EAAP meeting Ghent 2019

Generalized gametic relationships for flexible imprinting analyses

N. Reinsch, M. Mayer and I. Blunk
Institute of Genetics and Biometry
Leibniz-Institute for Farm Animal Biology, 18196 Dummerstorf, Germany

Outline

reduced and gametic imprinting models

relationship matrix combining transmitting abilities and gametic effects

direct inversion, saved equations, genomic counterpart

Models available for comprehensive imprinting analyses

reduced imprinting model (Neugebauer et al., 2010; Blunk et al., 2018)

- uses numerator relationship matrix for parents of final progeny
- parents and ancestors do not have phenotypic records (slaughter traits)
- number of genetic effects: twice the no. of parents and ancestors (excluding final progeny)

$$y_i = \mu + ta_{fs} + ta_{md} + (m_{fs,i} + m_{md,i} + e_i)$$

gametic imprinting model (Tier and Meyer, 2018)

- parent animals may have phenotypic records
- requires gametic relationship matrix for all individuals
- number of genetic effects: fourfold the no. of all individuals (including final progeny)

$$y_i = \mu + g_{i,s}^{p} + g_{i,d}^{m} + e_i$$

Reduced imprinting model: genetic effects for ancestors only

Phenotypes from non-parents only, two genetic effects for each ancestor

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_s\mathbf{a}_s + \mathbf{Z}_d\mathbf{a}_d + \mathbf{e}$$

two transmitting abilities
for parents and ancestors

TA „as sire“ a_s
TA „as dam“ a_d

both are <u>correlated</u>

Covariances

$$Var\begin{bmatrix}\mathbf{a}_s\\ \mathbf{a}_d\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\,\tfrac{1}{2}\sigma_s^2 & \mathbf{A}\,\tfrac{1}{2}\sigma_{sd} & 0\\ \mathbf{A}\,\tfrac{1}{2}\sigma_{sd} & \mathbf{A}\,\tfrac{1}{2}\sigma_d^2 & 0\\ 0 & 0 & \mathbf{W}\sigma_e^2\end{bmatrix}$$

Weights

$$w_i = \left[\frac{\tfrac{1}{2}\sigma_s^2(1-F_{si}) + \tfrac{1}{2}\sigma_d^2(1-F_{di}) + \sigma_e^2}{\sigma_e^2}\right]^{-1}$$

Mixed model equations

$$\begin{bmatrix}\mathbf{X'W^{-1}X} & \mathbf{X'W^{-1}Z_s} & \mathbf{X'W^{-1}Z_d}\\ \mathbf{Z'_sW^{-1}X} & \mathbf{Z'_sW^{-1}Z_s} + \bar{\mathbf{A}}^{-1}\alpha_1 & \mathbf{Z'_sW^{-1}Z_d} + \bar{\mathbf{A}}^{-1}\alpha_2\\ \mathbf{Z'_dW^{-1}X} & \mathbf{Z'_dW^{-1}Z_s} + \bar{\mathbf{A}}^{-1}\alpha_2 & \mathbf{Z'_dW^{-1}Z_d} + \bar{\mathbf{A}}^{-1}\alpha_3\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{a}_s\\ \mathbf{a}_d\end{bmatrix} = \begin{bmatrix}\mathbf{X'W^{-1}y}\\ \mathbf{Z'_sW^{-1}y}\\ \mathbf{Z'_dW^{-1}y}\end{bmatrix}$$

imprinting variance

$$\sigma_i^2 = \sigma_s^2 + \sigma_d^2 - 2\sigma_{sd}$$

in terms of gametic variances

Gametic imprinting model: genetic effects for all individuals

Tier and Meyer, 2018

four genetic effects for each animal

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_s\mathbf{g}_s + \mathbf{Z}_d\mathbf{g}_d + \mathbf{e}$$

Covariances

$$Var\begin{bmatrix}\mathbf{g}_s\\ \mathbf{g}_d\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}\sigma_s^2 & \mathbf{G}\sigma_{sd} & 0\\ \mathbf{G}\sigma_{sd} & \mathbf{G}\sigma_d^2 & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

four gametic effects

gametic effects „as sire“ $\mathbf{g}_s$

gametic effects „as dam“ $\mathbf{g}_d$

both are <u>correlated</u>

Mixed model equations

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z}_s & \mathbf{X'Z}_d\\ \mathbf{Z'}_s\mathbf{X} & \mathbf{Z'}_s\mathbf{Z}_s + \mathbf{G}^{-1}\alpha_1 & \mathbf{Z'}_s\mathbf{Z}_d + \mathbf{G}^{-1}\alpha_2\\ \mathbf{Z'}_d\mathbf{X} & \mathbf{Z'}_d\mathbf{Z}_s + \mathbf{G}^{-1}\alpha_2 & \mathbf{Z'}_d\mathbf{Z}_d + \mathbf{G}^{-1}\alpha_3\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{g}_s\\ \mathbf{g}_d\end{bmatrix} = \begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'}_s\mathbf{y}\\ \mathbf{Z'}_d\mathbf{y}\end{bmatrix}$$

imprinting variance

$$\sigma_i^2 = \sigma_s^2 + \sigma_d^2 - 2\sigma_{sd}$$

in terms of gametic variances

if reduced model is not applicable: can equations be saved?

reduced imprinting model not applicable if

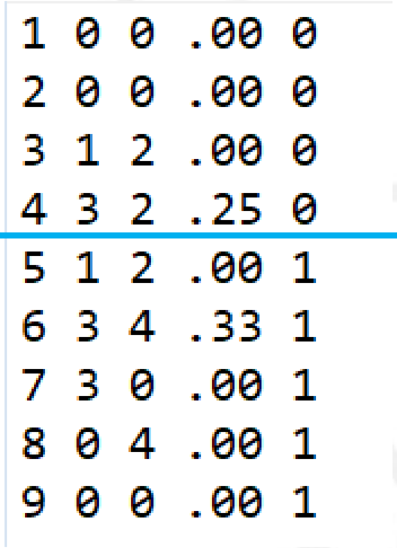
- on principle: if parents have own phenotypic records
- on principle: if maternal effects are included
- technically: if MCMC is to be applied (left hand side of MME had to be set up in each iteration)
- technically: in multivariate analyses (bloc-diagonal weight matrix not supported by all REML-packages)

solution: relationship matrix with mixed kinds of genetic effects

- animals without own phenotype: represented by their transmitting ability
- animals with an own phenotype: represented by two gametic effects

$$y_i = \mu + g_{i,s}^{p} + g_{i,d}^{m} + e_i$$

models with generalized gametic relationship matrix



```
1 0 0 .00 0
2 0 0 .00 0
3 1 2 .00 0
4 3 2 .25 0
5 1 2 .00 1
6 3 4 .33 1
7 3 0 .00 1
8 0 4 .00 1
9 0 0 .00 1
```

parents (including their ancestors) of meat animals

non-parents (meat animals with trait records)

Information in pedigree file:

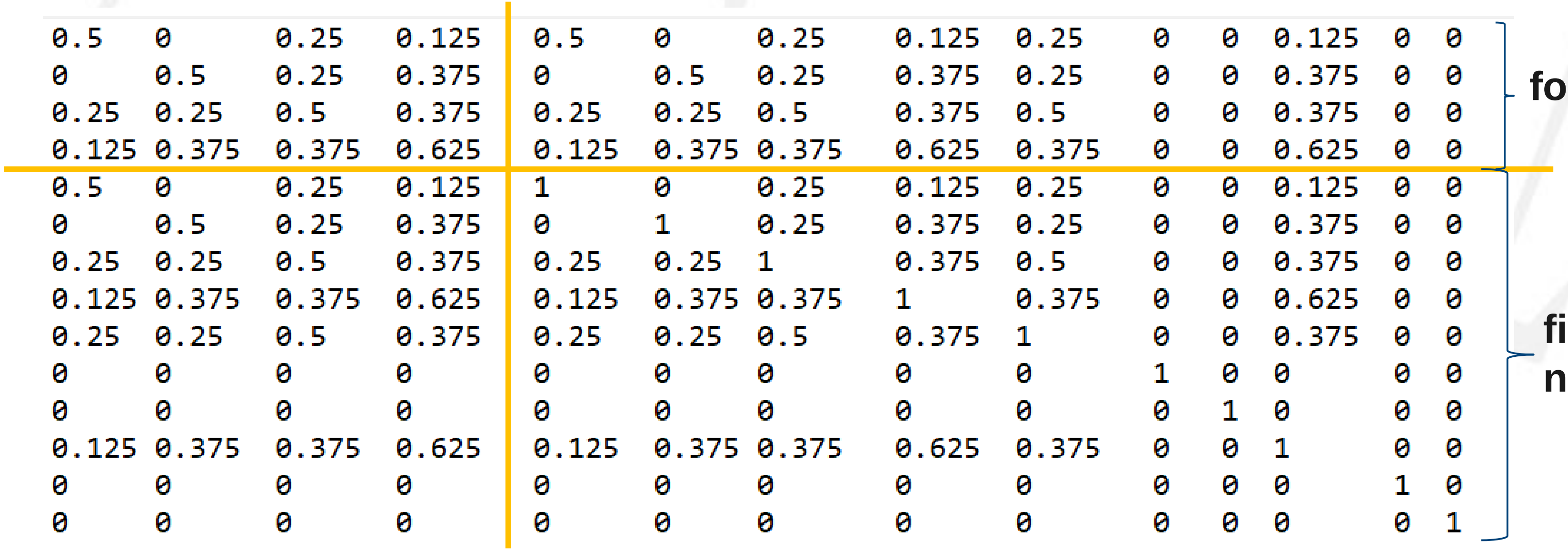
- animal-ID, sire-ID, dam-ID,
- inbreeding coefficient,
- indicator variable

Indicator variable tells if animal is to be represented by its average gametic effect (transmitting ability) or by two gametic effects

Inbreeding coefficient is needed for computing the inverse rel. matrix

generalized gametic relationship matrix

Single effects (TAs) for parents, two gametic effects for non-parents



```
0.5   0     0.25  0.125  0.5   0     0.25  0.125  0.25  0  0  0.125  0  0
0     0.5   0.25  0.375  0     0.5   0.25  0.375  0.25  0  0  0.375  0  0
0.25  0.25  0.5   0.375  0.25  0.25  0.5   0.375  0.5   0  0  0.375  0  0
0.125 0.375 0.375 0.625  0.125 0.375 0.375 0.625  0.375 0  0  0.625  0  0
0.5   0     0.25  0.125  1     0     0.25  0.125  0.25  0  0  0.125  0  0
0     0.5   0.25  0.375  0     1     0.25  0.375  0.25  0  0  0.375  0  0
0.25  0.25  0.5   0.375  0.25  0.25  1     0.375  0.5   0  0  0.375  0  0
0.125 0.375 0.375 0.625  0.125 0.375 0.375 1      0.375 0  0  0.625  0  0
0.25  0.25  0.5   0.375  0.25  0.25  0.5   0.375  1     0  0  0.375  0  0
0     0     0     0      0     0     0     0      0     1  0  0      0  0
0     0     0     0      0     0     0     0      0     0  1  0      0  0
0.125 0.375 0.375 0.625  0.125 0.375 0.375 0.625  0.375 0  0  1      0  0
0     0     0     0      0     0     0     0      0     0  0  0      1  0
0     0     0     0      0     0     0     0      0     0  0  0      0  1
```

four parents

five non-parents

Diagonal elements:
- 0.5 for transmitting abilities (non-inbred animals)
- 1.0 for gametic effects

imprinting model and generalized gametic relationship

Generalized rel. matrix $\bar{\mathrm{A}}$ with TA for parents and gametic effects for non-parents

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{a}_s + \mathbf{Z}_d\mathbf{a}_d + \mathbf{e}$$

Covariances

$$Var\begin{bmatrix}\mathbf{a}_s\\ \mathbf{a}_d\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{A}}\sigma_s^2 & \bar{\mathbf{A}}\sigma_{sd} & 0\\ \bar{\mathbf{A}}\sigma_{sd} & \bar{\mathbf{A}}\sigma_d^2 & 0\\ 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

for each parent there are two transmitting abilities

TA „as sire“ a_s and TA „as dam“ a_d

plus four gametic effects for each non-parent: $\mathbf{a}_s^1$, $\mathbf{a}_s^2$, $\mathbf{a}_d^1$, $\mathbf{a}_d^2$
(paternal and maternal gametic effects „as sire“ and „as dam“)

no weights required

mixed model equations

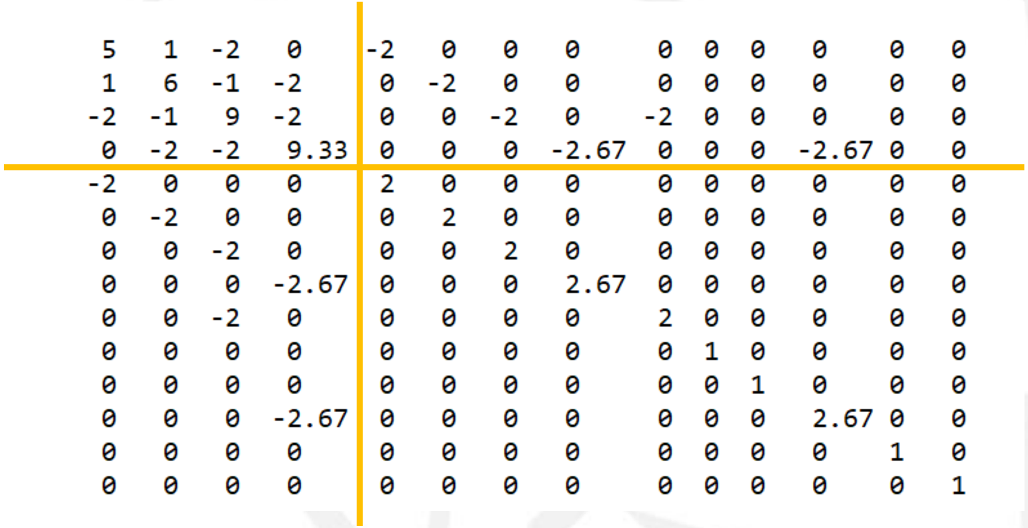
$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z}_s & \mathbf{X'Z}_d\\ \mathbf{Z'}_s\mathbf{X} & \mathbf{Z'}_s\mathbf{Z}_s + \bar{\mathbf{A}}^{-1}\alpha_1 & \mathbf{Z'}_s\mathbf{Z}_d + \bar{\mathbf{A}}^{-1}\alpha_2\\ \mathbf{Z'}_d\mathbf{X} & \mathbf{Z'}_d\mathbf{Z}_s + \bar{\mathbf{A}}^{-1}\alpha_2 & \mathbf{Z'}_d\mathbf{Z}_d + \bar{\mathbf{A}}^{-1}\alpha_3\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{a}_s\\ \mathbf{a}_d\end{bmatrix} = \begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'}_s\mathbf{y}\\ \mathbf{Z'}_d\mathbf{y}\end{bmatrix}$$

imprinting variance:

$$\sigma_i^2 = \sigma_s^2 + \sigma_d^2 - 2\sigma_{sd}$$

in terms of gametic variances

Inverse generalized relationship matrix



```
  5   1  -2   0     -2   0   0   0      0  0  0   0     0   0
  1   6  -1  -2      0  -2   0   0      0  0  0   0     0   0
 -2  -1   9  -2      0   0  -2   0     -2  0  0   0     0   0
  0  -2  -2   9.33   0   0   0  -2.67   0  0  0  -2.67  0   0
 -2   0   0   0      2   0   0   0      0  0  0   0     0   0
  0  -2   0   0      0   2   0   0      0  0  0   0     0   0
  0   0  -2   0      0   0   2   0      0  0  0   0     0   0
  0   0   0  -2.67   0   0   0   2.67   0  0  0   0     0   0
  0   0  -2   0      0   0   0   0      2  0  0   0     0   0
  0   0   0   0      0   0   0   0      0  1  0   0     0   0
  0   0   0   0      0   0   0   0      0  0  1   0     0   0
  0   0   0  -2.67   0   0   0   0      0  0  0   2.67  0   0
  0   0   0   0      0   0   0   0      0  0  0   0     1   0
  0   0   0   0      0   0   0   0      0  0  0   0     0   1
```

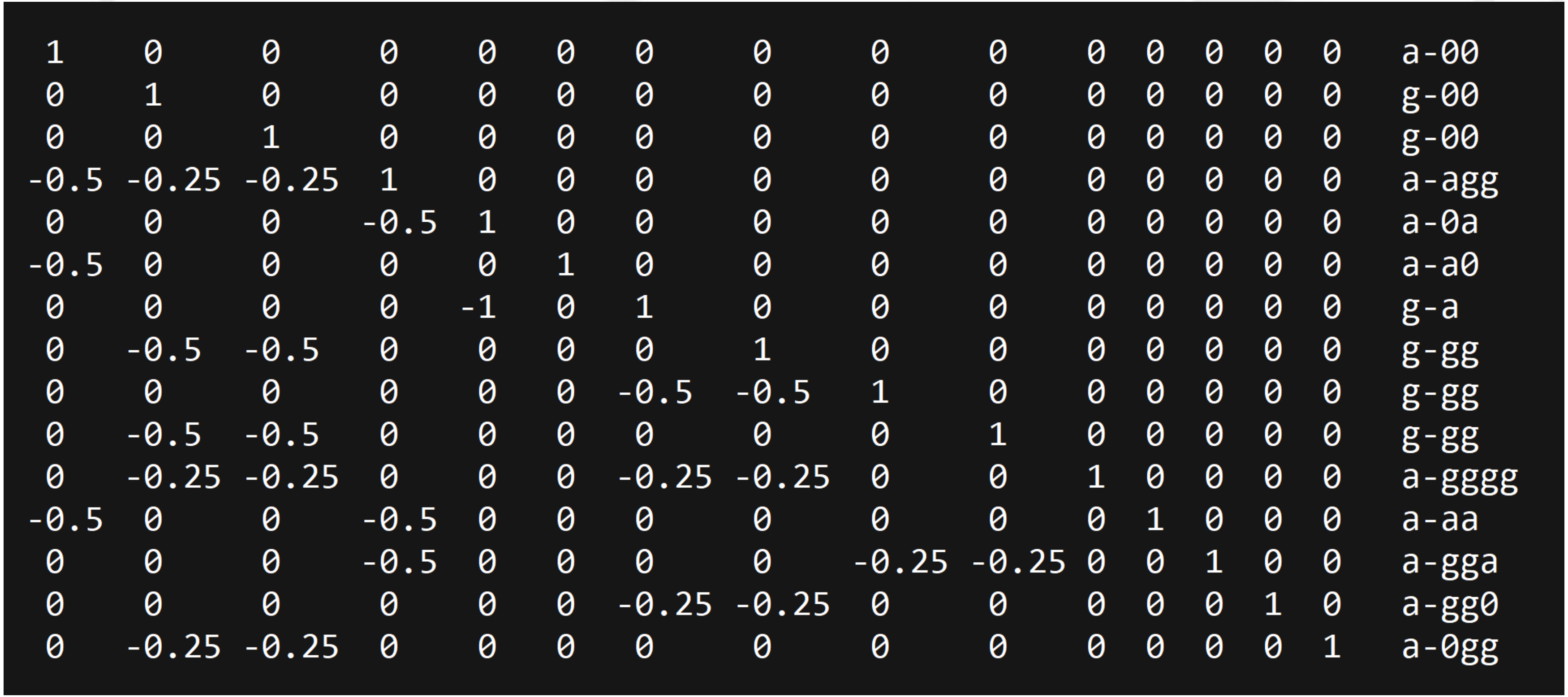
very sparse,
one diagonal element plus two off-diagonal elements per gametic effect (if parent is known)

at most three elements per gametic effect to be stored in inverse generalized relationship matrix

four equations saved in comparison to gametic model

Lower triangle of decomposed inverse

One row per effect. Last column indicates case.



```
 1    0     0     0    0   0   0     0     0     0     0  0  0  0  0   a-00
 0    1     0     0    0   0   0     0     0     0     0  0  0  0  0   g-00
 0    0     1     0    0   0   0     0     0     0     0  0  0  0  0   g-00
-0.5 -0.25 -0.25  1    0   0   0     0     0     0     0  0  0  0  0   a-agg
 0    0     0    -0.5  1   0   0     0     0     0     0  0  0  0  0   a-0a
-0.5  0     0     0    0   1   0     0     0     0     0  0  0  0  0   a-a0
 0    0     0     0   -1   0   1     0     0     0     0  0  0  0  0   g-a
 0   -0.5  -0.5   0    0   0   0     1     0     0     0  0  0  0  0   g-gg
 0    0     0     0    0   0  -0.5  -0.5   1     0     0  0  0  0  0   g-gg
 0   -0.5  -0.5   0    0   0   0     0     0     1     0  0  0  0  0   g-gg
 0   -0.25 -0.25  0    0   0  -0.25 -0.25  0     0     1  0  0  0  0   a-gggg
-0.5  0     0    -0.5  0   0   0     0     0     0     0  1  0  0  0   a-aa
 0    0     0    -0.5  0   0   0     0    -0.25 -0.25  0  0  1  0  0   a-gga
 0    0     0     0    0   0  -0.25 -0.25  0     0     0  0  0  1  0   a-gg0
 0   -0.25 -0.25  0    0   0   0     0     0     0     0  0  0  0  1   a-0gg
```

Inversion: twelve different cases to be distinguished

a-00
g-00 g-00
a-agg
a-0a
a-a0
g-a g-gg
g-gg g-gg
a-gggg
a-aa
a-gga
a-gg0
a-0gg

An **a-** as the first character indicates that a certain genetic effect is a **transmitting ability (TA)**, while a **g-** stands for a **gametic effect**.

An individual that is represented by its transmitting ability may have

- two unknown parents **a-00**
- a single known parent represented by TA **a-0a**, **a-a0**
- a single known parent represented by gametes **a-0gg**, **a-gg0**
- two parents represented by their TAs **a-aa**
- one parent represented by TA, one by gametes **a-agg**, **a-gga**
- both parents represented by gametes **a-gggg**

A gametic effect may be derived from

- the base population (unknown parent) **g-00**
- an animal represented by two gametes **g-gg**
- an animal represented by its TA **g-a**

inverse Mendelian sampling variances

a-00	$\delta = 2$
a-0a, a-a0, a-0gg, a-gg0	$\delta = \left[\tfrac{1}{2}\left(1 - F_{known\ parent}\right)\right]^{-1}$
a-aa, a-agg, a-gga, a-gggg	$\delta = 2\left(\tfrac{1}{2}\left[\tfrac{1}{2}\left(1 - F_{sire}\right) + \tfrac{1}{2}\left(1 - F_{dam}\right)\right]\right)^{-1}$
g-a, g-gg	$\delta = 2\left[\tfrac{1}{2}\left(1 - F_{parent}\right)\right]^{-1}$
g-00	$\delta = 1$

five variants, depending on the case

application example: brown swiss slaughterhouse data

Brown-Swiss data (Blunk et al., 2018 Animal; 2019 SciRep) : total pedigree size **663,515** including 173,051 non-parents with records

	gametic rel. G	generalized rel. A_g
dimension	1,327,030	836,566
non-zero elements below diagonal	2,864,694	1,592,843
total number of non-zero elements	4,191,724	2,429,409

37% of equations saved

42% of non-zero elements saved

application example: litter size in a mouse line

mouse example data (unpublished): total pedigree size **15,222**
2,137 female parents with records for first-parity litter size

	gametic rel. G	generalized rel. A_g
dimension	30.444	17,359
non-zero elements below diagonal	66.900	47,761
total number of non-zero elements	97.344	65,120

42% of equations saved

32% of non-zero elements saved

ordered genotypes needed for genomic covariances of gametes

Meat animal pedigree and number of effects

1	0	0	1
2	0	0	1
3	0	0	1
4	1	2	2
5	1	2	2
6	1	3	2

Gene counts can be summarized in a matrix of centered gametic gene counts **Z**.

Parental origin has to be known for imprinting analyses!

This is actually not the case for the base animals of the example.

$$\mathbf{Z} = \left[\begin{array}{cc} 1-p_1 & 1-p_2 \\ 1-p_1 & 0-p_2 \\ 1-p_1 & 0-p_2 \\ 1-p_1 & 1-p_2 \\ 0-p_1 & 0-p_2 \\ 0-p_1 & 0-p_2 \\ \hline 1-p_1 & 1-p_2 \\ 1-p_1 & 0-p_2 \\ 0-p_1 & 0-p_2 \\ 1-p_1 & 1-p_2 \\ 0-p_1 & 1-p_2 \\ 1-p_1 & 0-p_2 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\end{array}}\right\}\text{parents} \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\\1\end{array}}\right\}\text{non-parents} \end{array}$$

Matrix **Z** can be converted into a genomic gametic relationship matrix

Scaling parameter s for the genomic relationship matrix fits to a gametic covariance: $s = \sum_i p_i(1-p_i)$

Finally we get: $\mathbf{G}_g = \mathbf{ZZ}'/s$

gametic model requires all genomic information to be ordered in case of imprinting

generalized relationships account for un-ordered genotypes

Meat animal pedigree and number of effects

1	0	0	1
2	0	0	1
3	0	0	1
4	1	2	2
5	1	2	2
6	1	3	2

The centered matrix of gametic gene counts can be transformed by taking average gene counts of two gametes if parental origin cannot be traced back.
A genomic counterpart of the generalized genomic relationship marix exists.

$$\mathbf{K}'\mathbf{Z} = \left[\begin{array}{cc} \tfrac{1}{2}(2-2p_1) & \boxed{\tfrac{1}{2}(1-2p_2)} \\ \tfrac{1}{2}(2-2p_1) & \tfrac{1}{2}(0-2p_2) \\ \tfrac{1}{2}(0-2p_1) & \tfrac{1}{2}(1-2p_2) \\ \hline (1-p_1) & (1-p_2) \\ (1-p_1) & (0-p_2) \\ (0-p_1) & (0-p_2) \\ (1-p_1) & (1-p_2) \\ (0-p_1) & (1-p_2) \\ (1-p_1) & (0-p_2) \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \text{parents} \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\\a\\a\end{matrix}}\right\} \text{non-parents} \end{array}$$

Scaling parameter S for the genomic relationship matrix fits to a gametic covariance: $s = \sum_i p_i(1-p_i)$

The result is: $$\bar{\mathbf{G}}_g = \mathbf{K}'\mathbf{Z}(\mathbf{K}'\mathbf{Z})'/s = \mathbf{K}'\mathbf{Z}\mathbf{Z}'\mathbf{K}/s = \mathbf{K}'\mathbf{G}_g\mathbf{K}/s$$

generalized gametic relationship model makes it possible to integrate un-ordered genomic information

Summary

- generalized gametic relationship model leads to considerable savings in terms of
  - number of equations
  - storage requirements for non-zero elements of the inverse

actual savings depend on trait and population structure

- generalized gametic relationships have useful genomic counterpart
  - un-ordered genomic information can be integrated in gBLUP models and in
  - single-step genomic prediction

- Fortran-program for inversion available upon request
  - inverse and cross-reference table (animals – effects)

Thank you for attention !