Chapter 1: Linear Algebra

1.1 *Introduction to Vector and Matrix Notation*

Much of the mathematical reasoning in all of the sciences that pertain to humans is linear in nature, and linear equations can be greatly condensed by matrix notation and matrix algebra. In fact, were it not for matrix notation, some equations could fill entire pages and defy our understanding. The first step in creating easier-to-grasp linear equations is to define the *vector*. A vector is defined as an ordered set of numbers. Vectors are classified as either *row vectors* or *column vectors*. Note that a vector with one element is called a *scalar*. Here are two examples. The vector **a** is a column vector with m elements,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix},$$

and the vector **b** is a row vector with q elements:

$$\mathbf{b} = [b_1 b_2 \cdots b_q] .$$

You should notice that in this text vectors are generally represented with lower case letters in bold.

There are a variety of ways that we can operate on vectors, but one of the simplest is the *transpose operator*, which, when applied to a vector, turns a row into a column and vice versa. For example,

$$\mathbf{a}' = [a_1 \quad a_2 \quad \cdots \quad a_m].$$

By convention, in this book, a vector with a transpose will generally imply that we are dealing with a row. The implication is that by default, all vectors are columns.

A *matrix* is defined as a collection of vectors, for example

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

$$= \{x_{ij}\}.$$

In this text, matrices are typically represented with an upper case bold letter.

The square brackets are used to list all of the elements of a matrix while the curly brackets are sometimes used to show a typical element of the matrix and thereby symbolize the entire matrix in that manner. Note that the first subscript of **X** indexes the row, while the second indexes columns. Matrices are characterized by their *order*, that is to say, the number of rows and columns that they have. The above matrix **X** is of order n by m, sometimes written n · m. From time to time we may see a matrix like **X** written with its order like so: ${}_n\mathbf{X}_m$. It is semantically appropriate to say that a row vector is a matrix of but one row, and a column vector is a matrix of one column. Of course, a scalar can be thought of as the special case when we have a 1 by 1 matrix.

At times it will prove useful to keep track of the individual vectors that comprise a matrix. Suppose, for example that we defined each of the rows of **X** as

$$\begin{aligned}\mathbf{x}'_{1\cdot} &= [\mathrm{x}_{11} \quad \mathrm{x}_{12} \quad \cdots \quad \mathrm{x}_{1m}] \\ \mathbf{x}'_{2\cdot} &= [\mathrm{x}_{21} \quad \mathrm{x}_{22} \quad \cdots \quad \mathrm{x}_{2m}] \\ \cdots &= \quad \cdots \\ \mathbf{x}'_{n\cdot} &= [\mathrm{x}_{n1} \quad \mathrm{x}_{n2} \quad \cdots \quad \mathrm{x}_{nm}]\end{aligned}$$

and then defined each column of **X**:

$$\mathbf{x}_{\cdot 1} = \begin{bmatrix} \mathrm{x}_{11} \\ \mathrm{x}_{21} \\ \cdots \\ \mathrm{x}_{n1} \end{bmatrix}, \mathbf{x}_{\cdot 2} = \begin{bmatrix} \mathrm{x}_{12} \\ \mathrm{x}_{22} \\ \cdots \\ \mathrm{x}_{n2} \end{bmatrix}, \cdots, \mathbf{x}_{\cdot m} = \begin{bmatrix} \mathrm{x}_{1m} \\ \mathrm{x}_{2m} \\ \cdots \\ \mathrm{x}_{nm} \end{bmatrix} \tag{1.1}$$

so that **X** could be represented as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_{1\cdot} \\ \mathbf{x}'_{2\cdot} \\ \cdots \\ \mathbf{x}'_{n\cdot} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{\cdot 1} & \mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}_{\cdot m} \end{bmatrix}. \tag{1.2}$$

In this context, the dot is known as a *subscript reduction operator* since it allows us to aggregate over the subscript replaced by the dot. So for example, the dot in $\mathbf{x}'_{i\cdot}$ summarizes all of the columns in the ith row of **X**.

Every so often a matrix will have exactly as many rows as columns, in which case it is a *square matrix*. Many matrices of importance in statistics are in fact square.

1.2 *The First Steps Towards an Algebra for Matrices*

One of the first steps we need to make to create an algebra for matrices is to define equality. We now do so defining two matrices

$$\mathbf{A} = \mathbf{B} \text{ iff } a_{ij} = b_{ij} \text{ for all i, j.} \tag{1.3}$$

Every element of **A** and **B** needs to be identical. For this to be possible, obviously both **A** and **B** must have the same order!

Just as one can transpose a vector, a matrix can be transposed as well. *Matrix transposition* takes all rows into columns and vice versa. For example,

$$\begin{bmatrix} 3 & 2 \\ 4 & 1 \\ 4 & 5 \end{bmatrix}' = \begin{bmatrix} 3 & 4 & 4 \\ 2 & 1 & 5 \end{bmatrix}.$$

Bringing our old friend **X** back, we could say that

$$\mathbf{X}' = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{x}'_{\cdot 1} \\ \mathbf{x}'_{\cdot 2} \\ \cdots \\ \mathbf{x}'_{\cdot m} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1\cdot} & \mathbf{x}_{2\cdot} & \cdots & \mathbf{x}_{n\cdot} \end{bmatrix}.$$

We might add that

$$(\mathbf{X}')' = \mathbf{X}. \tag{1.4}$$

A square matrix **S** is called *symmetric* if

$$\mathbf{S} = \mathbf{S}'. \tag{1.5}$$

Of course, a scalar, being a 1 by 1 matrix, is always symmetric.

Now we are ready to define *matrix addition*. For two matrices **A** and **B** of the same order, their sum is defined as the addition of each corresponding element as in

$$\begin{aligned} \mathbf{C} &= \mathbf{A} + \mathbf{B} \\ \{c_{ij}\} &= a_{ij} + b_{ij}. \end{aligned} \tag{1.6}$$

That is to say, we take each element of **A** and **B** and add them to produce the corresponding element of the sum. Here it must be emphasized that matrix addition is only possible if the components are *conformable for addition*. In order to be conformable for addition, they must have the same number of rows and columns.

It is possible to multiply a scalar times a matrix. This is called, appropriately enough, *scalar multiplication*. If c is a scalar, we could have

$$c\mathbf{A} = \mathbf{B}.$$

For example we might have

$$c \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} c \cdot a_{11} & c \cdot a_{12} \\ c \cdot a_{21} & c \cdot a_{22} \\ c \cdot a_{31} & c \cdot a_{32} \end{bmatrix}.$$

Assuming that c_1 and c_2 are scalars, we can outline some properties of scalar multiplication:

Associative: $c_1(c_2\mathbf{A}) = (c_1c_2)\mathbf{A}$ (1.7)

Distributive: $(c_1 + c_2)\,\mathbf{A} = c_1\,\mathbf{A} + c_2\,\mathbf{A}$ (1.8)

Now that we have defined matrix addition and scalar multiplication, we can define *matrix subtraction* as

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}\ .$$

There are a few special matrices that will be of use later that have particular names. For example, an n by m matrix filled with zeroes is called a *null matrix*,

$$ {}_{\mathrm{n}}\mathbf{0}_{\mathrm{m}} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{1.9}$$

and an n · m matrix of ones is called a *unit matrix*:

$$ {}_{\mathrm{n}}\mathbf{1}_{\mathrm{m}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{1.10}$$

We have already seen that a matrix that is equal to its transpose ($\mathbf{S} = \mathbf{S}'$) is referred to as symmetric. A *diagonal matrix*, such as **D**, is a special case of a symmetric matrix such that

$$\mathbf{D} = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_{mm} \end{bmatrix}. \tag{1.11}$$

i. e. the matrix consists of zeroes in all of the *off-diagonal* positions. In contrast, the *diagonal* positions hold elements for which the subscripts are identical.

A special case of a diagonal matrix is called a *scalar matrix*, a typical example of which appears below:

$$\begin{bmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & c \end{bmatrix}. \tag{1.12}$$

And finally, a special type of scalar matrix is called the *identity matrix*. As we will soon see, the identity matrix serves as the identity element of matrix multiplication. For now, note that we generally use the symbol **I** to refer to such a matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Having defined the identity matrix, we can think of a scalar matrix as being expressible as c**I** where c is a scalar.

We can now define some properties of matrix addition.

Commutative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (1.13)

Associative: $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$ (1.14)

Identity: $\mathbf{A} + \mathbf{0} = \mathbf{A}$ (1.15)

Note that in the definitions above we have assumed that all matrices are conformable for addition.

At this point we are ever closer to having all of the tools we need to create an algebra with vectors and matrices. We are only missing a way to multiply vectors and matrices. We now turn to that task. Assume we have a 1 by m row vector, **a**', and an m by 1 column vector, **b**. In that case, we can have

$$\mathbf{a}'\mathbf{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix}$$

$$= a_1 b_1 + a_2 b_2 + \cdots + a_m b_m \tag{1.16}$$

$$= \sum_{i=1}^{m} a_i b_i .$$

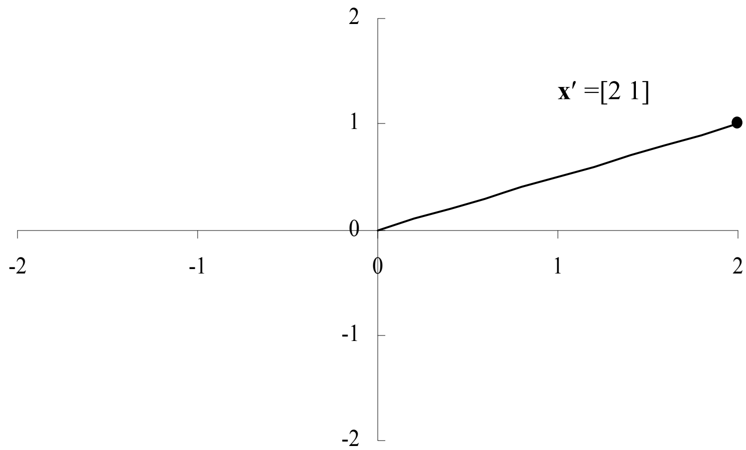
This operation is called taking a *linear combination*, but it is also known as the *scalar product*, the *inner product*, and the *dot product*. This is an extremely useful operation and a way to express a linear function with a very dense notation. For example, to sum the elements of a vector, we need only write

$$\mathbf{1}'\mathbf{a} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix} = \sum_{i}^{n} a_i .$$

When a linear combination of two non-null vectors equals zero, we say that they are *orthogonal* as $\mathbf{x}'$ and $\mathbf{y}$ below:

$$\mathbf{x}'\mathbf{y} = 0 . \tag{1.17}$$

Geometrically, this is equivalent to saying that they are at right angles in a space with as many axes as there are elements in the vector. Assume for example that we have a 2 element vector. This can be interpreted as a point, or a vector with a terminus, in a plane (a two space). Consider the graph below:



Note that the vector $\mathbf{x}' = [2 \quad 1]$ is represented in the graph. Can you picture an orthogonal vector? The *length of a vector* is given by $\sqrt{\mathbf{x}'\mathbf{x}} = \sum x_i^2$.

1.3 *Matrix Multiplication*

The main difference between scalar and matrix multiplication, a difference that can really throw off students, is that the commutative property does not apply in matrix multiplication. In general, **AB** ≠ **BA**, but what's more, **BA** may not even be possible. We shall see why in a second. For now, note that in the product **AB**, where **A** is $m \cdot n$ and **B** is $n \cdot p$, we would call **A** the *premultiplying* matrix and **B** the *postmultiplying* matrix. Each row of **A** is combined with each column of **B** in vector multiplication. An element of the product matrix, c_{ij}, is produced from the ith row of **A** and the jth column of **B**. In other words,

$$c_{ij} = \mathbf{a}'_{i\cdot}\mathbf{b}_{\cdot j} = \sum_k^n a_{ik} b_{kj} \tag{1.18}$$

The first thing we should note here is that the row vectors of **A** must be of the same order as the column vectors of **B**, in our case of order n. If not, **A** and **B** would not be *conformable for multiplication.* We could diagram things like this:

$${}_m\mathbf{C}_p = {}_m\mathbf{A}_n\,{}_n\mathbf{B}_p$$

Here the new matrix **C** takes on the number of rows of **A** and the number of columns of **B**. The number of columns of **A** must match the number of rows of **B**. OK, now let's look at a quick example. Say

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

$$= \begin{bmatrix} (-1)2 + (3)1 + (2)1 & (-1)3 + (3)4 + (2)2 \\ (2)2 + (0)1 + (1)1 & (2)3 + (0)4 + (1)2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 13 \\ 5 & 8 \end{bmatrix}.$$

A particular triple product, with a premultiplying row vector, a square matrix, and a postmultiplying column vector, is known as a *bilinear form*:

$$_1\mathbf{c}_1 = {}_1\mathbf{a}'_m \; {}_m\mathbf{B}_m \; {}_m\mathbf{d}_1 \tag{1.19}$$

A very important special case of the bilinear form is the *quadratic form*, in which the vectors **a** and **d** above are the same:

$$_1\mathbf{c}_1 = {}_1\mathbf{a}'_m \; {}_m\mathbf{B}_m \; {}_m\mathbf{a}_1 \tag{1.20}$$

The quadratic form is widely used because it represents the variance of a linear transformation.

For completion, we now present a *vector outer product*, in which an m by 1 vector, say **a**, is postmultiplied by a row vector, $\mathbf{b}'$:

$$_m\mathbf{C}_n = {}_m\mathbf{a}_1 \; {}_1\mathbf{b}'_n$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \tag{1.21}$$

$$= \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \cdots & \cdots & \cdots & \cdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

The matrix **C** has m · n elements, but yet it was created from only m + n elements. Obviously, some elements in **C** must be redundant in some way. It is possible to have a matrix outer product as well - for example a 4 by 2 multiplied by a 2 by 4 would also be considered an outer product.

1.4 *Partitioned Matrices*

It is sometimes desirable to keep track of parts of matrices other than either individual rows or columns as we did with the dot subscript reduction operator. For example, lets say that the matrix **A**, which is m by p consists of two partitions, $\mathbf{A_1}$ which is m by p_1 and $\mathbf{A_2}$ which is m by p_2, where $p_1 + p_2 = p$. Thus both $\mathbf{A_1}$ and $\mathbf{A_2}$ have the same number of rows and when stacked horizontally, as they will be below, their columns will add up to the number of columns of **A**. Then lets say we have the matrix **B**, which is of the order p by r, has two partitions $\mathbf{B_1}$ and $\mathbf{B_2}$ with $\mathbf{B_1}$ being p_1 by r and $\mathbf{B_2}$ being p_2 by r. The partitions $\mathbf{B}_1$ and $\mathbf{B}_2$ both have the same number of columns, namely r, so that when they are stacked vertically they match perfectly and their rows add up to the number of rows in **B**. In that case,

$$\mathbf{AB} = \left[\mathbf{A}_1 \;\vdots\; \mathbf{A}_2\right]\begin{bmatrix}\mathbf{B}_1 \\ --- \\ \mathbf{B}_2\end{bmatrix} = \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2 \tag{1.22}$$

We note that the product $\mathbf{A}_1\mathbf{B}_1$ and the product $\mathbf{A}_2\mathbf{B}_2$ are both conformable with order of m by r, precisely the order of **AB**.

1.5 *Cross-Product Matrices*

The cross product matrix is one of the most useful and common matrices in statistics. Assume we have a sample of n cases and that we have m variables. We define x_{ij} as the observation on consumer i (or store i or competitor i or segment i, etc.) with variable or measurement j. We can say that $\mathbf{x}'_{i\cdot}$ is a 1 · m row vector that contains all of the measurements on case i and that $\mathbf{x}_{\cdot j}$ is the n · 1 column vector containing all cases' measurements on variable j. The matrix **X** can then be expressed as a partitioned matrix, either as a series of row vectors, one per case, or as a series of columns, one per variable:

$$\mathbf{X} = \begin{bmatrix}\mathbf{x}'_{1\cdot} \\ \mathbf{x}'_{2\cdot} \\ \cdots \\ \mathbf{x}'_{n\cdot}\end{bmatrix} = \begin{bmatrix}\mathbf{x}_{\cdot 1} & \mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}_{\cdot m}\end{bmatrix} \tag{1.23}$$

What happens when we transpose X? All the rows become columns and all the columns become rows, as we can see below:

$$\mathbf{X}' = \begin{bmatrix}\mathbf{x}_{1\cdot} & \mathbf{x}_{2\cdot} & \cdots & \mathbf{x}_{n\cdot}\end{bmatrix} = \begin{bmatrix}\mathbf{x}'_{\cdot 1} \\ \mathbf{x}'_{\cdot 2} \\ \cdots \\ \mathbf{x}'_{\cdot m}\end{bmatrix} \tag{1.25}$$

In the right piece, a typical row would be $\mathbf{x}'_{\cdot j}$ which holds the data on variable j, but now in row format. This row has n columns. In the left piece, $\mathbf{x}_{i\cdot}$ is an m by 1 column holding all of the variables for case i. Now we have two possible ways to express the cross product, $\mathbf{X}'\mathbf{X}$. In the first approach, we show the columns of **X** which are now the rows of **X'**:

$$\mathbf{B} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{x}'_{\cdot 1} \\ \mathbf{x}'_{\cdot 2} \\ \cdots \\ \mathbf{x}'_{\cdot \mathbf{m}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\cdot 1} & \mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}_{\cdot \mathbf{m}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{x}'_{\cdot 1}\mathbf{x}_{\cdot 1} & \mathbf{x}'_{\cdot 1}\mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}'_{\cdot 1}\mathbf{x}_{\cdot \mathbf{m}} \\ \mathbf{x}'_{\cdot 2}\mathbf{x}_{\cdot 1} & \mathbf{x}'_{\cdot 2}\mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}'_{\cdot 2}\mathbf{x}_{\cdot \mathbf{m}} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{x}'_{\cdot \mathbf{m}}\mathbf{x}_{\cdot 1} & \mathbf{x}'_{\cdot \mathbf{m}}\mathbf{x}_{\cdot 2} & \cdots & \mathbf{x}'_{\cdot \mathbf{m}}\mathbf{x}_{\cdot \mathbf{m}} \end{bmatrix} \tag{1.26}$$

$$= \{\mathbf{x}'_{\cdot j}\mathbf{x}_{\cdot k}\} = \{b_{jk}\}$$

The above method of describing $\mathbf{X}'\mathbf{X}$ shows each element of the m by m matrix being created, one at a time. Each element of **X'X** is comprised of an inner product created by multiplying two n element vectors together. But now lets keep track of the rows of **X**, which are columns of **X**' which is just the opposite of what we did above. In this case, we have

$$\mathbf{B} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1\cdot} & \mathbf{x}_{2\cdot} & \cdots & \mathbf{x}_{n\cdot} \end{bmatrix} \begin{bmatrix} \mathbf{x}'_{1\cdot} \\ \mathbf{x}'_{2\cdot} \\ \cdots \\ \mathbf{x}'_{\mathbf{n}\cdot} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{x}_{1\cdot}\mathbf{x}'_{1\cdot} + \mathbf{x}_{2\cdot}\mathbf{x}'_{2\cdot} + \cdots + \mathbf{x}_{n\cdot}\mathbf{x}'_{n\cdot} \end{bmatrix} \tag{1.27}$$

$$= \sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{x}_{\mathbf{i}\cdot}\mathbf{x}'_{\mathbf{i}\cdot}$$

and the m · m outer products, $\mathbf{x}_{i\cdot}\mathbf{x}'_{i\cdot}$, are summed across all n cases to build up the cross product matrix, **B**.

1.6 *Properties of Matrix Multiplication*

In what follows, c is a scalar, and **A**, **B**, **C**, **D**, **E** are matrices. Note that we are assuming in all instances below that the matrices are conformable for multiplication.

Commutative: $$c\mathbf{A} = \mathbf{A}c \tag{1.28}$$

Associative: $$\mathbf{A}(c\mathbf{B}) = (c\mathbf{A})\mathbf{B} = c(\mathbf{AB}) \tag{1.29}$$

Looking at the above associative property for scalar multiplication, we can say that a scalar can *pass through* a matrix or a parenthesis.

Associative: $$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{1.30}$$

Right Distributive: $\mathbf{A}[\mathbf{B} + \mathbf{C}] = \mathbf{AB} + \mathbf{AC}$ (1.31)

Left Distributive: $[\mathbf{B} + \mathbf{C}]\mathbf{A} = \mathbf{BA} + \mathbf{CA}$ (1.32)

It is important to note here that unlike scalar algebra, we must distinguish between the left and right distributive properties. Again, note that these properties only hold when the symbols represent matrices that are conformable to the operations used in the equation.

From Equation (1.31) and (1.32) we can deduce the following

$$(\mathbf{A} + \mathbf{B})'(\mathbf{A} + \mathbf{B}) = \mathbf{A}'\mathbf{A} + \mathbf{A}'\mathbf{B} + \mathbf{B}'\mathbf{A} + \mathbf{B}'\mathbf{B} \,. \tag{1.33}$$

To multiply out an equation like Equation (1.33), students sometimes remember the mnemonic FOIL = first, outside, inside, last, which gives the sequence of terms to be multiplied.

Transpose of a Product: $[\mathbf{AB}]' = \mathbf{B}'\mathbf{A}'$ (1.34)

In words, the above theorem states that the transpose of a product is the product of the transposes in reverse order. And finally, the *identity element of matrix multiplication* is the previously defined matrix **I**:

Identity: $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$ (1.35)

1.7 *The Trace of a Square Matrix*

With a square matrix, from time to time we will have occasion to add up the diagonal elements, a sum known as *the trace of a matrix*. For example for the p by p matrix **S**, the trace of **S** is defined as

$$\mathrm{Tr}\,\mathbf{S} = \sum_i s_{ii} \,. \tag{1.36}$$

A scalar is equal to its own trace. We can also say that with conformable matrices **A** and **B**, such that **AB** and **BA** both exist, it can be shown that the

$$\mathrm{Tr}[\mathbf{AB}] = \mathrm{Tr}[\mathbf{BA}] \,. \tag{1.37}$$

The theorem is applicable if both **A** and **B** are square, or if **A** is $m \cdot n$ and **B** is $n \cdot m$.

1.8 *The Determinant of a Matrix*

While a square matrix of order m contains m^2 elements, one way to summarize all these numbers with one quantity is the *determinant*. The determinant has a key role in solving systems of linear equations. Consider the following two equations in two unknowns, x_1 and x_2.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= y_1 \\ a_{21}x_1 + a_{22}x_2 &= y_2 \end{aligned}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\mathbf{Ax} = \mathbf{y}$$

In a little while we will solve for the unknowns in the vector **x** using matrix notation. But for now, sticking with scalars, we can solve this using the following formula for x_1:

$$x_1 = \frac{y_1 a_{22} - y_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}} \tag{1.38}$$

The denominator of this formula is the determinant of the 2 by 2 matrix **A**. The determinant of a square matrix like **A** is usually written $|\mathbf{A}|$. Being in the denominator, the system cannot be solved when the determinant is zero. Whether the determinant is zero depends on how much information is in **A**. If rows or columns are redundant, then $|\mathbf{A}| = 0$ and there is no unique solution to the system of equations.

The determinant of a scalar is simply that scalar. Rules for determining the determinant of 3 by 3 and larger matrices can be found in Bock (1975, p. 62), Johnson and Wichern (2002, pp. 94-5) and other books on the linear model.

1.9 *The Inverse of a Matrix*

In scalar algebra we implicitly take the inverse to solve multiplication problems. If our system above was one equation in one unknown, it would be

$$ax = y$$

$$a^{-1}ax = a^{-1}y$$

$$1x = a^{-1}y$$

$$x = a^{-1}y$$

With a system of equations, the analog of $a^{-1} = 1/a$ is the inverse of a matrix, $\mathbf{A}^{-1}$.

$$\mathbf{Ax} = \mathbf{y}$$

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{y}$$

$$\mathbf{Ix} = \mathbf{A}^{-1}\mathbf{y}$$

To solve the system, you must find a matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. You can only do so when $|\mathbf{A}| \neq 0$. In fact, we have now just officially defined the inverse of a matrix. The inverse of a square matrix **A** is simply that matrix, which when pre- or post-multiplied by **A**, yields the identity matrix, i. e. $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. One property of inverses is that the inverse of a product is equal to the product of the inverses in reverse order:

Inverse of a Product: $$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{1.39}$$

For proof, consider that

$$\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B}$$

$$= \mathbf{B}^{-1}\mathbf{IB}$$

$$= \mathbf{I}$$

The inverse of the transpose of a square matrix is equal to the transpose of the inverse of that matrix. In other words, if $\mathbf{A}^{-1}$ is the inverse of **A**, then

$$\mathbf{A}^{-1}{}'\mathbf{A}' = \mathbf{I}\,. \tag{1.40}$$

1.10 *Kronecker Product*

The Kronecker Product with operator ⊗, is defined as

$${}_{mp}\mathbf{C}_{nq} = {}_{m}\mathbf{A}_{n} \otimes {}_{p}\mathbf{B}_{q} = \{a_{ij}\mathbf{B}\}\,. \tag{1.41}$$

For example,

$$\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}\mathbf{B} \\ \hdashline a_{21}\mathbf{B} \end{bmatrix}.$$

References

R. Darrell Bock (1975) *Multivariate Statistical Methods in Behavioral Research.* New York: McGraw-Hill.

Green, Paul E. (1976) *Mathematical Tools for Applied Multivariate Analysis.* New York: Academic

Johnson, Richard A. and Dean W. Wichern (2002) *Applied Multivariate Statistical Analysis, Fifth Edition.* Upper Saddle River, NJ: Prentice-Hall.