Matrix Introduction and Operations

Janna Fierst and Peter Beerli

January 23, 2006

1 Matrix operations

1.1 Introduction

Matrix notation is a way of representing data and equations. An example from Bronson (1995):
T-shirts
Nine teal small and five teal medium; eight plum small and six plum medium; large sizes- three sand, one rose, two peach; also three medium rose, five medium sand, one peach medium, and seven peach small.

In a matrix the same information looks like this:

$$\mathbf{S} = \begin{array}{c} \\ small \\ medium \\ large \end{array} \begin{array}{c} \begin{array}{ccccc} Rose & Teal & Plum & Sand & Peach \end{array} \\ \begin{pmatrix} 0 & 9 & 8 & 0 & 7 \\ 3 & 5 & 6 & 5 & 1 \\ 1 & 0 & 0 & 3 & 2 \end{pmatrix} \end{array}$$

In this format, it is easy to understand and work with the information. By summing each column we can tell how many of each color there are, and by summing each row we can tell how many of each size there are.

Matrices are made up of elements arranged in horizontal rows and vertical columns. The size of a matrix is given as rows $\times$ columns, for example the t-shirt matrix is 3×5. A vector is a matrix with 1 row $\times$ any number of columns or 1 column $\times$ any number of rows.

$$\mathbf{q} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix}$$

The matrix itself is usually capitalized, and the elements of the matrix are referred to in lower case letters with row and then column in subscript. In the above matrix **S** each element s_{ij} corresponds to the element in the ith row and jth column, with s_{ij} representing the number of small teal shirts. Matrices are often expressed as capital and bold typed latin letters, whereas vectors most often are expressed as bold lower case latin letters.

1.2 Matrix Addition

Matrices can be added together or subtracted from one another if the dimensions of the two matrices are the same. If this is true, then each element of the matrix is added to its corresponding element in the other matrix. Subtraction works similarly. For example, if matrix **A** is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix}$$

and matrix **B** is

$$\mathbf{B} = \begin{bmatrix} 4 & 2 & 1 \\ 1 & 0 & 3 \\ 0 & 9 & 7 \end{bmatrix}$$

then matrix $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is defined as

$$\mathbf{C} = \begin{bmatrix} 0+4 & 1+2 & 2+1 \\ 3+1 & 1+0 & 0+3 \\ 2+0 & 0+9 & 0+7 \end{bmatrix}$$

A natural extension of matrices to programming is through arrays. 2-dimensional arrays are analogous the matrices shown here, but because arrays are not actually matrices, matrix operations have to be specified in most general programming languages.

Algorithm 1 Matrix Addition

```
if A and B are both n × p matrices then
  for i = 1 to n do
    for j = 0 to p do
      C_ij = A_ij + B_ij
    end for
  end for
end if
```

1.3 Matrix Multiplication

Matrix multiplication is less intuitive than matrix addition. First, the matrices have to be of defined proportions. Matrices of different sizes can be multipled as long as the number of columns in the first matrix equals the number of rows in the second matrix. The formal definition of this is that any matrix of n rows $\times p$ columns can be multiplied by any matrix of p rows $\times r$ columns, where the resulting matrix is n rows $\times r$ columns. Second, the actual operations are defined differently. If

$$\mathbf{D} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$$

and matrix **E** is

$$\mathbf{E} = \begin{bmatrix} 3 & -1 \\ 4 & -3 \end{bmatrix}$$

then matrix

$$\mathbf{F} = \mathbf{DE}$$

is defined as

$$\mathbf{F} = \begin{bmatrix} (2 \times 3) + (1 \times 4) & (2 \times -1) + (1 \times -3) \\ (3 \times 3) + (2 \times 4) & (3 \times -1) + (2 \times -3) \end{bmatrix}$$

This becomes more clear when we are working with linear equations. Linear equations can be written as:

$$2x + 4y = 10$$
$$7x + 2y = 12$$

Or, in matrix-vector notation:

$$\begin{bmatrix} 2 & 4 \\ 7 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 10 \\ 12 \end{bmatrix}$$

Now it is a little easier to see how to multiply the matrices and vectors, and what the result is.

1.4 The Lande Equation

An example of where this is used in evolution is the Lande equation and G-matrices. Lande (1979) defined the phenotypic response to selection as $\mathbf{\Delta z} = \mathbf{G}\beta$. This means that the per-generation change ($\mathbf{\Delta z}$) in a phenotypic trait (z) is equal to the additive genetic variance (G) multiplied by

Algorithm 2 Matrix Multiplication $\mathbf{C} = \mathbf{AB}$

```
if A is an n × p and B is p × r then
  for i = 1 to n do
    for j = 1 to r do
      for z = 1 to p do
        c_ij = a_iz b_zj + c_ij
      end for
    end for
  end for
end if
```

the selection on that trait (β), or the partial derivative of mean fitness with respect to the trait ($\frac{\partial \bar{w}}{\partial \bar{z}}$) (Lande and Arnold 1983). If we expand this to multiple phenotypic traits we can write the response to selection as a vector of changes in phenotypic traits. The matrix **G** contains the additive genetic variances for each trait on the diagonal and the genetic covariances between traits are the off-diagonal elements. To find the response to selection, we multiply this **G** matrix by the selection vector.

$$\begin{bmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta z_3 \end{bmatrix} = \begin{bmatrix} G_1 & CV_{12} & CV_{13} \\ CV_{12} & G_2 & CV_{23} \\ CV_{13} & CV_{23} & G_3 \end{bmatrix} \begin{bmatrix} \frac{\partial \bar{w}}{\partial \bar{z_1}} \\ \frac{\partial \bar{w}}{\partial \bar{z_2}} \\ \frac{\partial \bar{w}}{\partial \bar{z_3}} \end{bmatrix}$$

The system can be represented in matrix-vector notation, but can also be split into three separate equations.

$$\Delta z_1 = \frac{\partial \bar{w}}{\partial \bar{z_1}} \times G_1 + \frac{\partial \bar{w}}{\partial \bar{z_2}} \times CV_{12} + \frac{\partial \bar{w}}{\partial \bar{z_2}} \times CV_{13}$$

Looking at it this way, we can see that the change in each trait is found by summing the selective forces caused by selection on the trait itself and correlated effects from selection on other traits.

1.5 Matrix Transposition

A matrix can be transposed from $\mathbf{A}$ to $\mathbf{A}^T$ by converting all the columns of matrix $\mathbf{A}$ to the rows of matrix $\mathbf{A}^T$ and the rows of matrix $\mathbf{A}$ to the columns of matrix $\mathbf{A}^T$. The first row of $\mathbf{A}$ becomes the first column of $\mathbf{A}^T$, and so on. The definition of this is if $\mathbf{A}$ is an $n \times p$ matrix, then the transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^T$ and is defined as $\mathbf{A}^T(j, i) = \mathbf{A}(i, j)$.

Algorithm 3 Matrix Transposition $\mathbf{A} \rightarrow \mathbf{A}^T$

```
for i = 1 to n do
  for j = 1 to m do
    A^T(j,i) = A(i,j)
  end for
end for
```

1.6 Matrix Inversion

If we have a system of linear equations $\mathbf{Ax} = \mathbf{b}$, we might like to solve for $\mathbf{x}$. In an algebraic equation this would happen by dividing both sides by $\mathbf{A}$ but in matrix algebra division is undefined. Instead, we use matrix inversion. A matrix $\mathbf{A}^{-1}$ is defined as the inverse of matrix $\mathbf{A}$ if

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix. $\mathbf{I}$ is defined as a square matrix with all diagonal elements equal to one and all off-diagonal elements equal to 0. In order to satisfy this, both matrices must be square and of the same order. If there is no matrix $\mathbf{A}^{-1}$ that satisfies this condition the matrix is *singular*. Multiplication of a square matrix with its inverse is commutative

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I},$$

but multiplication of two different (square) matrices $\mathbf{A}$ and $\mathbf{B}$ is not

$$\mathbf{AB} \neq \mathbf{BA}$$

here an example of an inversion:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix inversion can be used in the Lande example to solve for β, the selection vector. The inverse of a 2×2 matrix is found with the determinant, defined as a11×a22 - a12×a21. The inverse is:

$$A^{-1} = \frac{1}{detA} \begin{bmatrix} a22 & -a12 \\ -a21 & a11 \end{bmatrix}$$

The Lande equation specifies that at equilibrium the change in the trait will be zero, or Δz=0. If we have biased mutation (w) or some other force acting on the system, we can express it as:

$$\begin{bmatrix} 0 \\ w \end{bmatrix} = \begin{bmatrix} G_1 & CV_{12} \\ CV_{12} & G_2 \end{bmatrix} \begin{bmatrix} \frac{\partial \bar{w}}{\partial \bar{z}_1} \\ \frac{\partial \bar{w}}{\partial \bar{z}_2} \end{bmatrix}$$

To solve for β at equilibrium, we multiply both sides by the inverse matrix

$$\begin{bmatrix} 0 \\ w \end{bmatrix} \begin{bmatrix} G_2/(detA) & -CV_{12}/(detA) \\ -CV_{12}/(detA) & G_1/(detA) \end{bmatrix} = \begin{bmatrix} \frac{\partial \bar{w}}{\partial \bar{z_1}} \\ \frac{\partial \bar{w}}{\partial \bar{z_2}} \end{bmatrix}$$

And we can now solve for each β_i.

1.7 Vector and matrix norms

A Norm is a measure of distance. We can apply norms to vectors and matrices.

1.7.1 Vector norms

Requirements for vector norms are:

- $f(x) \geq 0 \quad (f(x) = 0 \iff x = 0)$
- $f(x + y) \leq f(x) + f(y)$
- $f(\alpha x) = |\alpha| f(x)$

$f(x)$ is the vector norm and is expressed typically as $||\mathbf{x}||$. Several vector norms are often used. The general expression is sthe *p-norm*. It is expressed as

$$||x||_p = (|x_1|^p + |x_2|^p + ... + |x_n|^p)^{\frac{1}{p}}$$

Of these the 1,2, and ∞ norms are the most commonly used ones:

$$\begin{aligned} ||x||_1 =& |x_1| + |x_2| + ... + |x_n| \\ ||x||_2 =& |x_1|^2 + |x_2|^2 + ... + |x_n|^2)^{\frac{1}{2}} \\ ||x||_\infty =& |x_1|^\infty + |x_2|^\infty + ... + |x_n|^\infty)^{\frac{1}{\infty}} = \max_{1 \leq i \leq n} |x_i| \end{aligned}$$

the 1-norm is also called Manhattan distance or city block distance and the 2-norm is the Euclidian distance. Vector norms have some cool properties for example the *Hölder inequality:*

$$|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||_p||\mathbf{y}||_q \quad \frac{1}{p} + \frac{1}{q} = 1$$

A special case is the *Cauchy-Schwartz inequality*

$$|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||_2||\mathbf{y}||_2$$

Several more inequalities can be used to approximate or bound norms, but you might want to look into the book by Golub and van Loan (1996).

1.7.2 Matrix norms

Matrix norms are an important measure to assess whether they are fit for some operations, matrix norm can measure whether a matrix is near singularity. Matrix norms need the same requirements as the vector norms. Examples for matrix norms are the *Frobenius-norm*

$$||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

and the *p-norm*

$$||\mathbf{A}||_p = \sup_{x\neq 0}\frac{||\mathbf{Ax}||_p}{||\mathbf{x}||_p}.$$

Think of sup as the maximum, at least for real numbers. the matrix norms often can be broken down into vector norms.

1.8 Summary of matrix operations not explicitely discussed

$$\begin{aligned}
\mathbf{A}\alpha &= \alpha\mathbf{A} \\
\mathbf{A}_{m\times m,\text{symmetric}} &= \mathbf{A}^T_{m\times m,\text{symmetric}} \\
\mathbf{AB} &\neq \mathbf{BA} \\
(\mathbf{AB})^T &= \mathbf{B}^T\mathbf{A}^T \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1}
\end{aligned}$$

1.9 Sources and Additional Reading

Bronson, R. 1995. *Linear Algebra: An Introduction.* San Diego, CA, Academic Press.

Golub, G. H., and C. F van Loan 1996. *Matrix computations.* 3rd edition. John Hopkins University Press, Baltimore and London.

Lande, R. 1979. Quantitative-genetic analysis of multivariate evolution, applied to brain-body size allometry. *Evolution* 33: 402416.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:12101226.

Trefethen, L.N. and D. Bau, III. 1997. *Numerical Linear Algebra.* Philadelphia, PA, Society for Industrial and Applied Mathematics.