

Genetic Relationships

1 Pedigree Preparation

Pedigrees of animals need to be arranged in chronological order. Parents should appear in a list before (ahead of) their progeny. Ordering a pedigree is most easily accomplished by sorting animals by birthdate. Birthdates can be incorrectly recorded or entered, or for many individuals may not be available. One approach is to assume that all birthdates are incorrect. Animals can be arranged by assigning generation numbers to animals, then iterate through the pedigrees modifying the generation numbers of the sire and dam to be at least one greater than the generation number of the offspring. The number of iterations depends on the number of generations of animals in the list. Probably 20 or less iterations are needed for most situations.

If the number of iterations reaches 50 or more, then there is an increased likelihood that there is a loop in the pedigrees. That means that an animal is its own ancestor, somewhere back in the pedigree. For example, A might be the parent of B, and B is the parent of C, and C is the parent of A. In this case the generation numbers will keep increasing in each iteration. Thus, if more than 50 iterations are used, then look at the animals with the highest generation numbers and try to find the loop. A loop is an error in the pedigrees and must be repaired. Either correct the parentage, or remove the parent of the older animal.

1.1 Example Pedigree to Sort

Animal	Sire	Dam	Generation Number
BF	DD	HE	1
DD	GA	EC	1
GA			1
EC	GA	FB	1
FB			1
AG	BF	EC	1
HE	DD	FB	1

All animals begin with generation number 1. Proceed through the pedigrees one animal at a time.

1. Take the current generation number of the animal and increase it by one (1), call it m . The first animal is BF, for example, and its generation number is 1, increased by 1 becomes $m=2$.
2. Compare m to the generation numbers of the animal's sire and dam. In the case of BF, the sire is DD and DD's generation number is 1. That is less than 2 so DD's generation

number has to be changed to 2 (m). The dam is HE, and HE's generation number is also changed to 2.

Repeat for each animal in the pedigree list. Keep modifying the generation numbers until no more need to be changed. The animal with the highest generation number is the oldest animal.

The end result after four iterations of the example pedigree is shown below.

Animal	Sire	Dam	Generation Number
BF	DD	HE	2
DD	GA	EC	4
GA			6
EC	GA	FB	5
FB			6
AG	BF	EC	1
HE	DD	FB	3

Now sort the list by decreasing order of the generation number.

Animal	Sire	Dam	Generation Number
GA			6
FB			6
EC	GA	FB	5
DD	GA	EC	4
HE	DD	FB	3
BF	DD	HE	2
AG	BF	EC	1

The order of animals GA or FB is not important. The order of animals with the same generation number is not critical.

Once the pedigree is sorted, then the birthdates can be checked. Errors can be spotted more readily. Once the errors are found and corrected, then the generation numbers could be checked again. Animals should then be numbered consecutively according to the last list from 1 to the total number of animals in the list. That means that parent numbers should always be smaller than progeny ID numbers. Having animals in this order facilitates calculation of inbreeding coefficients, assignment of animals with unknown parents to groups, and utilization of the inverse of the relationship matrix in the solution of mixed model equations.

2 Genomic Relationships

Genomic relationships are constructed by identifying the genomic sources for each animal. One half of the alleles, genomic effects, are from the male parent and the other half of alleles are from the female parent. Let \mathbf{g} be a vector of the genomic effects for all animals, of length $2N$ where N is the number of animals, then

$$\text{Var}(\mathbf{g}) = \mathbf{G}\sigma_g^2.$$

The genomic relationship matrix, \mathbf{G} , can be viewed as the average of an infinite number of gametic relationship matrices, \mathbf{G}_i , for the i^{th} gene. The genomic relationship matrix can be constructed using simple rules.

2.1 Example Pedigree

Parentage of five animals are given below.

Example Pedigree.

Animal	Sire	Dam
A	-	-
B	-	-
C	A	B
D	A	C
E	D	B

Expand this table to identify the genomic structure. Parent1 and Parent2 indicate the genomic sources for the male or female parents of the sire of an animal, respectively. For example, for animal C, the male source of alleles is C1, and the source of alleles for C1 comes from animal A's genes. The source of alleles for C2 is from the female parent, B.

Example Genomic Pedigree.

Animal	Genome	Parent1	Parent2
A	A1	-	-
A	A2	-	-
B	B1	-	-
B	B2	-	-
C	C1	A1	A2
C	C2	B1	B2
D	D1	A1	A2
D	D2	C1	C2
E	E1	D1	D2
E	E2	B1	B2

This genomic relationship matrix will be of order 10. The diagonals of all genomic relationship matrices are always equal to 1. The quantities in the off-diagonals of the matrix are probabilities of genes being identical by descent (an average probability across all genes).

		A		B		C		D		E	
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
A	A1	1	0	0	0						
	A2	0	1	0	0						
B	B1	0	0	1	0						
	B2	0	0	0	1						
C	C1					1					
	C2						1				
D	D1							1			
	D2								1		
E	E1									1	
	E2										1

Because the parents of A and B are unknown, then they are assumed to be randomly drawn from a large random mating population and assumed to have no genes identical by descent between them.

Let (A1,C1) indicate an element in the above table between the A1 male parent contribution of animal A and the C1 male parent contribution of animal C, then the value that goes into that location is

$$(A1,C1) = 0.5 * [(A1,A1) + (A1,A2)] = 0.5.$$

Similarly, for the rest of the A1 row,

$$\begin{aligned} (A1,C2) &= 0.5 * [(A1,B1) + (A1,B2)] = 0, \\ (A1,D1) &= 0.5 * [(A1,A1) + (A1,A2)] = 0.5, \\ (A1,D2) &= 0.5 * [(A1,C1) + (A1,C2)] = 0.25, \\ (A1,E1) &= 0.5 * [(A1,D1) + (A1,D2)] = 0.375, \\ (A1,E2) &= 0.5 * [(A1,B1) + (A1,B2)] = 0. \end{aligned}$$

This recursive pattern follows through the entire table. Relationships should be determined row-wise, and when a row is completed, the values are transcribed down the corresponding

column. Thus, if (X,Y) corresponds to the relationship between two genomic contributions, then X should always chronologically precede Y. If this is not the case, then errors in relationship calculations can result. The completed table is shown below.

		A		B		C		D		E	
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
A	A1	1	0	0	0	.5	0	.5	.25	.375	0
	A2	0	1	0	0	.5	0	.5	.25	.375	0
B	B1	0	0	1	0	0	.5	0	.25	.125	.5
	B2	0	0	0	1	0	.5	0	.25	.125	.5
C	C1	.5	.5	0	0	1	0	.5	.5	.5	0
	C2	0	0	.5	.5	0	1	0	.5	.25	.5
D	D1	.5	.5	0	0	.5	0	1	.25	.625	0
	D2	.25	.25	.25	.25	.5	.5	.25	1	.625	.25
E	E1	.375	.375	.125	.125	.5	.25	.625	.625	1	.125
	E2	0	0	.5	.5	0	.5	0	.25	.125	1

Animals D and E are inbred and the offdiagonals between D1 and D2 and between E1 and E2 show the inbreeding coefficient.

2.2 Additive Genetic Relationships

Additive and dominance relationships may be obtained from this genomic relationship table. The **additive relationship** between animals A and C is given by

$$0.5 * [(A1,C1) + (A1,C2) + (A2,C1) + (A2,C2)] = 0.5.$$

Add the four numbers in each square of the table and divide by 2. Then the matrix of additive relationships is

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & .5 & .75 & .375 \\ 0 & 1 & .5 & .25 & .625 \\ .5 & .5 & 1 & .75 & .625 \\ .75 & .25 & .75 & 1.25 & .75 \\ .375 & .625 & .625 & .75 & 1.125 \end{pmatrix}.$$

2.3 Dominance Genetic Relationships

The **dominance genetic relationship** between animals X and Y, in general, is given by

$$(X1, Y1) * (X2, Y2) + (X1, Y2) * (X2, Y1) .$$

The complete dominance relationship matrix is

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & .25 & 0 \\ 0 & 1 & 0 & 0 & .125 \\ 0 & 0 & 1 & .25 & .25 \\ .25 & 0 & .25 & 1.0625 & .15625 \\ 0 & .125 & .25 & .15625 & 1.015625 \end{pmatrix} .$$

3 Example Genomic Model

Assume the five animals (A through E) had records equal to 5, 7, 9, 2, and 4, respectively. The process is to define \mathbf{y} , \mathbf{X} , \mathbf{Z} , \mathbf{G} , and \mathbf{R} .

$$\mathbf{y} = \begin{pmatrix} 50 \\ 70 \\ 90 \\ 20 \\ 40 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

$$\mathbf{G} = \frac{1}{8} \begin{pmatrix} 8 & 0 & 0 & 0 & 4 & 0 & 4 & 2 & 3 & 0 \\ 0 & 8 & 0 & 0 & 4 & 0 & 4 & 2 & 3 & 0 \\ 0 & 0 & 8 & 0 & 0 & 4 & 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 8 & 0 & 4 & 0 & 2 & 1 & 4 \\ 4 & 4 & 0 & 0 & 8 & 0 & 4 & 4 & 4 & 0 \\ 0 & 0 & 4 & 4 & 0 & 8 & 0 & 4 & 2 & 4 \\ 4 & 4 & 0 & 0 & 4 & 0 & 8 & 2 & 5 & 0 \\ 2 & 2 & 2 & 2 & 4 & 4 & 2 & 8 & 5 & 2 \\ 3 & 3 & 1 & 1 & 4 & 2 & 5 & 5 & 8 & 1 \\ 0 & 0 & 4 & 4 & 0 & 4 & 0 & 2 & 1 & 8 \end{pmatrix},$$

and

$$\mathbf{R} = \mathbf{I}.$$

The variances are $\sigma_g^2 = \sigma_e^2$, so that the ratio of residual to genomic variances is equal to 1.

Solving the MME for this model can be done using the function given in the notes on prediction theory.

R Statements

```
GI = ginv(G)
RI = ginv(R)
genom = MME(X,Z,GI,RI,y)
```

The solutions to the equations are

$$\begin{pmatrix} \hat{\mu} \\ \hat{A}1 \\ \hat{A}2 \\ \hat{B}1 \\ \hat{B}2 \\ \hat{C}1 \\ \hat{C}2 \\ \hat{D}1 \\ \hat{D}2 \\ \hat{E}1 \\ \hat{E}2 \end{pmatrix} = \begin{pmatrix} 56.4194 \\ -3.0015 \\ -3.0015 \\ 3.0015 \\ 3.0015 \\ -0.7295 \\ 5.2736 \\ -8.1364 \\ -2.8628 \\ -6.8468 \\ 1.2053 \end{pmatrix}.$$

The total additive genetic merit of an animal is equal to the sum of the two genomic contributions. Thus, for animal E, an estimate of the total additive genetic merit is called an *Estimated Breeding Value*.

$$EBV_E = -6.8468 + 1.2053 = -5.6415.$$

Animal E received the more favourable alleles from its female parent.

4 Inverse of the Genomic Matrix

The methods of Henderson(1975), Quaas(1976), and Meuwissen and Luo (1992) were combined to find a fast way of inverting the genomic relationship matrix.

Partition the genomic relationship matrix as

$$\mathbf{G} = \mathbf{TDT}'$$

where \mathbf{T} is a lower triangular matrix and \mathbf{D} is a diagonal matrix. The diagonals of \mathbf{D} are obtained while forming a row of \mathbf{T} . Animal genomes are processed in order from oldest to youngest (i.e. parents before progeny).

For animal genomes with unknown parent genomes, the diagonals of \mathbf{D} are equal to 1. Therefore, the diagonals of \mathbf{D} for A1, A2, B1, and B2 are equal to 1.

Begin with C1, the parent genomes are A1 and A2. Form a table as follows:

Genome	t	D
C1	1	x
A1	.5	1
A2	.5	1

The diagonal element for (C1,C1) in \mathbf{G} is equal to 1, which is equal to $\mathbf{t}'\mathbf{D}\mathbf{t}$, which is

$$(1)^2x + (.5)^2(1) + (.5)^2(1) = 1,$$

which can be re-arranged and solved for x ,

$$x = 1 - .25 - .25 = .5.$$

A similar table and calculations can be made for C2, D1, and E2. Thus, the diagonal elements of \mathbf{D} for these genomic contributions are also equal to .5.

The table for D2 is a little longer. Start with parent genomes C1 and C2

Genome	t	D
D2	1	x
C1	.5	.5
C2	.5	.5

Now add the parent genomes of C1 and C2, as follows:

Genome	t	D
D2	1	x
C1	.5	.5
C2	.5	.5
A1	.25	1
A2	.25	1
B1	.25	1
B2	.25	1

The next step would be to add the 'parents' of A1 and A2, then B1 and B2, but these 'parents' are unknown, and so no further additions to the table are made. Now compute $\mathbf{t}'\mathbf{D}\mathbf{t}$ as

$$x + (.5)^2(.5) + (.5)^2(.5) + 4(.25)^2(1) = 1,$$

or

$$x = 1 - .125 - .125 - 4(.0625) = .5.$$

The table of E1 is more complex. The parent genomes are D1 and D2. As the animals become younger, the length of these tables can become greater, and with n generations there can be up to $2^n + 1$ elements in a table.

Genome	t	D
E1	1	x
D1	.5	.5
D2	.5	.5
A1	.25	1
A2	.25	1
C1	.25	.5
C2	.25	.5
A1	.125	1
A2	.125	1
B1	.125	1
B2	.125	1

Note that A1 and A2 appear twice in the table. Their coefficients in \mathbf{t} must be added together before computing $\mathbf{t}'\mathbf{D}\mathbf{t}$. The new table, after adding coefficients is

Genome	t	D
E1	1	x
D1	.5	.5
D2	.5	.5
A1	.375	1
A2	.375	1
C1	.25	.5
C2	.25	.5
B1	.125	1
B2	.125	1

Then

$$x = 1 - 2(.5)^2(.5) - 2(.375)^2(1) - 2(.25)^2(.5) - 2(.125)^2(1) = .375.$$

The complete results for the diagonals of \mathbf{D} are given in the next table.

Diagonals of **D**

Animal	Genome	Parent1	Parent2	D
A	A1	-	-	1
A	A2	-	-	1
B	B1	-	-	1
B	B2	-	-	1
C	C1	A1	A2	.5
C	C2	B1	B2	.5
D	D1	A1	A2	.5
D	D2	C1	C2	.5
E	E1	D1	D2	.375
E	E2	B1	B2	.5

The inverse of **G** is

$$\mathbf{G}^{-1} = \mathbf{T}^{-T} \mathbf{D}^{-1} \mathbf{T}^{-1},$$

and as Henderson (1975) discovered, the elements in \mathbf{T}^{-1} are all 1's on the diagonals, and each row has a -.5 in the columns corresponding to the two parent genomes. All other elements are equal to 0. This structure leads to a simple set of rules for creating the inverse of **G**, which can be accomplished by going through the pedigrees, one genome at a time.

Let d^i be equal to one over the diagonal of **D** for the i^{th} genome, and let $p1$ and $p2$ be the parent genomes, then the contributions to the inverse of **G** from this genome would be to add the following values:

	i	$p1$	$p2$
i	d^i	$.5d^i$	$.5d^i$
$p1$	$.5d^i$	$.25d^i$	$.25d^i$
$p2$	$.5d^i$	$.25d^i$	$.25d^i$

Applying these rules, then the complete inverse is shown in the table below.

		A		B		C		D		E	
		A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
A	A1	2	1	0	0	-1	0	-1	0	0	0
	A2	1	2	0	0	-1	0	-1	0	0	0
B	B1	0	0	2	1	0	-1	0	0	0	-1
	B2	0	0	1	2	0	-1	0	0	0	-1
C	C1	-1	-1	0	0	2.5	.5	0	-1	0	0
	C2	0	0	-1	-1	.5	2.5	0	-1	0	0
D	D1	-1	-1	0	0	0	0	2.6667	.6667	-1.3333	0
	D2	0	0	0	0	-1	-1	.6667	2.6667	-1.3333	0
E	E1	0	0	0	0	0	0	-1.3333	-1.3333	2.6667	0
	E2	0	0	-1	-1	0	0	0	0	0	2

5 Additive Relationship Matrix

The additive genetic relationships between animals were obtained from the genomic relationship matrix. The order of the additive genetic relationship matrix, \mathbf{A} , equals the number of animals (N) in the pedigree. However, elements of \mathbf{A} can be determined by the tabular method directly, and its inverse can be derived directly using the methods of Henderson (1975) and Meuwissen and Luo (1992).

Sewell Wright, in his work on genetic relationships and inbreeding, defined the relationship between two animals to be a correlation coefficient. That is, the genetic covariance between two animals divided by the square root of the product of the genetic variances of each animal. The genetic variance of an animal was equal to $(1 + F_i)\sigma_a^2$, where F_i is the inbreeding coefficient of that animal, and σ_a^2 is the population additive genetic variance. Correlations range from -1 to +1, and therefore, represented a percentage relationship between two individuals, usually positive only.

The elements of the additive relationship matrix are the numerators of Wright's correlation coefficients. Consequently, the diagonals of \mathbf{A} can be as high as 2, and relationships between two individuals can be greater than 1. The \mathbf{A} is a matrix that represents the relative genetic variances and covariances among individuals.

5.1 The Tabular Method

Additive genetic relationships among animals may be calculated using a recursive procedure called the Tabular Method (attributable to Henderson and perhaps to Wright before him). To begin, make a list of all animals that have observations in your data, and for each of these determine their parents (called the sire and dam). An example list is shown below.

Animal	Sire	Dam
A	-	-
B	-	-
C	-	-
D	A	B
E	A	C
F	E	D

The list should be in chronological order so that parents appear before progeny. The sire and dam of animals A, B, and C are assumed to be unknown, and consequently animals A, B, and C are assumed to be genetically unrelated. In some instances the parentage of animals may be traced for several generations, and for each animal the parentage should be traced to a common *base* generation.

Using the completed list of animals and pedigrees, form a two-way table with n rows and columns, where n is the number of animals in the list, in this case $n = 6$. Label the rows and columns with the corresponding animal identification and above each animal ID write the ID of its parents as shown below.

Tabular Method Example,
Starting Values.

	-,- A	-,- B	-,- C	A,B D	A,C E	E,D F
A	1	0	0			
B	0	1	0			
C	0	0	1			
D						
E						
F						

For each animal whose parents were unknown a one was written on the diagonal of the table (i.e for animals A, B, and C), and zeros were written in the off-diagonals between these three animals, assuming they were unrelated. Let the elements of this table (referred to as matrix **A**) be denoted as a_{ij} . Thus, by putting a 1 on the diagonals for animals with unknown parents, the additive genetic relationship of an animal with itself is one. The additive genetic relationship to animals without common parents or whose parents are unknown is assumed to be zero.

The next step is to compute relationships between animal A and animals D, E, and F. The relationship of any animal to another is equal to the average of the relationships of that animal with the parents of another animal. For example, the relationship between A and D is the average of the relationships between A and the parents of D, who are A and B. Thus,

$$\begin{aligned} a_{AD} &= .5 (a_{AA} + a_{AB}) = .5(1 + 0) = .5 \\ a_{AE} &= .5 (a_{AA} + a_{AC}) = .5(1 + 0) = .5 \\ a_{AF} &= .5 (a_{AE} + a_{AD}) = .5(.5 + .5) = .5 \end{aligned}$$

The relationship table, or **A** matrix, is symmetric, so that $a_{AD} = a_{DA}$, $a_{AE} = a_{EA}$, and $a_{AF} = a_{FA}$. Continue calculating the relationships for animals B and C to give the following table.

Tabular Method Example,
Partially Completed.

	-,- A	-,- B	-,- C	A,B D	A,C E	E,D F
A	1	0	0	.5	.5	.5
B	0	1	0	.5	0	.25
C	0	0	1	0	.5	.25
D	.5	.5	0			
E	.5	0	.5			
F	.5	.25	.25			

Next, compute the diagonal element for animal D. By definition this is one plus the inbreeding coefficient, i.e.

$$a_{DD} = 1 + F_D.$$

The inbreeding coefficient, F_D is equal to one-half the additive genetic relationship between the parents of animal D, namely,

$$F_D = .5a_{AB} = 0.$$

When parents are unknown, the inbreeding coefficient is zero assuming the parents of the individual were unrelated. After computing the diagonal element for an animal, like D, then the remaining relationships to other animals in that row are calculated as before. The completed matrix is given below. Note that only animal F is inbred in this example. The inbreeding coefficient is a measure of the percentage of loci in the genome of an animal that has become homogeneous, that is, the two alleles at a locus are the same (identical by descent). Sometimes these alleles may be lethal and therefore, inbreeding is generally avoided.

Tabular Method Example
Completed Table.

	-, A	-, B	-, C	A,B D	A,C E	E,D F
A	1	0	0	.5	.5	.5
B	0	1	0	.5	0	.25
C	0	0	1	0	.5	.25
D	.5	.5	0	1	.25	.625
E	.5	0	.5	.25	1	.625
F	.5	.25	.25	.625	.625	1.125

Generally, the matrix \mathbf{A} is nonsingular, but if the matrix includes two animals that are identical twins, then two rows and columns of \mathbf{A} for these animals would be identical, and therefore, \mathbf{A} would be singular. In this situation assume that the twins are genetically equal and treat them as one animal (by giving them the same registration number or identification) (see Kennedy and Schaeffer, 1989).

6 Inbreeding Calculations

The inbreeding coefficients and the inverse of \mathbf{A} for inbred animals are generally required for BLUP analyses of animal models. Thus, fast methods of doing both of these calculations, and for very large populations of animals are necessary.

$$\mathbf{A} = \mathbf{T}\mathbf{B}\mathbf{T}',$$

where \mathbf{T} is a lower triangular matrix and \mathbf{B} is a diagonal matrix. Quaas (1976) showed that the diagonals of \mathbf{B} , say b_{ii} were

$$b_{ii} = (.5 - .25(F_s + F_d)),$$

where F_s and F_d are the inbreeding coefficients of the sire and dam, respectively, of the i^{th} individual. If one parent is unknown, then

$$b_{ii} = (.75 - .25F_p),$$

where F_p is the inbreeding coefficient of the parent that is known. Lastly, if neither parent is known then $b_{ii} = 1$.

One of the more efficient algorithms for calculating inbreeding coefficients is that of Meuwissen and Luo (1992). Animals should be in chronological order, as for the Tabular Method. To illustrate consider the example given in the Tabular Method section. The corresponding elements of \mathbf{B} for animals A to F would be

$$\left(\begin{array}{cccccc} 1 & 1 & 1 & .5 & .5 & .5 \end{array} \right).$$

Now consider a new animal, G, with parents F and B. The first step is to set up three vectors, where the first vector contains the identification of animals in the pedigree of animal G, the second vector will contain the elements of a row of matrix \mathbf{T} , and the third vector will contain the corresponding b_{ii} for each animal.

Step 1 Add animal G to the ID vector, a 1 to the T-vector, and

$$b_{GG} = .5 - .25(.125 + 0) = 15/32$$

to the B-vector, giving

ID vector	T-vector	B-vector
G	1	15/32

Step 2 Add the parents of G to the ID vector, and because they are one generation back, add .5 to the T-vector for each parent. In the D-vector, animal B has $b_{BB} = 1$, and animal F has $b_{FF} = .5$. The vectors now appear as

ID vector	T-vector	B-vector
G	1	15/32
F	.5	.5
B	.5	1

Step 3 Add the parents of F and B to the ID vector, add .25 (.5 times the T-vector value of the individual (F or B)) to the T-vector, and their corresponding b_{ii} values. The parents of F were E and D, and the parents of B were unknown. These give

ID vector	T-vector	B-vector
G	1	15/32
F	.5	.5
B	.5	1
E	.25	.5
D	.25	.5

Step 4 Add the parents of E and D to the ID vector, .125 to the T-vector, and the appropriate values to the B-vector. The parents of E were A and C, and the parents of D were A and B.

ID vector	T-vector	B-vector
G	1	15/32
F	.5	.5
B	.5	1
E	.25	.5
D	.25	.5
A	.125	1
C	.125	1
A	.125	1
B	.125	1

The vectors are complete because the parents of A, B, and C are unknown and no further ancestors can be added to the pedigree of animal G.

Step 5 Accumulate the values in the T-vector for each animal ID. For example, animals A and B appear twice in the ID vector. Accumulating their T-vector values gives

ID vector	T-vector	B-vector
G	1	15/32
F	.5	.5
B	.5+.125=.625	1
E	.25	.5
D	.25	.5
A	.125+.125=.25	1
C	.125	1

Do not accumulate quantities until all pathways in the pedigree have been processed, otherwise a coefficient may be missed and the wrong inbreeding coefficient could be calculated.

Step 6 The diagonal of the **A** matrix for animal G is calculated as the sum of the squares of the values in the T-vector times the corresponding value in the B-vector, hence

$$\begin{aligned}
 a_{GG} &= (1)^2(15/32) + (.5)^2(.5) + (.625)^2 \\
 &\quad + (.25)^2(.5) + (.25)^2(.5) + (.25)^2 + (.125)^2 \\
 &= 72/64 \\
 &= 1\frac{1}{8}
 \end{aligned}$$

The inbreeding coefficient for animal G is one-eighth.

The efficiency of this algorithm depends on the number of generations in each pedigree. If each pedigree is 10 generations deep, then each of the vectors above could have over 1000

elements for a single animal. To obtain greater efficiency, animals with the same parents could be processed together, and each would receive the same inbreeding coefficient, so that it only needs to be calculated once. For situations with only 3 or 4 generation pedigrees, this algorithm would be very fast and the amount of computer memory required would be low compared to other algorithms (Golden et al. (1991), Tier(1990)).

6.1 Example Additive Matrix

Consider the pedigrees in the table below:

Animal	Sire	Dam
1	-	-
2	-	-
3	1	-
4	1	2
5	3	4
6	1	4
7	5	6

Animals with unknown parents may or may not be selected individuals, but their parents (which are unknown) are assumed to belong to a em base generation of animals, i.e. a large, random mating population of unrelated individuals. Animal 3 has one parent known and one parent unknown. Animal 3 and its sire do not belong to the base generation, but its unknown dam is assumed to belong to the base generation. If these assumptions are not valid, then the concept of phantom parent groups needs to be utilized (covered later in these notes). Using the tabular method, the **A** matrix for the above seven animals is given below.

	-,- 1	-,- 2	1,- 3	1,2 4	3,4 5	1,4 6	5,6 7
1	1	0	.5	.5	.5	.75	.625
2	0	1	0	.5	.25	.25	.25
3	.5	0	1	.25	.625	.375	.5
4	.5	.5	.25	1	.625	.75	.6875
5	.5	.25	.625	.625	1.125	.5625	.84375
6	.75	.25	.375	.75	.5625	1.25	.90625
7	.625	.25	.5	.6875	.84375	.90625	1.28125

Now partition **A** into **T** and **B** giving:

Sire	Dam	Animal	1	2	3	4	5	6	7	B
		1	1	0	0	0	0	0	0	1.0
		2	0	1	0	0	0	0	0	1.0
1		3	.5	0	1	0	0	0	0	.75
1	2	4	.5	.5	0	1	0	0	0	.50
3	4	5	.5	.25	.5	.5	1	0	0	.50
1	4	6	.75	.25	0	.5	0	1	0	.50
5	6	7	.625	.25	.25	.5	.5	.5	1	.40625

Note that the rows of **T** account for the direct relationships, that is, the direct transfer of genes from parents to offspring.

6.2 The Inverse of Additive Relationship Matrix

The inverse of the relationship matrix can be constructed similarly to the inverse of the genomic relationship matrix by a set of rules. Recall the previous example of seven animals with the following values for b_{ii} .

Animal	Sire	Dam	b_{ii}	b_{ii}^{-1}
1	-	-	1.00	1.00
2	-	-	1.00	1.00
3	1	-	0.75	1.33333
4	1	2	0.50	2.00
5	3	4	0.50	2.00
6	1	4	0.50	2.00
7	5	6	0.40625	2.4615385

Let $\delta = b_{ii}^{-1}$, then if both parents are known the following constants are added to the appropriate elements in the inverse matrix:

	animal	sire	dam
animal	δ	$-.5\delta$	$-.5\delta$
sire	$-.5\delta$	$.25\delta$	$.25\delta$
dam	$-.5\delta$	$.25\delta$	$.25\delta$

If one parent is unknown, then delete the appropriate row and column from the rules above, and if both parents are unknown then just add δ to the animal's diagonal element of the inverse.

Each animal in the pedigree is processed one at a time, but any order can be taken. Let's start with animal 6 as an example. The sire is animal 1 and the dam is animal 4. In this case,

$\delta = 2.0$. Following the rules and starting with an inverse matrix that is empty, after handling animal 6 the inverse matrix should appear as follows:

	1	2	3	4	5	6	7
1	.5			.5		-1	
2							
3							
4	.5			.5		-1	
5							
6	-1			-1		2	
7							

After processing all of the animals, then the inverse of the relationship matrix for these seven animals should be as follows:

	1	2	3	4	5	6	7
1	2.33333	.5	-.66667	-.5	0	-1	0
2	.5	1.5	0	-1.00000	0	0	0
3	-.66667	0	1.83333	.5	-1	0	0
4	-.5	-1	.5	3.0000	-1	-1	0
5	0	0	-1	-1	2.61538	.61538	-1.23077
6	-1	0	0	-1	.61538	2.61538	-1.23077
7	0	0	0	0	-1.23077	-1.23077	2.46154

The product of the above matrix with the original relationship matrix, \mathbf{A} , gives an identity matrix.

6.3 R function To Construct \mathbf{A} -inverse

Given a pedigree list and the corresponding b_{ii} values for each animal in the pedigree, then the inverse of \mathbf{A} can be written. Below is a function in R that will do those calculations. Animals should be numbered consecutively from 1 to N .

R Statements

```
# sid is a list of the sire IDs
# did is a list of the dam IDs
AINV = function(sid,did,bi) {
  nam = length(sid)
  np = nam + 1
  roul = matrix(data=c(1,-0.5,-0.5,
-0.5,0.25,0.25,-0.5,0.25,0.25),ncol=3)
  ss = sid + 1
  dd = did + 1
  LAI = matrix(data=c(0),nrow=np,ncol=np)
  for(i in 1:nam) {
    ip = i + 1
    k = cbind(ip,ss[i],dd[i])
    x = 1/bi[i]
    LAI[k,k] = LAI[k,k] + roul*x
  }
  k = c(2:np)
  AI = LAI[k,k]
  return(AI) }
```

6.4 Phantom Parent Groups

Westell (1984) and Robinson (1986) assigned *phantom* parents to animals with unknown parents. Each phantom parent was assumed to have only one progeny. Phantom parents were assumed to be unrelated to all other real or phantom animals.

Phantom parents whose first progeny were born in a particular time period probably underwent the same degree of selection intensity to become a breeding animal. However, male phantom parents versus female phantom parents might have been selected differently. Phantom parents were assigned to phantom parent *groups* depending on whether they were sires or dams and on the year of birth of their first progeny.

Genetic groups may also be formed depending on breed composition and/or regions within a country. The basis for further groups depends on the existence of different selection intensities involved in arriving at particular phantom parents.

Phantom parent groups are best handled by considering them as additional animals in the pedigree. Then the inverse of the relationship matrix can be constructed using the same rules as before. These results are due to Quaas (1984). To illustrate, use the same seven animals as before. Assign the unknown sires of animals 1 and 2 to phantom group 1 (*P1*) and the unknown dams to phantom group 2 (*P2*). Assign the unknown dam of animal 3 to phantom group 3 (*P3*).

The resulting matrix will be of order 10 by 10 :

$$\mathbf{A}_*^{-1} = \begin{pmatrix} \mathbf{A}^{rr} & \mathbf{A}^{rp} \\ \mathbf{A}^{pr} & \mathbf{A}^{pp} \end{pmatrix},$$

where \mathbf{A}^{rr} is a 7 by 7 matrix corresponding to the elements among the real animals; \mathbf{A}^{rp} and its transpose are of order 7 by 3 and 3 by 7, respectively, corresponding to elements of the inverse between real animals and phantom groups, and \mathbf{A}^{pp} is of order 3 by 3 and contains inverse elements corresponding to phantom groups. \mathbf{A}^{rr} will be exactly the same as \mathbf{A}^{-1} given in the previous section. The other matrices are

$$\mathbf{A}^{rp} = \begin{pmatrix} -.5 & -.5 & .33333 \\ -.5 & -.5 & 0 \\ 0 & 0 & -.66667 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{A}^{pp} = \begin{pmatrix} .5 & .5 & 0 \\ .5 & .5 & 0 \\ 0 & 0 & .33333 \end{pmatrix}$$

In this formulation, phantom groups (according to Quaas (1984)) are additional fixed factors and there is a dependency between phantom groups 1 and 2. This singularity can cause problems in deriving solutions for genetic evaluation. The dependency can be removed by adding an identity matrix to \mathbf{A}^{pp} . When genetic groups have many animals assigned to them, then adding the identity matrix to \mathbf{A}^{pp} does not result in any significant re-ranking of animals in genetic evaluation and aids in getting faster convergence of the iterative system of equations.

Phantom groups are used in many genetic evaluation systems today. The phantom parents assigned to a genetic group are assumed to be the outcome of non random mating and similar selection differentials on their parents. This assumption, while limiting, is not as severe as assuming that all phantom parents belong to one base population.

7 Identical Genotypes

Occasionally genetically identical twins are born, arising from a single embryo. These individuals share all of the same genetic material, both nuclear and non-nuclear DNA. In an additive genetic relationship matrix the rows for those two animals will be identical, and therefore, a dependency exists in the relationship matrix and an inverse is not possible.

Clones are individuals that tend to share only the nuclear DNA, and the assumption is that the non-nuclear DNA can cause genetic and phenotypic differences between the animals in their development. The additive genetic portion, which is passed on to progeny, is in the nuclear

DNA, and therefore, the additive relationship matrix will have identical rows of numbers for clones from the same animal. The additive relationship matrix would be singular.

Kennedy and Schaeffer (1989) suggested that the relationship matrix be constructed for “genotypes” rather than for individuals in the case of these identical animals. If there were five clones of one animal, then the animal and its clones would represent one genotype, and so there would be only one row of the additive relationship matrix for all six animals. If the animals were measured for some trait, then that “genotype” would have repeated observations (without permanent environmental effects). They would all receive the same estimated breeding value.

One could also treat them as full-sibs, all having the same parents, but not sharing the exact same DNA. If there were many of these in the data file, then it could cause an overestimation of the additive genetic variance. Therefore, this approach would not be suitable.

8 Unknown Sires

In some situations a female is exposed to more than one possible mate. For example, a group of cows in a beef herd may have access to 3 or 4 males during the breeding season. Another example occurs in mink breeding where the success of having conception requires mating a female three times with different males at each mating. Progeny born from one female are a mixture of progeny of those three males. That is, different eggs could be fertilized by different males. Through genetic tests, the probabilities that a progeny is from either the first, second, or third male are known.

The additive relationship matrix can be constructed using the probabilities that a specific male is the sire of a given progeny. An example is as follows: Animals A and B are males, C is a female, and D is a progeny of C with 0.25 probability that the sire was A and 0.75 probability that the sire was B. Construct the additive genetic relationship matrix for this pedigree.

	A	B	C	D
A	1	0	0	.125
B	0	1	0	.375
C	0	0	1	.5
D	.125	.375	.5	1

Normally the relationship of the sire to its progeny (if unrelated to the dam) is 0.5, but in this case, for the relationship between A and D, the value has to be multiplied times the probability of A being the sire of D. Between B and D, the relationship of .5 has to be multiplied times .75.

The inverse of this matrix is derived in a similar manner as the regular additive relationship matrix. The b_i values for animals A, B, and C are equal to 1. Because D has two possible sires, its b_i value needs to be calculated differently.

ID vector	T-vector	B-vector
D	1	x
C	.5	1
B	.375	1
A	.125	1

The diagonal of \mathbf{A} for animal D has to be assumed to be known, as one plus one half the relationship between the sire(s) and dam. In this case, A, B, and C are all unrelated, and therefore, D will not be inbred, so that the diagonal of \mathbf{A} will be 1.

$$a_{DD} = 1 = x + (.5)^2(1) + (.375)^2(1) + (.125)^2(1)$$

Solving for x gives $.59375 = b_D$. The inverse elements added to \mathbf{A} for animal D are given by

$$\begin{pmatrix} -.125 \\ -.375 \\ -.5 \\ 1 \end{pmatrix} \frac{1}{.59375} \begin{pmatrix} -.125 & -.375 & -.5 & 1 \end{pmatrix} = \begin{pmatrix} .0263 & .0789 & .1053 & -.2105 \\ .0789 & .2368 & .3158 & -.6316 \\ .1053 & .3158 & .4211 & -.8421 \\ -.2105 & -.6316 & -.8421 & 1.6842 \end{pmatrix}.$$

To complete the inverse, add 1 to the diagonals for animals A, B, and C.

9 EXERCISES

1. Use the following data for this problem.

Treatment	Animal	Sire	Dam	Observations
1	1			15
2	2			73
1	3			44
2	4	1	3	56
1	5	2	4	55
2	6	1	5	61
1	7	6	4	32
2	8	7	5	47

Let the model equation be

$$y_{ij} = T_i + p_j + m_j + e_{ij}$$

where T_i is a fixed treatment effect, p_j is a random, paternal gamete effect of animal j , m_j is a random, maternal gamete effect of animal j , and e_{ijk} is a random, residual effect. Assume that

$$\sigma_e^2 = 3.2\sigma_G^2.$$

- (a) Complete the genomic relationship matrix for the eight animals, and the inverse of it.
- (b) Construct the MME and solve.
- (c) Predict the breeding values of each animal and obtain the standard errors of prediction.
- (d) Test the difference between the treatments.
- (e) Form the additive relationship matrix for this pedigree.
- (f) Calculate the inverse of \mathbf{A} .
- (g) Assume the model

$$y_{ij} = T_i + a_j + e_{ij}$$

where a_j is the animal additive genetic effect with covariance matrix $\mathbf{A}\sigma_a^2$, and where

$$\sigma_e^2 = 1.6\sigma_a^2.$$

Construct the MME for this model and solve. Compare EBVs from this model and the previous model.

- 2. Use the following data for this problem.

Animal	Sire	Dam	CG	Observations
1				
2				
3				
4	1	3	1	6
5	2	4	1	15
6	1		1	9
7	2	6	2	11
8	1	3	2	8
9	2	4	2	13
10	5	8	3	7
11	5	3	3	10
12	2	7	3	5

Let the model equation be

$$y_{ij} = CG_i + a_j + e_{ij}$$

where CG_i is a fixed contemporary group effect, a_j is a random, animal additive genetic effect, and e_{ijk} is a random residual effect. Assume that

$$\sigma_e^2 = 1.2 \sigma_a^2.$$

- (a) Complete the additive relationship matrix and the inverse of it.
- (b) Construct the MME and solve for this model.
- (c) Compute SEP for the EBV and also reliabilities.

3. You are given the following pedigree information and values of b_i . Determine the b_i value and inbreeding coefficient of animal H which is a progeny of animals G and F.

Animal	Sire	Dam	F_i	b_i
A			0	1
B	A		0	3/4
C	A	B	1/4	1/2
D	C	B	3/8	7/16
E	A	D	5/16	13/32
F	C	D	1/2	11/32
G	E	B	11/32	27/64
H	G	F		

Write out \mathbf{A}^{-1} for this pedigree using Henderson's rules.

4. The following boxes are from a larger genomic relationship matrix.

	Km	Kf	Lm	Lf	L	K
					Mm	Mf
Gm	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{3}{8}$	w	x
Gf	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	y	z

- (a) Calculate w, x, y, and z.
 (b) Calculate a_{GL} and d_{GL}
5. Assign phantom groups to replace missing sire and dam identifications in the following pedigree.

Animal	Sire	Dam	Sex	Year of birth
1			M	1970
2			F	1970
3			M	1971
4			F	1971
5			F	1971
6			M	1970
7	1		F	1972
8		2	M	1972
9			F	1972
10			M	1972

Form the inverse of the relationship matrix, including the phantom groups.

6. For the example pedigree in the section on **Unknown Sires**, compute the inverse of the **A** matrix assuming the probability that the sire is animal A of 0.3, and for animal B is 0.7.
7. In the section on **Unknown Sires**, let three unrelated sires be animals G, H, and K. Female M was exposed to all three sires (as in mink). The probabilities of the three sires being the sire of a progeny are .2, .5, and .3, for animals G, H, and K, respectively. Also assume that animal M is related to animal H by .25. Construct the **A** matrix and derive the inverse (without inverting **A** directly).