Appendix M:

The Appendix of Matrices

Overview:

Before we can start talking about regression, we need to cover the necessary mathematical background. For this book, I expect that you have successfully complete a course in matrix algebra.

This means that you can add and multiply matrices, that you understand matrices are linear transformations, that you have worked with eigenvalues and eigenvectors, and that you can calculate the rank and the trace of a matrix.

All of these topics are important in better understanding the mathematics underlying linear models. Thus, this appendix reviews some of the important parts from your matrix algebra class, and adds some items that you may not have had.

Forsberg, Ole J. (November 18, 2023). "The Appendix of Matrices." In *Linear Models and Řurità Kràlovstvì*. Version 0.704442ζ.

Chapter Contents

The purpose of this appendix is to provide the necessary matrix background for this book. Everything here is important at some point in this test. There will be new things here, in which case you need to learn them. If nothing here is new, then you merely need to review them to keep them fresh in your mind.

You should treat this appendix as necessary background knowledge. Gain it now, if necessary. At the very least, know what is here so that you can refer to it as you work through the main part of the text.

M.1: Matrix Basics

matrix

A matrix is just a rectangular array of scalars. It is used to simplify many mathematical calculations. Throughout this book, I will use it in such a sense. The following is an example of a matrix:

$$\mathbf{A} = \begin{bmatrix} 3 & 5 & 2 \\ a & 1 & 18 \end{bmatrix} \tag{M.1}$$

size

Because a matrix is a *rectangular* array, it has a dimension. The matrix $\mathbf{A}$ above has dimension 2×3 because there are 2 rows and 3 columns. We could also write this as

$$\mathbf{A} \in \mathcal{M}_{2\times 3} \tag{M.2}$$

This can be read as "$\mathbf{A}$ is a matrix with dimension 2×3" or as "$\mathbf{A}$ is an element of the set of 2×3 matrices." Note that the symbol $\in$ means "is an element of" and $\mathcal{M}_{2\times 3}$ is "the set of all matrices of dimension 2×3."

Also note that the dimension order is very important and is always written as rows × columns. $\mathcal{M}_{2\times3}$ and $\mathcal{M}_{2\times3}$ are entirely different sets of matrices.

A matrix is square if the number of rows equals the number of columns. That is, **B** is square if **square**

$$\mathbf{B} \in \mathcal{M}_{n\times n} \tag{M.3}$$

for some number $n \in \mathbb{Z}^+$. If a matrix is square, the set is often denoted simply by $\mathcal{M}_n$. The matrix **A** above is not square because the number of rows does not equal the number of columns.

M.1.1 Representation

The next sections cover the algebra of matrices. To ease the notation, let me show you two ways of representing matrices. First, here is matrix **A** written out.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1c} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2c} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & a_{r3} & \cdots & a_{rc} \end{bmatrix} \tag{M.4}$$

Note that the subscripts can also use a comma to separate the values. That is only done, however, when you get to double digits and ambiguity ensues. For instance, does a_{242} represent $a_{2,42}$ or $a_{24,2}$? Perhaps it actually represents $a_{2,4,2}$ in a tensor. Who knows when ambiguity enuses?

When we stick to single digits for the indices, commas are dropped.

Note that every element in the **A** matrix is represented by a lowercase a and its r, c position in the matrix. This allows us to simplify representation at times:

$$\mathbf{A} = \left[a_{ij}\right] \tag{M.5}$$

Here, i is the row index and j is the column index.

Now, with these two representations, we can enter the realm of the algebra of matrices.

M.2: Addition

Matrix addition is closed. This means that the sum of two matrices will always give you another matrix... as long as it makes sense to add two matrices. Two matrices are can be added if they have the same dimension.[1]

Let $\mathbf{A} \in \mathcal{M}_{r\times c}$ and $\mathbf{B} \in \mathcal{M}_{r\times c}$ for some values of r and c. $\mathbf{A}$ and $\mathbf{B}$ are commensurate and can be summed. Matrix addition is element-by-element (elementwise) addition. Thus,

elementwise

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} & \cdots & a_{1c}+b_{1c} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} & \cdots & a_{2c}+b_{2c} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} & \cdots & a_{3c}+b_{3c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{r1}+b_{r1} & a_{r2}+b_{r2} & a_{r3}+b_{r3} & \cdots & a_{rc}+b_{rc} \end{bmatrix} \tag{M.6}$$

This can also be symbolized (shortened) as

$$\mathbf{A}+\mathbf{B} = \left[a_{ij}+b_{ij}\right] \tag{M.7}$$

zero

Matrix addition has a zero (additive identity). It is the commensurate matrix with all elements equal to zero:

$$\mathbf{0} = \left[0_{ij}\right] \tag{M.8}$$

How does it work in addition? Just as you would expect:

$$\mathbf{A}+\mathbf{0} = \begin{bmatrix} a_{11}+0 & a_{12}+0 & a_{13}+0 & \cdots & a_{1c}+0 \\ a_{21}+0 & a_{22}+0 & a_{23}+0 & \cdots & a_{2c}+0 \\ a_{31}+0 & a_{32}+0 & a_{33}+0 & \cdots & a_{3c}+0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{r1}+0 & a_{r2}+0 & a_{r3}+0 & \cdots & a_{rc}+0 \end{bmatrix} = \mathbf{A} \tag{M.9}$$

This can also be symbolized (shortened) as

$$\mathbf{A}+\mathbf{0} = \left[a_{ij}+0\right] = \left[a_{ij}\right] \tag{M.10}$$

exercise

I leave it as an exercise to prove $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$.

[1] When the matrices have the correct dimension to perform the mathematical operation, they are called "commensurate." For addition, commensurate matrices have the same dimension. For multiplication, the requirement is much different (see M.3).

Matrices also have an additive inverse.As with scalar arithmetic, a matrix plus its additive inverse equals the zero matrix; that is, if $\mathbf{B}$ is the additive inverse of $\mathbf{A}$, then $\mathbf{A} + \mathbf{B} = \mathbf{0}$.

Two things about the additive inverse: First, it is commensurate with the original matrix. Second, it is unique (just as in scalar arithmetic).

To calculate the additive inverse of $\mathbf{A}$, just negate each element of $\mathbf{A}$. Thus, if $\mathbf{B} = \left[-a_{ij}\right]$ then $\mathbf{B}$ is the additive inverse of $\mathbf{A}$.

Finally, as with all elementwise operations, matrix addition is both commutative and associative:

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

In summation, matrix addition behaves like scalar addition, as long as the matrices are commensurate.

commensurate

M.3: Multiplication

There are many, many, *many* types of multiplication with matrices. The one selected depends on the intention (the need). They include scalar product, matrix product, Hadamard product (a.k.a. Schur product), and Kronecker product. As only the first two are typically seen in an undergraduate linear models course, we will only discuss those two here.

M.3.1 Scalar Product

As in arithmetic, the scalar product arose from needing to repeatedly add a matrix to itself. Thus, instead of writing $\mathbf{A} + \mathbf{A} + \mathbf{A} + \mathbf{A} + \mathbf{A} + \mathbf{A} + \mathbf{A} + \mathbf{A}$, one could write $8\mathbf{A}$, where 8 is a scalar. This was quickly generalized to non-integer values for the scalar multiple, just as $3 \times a$ was quickly generalized to things like $4.25 \times a$.

Scalar multiplication is defined as

$$c\mathbf{A} = \left[ca_{ij}\right] \tag{M.11}$$

Scalar products are commutative. That is, if c is a scalar and $\mathbf{A}$ is a matrix, then $c\mathbf{A} = \mathbf{A}c$. This will come in handy later, so be aware of it. Note that c does not need to be a natural number.

commutative

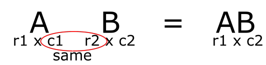

Figure M.1: *A schematic designed to illustrate commensurability with matrix multiplication. Note that the "inner" dimensions of the factors must be equal and that the product dimension is the "outers" of the two factors.*

associative

Scalar products are also associative. That is, if c is a scalar, then the following are equivalent:

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

distributive

Scalar multiplication is also distributive over matrix addition. Thus, $c(\mathbf{A}+\mathbf{B}) = c\mathbf{A}+c\mathbf{B}$.

M.3.2 Matrix Product

The matrix product is the multiplication that is (usually) meant when one just says "matrix multiplication." Its definition arises from linear algebra and repeated linear transformations. It has many nice properties. Calculation is not one of them.

Let us define two matrices **A** and **B** such that the number of columns of **A** equals the number of rows of **B**. Their product is defined as

$$\mathbf{AB} = \left[ab_{ij}\right] = \left[\sum_k a_{ik} b_{kj}\right] \tag{M.12}$$

where k ranges between 1 and the number of columns of **A**. The dimension of the product is the number of rows of **A** by the number of columns of **B**. That is, let $\mathbf{A} \in \mathcal{M}_{r1\times c1}$ and $\mathbf{B} \in \mathcal{M}_{r2\times c2}$. Then, one can multiply **A** and **B** if $c1 = r2$. The dimension of the product is $r1 \times c2$. Figure M.1 illustrates this.

Example 13.1: Here is an example of matrix multiplication. Let us define our two matrices as

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \tag{M.13}$$

and

$$\mathbf{B} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & m \end{bmatrix} \tag{M.14}$$

Let us find the product **AB**. •

Solution: The first step is to check that multiplying these matrices can be done. Do the number of columns of **A** equal the number of rows of **B**? Note that $\mathbf{A} \in \mathcal{M}_{2\times 3}$ and $\mathbf{B} \in \mathcal{M}_{3\times 3}$. Because the "inners" of **AB** are equal to each other, the matrix multiplication **AB** makes sense. **commensurate**

Second, we determine the dimension of the product. It is the number of rows of **A** by the number of columns of **B**: 2×3, the "outers."

Third, since we know the dimension of the product, we just have to fill in the blanks in the matrix:

$$\mathbf{AB} = \begin{bmatrix} - & - & - \\ - & - & - \end{bmatrix} \tag{M.15}$$

The top-right element in the product matrix is element 1,1. Thus, by our definition, it equals

$$ab_{11} = \left[\sum_k a_{1k} b_{k1}\right] \tag{M.16}$$

$$= [a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}] \tag{M.17}$$

$$= [1a + 2d + 3g] \tag{M.18}$$

The top-middle element, ab_{12}, is

$$ab_{12} = \left[\sum_k a_{1k} b_{k2}\right] \tag{M.19}$$

$$= [a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}] \tag{M.20}$$

$$= [1b + 2e + 3h] \tag{M.21}$$

inner product

Note what is happening here. The elements of the "top-right" cell is the inner product of the top row and the right column. Similarly, the bottom-center element is the inner product of the bottom row and the center column.

exercise

I leave it as an exercise for you to finish the multiplication. Here is the final product:

$$\mathbf{AB} = \begin{bmatrix} 1a + 2d + 3g & 1b + 2e + 3h & 1c + 2f + 3m \\ 4a + 5d + 6g & 4b + 5e + 6h & 4c + 5f + 6m \end{bmatrix} \tag{M.22}$$

♦

except

In scalar arithmetic, we have a multiplicative identity, multiplicative inverse, and multiplication is commutative, associative, and distributes over addition. All of these hold for matrix multiplication — *except* commutativity. In general, $\mathbf{AB} \neq \mathbf{BA}$, even when the multiplications both make sense.

identity

The multiplicative identity, which we will symbolize by $\mathbf{I}$, has the property that $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$.[2] Note that $\mathbf{I}$ is a square matrix.

inverse

If a multiplicative identity exists, then a multiplicative inverse also exists. The inverse of a matrix $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, is a matrix that satisfies these two requirements:

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{M.23}$$

invertible

Not all matrices have inverses. Those that do *not* are called singular. Those that *do* are called invertible.

full rank

Only square matrices can be invertible. However, not all square matrices *are* invertible. From linear algebra, a matrix is invertible if and only if it is square and is of full rank.

[2] Technically, this statement is only true if $\mathbf{A}$ is square. If it is not square, then the two $\mathbf{I}$ matrices will have different dimension. We will restrict ourselves to square matrices. The "generalized inverse" is beyond the scope of this text.

A consequence of this is that a matrix is invertible if its determinant is non-zero. In general, the calculation of the determinant and the inverse are computationally intensive. However, they are rather straight-forward for 2×2 matrices. **determinant**

Let $\mathbf{A} \in \mathcal{M}_{2\times 2}$. Then, the determinant of $\mathbf{A}$ is $\det \mathbf{A} = a_{11}a_{22} - a_{12}a_{21}$. Note that the determinant is a *scalar*, not a matrix.

Theorem M.1. *The inverse of* $\mathbf{A}$ *is*

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{M.24}$$

Proof. To prove this, we will show $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{M.25}$$

$$= \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{M.26}$$

$$= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & -a_{11}a_{12} + a_{11}a_{12} \\ a_{21}a_{22} - a_{21}a_{22} & -a_{12}a_{21} + a_{11}a_{22} \end{bmatrix} \tag{M.27}$$

$$= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & 0 \\ 0 & -a_{12}a_{21} + a_{11}a_{22} \end{bmatrix} \tag{M.28}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{M.29}$$

$$= \mathbf{I} \tag{M.30}$$

Thus, we have shown that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. This is one-half of the proof. I leave it as an exercise for you to prove the second half: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. □ **exercise**

> ***Note***: The reason I provide the mathematics for 2×2 matrices is that in our study of simple linear regression, many of the important calculations will be done with 2×2 matrices. See Chapter 2.

Note: Formulas exist for 3×3 matrices, too. However, once we move beyond that, hand calculations are time-prohibitive. At the end of this section, I provide some examples of performing these calculations in `R`.

Now that we have a mechanism to calculate a multiplicative inverse, let us see that not all square matrices have one.

Example 13.2: Let $\mathbf{A} \in \mathcal{M}_{2\times 2}$ be defined as

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \tag{M.31}$$

Calculate $\mathbf{A}^{-1}$. •

Solution: Let us calculate $\mathbf{A}^{-1}$ using our formula,

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{M.32}$$

Applying this formula is straight-forward:

$$\mathbf{A}^{-1} = \frac{1}{0} \begin{bmatrix} 6 & -3 \\ -2 & 1 \end{bmatrix} \tag{M.33}$$

singular

Yes, the determinant of $\mathbf{A}$ is $\det \mathbf{A} = 1 \times 6 - 3 \times 2 = 6 - 6 = 0$. Since the determinant is 0, the inverse does not exist (one cannot divide by 0). ♦

independent

Note: What is it about the $\mathbf{A}$ matrix that makes it singular? Note that the second column is just 3 times the first column (or the second row is twice the first). This means the matrix is not full rank. The columns are not linearly independent. When we get to using these matrices with real data, we will see this as the second column gives us no knowledge about the world that is not already contained in the first column. The second column is redundant.

redundant

❧❧❧

So far, we have seen the multiplicative identity and the multiplicative inverse. It is time to note that matrix multiplication is not commutative.

Theorem M.2. *Matrix multiplication is not commutative. That is, there exist* $\mathbf{A}$ *and* $\mathbf{B}$ *such that* $\mathbf{AB} \neq \mathbf{BA}$.

Note: If $\mathbf{A}$ and $\mathbf{B}$ are not square with the same dimension, then this statement is trivially true.

Proof. The proof is simple. It is simply a counter-example. Let

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 7 \end{bmatrix} \tag{M.34}$$

and

$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{M.35}$$

Note that $\mathbf{AB} \neq \mathbf{BA}$.

Since we have found a counter-example, we have shown that matrix multiplication is not commutative, in general. □

While technically correct, this proof leaves us feeling a little empty. So we found one counter-example. Cool beans. But we learned precious little about commutativity with matrix multiplication. Let us explore a bit and see if we can determine *when* matrix multiplication is commutative. We may learn something interesting.

First, let us assume **A** and **B** are square and commensurate. This ensures that **AB** and **BA** can be calculated. For instance, let both be 2×2 matrices. Then, **AB** is

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \tag{M.36}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} \tag{M.37}$$

and **BA** is

$$\mathbf{BA} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{M.38}$$

$$= \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{bmatrix} \tag{M.39}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{21}b_{12} & a_{12}b_{11} + a_{22}b_{12} \\ a_{11}b_{21} + a_{21}b_{22} & a_{12}b_{21} + a_{22}b_{22} \end{bmatrix} \tag{M.40}$$

By comparing the two product matrices, **AB** and **BA**, we can determine an instance where multiplication is commutative.

For instance, if $a_{12} = a_{21} = 0$ and $b_{12} = b_{21} = 0$, then the two product matrices are the same. In other words, if both **A** and **B** are diagonal matrices, multiplication will be commutative. That's an interesting consequence we would have missed if we just stopped with our proof.

Cool beans!

In fact, it can be proven that multiplication of diagonal matrices is commutative in general. Kewl!

Theorem M.3. *Let* **A** *and* **B** *be diagonal matrices of the same size. The product is commutative; that is,* $\mathbf{AB} = \mathbf{BA}$.

Proof. Since **A** and **B** are diagonal and of the same shape, they can be represented as

$$\mathbf{A} = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix} \tag{M.41}$$

and

$$\mathbf{B} = \begin{bmatrix} b_1 & 0 & 0 & \cdots & 0 \\ 0 & b_2 & 0 & \cdots & 0 \\ 0 & 0 & b_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_n \end{bmatrix} \tag{M.42}$$

Their product is

$$\mathbf{AB} = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 & 0 & 0 & \cdots & 0 \\ 0 & b_2 & 0 & \cdots & 0 \\ 0 & 0 & b_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_n \end{bmatrix} \tag{M.43}$$

$$= \begin{bmatrix} a_1 b_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 b_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 b_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n b_n \end{bmatrix} \tag{M.44}$$

Similarly, the product **BA** is

$$\mathbf{BA} = \begin{bmatrix} b_1 & 0 & 0 & \cdots & 0 \\ 0 & b_2 & 0 & \cdots & 0 \\ 0 & 0 & b_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_n \end{bmatrix} \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \end{bmatrix} \tag{M.45}$$

$$= \begin{bmatrix} b_1 a_1 & 0 & 0 & \cdots & 0 \\ 0 & b_2 a_2 & 0 & \cdots & 0 \\ 0 & 0 & b_3 a_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_n a_n \end{bmatrix} \tag{M.46}$$

Since scalar multiplication (in the cells) is commutative, we have

$$= \begin{bmatrix} a_1 b_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 b_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 b_3 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n b_n \end{bmatrix} \tag{M.47}$$

$$= \mathbf{AB} \tag{M.48}$$

Thus, we have shown $\mathbf{AB} = \mathbf{BA}$ for two diagonal matrices of the same size. □

❧ ❧ ❧

exercise

Finally, I leave it as an exercise for you to prove that matrix multiplication is associative (when the multiplication can be done). That is, if $\mathbf{ABC}$ can be calculated, then it can be calculated as either $(\mathbf{AB})\mathbf{C}$ or as $\mathbf{A}(\mathbf{BC})$.

M.4: Other Matrix Terms

There exist other helpful operations on matrices. Already, we have come across the determinant as being especially helpful in determining if a matrix is invertible or singular.

trace

Another useful function is the trace. It is just the sum of the diagonal elements. That is:

$$\text{tr } \mathbf{A} = \sum_{i=1}^{r} a_{ii} \tag{M.49}$$

The formula is simple... deceptively so. In fact, one may wonder what the trace actually tells us about a matrix. Well, in general, you will need to revisit your matrix algebra class notes. In linear models, however, the trace is used to calculate the number of degrees of freedom (see Section 2.3).

transpose

The transpose of a matrix is just the matrix where the rows and columns are switched. Thus, if $\mathbf{B}$ is the transpose of $\mathbf{A}$, then $b_{ij} = a_{ji}$. In symbols, we indicate the transpose as

$$\mathbf{B} = \mathbf{A}' \tag{M.50}$$

In all cases, $\mathbf{A}'\mathbf{A}$ and $\mathbf{AA}'$ both exist and are symmetric. Also, the rank of $\mathbf{A}'\mathbf{A} = rank\mathbf{AA}'$.

symmetric

A matrix is symmetric if it is equal to its transpose, $\mathbf{A} = \mathbf{A}'$.

$$\left[a_{ij}\right] = \left[a_{ji}\right] \tag{M.51}$$

Note that only square matrices can be symmetric. Symmetric matrices have some nice properties with respect to calculations.

symmetrize

Also note that one can "symmetrize" any square matrix. That is, one can form symmetric matrix $\mathbf{X}$ from a square matrix $\mathbf{A}$ as

$$\mathbf{X} = \frac{\mathbf{A} + \mathbf{A}'}{2} \tag{M.52}$$

I leave it as an exercise to prove that $\mathbf{X}$ is symmetric. **exercise**

One important feature of symmetric matrices is that they can be transformed into a diagonal matrix. If $\mathbf{A}$ is symmetric, then there exists a $\mathbf{Q}$ such that $\mathbf{AQ}$ is diagonal. Why is this helpful? First, remember that multiplication of diagonal matrices is commutative. Second, as you will see in the text, diagonal covariance matrices indicate independence.

This means that any set of variables can be linearly transformed into a set of independent variables. This fact is the basis for a procedure called "principal component analysis."

j vector The $\mathbf{j}$ vector is a vector of 1s. It is used to calculate row sums (if pre-multiplying) or column sums (if post-multiplying). The matrix $\mathbf{J}$ is a matrix of 1s. It does what $\mathbf{j}$ does, but puts the sums in a matrix.

$\mathbf{e_i}$ vector The $\mathbf{e_i}$ vector is a vector of 0s, with a 1 in the i^{th} position. It is used in proofs, as it can be used to select an individual row, column, or element of a matrix.

eigenvalue The eigenvalues of a matrix $\mathbf{A}$ are those values λ that solve the equation $\mathbf{Av} = \lambda\mathbf{v}$. The vectors $\mathbf{v}$ corresponding to each of the eigenvalues are called the eigenvectors.

idempotent A matrix $\mathbf{A}$ is idempotent if $\mathbf{AA} = \mathbf{A}$. The trace of an idempotent matrix equals its rank. This is rather important in studying linear models, since the rank is also the degrees of freedom.

Matrices $\mathbf{A}$ and $\mathbf{B}$ are orthogonal, $\mathbf{A} \perp \mathbf{B}$, if $\mathbf{A}'\mathbf{B} = \mathbf{0}$. That is, if the inner products of the columns of $\mathbf{A}$ and $\mathbf{B}$ are orthogonal, then the matrices themselves are orthogonal.

projection matrix A matrix $\mathbf{P}$ is a projection matrix if it is idempotent. The purpose of projection matrices is to project a higher space onto a subspace. If $\mathbf{P}$ is also symmetric, it is called an orthogonal projection matrix. This means it projects the larger space orthogonally (perpendicularly) onto the subspace. Think of shining a flashlight on a plant. If you put the flashlight directly over the plant, it will project the plant orthogonally onto the floor. If you do it at an angle, then the projection is called oblique. **oblique projection**

The key in both instances is that you are simplifying a complicated reality (3-D object) onto a simpler model (2-D shadow).

positive definite

M.4.1 Positive Definite Matrices A matrix $\mathbf{A}$ is a positive definite (pd) if $\mathbf{q}'\mathbf{A}\mathbf{q} > 0$ for all non-zero vectors $\mathbf{q}$. It is usually difficult to determine if a matrix is positive definite (pd). However, once you know it is, there are some important properties, which we look at in the next section (Sections M.5 and M.5.1).

M.5: Consequences

With these definitions, there are a lot of consequences. Many of which are important in the study of linear models. This section covers many of them.

First, when taking the transpose of a product, you switch the order of the multiplication: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. A similar result holds with inverses. The only difference is that all three inverses must exist: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Lemma M.4. *For any matrix* $\mathbf{X}$, *the matrix* $\mathbf{X}'\mathbf{X}$ *is symmetric.*

If you know the determinant of a matrix, you can easily calculate the determinant of a scalar multiple of that matrix.

Lemma M.5. *If* $c \in \mathbb{R}$ *and* $\mathbf{A} \in \mathcal{M}_n$, *then* $\det c\mathbf{A} = c^n \det \mathbf{A}$.

There is a similar result with the trace of a matrix.

Lemma M.6. *If* $c \in \mathbb{R}$ *and* $\mathbf{A} \in \mathcal{M}_n$, *then* $tr\ c\mathbf{A} = c\ tr\ \mathbf{A}$.

pd

M.5.1 Positive Definite Matrices There are a lot of interesting properties of positive definite matrices. So, let us break these into a separate subsection.

Lemma M.7. *The diagonal elements of a pd matrix are all positive. That is, let* $\mathbf{A} \in \mathcal{M}_n$ *be positive definite, then* $a_{ii} > 0,\ \forall i \in \left\{1, 2, \ldots, n\right\}$.

This can easily be shown by letting the $\mathbf{q}'$ row vector be $\mathbf{e_i}$. The quadratic form $\mathbf{q}'\mathbf{A}\mathbf{q}$ would therefore equal the diagonal element at position i. Since $\mathbf{A}$ is positive definite, that element must be greater than 0.

Note that the converse is not true. Just because the diagonal elements are all positive does not mean that the matrix is positive definite. For an example, note

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{M.53}$$

is not positive definite. To see this, let $\mathbf{q}'$ be the vector $[1\ -1]$. Then

$$\mathbf{q}'\mathbf{A}\mathbf{q} = [1\ -1] \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{M.54}$$

$$= [0\ 0] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{M.55}$$

$$= 0 \tag{M.56}$$

Since this is not greater than 0, **A** is not positive definite.

Note that the determinant of this **A** is 0. This suggests a second consequence: The determinant of a positive definite matrix is positive. This means all pd matrices are invertible. Similarly:

Lemma M.8. *The inverse of a pd matrix is also pd. That is, if* **A** *is positive definite, then so is* $\mathbf{A}^{-1}$.

To see this, ask yourself: What is the determinant of the inverse of a matrix? How can you use that to show that the inverse of a positive definite matrix is also positive definite? **thoughts**

Since the determinant of a pd matrix is positive, all of the eigenvalues are positive. And, since all of the diagonal elements of a pd matrix are positive, then the trace is positive. Since the trace is used to calculate the degrees of freedom, the matrix must be positive definite.

If **X** is a full rank matrix, even if not square, $\mathbf{X}'\mathbf{X}$ is positive definite.

The converses of the above statements also require that the matrix is symmetric. Thus, if **A** is symmetric *and* all of its eigenvalues are positive, then **A** is positive definite.

And, *most importantly*, the covariance matrix is positive definite if the design matrix, **X**, is full rank. Otherwise, it is positive semi-definite and has a determinant of zero. **covariance**

M.6: Statistics in Matrices

In this section, we rewrite some of the equations you learned in your previous statistics course in terms of matrices. This section is useful for matrix practice. In all of the following, let **Y** be a column vector of length n.

Lemma M.9. *The sample mean using matrices:* $\overline{Y} = \frac{1}{n}\mathbf{j}'\mathbf{Y}$.

Proof.

$$\frac{1}{n}\mathbf{j}'\mathbf{Y} = \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \tag{M.57}$$

$$= \frac{1}{n}\left(1Y_1 + 1Y_2 + 1Y_3 + \cdots + 1Y_n\right) \tag{M.58}$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{M.59}$$

$$= \overline{Y} \tag{M.60}$$

□

Lemma M.10. *The sum of squared values using matrices:* $\mathbf{Y}'\mathbf{Y}$.

Proof.

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \; \vdots \\ y_n \end{bmatrix} \tag{M.61}$$

$$= \left(y_1y_1 + y_2y_2 + y_3y_3 + \cdots + y_ny_n\right) \tag{M.62}$$

$$= \sum_{i=1}^{n} y_i^2 \tag{M.63}$$

□

Lemma M.11. *The sample variance using matrices:* $s_y^2 = \frac{1}{n-1}(\mathbf{Y} - \overline{y}\mathbf{j})'(\mathbf{Y} - \overline{y}\mathbf{j})$.

Proof.

$$\frac{1}{n-1}(\mathbf{Y} - \overline{y}\mathbf{j})'(\mathbf{Y} - \overline{y}\mathbf{j}) \tag{M.64}$$

$$= \frac{1}{n-1}\ [y_1 - \overline{y},\ y_2 - \overline{y},\ y_3 - \overline{y},\ \cdots,\ y_n - \overline{y},\ ] \begin{bmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \\ y_3 - \overline{y} \\ \cdots \\ y_n - \overline{y} \end{bmatrix} \tag{M.65}$$

$$= \frac{1}{n-1}\ (y_1 - \overline{y})(y_1 - \overline{y}) + (y_2 - \overline{y})(y_2 - \overline{y}) + \cdots + (y_n - \overline{y})(y_n - \overline{y}) \tag{M.66}$$

$$= \frac{1}{n-1}\ \sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{M.67}$$

□

Lemma M.12. *The sample covariance using matrices:* $s_{xy} = \frac{1}{n-1}(\mathbf{Y} - \overline{y}\mathbf{j})'(\mathbf{X} - \overline{x}\mathbf{j})$.

Proof. This proof echoes the previous proof.

$$\frac{1}{n-1}(\mathbf{Y} - \overline{y}\mathbf{j})'(\mathbf{X} - \overline{x}\mathbf{j}) \tag{M.68}$$

$$= \frac{1}{n-1}\ [y_1 - \overline{y},\ y_2 - \overline{y},\ y_3 - \overline{y},\ \cdots,\ y_n - \overline{y},\ ] \begin{bmatrix} x_1 - \overline{x} \\ x_2 - \overline{x} \\ x_3 - \overline{x} \\ \cdots \\ x_n - \overline{x} \end{bmatrix} \tag{M.69}$$

$$= \frac{1}{n-1}\ (y_1 - \overline{y})(x_1 - \overline{x}) + (y_2 - \overline{y})(x_2 - \overline{x}) + \cdots + (y_n - \overline{y})(x_n - \overline{x}) \tag{M.70}$$

$$= \frac{1}{n-1}\ \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) \tag{M.71}$$

□

Note that this notation leads to the synonym $s_y^2 = s_{yy}$. It also leads to a nice proof that the covariance matrix is symmetric. **synonym**

Definition M.13. *Let* **Y** *be a random vector (a column vector whose elements are random variables). The quantity* $\mathbb{V}[\mathbf{Y}]$ *is called the variance-covariance matrix of* **Y**. *It is often called just the covariance matrix of* **Y**.

Lemma M.14. *Let* $\mathbf{Y} \in \mathcal{M}_{r,c}$. *If* $\mathbf{Y}' = [\mathbf{Y_1}, \mathbf{Y_2}, \mathbf{Y_3}, \cdots, \mathbf{Y_r}]$, *then the elements of* $\mathbb{V}[\mathbf{Y}]$ *are* $\left[\sigma_{ij}\right]$, *where* σ_{ij} *is the covariance between* Y_i *and* Y_j *and* $\sigma_{i,i}$ *is the variance of* Y_i.

Lemma M.15. *Covariance matrices are symmetric.*

Lemma M.16. *If* $\mathbf{Y}$ *is a random vector and* $\mathbf{X}$ *is not, then* $\mathbb{V}[\mathbf{X}'\mathbf{Y}] = \mathbf{X}'\mathbb{V}[\mathbf{Y}]\mathbf{X}$, *assuming the multiplication makes sense (the matrices are commensurate).*

M.7: End-of-Appendix Materials

M.7.1 Exercises

1. Prove $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$.
2. Prove matrix addition is commutative.
3. Prove matrix addition is associative.
4. Prove that scalar multiplication is associative.
5. Prove that scalar multiplication is distributive over addition.
6. Using a counterexample, prove that matrix multiplication is not commutative when only one of the matrices is diagonal (thus showing that *both* must be diagonal).
7. Let $\mathbf{A}$ be any square matrix. Show that $\frac{1}{2}(\mathbf{A} + \mathbf{A}')$ is symmetric.
8. Determine the determinant of $\mathbf{J}_{10}$.
9. Determine the rank of $\mathbf{J}_{10}$.
10. Prove that $\mathbf{j}_{10}\,\mathbf{j}_{10}'$ is not positive definite.
11. Prove that $\mathbf{j}_{10}'\,\mathbf{j}_{10}$ *is* positive definite.
12. Prove Lemma M.4.
13. Prove Lemma M.5.
14. Prove Lemma M.6.
15. Prove Lemma M.14.
16. Prove Lemma M.15.
17. Prove Lemma M.16.