

Chapter 12 The Numerator Relationship Matrix (A)

Julius van der Werf

Additive relationships are a measure of the proportion of genes, which are identical by descent, which are expected to be shared by two animals. Sewall Wright (1921) was responsible for the idea of tracing paths to establish the relationships among animals, although Malecot (1948) is given credit for the definition of relationships based on probabilities of individual genes at a locus being identical by descent.

The NRM is needed to account for the additive genetic covariances between records of related animals. In applications for breeding value estimation, use of the relationships matrix implies that information of related individuals will be used.

Accounting for genetic relationships also has proven to be very useful to account for selection. It is able to account for changes of means and variance after selection. The theory of deriving the relationships matrix can also be used for computing inbreeding coefficients for members of a population. This makes sense when we realize that inbreeding is computed from the additive genetic relationship between parents. The matrix is therefore essential for appropriate evaluation with data on complex pedigrees (more than one generation). Furthermore, the NRM can be used to optimize short-term selection decisions and predictions of response and inbreeding, since it can also in this respect give an account of the relationships among the current selection candidates. Understanding the structure of the relationships matrix helps in understanding genetic properties of the animal model.

The numerator relationships are equal to twice the coancestry, and they express the proportion of additive genetic variance that two individuals have in common. Ignoring epistatic effects and letting a equal the fraction of additive genetic variance, and d that of dominance variance, the generalized covariance for any sort of relationship is (Falconer, 1981):

$$\text{Cov} = aV_A + dV_D$$

For the covariance between P and Q (with parents A,B and C,D respectively), values for a and d are given by:

$$\begin{aligned} a_{PQ} &= 2f_{PQ} \\ d_{PQ} &= f_{AC}f_{BD} + f_{AD}f_{BC} \end{aligned} \quad \text{where } f \text{ is the coancestry}$$

but we'll ignore d (*dominance*) hereafter

The coefficients of coancestry of two individuals reflects the probability that two gametes taken at random, one from each, carry alleles that are identical by descent (= inbreeding coefficient of their progeny should they be mated together).

Furthermore, at the diagonals, the NRM contains the coefficient of inbreeding. That is the probability that two genes at any locus in an individual are identical by descent.

The following equivalencies hold (see also page 4.3):

$$a_{ii} = (1 + F_i) = 2f_{ii}$$

$$a_{ij} = 0.5(a_{i,sireofj} + a_{i,damofj}) = 2f_{ij} \text{ for } i \neq j$$

12: The numerator relationships matrix

$$F_i = 0.5a_{sd} = f_{sd}$$

The additive relationship (a_{ij}) is used as a measure of the covariance of breeding values between relatives. Wright's coefficient of relationship (R) is equivalent to the correlation between breeding values of two animals. For non-inbred animals these two measures are identical. R is given by:

$$\begin{aligned} R_{ij} &= a_{ij} / \sqrt{a_{ii}a_{jj}} \\ &= a_{ij} / \sqrt{(1 + F_i)(1 + F_j)} \end{aligned}$$

Elements of A are the numerator of Wright's coefficient of relationship!

The construction of the relationships matrix

Systematic recurrent rules that are based on the flow of genes from generation to generation with individual animals being specified.

- Path coefficient method
suitable for small pedigrees with few generations and little inbreeding
- Genomic table of probabilities
is useful if dominance genetic effects are to be included in the animal model
- Recursively using the tabular method.
where the pedigree is large and/or complicated, and/or where high levels of inbreeding are evident

For very large pedigrees it may be impossible to calculate A at all!

The algorithm to compute A using the tabular method is easy to implement. However, with a large number of animals, A may also be large and time consuming to compute (depending on the efficiency of your algorithm). Further, large A excludes direct inversion of A to obtain A^{-1} (needed for the MME) as a feasible option. Henderson (1976) developed a set of rules by which A^{-1} can be built directly. Quaas (1976) has generalized these rules for large pedigrees and inbreeding. These findings contributed to routine use of A in animal breeding applications.

The tabular method

Step 1: Order pedigree list chronologically so that parents precede offspring. Base parents are considered unrelated and non-inbred.

Step 2: Working one row at a time, compute elements of A using the following relationships: For individuals i or j , elements of A are:

$$\begin{aligned} a_{ii} &= (1 + F_i) \\ a_{ij} &= 0.5(a_{i'j''}, a_{j'i''}) \text{ for } i^1j \end{aligned}$$

$$F_i = 0.5a_{sd}$$

where F_i is the inbreeding coefficient of animal i , j' = sire of j and j'' = dam of j , and s and d refer to sire and dam. By knowing previous relationships, it is possible to calculate future relationships (given no missing information). The inbreeding coefficients (F) for base animals are zero. The following recursive function would calculate the additive genetic relationship between all animals.

```

RECURSIVE function xnumrelmat(i,j) RESULT (value)
INTEGER PED
common PED(10000,2)
  if(i.eq.0 .or. j.eq.0) then
    value=0.0
    return
  endif
  IF(i.eq.j)then
    value=1+.5*xnumrelmat(PED(i,1),PED(i,2))
  elseIF(i.lt.j)then
    value=.5*(xnumrelmat(PED(j,1),i) + xnumrelmat(PED(j,2),i))
  elseIF(j.lt.i)then
    value=.5*(xnumrelmat(ped(i,1),j) + xnumrelmat(ped(i,2),j))
  endif
end function xnumrelmat

```

Construction of the inverse of the relationships matrix

Consider the case that u is a vector with additive genetic values (breeding values) of animals in a population. The variance of u can be given as $\text{var}(u) = A\sigma_a^2$ where A is the matrix with additive genetic relationships between animals and σ_a^2 is the additive genetic variance. Now let us order the animals in u according to age, such that the oldest animals are first. A breeding value of an animal from which both parents are known in u can be written as $u_p = .5u_s + .5u_d + ms$, where u_p , u_s and u_d are breeding values of progeny, sire and dam, and ms is a part of the breeding value due to Mendelian sampling (within full-sib family variance).

We can define now a matrix P represents the transmission of genes. The vector of breeding values can be written as

$$u = P u + \phi$$

This can easiest be illustrated by dividing vector u into two parts, one being a part with breeding values of 'base' parents, which we define here as animals that have unknown pedigree, and the other part referring to animals that have both parents known. For simplicity we assume there are no animals with one parent known, but this could also be implemented. The vector of breeding values can be written as

$$\begin{pmatrix} u_b \\ u_p \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} u_b \\ u_p \end{pmatrix} + \begin{pmatrix} u_b \\ \mathbf{f} \end{pmatrix}$$

where P_{21} refers to the flow from base animals in u_b to progeny in u_p and P_{22} refers to the passing of genes to progeny from parents that are also progeny themselves. Each row of P has therefore at most two nonzero elements: a .5 for each parent.

We can rewrite (x) as

$$(I-P) u = \phi$$

12: The numerator relationships matrix

$$u = (I-P)\phi$$

The vector with breeding values u is now presented as a linear function of independent genetic values, which are in vector ϕ . This vector contains breeding values for base animals and Mendelian sampling values for offspring with two parents known. The additive genetic variance can be written as

$$\text{var}(u) = A = (I-P)^{-1} \text{var}(\phi) (I-P)^{-1} = (I-P)^{-1} D (I-P)^{-1} = TDT'$$

The matrix D is a diagonal matrix. If there is no inbreeding, then the values for each diagonal d_i are
 $d_i=1$ if no parents are known,
 $d_i=.75$ if one of the parents is known
 $d_i=.5$ if both parents are known.

With d_i referring to the variance (proportional to S_a^2) that has not been explained by known parents. We will first derive rules for A^{-1} because these rules are easiest and most often used in genetic evaluations. The inverse of the relationships matrix is

$$A^{-1} = [TDT']^{-1} = (I-P')D^{-1}(I-P)$$

P is a matrix with at most two elements of 0.5 in each row, one in the column for each parent, assuming there are no animals that have grandparents known but not their parents. This simple structure provides also simple rules to construct A^{-1}

The matrix $(I-P)$ is lower triangular with all diagonals equal to 1 and, if parents are known, for each row only two parent-progeny off-diagonals equal to -0.5 . Multiplication of $(I-P')$ and $(I-P)$ gives a symmetric matrix. It is constructed as: $(I-P')(I-P) = I - (P + P') + P' * P$.

This results is an identity matrix with a -0.5 on each known parent-progeny off-diagonal (created by subtraction of $(P'+P)$), and a value of 0.25 added to each parental diagonal as well as the off diagonal between 2 mates. In the A -inverse, these values are weighted by the diagonal matrix D^{-1} .

The matrix D^{-1} contains diagonals being (with no inbreeding) 1, 4/3, and 2 for animals with none, one, and two parents known, respectively.

The rules for constructing A^{-1} for this non-inbreeding case are then based on the following table (where the i, j and k refer to row and column numbers of an animal, its sire and its dam)

pedigree knowledge:		element added to	value
<i>no parents known</i> (i,0,0)	animals' diagonal	i,i	1.0
<i>one parent known</i> (i,j,0) or (i,0,k)	animals' diagonal	i,i	1.0 * 4/3
	parent-offspring off-diagonal	i,j and j,i or i,k and k,i	-0.5 * 4/3
	parents' diagonal	j,j or k,k	0.25 * 4/3
<i>two parents known</i> (i,j,k)	animals' diagonal	i,i	1.0 * 2
	parent-offspring off-diagonal	i,j; j,i; i,k and k,i	-0.5 * 2
	parents' diagonal	j,j and k,k	0.25 * 2
	mates' off-diagonal	j,k and k,j	0.25 * 2

12: The numerator relationships matrix

The P matrix has another nice property P identifies in each row i the parents of i, P² identifies in each row of i the grandparents of i, etc.

Therefore, P² contains in each row 4 nonzero elements, each equal to 0.25. When this does not hold, some grandparents have to be missing from the pedigree, or there has been inbreeding (e.g. when there are only three elements, and one of the elements is 0.5, then that grandparent has been used twice in the pedigree of the same animal). Note that the sum of each row of P is equal 0, 0.5 or 1 for none, one or two parents known.

In general P^t identifies ancestors back to t generations, hence P³ identifies great grand parents etc. If P^t=0, then the first ancestors date back to less than t generations.

Pedigree

1	0	0
2	0	0
3	0	0
4	0	0
5	1	0
6	3	4
7	5	6

P =	0	0	0	0	0	0	0
	0	0	0	0	0	0	0
	0	0	0	0	0	0	0
	0	0	0	0	0	0	0
	0.5	0	0	0	0	0	0
	0	0	0.5	0.5	0	0	0
	0	0	0	0	0.5	0.5	0

(I-P) =	1	0	0	0	0	0	0
	0	1	0	0	0	0	0
	0	0	1	0	0	0	0
	0	0	0	1	0	0	0
	-0.5	0	0	0	1	0	0
	0	0	-0.5	-0.5	0	1	0
	0	0	0	0	-0.5	-0.5	1

12: The numerator relationships matrix

P+P' =	0	0	0	0	0.5	0	0
	0	0	0	0	0	0	0
	0	0	0	0	0	0.5	0
	0	0	0	0	0	0.5	0
	0.5	0	0	0	0	0	0.5
	0	0	0.5	0.5	0	0	0.5
	0	0	0	0	0.5	0.5	0
<hr/>							
P*P	0.25	0	0	0	0	0	0
	0	0	0	0	0	0	0
	0	0	0.25	0.25	0	0	0
	0	0	0.25	0.25	0	0	0
	0	0	0	0	0.25	0.25	0
	0	0	0	0	0.25	0.25	0
	0	0	0	0	0	0	0
<hr/>							
(I-P')(I-P)	1.25	0	0	0	-0.5	0	0
	0	1	0	0	0	0	0
	0	0	1.25	0.25	0	-0.5	0
	0	0	0.25	1.25	0	-0.5	0
	-0.5	0	0	0	1.25	0.25	-0.5
	0	0	-0.5	-0.5	0.25	1.25	-0.5
	0	0	0	0	-0.5	-0.5	1
<hr/>							
D =	1	0	0	0	0	0	0
	0	1	0	0	0	0	0
	0	0	1	0	0	0	0
	0	0	0	1	0	0	0
	0	0	0	0	0.75	0	0
	0	0	0	0	0	0.5	0
	0	0	0	0	0	0	0.5
<hr/>							
A ⁻¹ = (I-P') D ⁻¹ (I-P)	1.33	0	0	0	-0.67	0	0
	0	1	0	0	0	0	0
	0	0	1.5	0.5	0	-1	0
	0	0	0.5	1.5	0	-1	0
	-0.67	0	0	0	1.83	0.5	-1
	0	0	-1	-1	0.5	2.5	-1
	0	0	0	0	-1	-1	2

Building the NRM while accounting for inbreeding

If there is inbreeding, there is only one element of the previous derivation that changes, which is the diagonal D matrix. The diagonal represents the variance due to Mendelian sampling within family. The within family variance is equal to $\frac{1}{2}\mathbf{s}_a^2$ with no inbreeding, but this variance is smaller if the parents are inbred. In practice, it becomes much more difficult to derive the appropriate coefficients of A when there is inbreeding, because to know the inbreeding, one should know the additive genetic relationship between parents. In the previous section we only derived simple rules for the inverse of A. Obtaining coefficients A for two individuals (e.g. two parents), which is more difficult. When an animal is inbred, it has an inbreeding coefficient F_i with $F_i = \frac{1}{2}a_{sd}$ where a_{sd} is equal to the additive genetic relationship between sire and dam. The variance of a breeding value of such an animal is then $\text{var}(u) = (1+F)\mathbf{s}_a^2 = \frac{1}{4}\text{var}(u_s) + \frac{1}{4}\text{var}(u_d) + \frac{1}{2}\text{cov}(u_s, u_d) + \text{var}(\phi)$ so that the within family variance is equal to

$$\text{var}(\phi) = (1+F_i)\mathbf{s}_a^2 - \frac{1}{4}(1+F_s)\mathbf{s}_a^2 - \frac{1}{4}(1+F_d)\mathbf{s}_a^2 - F_i\mathbf{s}_a^2 = \frac{1}{2}(1 - \frac{1}{2}(F_s + F_d))\mathbf{s}_a^2$$

Hence, the within family variance is not reduced if the parents are related (and the full sibs are inbred), but only when the parents are inbred. The variance of an inbred animal is $(1+F)$ and therefore the i^{th} diagonal of the relationships matrix is $1+F$.

The diagonal for D becomes $d_i = 1 - 0.25a_{ss} - 0.25a_{dd} = \frac{1}{2}(1 - \frac{1}{2}(F_s + F_d))$

And if only one parent known: $d_{ii} = 1 - 0.25a_{jj} = \frac{3}{4} - 0.25F_j$

Hence, to determine rules for A accounting for inbreeding, the diagonal elements of A have to be changed compared to the situation without inbreeding. It is a lot of work to determine A for large populations, but Quaas (1976) has given some rules to determine the diagonals of A efficiently, so that inbreeding coefficients and elements of A^{-1} with account for inbreeding can be derived from this.

We use again the structure of the relationships matrix and write

$$A = TDT = LL'$$

where $L = TD^{0.5}$ is a lower triangular matrix.

$T = (I-P)^{-1}$ is also lower triangular and it describes the flow of genes through the population. The $T(i,j)$ element of T indicates the fraction of the genes that animal i has received from animal j.

Diagonals of T are 1 and off-diagonals are nonzero between animals and their ancestors: equal to 0.5^n , where n is the number of generations between animal and ancestor.

This can be best seen in a numerical example.

12: The numerator relationships matrix

Pedigree with inbreeding (animal, sire, dam)

1	0	0
2	0	0
3	1	0
4	1	2
5	3	2
6	3	5
7	3	6

T = (I-P)⁻¹

1	0	0	0	0	0	0
0	1	0	0	0	0	0
0.5	0	1	0	0	0	0
0.5	0.5	0	1	0	0	0
0.25	0.5	0.5	0	1	0	0
0.375	0.25	0.75	0	0.5	1	0
0.4375	0.125	0.875	0	0.25	0.5	1

D =

1	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	0.75	0	0	0	0
0	0	0	0.5	0	0	0
0	0	0	0	0.5	0	0
0	0	0	0	0	0.5	0
0	0	0	0	0	0	0.4375

L = TD⁻⁵

1	0	0	0	0	0	0
0	1	0	0	0	0	0
0.5	0	0.866	0	0	0	0
0.5	0.5	0	0.7071	0	0	0
0.25	0.5	0.433	0	0.7071	0	0
0.375	0.25	0.6495	0	0.3536	0.7071	0
0.4375	0.125	0.7578	0	0.1768	0.3536	0.6614

A=T*D*T' = LL'

1	0	0.5	0.5	0.25	0.375	0.43
0	1	0	0.5	0.5	0.25	0.125
0.5	0	1	0.25	0.5	0.75	0.875
0.5	0.5	0.25	1	0.375	0.3125	0.281
0.25	0.5	0.5	0.375	1	0.75	0.625
0.375	0.25	0.75	0.3125	0.75	1.25	1
0.4375	0.125	0.875	0.2813	0.625	1	1.375

An algorithm to determine inbreeding coefficients for large pedigrees can be based on this principle (Quaas, 1976). Later, algorithms have been described, e.g. by Meuwissen and Luo (1992) as described in the book by Mrode (1996)

Quaas algorithm

The (i,j) element of A is now obtained as the multiplication of the ith row of L⁻¹ with the jth row of L⁻¹. The inbreeding coefficient of an animal can be computed as $F_i = 0.5a_{sd}$ where s and d are the rows in L of the sire and the dam. We do not have to store all rows of L to compute inbreeding coefficients, which would require a lot of memory space for large populations. Quaas (1976) has shown that the triangular structure of L can be efficiently exploited to calculate F efficiently. Therefore, each column of L will be computed and only one column at a time will be kept in memory (say in array V). The squared elements will be accumulated in another array (say array U) and working from columns 1 to N (N= number of animals), the accumulations in U will contain the sum of the squared elements for each row. Therefore, U contains after completion the diagonal elements of the relationship matrix A and the inbreeding coefficient for animal i is then

The i-th column of L can be made for each animal, and the sum of the squared elements can be accumulated giving each time that l_i is formed, the complete pedigree list has to be read, which is most of the work of this method. Even if a new generation of progeny is added to the pedigree list, we have to read also the old list of pedigree to create the columns of L for the new animals.

After the inbreeding coefficients have been determined, the elements of D can be computed as $d_i = 1 / [1 - (1 + F_s) - (1 + F_d)]$. We use then the rules of [1-5] to create the elements of the A matrix.

The algorithm to determine inbreeding coefficients the routine builds a column of L-matrix, the routine runs for an animal a column in array v and accumulates the sum of squared elements in array U= diagonals of A

```

!N=total number of animals
DO 10 I=1,N      ! for each animal
s=sire of i
d=dam of i
IF s NE 0 and d NE 0 THEN
XX= 1-0.25*U(s)-0.25*U(d)
IF s NE 0 and d = 0 THEN
XX= (1-0.25*U(s))
IF s = 0 and d NE 0 THEN
XX= (1-0.25*U(d))
IF s = 0 and d = 0 THEN
XX= 1
END IF
V(I)= sqrt(XX)      !V is column of L matrix
U(I)= U(I)+ XX*XX      !U accumulates
DO 2 J=I+1, N
s= sire of J,
then other off diagonals
d= dam of J
IF s GE I and d GE I THEN
XX= .5*(V(s)+ V(d))
IF s GE I and d LT I THEN
XX= .5*V(s)
IF s LT I and d GE I THEN
XX= .5*V(d)
IF s LT I and d LT I Then
XX= 0
END IF
V(J)=XX
U(J)= U(J)+ XX*XX
2 CONTINUE
10 CONTINUE
    
```

Exercise

- From following data set (Table 1), set up matrix A, and invert it. Construct a MME and solve it to estimate sex effect (as fixed) and predict the random effect of each animal (breeding value). You may have to draw a diagram of the pedigree to identify base animals and animals inbred, that helps you work out the Matrix A.
 For working out A, work also out the matrices P, D and T and L as in the steps above.
- Find the base population and calculate the average breeding value for the base population.

Table 1 Yearling weight in Herefords

animal	sire	dam	sex	YWT(kg)
1	0	0	M	-
2	0	0	F	-
3	0	0	M	-
4	0	0	F	-
5	0	0	M	-
6	1	2	M	400
7	3	4	F	300
8	5	4	M	312
9	6	7	M	405
10	8	7	F	298