[ T^{i,s} = -0.5 \quad ; \quad T^{i,d} = -0.5 \quad ; \quad T^{i,i} = 1 \]
  - $D^{-1}$: for any animal $i$, inverse of $D_{i,i}$
Inverse of the factorization of $A$

- Structure of matrix $T^{-1}$:
  
  $d, s$ are parents of $i$
• Inverse of $A$ can be viewed as a sum of sparse matrix:

$$A^{-1} = (A^{-1})_1 + (A^{-1})_2 + \ldots + (A^{-1})_n$$

• Each sparse matrix is created as:

$$(T^i)' \times D_{i,i} \times T^i$$

• Max. 3 non-zeros entries in $T^i$, each sparse matrix is thus a block of max 9 non-zeros entries (with known positions: animal, sire, dam)

• Time for creation is thus linearly related to size of the population
Direct creation of the inverse of A

- Three possible configurations (when no inbreeding)
  - 0:
    \[
    \begin{bmatrix}
      1 & 1 & 1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      -0.5 \\
      1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      1/3 & -2/3 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      1 \n    \end{bmatrix}
    \]
  - 1:
    \[
    \begin{bmatrix}
      1 & 4/3 & -0.5 & 1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      1 & -2/3 & 4/3 \\
    \end{bmatrix}
    \]
  - 2:
    \[
    \begin{bmatrix}
      -0.5 & 2 & -0.5 & -0.5 & 1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      1/2 & 1/2 & -1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      1/2 & 1/2 & -1 \\
    \end{bmatrix}
    \]
    \[
    \begin{bmatrix}
      -1 & -1 & 2 \\
    \end{bmatrix}
    \]
Iterative construction of inverses using the inverse factorization
Alternative way of direct creation

- Let us define:
  - \( Z_j = A \), from row 1 to \( i \) and from column 1 to \( i \)
  - \( y_j = A \), from row 1 to \( i \), at column \( j \)
  - \( b_j = (Z_j)^{-1} y_j \)

- For this particular case, computation of \( b_j \) is trivial

- Link with tabular method!
Alternative way of direct creation

• Let us now define $Z_k$ and $(Z_k)^{-1}$, $(k=j+1)$

$$Z_k = \begin{bmatrix} Z_j & y_j \\ y'_j & a_{jj} \end{bmatrix} \Rightarrow (Z_k)^{-1} = \begin{bmatrix} (Z_j)^{-1} + \alpha_j b_j b'_j & -\alpha_j b_j \\ -\alpha_j b'_j & \alpha_j \end{bmatrix}$$

where: $\alpha_j = (a_{jj} - b'_j y_j)^{-1}$

• Relation between $(Z_k)^{-1}$ and $(Z_k)^{-1}$

$$(Z_k)^{-1} = (Z_j)^{-1} 0 + \alpha_j \begin{bmatrix} -b_j \\ 1 \end{bmatrix} \begin{bmatrix} -b'_j & 1 \end{bmatrix}$$
As you can see ...

\[ b_j = - T^{-1} \text{ (at row } j, \text{ from column } 1 \text{ to } i) \]

Situation is very trivial for the case of \( A \)

For other matrices, the aim is thus to determine a set of \( b_j \) with as much advantages as in \( A \):

- easy to determine
- involving few computations
- sparse
- having linear computation cost with size of matrix
Application to the case of the inverse of G
(Faux, Gengler and Misztal, 2012)
The genomic relationship matrix

- A few words about genomic relationship...
  - Observed (or, at least, sampled) vs. expected
  - Some introduction with Bömcke and Gengler (2009)

\[
\begin{align*}
f_{M_{x,y;l}} &= \frac{1}{4} \left( S_{ac} + S_{ad} + S_{bc} + S_{bd} \right) ; \quad T_{A_{x,y;l}} = 2 f_{M} \\
T_{A_{x,y}} &= \frac{1}{m} \sum_{l=1}^{m} T_{A_{x,y;l}} ; \quad T_{A} = \begin{bmatrix}
T_{A_{1,1}} & \cdots & T_{A_{1,1}} \\
\vdots & \ddots & \vdots \\
T_{A_{1,1}} & \cdots & T_{A_{1,1}}
\end{bmatrix}
\end{align*}
\]
The genomic relationship matrix

• When working with a SNP matrix \((\mathbf{M}; n \times m)\):

\[
\mathbf{M} = \begin{bmatrix}
1 & 2 & 1 & 1 & 0 & \ldots \\
0 & 0 & 1 & 1 & 0 & \ldots \\
1 & 2 & 0 & 1 & 1 & \ldots \\
\end{bmatrix}; \quad \mathbf{Z} = \mathbf{M} - 1
\]

\[
\mathbf{T} \mathbf{A} = \frac{\mathbf{Z} \mathbf{Z}^\prime + \mathbf{m}}{\mathbf{m}}
\]

• Van Raden’s \(\mathbf{G}\):

\[
\mathbf{f} = \begin{bmatrix}
\text{MAF}_1 & \ldots & \text{MAF}_m
\end{bmatrix}; \quad \mathbf{d} = 2 \cdot \mathbf{f}^\prime \cdot (\mathbf{1} - \mathbf{f})
\]

\[
\mathbf{P} = \begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix} \otimes \begin{bmatrix}
2f_1 - 1 & \ldots & 2f_m - 1
\end{bmatrix}; \quad \mathbf{Z}_g = \mathbf{M} - \mathbf{P} \quad \mathbf{G} = \frac{\mathbf{Z} \mathbf{Z}^\prime}{\mathbf{d}}
\]
Approximation of the inverse of the factorization of G

- **Purpose**: decrease computation time
- **Solution**: create the $b_j$ using **OLS on close-related**
- **Why?** Go back to tabular method...

... Model used in tabular method is back in the inverse decomposition!

- **Why?** It ensures $D$ to be diagonal:

\[
A^{-1} = (T^{-1})'D^{-1}T^{-1} \quad \Rightarrow \quad D = T^{-1}A(T^{-1})'
\]
Approximation of the inverse of the factorization of $G$

- Solutions are computed by OLS using a simple model:

$$
G \approx \begin{bmatrix}
    s & \ldots & d & \ldots & a & \ldots \\
    s & & & & & \\
    \ldots & & & & & \\
    d & & & & & \\
    \ldots & & & & & \\
    a & & & & & \\
    \ldots & & & & & \\
\end{bmatrix} \cdot \begin{bmatrix}
    1 \\
    1 \\
    1 \\
    1 \\
    1 \\
\end{bmatrix} = \begin{bmatrix}
    b \\
\end{bmatrix} = y
$$
Approximation of the inverse of the factorization of $G$

• Solutions are computed by OLS using a simple model:

$$b = (Z'Z)^{-1}Z'y$$

$$b = (Z)^{-1}(Z)^{-1}(Z)y$$

$$\Rightarrow b = (Z)^{-1}y$$
Approximation of the inverse of the factorization of $G$

- Solutions are computed by OLS using a simple model:

$$b = (Z'Z)^{-1}Z'y$$
$$b = (Z)^{-1}(Z)^{-1}(Z)y$$
$$\Rightarrow b = (Z)^{-1}y$$

- Restriction to the close-family ($\Omega$) of an animal
  - $p$ is a « genomic » threshold that defines close-family

$$\Omega_i = \{k : k<i, g_{ik} \geq p\}$$
Computation of $D^{-1}$ and $G^{-1}$

- Use of the “backward” equation to get $D$:

$$D = T^{-1}G(T^{-1})'$$
Computation of $D^{-1}$ and $G^{-1}$
Computation of $D^{-1}$ and $G^{-1}$
Computation of $D^{-1}$ and $G^{-1}$
Computation of $D^{-1}$ and $G^{-1}$

- Use of the “backward” equation to get $D$:
  \[ D = T^{-1}G(T^{-1})' \]

- Next? 2 options:
  1) Take a diagonal element of $D$ and invert it
  2) Recursively process the remaining $D$ as $G$

- Final equation:
  \[ G^{-1} = (T^{-1})'D^{-1}T^{-1} \]
Acknowledgements

• Dr. G. Gorjanc (UL), Dr. F. Colinet and Mr. J. Vandenplas (ULg)
• Wallonie-Bruxelles International (WBI)
• Government of Republic of Slovenia

• Fonds National de la Recherche Luxembourg (FNR)
• Host institution: Gembloux Agro-Bio Tech - University of Liège
• Collaborating institutions:
  • CONVIS s.c. (Luxembourg)
  • UGA, Animal and Dairy Science Department (USA)