3

Linear Algebra: Vectors, Matrices, and Operations

3.1 Objectives

This chapter covers the *mechanics* of vector and matrix manipulation and the next chapter approaches the topic from a more theoretical and conceptual perspective. The objective for readers of this chapter is not only to learn the mechanics of performing algebraic operations on these mathematical forms but also to start seeing them as organized collections of numerical values where the manner in which they are put together provides additional mathematical information. The best way to do this is to perform the operations oneself. Linear algebra is fun. Really! In general, the mechanical operations are nothing more than simple algebraic steps that anybody can perform: addition, subtraction, multiplication, and division. The only real abstraction required is "seeing" the rectangular nature of the objects in the sense of visualizing operations at a high level rather than getting buried in the algorithmic details.

When one reads high visibility journals in the social sciences, matrix algebra (a near synonym) is ubiquitous. Why is that? Simply because it lets us express extensive models in quite readable notation. Consider the following linear statistical model specification [from real work, Powers and Cox (1997)]. They are relating political blame to various demographic and regional political variables:

$$\begin{aligned}
\text{for } i = 1 \text{ to } n,\ (BLAMEFIRST)Y_i = \\
&\beta_0 + \beta_1 CHANGELIV + \beta_2 BLAMECOMM + \beta_3 INCOME \\
&+ \beta_4 FARMER + \beta_5 OWNER + \beta_6 BLUESTATE \\
&+ \beta_7 WHITESTATE + \beta_8 FORMMCOMM + \beta_9 AGE \\
&+ \beta_{10} SQAGE + \beta_{11} SEX + \beta_{12} SIZEPLACE \\
&+ \beta_{13} EDUC + \beta_{14} FINHS + \beta_{15} ED * HS \\
&+ \beta_{16} RELIG + \beta_{17} NATION + E_i
\end{aligned}$$

This expression is way too complicated to be useful! It would be easy for a reader interested in the political argument to get lost in the notation. In matrix algebra form this is simply $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. In fact, even for very large datasets and very large model specifications (many data variables of interest), this form is exactly the same; we simply indicate the size of these objects. This is not just a convenience (although it *really is* convenient). Because we can notate large groups of numbers in an easy-to-read structural form, we can concentrate more on the theoretically interesting properties of the analysis.

While this chapter provides many of the foundations for working with matrices in social sciences, there is one rather technical omission that some readers may want to worry about later. All linear algebra is based on properties that define a **field**. Essentially this means that logical inconsistencies that could have otherwise resulted from routine calculations have been precluded. Interested readers are referred to Billingsley (1995), Chung (2000), or Grimmett and Stirzaker (1992).

3.2 Working with Vectors

Vector. A vector is just a serial listing of numbers where the order matters. So we can store the first four positive integers in a single vector, which can be

$$\text{a row vector: } \mathbf{v} = [1, 2, 3, 4], \qquad \text{or a column vector: } \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix},$$

where $\mathbf{v}$ is the name for this new object. Order matters in the sense that the two vectors above are different, for instance, from

$$\mathbf{v}^* = [4, 3, 2, 1], \qquad \mathbf{v}^* = \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix}.$$

It is a convention that vectors are designated in bold type and individual values, *scalars*, are designated in regular type. Thus $\mathbf{v}$ is a vector with elements v_1, v_2, v_3, v_4, and v would be some *other* scalar quantity. This gets a little confusing where vectors are themselves indexed: $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ would indicate four *vectors*, not four scalars. Usually, however, authors are quite clear about which form they mean.

Substantively it does not matter whether we consider a vector to be of column or row form, but it does matter when performing some operations. Also, some disciplines (notably economics) tend to default to the column form. In the row form, it is equally common to see spacing used instead of commas as delimiters: $[1\ \ 2\ \ 3\ \ 4]$. Also, the contents of these vectors can be integers, rational or irrational numbers, and even complex numbers; there are no restrictions.

So what kinds of operations can we do with vectors? The basic operands are very straightforward: addition and subtraction of vectors as well as multiplication and division by a scalar. The following examples use the vectors $\mathbf{u} = [3, 3, 3, 3]$ and $\mathbf{v} = [1, 2, 3, 4]$

★ Example 3.1: Vector Addition Calculation.

$$\mathbf{u} + \mathbf{v} = [u_1 + v_1, u_2 + v_2, u_3 + v_3, u_4 + v_4] = [4, 5, 6, 7].$$

★ Example 3.2: Vector Subtraction Calculation.

$$\mathbf{u} - \mathbf{v} = [u_1 - v_1, u_2 - v_2, u_3 - v_3, u_4 - v_4] = [2, 1, 0, -1].$$

★ Example 3.3: Scalar Multiplication Calculation.

$$3 \times \mathbf{v} = [3 \times v_1, 3 \times v_2, 3 \times v_3, 3 \times v_4] = [3, 6, 9, 12].$$

★ Example 3.4: Scalar Division Calculation.

$$\mathbf{v} \div 3 = [v_1/3, v_2/3, v_3/3, v_4/3] = \left[\frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}\right].$$

So operations with scalars are performed on every vector element in the same way. Conversely, the key issue with addition or subtraction between two vectors is that the operation is applied only to the corresponding vector elements as pairs: the first vector elements together, the second vector elements together, and so on. There is one concern, however. With this scheme, *the vectors have to be exactly the same size* (same number of elements). This is called **conformable** in the sense that the first vector must be of a size that conforms with the second vector; otherwise they are (predictably) called **nonconformable**. In the examples above both $\mathbf{u}$ and $\mathbf{v}$ are 1×4 (row) vectors (alternatively called length $k = 4$ vectors), meaning that they have one row and four columns. Sometimes size is denoted beneath the vectors:

$$\underset{1\times 4}{\mathbf{u}} + \underset{1\times 4}{\mathbf{v}}.$$

It should then be obvious that there is no logical way of adding a 1×4 vector to a 1×5 vector. Note also that this is not a practical consideration with scalar multiplication or division as seen above, because we apply the scalar identically to each element of the vector when multiplying: $s(u_1, u_2, \ldots, u_k) = (su_1, su_2, \ldots, su_k)$.

There are a couple of "special" vectors that are frequently used. These are $\mathbf{1}$ and $\mathbf{0}$, which contain all 1's or 0's, respectively. As well shall soon see, there are a larger number of "special" matrices that have similarly important characteristics.

It is easy to summarize the formal properties of the basic vector operations. Consider the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, which are identically sized, and the scalars s and t. The following intuitive properties hold.

Elementary Formal Properties of Vector Algebra

$\rightarrowtail$	Commutative Property	$\mathbf{u}+\mathbf{v}=(\mathbf{v}+\mathbf{u})$
$\rightarrowtail$	Additive Associative Property	$(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
$\rightarrowtail$	Vector Distributive Property	$s(\mathbf{u}+\mathbf{v})=s\mathbf{u}+s\mathbf{v}$
$\rightarrowtail$	Scalar Distributive Property	$(s+t)\mathbf{u}=s\mathbf{u}+t\mathbf{u}$
$\rightarrowtail$	Zero Property	$\mathbf{u}+\mathbf{0}=\mathbf{u}\Longleftrightarrow\mathbf{u}-\mathbf{u}=\mathbf{0}$
$\rightarrowtail$	Zero Multiplicative Property	$\mathbf{0}\mathbf{u}=\mathbf{0}$
$\rightarrowtail$	Unit Rule	$\mathbf{1}\mathbf{u}=\mathbf{u}$

★ **Example 3.5: Illustrating Basic Vector Calculations**. Here is a numerical case that shows several of the properties listed above. Define $s=3$, $t=1$, $\mathbf{u}=[2,4,8]$, and $\mathbf{v}=[9,7,5]$. Then:

$(s+t)(\mathbf{v}+\mathbf{u})$	$s\mathbf{v}+t\mathbf{v}+s\mathbf{u}+t\mathbf{u}$
$(3+1)([9,7,5]+[2,4,8])$	$3[9,7,5]+1[9,7,5]+3[2,4,8]+1[2,4,8]$
$4[11,11,13]$	$[27,21,15]+[9,7,5]+[6,12,24]+[2,4,8]$
$[44,44,52]$	$[44,44,52]$

Multiplication of vectors is not quite so straightforward, and there are actually different forms of multiplication to make matters even more confusing. We will start with the two most important and save some of the other forms for the last section of this chapter.

Vector Inner Product. The *inner product*, also called the **dot product**, of two vectors, results in a scalar (and so it is also called the **scalar product**). The inner product of two conformable vectors of arbitrary length k is the sum of the item-by-item products:

$$\mathbf{u}\cdot\mathbf{v} = [u_1v_1 + u_2v_2 + \cdots u_kv_k] = \sum_{i=1}^{k} u_iv_i.$$

It might be somewhat surprising to see the return of the summation notation here (Σ, as described on page 11), but it makes a lot of sense since running through the two vectors is just a mechanical additive process. For this reason, it is relatively common, though possibly confusing, to see vector (and later matrix) operations expressed in summation notation.

★ **Example 3.6: Simple Inner Product Calculation**. A numerical example of an inner product multiplication is given by

$$\mathbf{u}\cdot\mathbf{v} = [3,3,3]\cdot[1,2,3] = [3\cdot 1 + 3\cdot 2 + 3\cdot 3] = 18.$$

When the inner product of two vectors is zero, we say that the vectors are **orthogonal**, which means they are at a right angle to each other (we will be more visual about this in Chapter 4). The notation for the orthogonality of two vectors is $\mathbf{u}\perp\mathbf{v}$ iff $\mathbf{u}\cdot\mathbf{v}=0$. As an example of orthogonality, consider $\mathbf{u}=[1,2,-3]$, and $\mathbf{v} = [1, 1, 1]$. As with the more basic addition and subtraction or scalar operations, there are formal properties for inner products:

Inner Product Formal Properties of Vector Algebra

$\rightarrowtail$	Commutative Property	$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
$\rightarrowtail$	Associative Property	$s(\mathbf{u} \cdot \mathbf{v}) = (s\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (s\mathbf{v})$
$\rightarrowtail$	Distributive Property	$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
$\rightarrowtail$	Zero Property	$\mathbf{u} \cdot \mathbf{0} = 0$
$\rightarrowtail$	Unit Rule	$\mathbf{1u} = \mathbf{u}$
$\rightarrowtail$	Unit Rule	$\mathbf{1u} = \sum_{i=1}^{k} \mathbf{u}_i$, for $\mathbf{u}$ of length k

★ **Example 3.7: Vector Inner Product Calculations**. This example demonstrates the first three properties above. Define $s = 5$, $\mathbf{u} = [2, 3, 1]$, $\mathbf{v} = [4, 4, 4]$, and $\mathbf{w} = [-1, 3, -4]$. Then:

$s(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w}$	$s\mathbf{v} \cdot \mathbf{w} + s\mathbf{u} \cdot \mathbf{w}$
$5([2, 3, 1] + [4, 4, 4]) \cdot [-1, 3, -4]$	$5[4, 4, 4] \cdot [-1, 3, -4]$
	$+5[2, 3, 1] \cdot [-1, 3, -4]$
$5([6, 7, 5]) \cdot [-1, 3, -4]$	$[20, 20, 20] \cdot [-1, 3, -4]$
	$+[10, 15, 5] \cdot [-1, 3, -4]$
$[30, 35, 25] \cdot [-1, 3, -4]$	$-40 + 15$
-25	-25

Vector Cross Product. The *cross product* of two vectors (sometimes called the **outer product**, although this term is better reserved for a slightly different operation; see the distinction below) is slightly more involved than the inner product, in both calculation and interpretation. This is mostly because the result is a vector instead of a scalar. Mechanically, the cross product of two conformable vectors of length $k = 3$ is

$$\mathbf{u} \times \mathbf{v} = \left[u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1\right],$$

meaning that the first element is a difference equation that leaves out the first elements of the original two vectors, and the second and third elements proceed accordingly. In the more general sense, we perform a series of "leave one out" operations that is more extensive than above because the suboperations are themselves cross products.

Fig. 3.1. Vector Cross Product Illustration

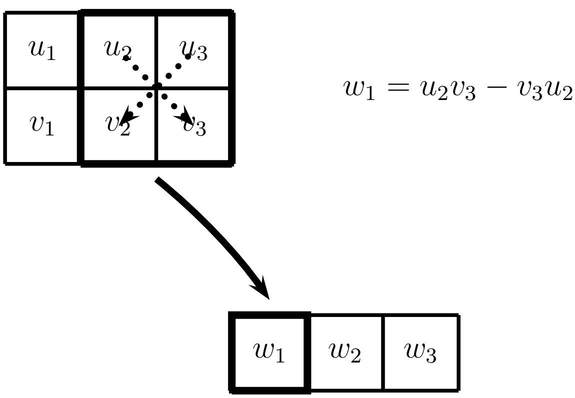
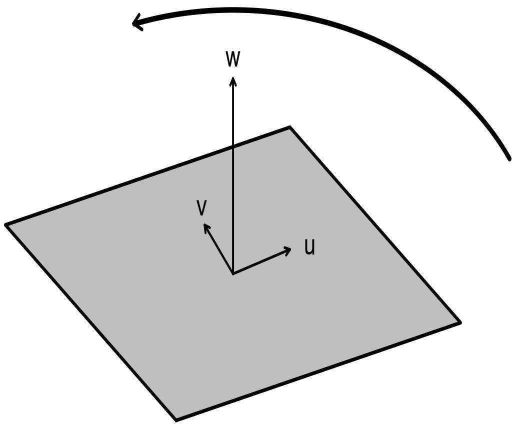


Figure 3.1 gives the intuition behind this product. First the $\mathbf{u}$ and $\mathbf{v}$ vectors are stacked on top of each other in the upper part of the illustration. The process of calculating the first vector value of the cross product, which we will call w_1, is done by "crossing" the elements in the solid box: u_2v_3 indicated by the first arrow and u_3v_2 indicated by the second arrow. Thus we see the result for

Fig. 3.2. The Right-Hand Rule Illustrated



w_1 as a difference between these two individual components. This is actually the ***determinant*** of the 2×2 submatrix, which is an important principle considered in some detail in Chapter 4.

Interestingly, the resulting vector from a cross product is orthogonal to both of the original vectors in the direction of the so-called "right-hand rule." This handy rule says that if you hold your hand as you would when hitchhiking, the curled fingers make up the original vectors and the thumb indicates the direction of the orthogonal vector that results from a cross product. In Figure 3.2 you can imagine your right hand resting on the plane with the fingers curling to the left ($\circlearrowleft$) and the thumb facing upward.

For vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, the cross product properties are given by

Cross Product Formal Properties of Vector Algebra		
$\rightarrowtail$	Commutative Property	$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$
$\rightarrowtail$	Associative Property	$s(\mathbf{u} \times \mathbf{v}) = (s\mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (s\mathbf{v})$
$\rightarrowtail$	Distributive Property	$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w})$
$\rightarrowtail$	Zero Property	$\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$
$\rightarrowtail$	Self-Orthogonality	$\mathbf{u} \times \mathbf{u} = \mathbf{0}$

★ **Example 3.8: Cross Product Calculation.** Returning to the simple numerical example from before, we now calculate the cross product instead of the inner product:

$$\begin{aligned}\mathbf{u} \times \mathbf{v} &= [3, 3, 3] \times [1, 2, 3] \\ &= [(3)(3) - (3)(2), (3)(1) - (3)(3), (3)(2) - (3)(1)] = [3, -6, 3].\end{aligned}$$

We can then check the orthogonality as well:

$$[3, 3, 3] \cdot [3, -6, 3] = 0 \qquad [1, 2, 3] \cdot [3, -6, 3] = 0.$$

Sometimes the distinction between row vectors and column vectors is important. While it is often glossed over, vector multiplication should be done in a conformable manner with regard to multiplication (as opposed to addition discussed above) where a row vector multiplies a column vector such that their adjacent "sizes" match: a $(1 \times k)$ vector multiplying a $(k \times 1)$ vector for k

elements in each. This operation is now an inner product:

$$\underset{1\times k}{[v_1, v_2, \ldots, v_k]} \times \underset{k\times 1}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}}.$$

This adjacency above comes from the k that denotes the columns of $\mathbf{v}$ and the k that denotes the rows of $\mathbf{u}$ and manner by which they are next to each other. Thus an inner product multiplication operation is implied here, even if it is not directly stated. An outer product would be implied by this type of adjacency:

$$\underset{k\times 1}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}} \times \underset{1\times k}{[v_1, v_2, \ldots, v_k]},$$

where the 1's are next to each other. So the cross product of two vectors is a vector, and the outer product of two conformable vectors is a matrix: a rectangular grouping of numbers that generalizes the vectors we have been working with up until now. This distinction helps us to keep track of the objective. Mechanically, this is usually easy. To be completely explicit about these operations we can also use the **vector transpose**, which simply converts a row vector to a column vector, or vice versa, using the apostrophe notation:

$$\underset{k\times 1}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}}' = \underset{1\times k}{[u_1, u_2, \ldots, u_k]}, \qquad \underset{1\times k}{[u_1, u_2, \ldots, u_k]}' = \underset{k\times 1}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}}.$$

This is essentially book-keeping with vectors and we will not worry about it extensively in this text, but as we will see shortly it is important with matrix operations. Also, note that the order of multiplication now matters.

★ **Example 3.9: Outer Product Calculation**. Once again using the simple numerical forms, we now calculate the outer product instead of the cross product:

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} [3, 3, 3] = \begin{bmatrix} 3 & 3 & 3 \\ 6 & 6 & 6 \\ 9 & 9 & 9 \end{bmatrix}.$$

And to show that order matters, consider:

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} [1, 2, 3] = \begin{bmatrix} 3 & 6 & 9 \\ 3 & 6 & 9 \\ 3 & 6 & 9 \end{bmatrix}.$$

3.2.1 Vector Norms

Measuring the "length" of vectors is a surprisingly nuanced topic. This is because there are different ways to consider Cartesian length in the dimension implied by the size (number of elements) of the vector. It is obvious, for instance, that $(5, 5, 5)$ should be considered longer than $(1, 1, 1)$, but it is not clear whether $(4, 4, 4)$ is longer than $(3, -6, 3)$. The standard version of the **vector norm** for an n-length vector is given by

$$\|\mathbf{v}\| = (v_1^2 + v_2^2 + \cdots + v_n^2)^{\frac{1}{2}} = (\mathbf{v}' \cdot \mathbf{v})^{\frac{1}{2}}.$$

In this way, the vector norm can be thought of as the distance of the vector from the origin. Using the formula for $\|\mathbf{v}\|$ we can now calculate the vector norm for $(4, 4, 4)$ and $(3, -6, 3)$:

$$\|(4, 4, 4)\| = \sqrt{4^2 + 4^2 + 4^2} = 6.928203$$
$$\|(3, -6, 3)\| = \sqrt{3^2 + (-6)^2 + 3^2} = 7.348469.$$

So the second vector is actually longer by this measure. Consider the following properties of the vector norm (notice the reoccurrence of the dot product in the Multiplication Form):

Properties of the Standard Vector Norm

$\rightarrowtail$	Vector Norm	$\|\|\mathbf{u}\|\|^2 = \mathbf{u} \cdot \mathbf{u}$
$\rightarrowtail$	Difference Norm	$\|\|\mathbf{u} - \mathbf{v}\|\|^2 = \|\|\mathbf{u}\|\|^2 - 2(\mathbf{u} \cdot \mathbf{v}) + \|\|\mathbf{v}\|\|^2$
$\rightarrowtail$	Multiplication Norm	$\|\|\mathbf{u} \times \mathbf{v}\|\| = \|\|\mathbf{u}\|\|^2\|\|\mathbf{v}\|\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$

★ **Example 3.10: Difference Norm Calculation**. As an illustration of the second property above we now include a numerical demonstration. Suppose $\mathbf{u} = [-10, 5]$ and $\mathbf{v} = [3, 3]$. Then:

$\|\|\mathbf{u} - \mathbf{v}\|\|^2$	$\|\|\mathbf{u}\|\|^2 - 2(\mathbf{u} \cdot \mathbf{v}) + \|\|\mathbf{v}\|\|^2$
$\|\|[-10, 5] - [3, 3]\|\|^2$	$\|\|[-10, 5]\|\|^2 - 2([-10, 5] \cdot [3, 3]) + \|\|[3, 3]\|\|^2$
$\|\|[-13, 2]\|\|^2$	$(100) + (25) - 2(-30 + 15) + (9) + (9)$
$169 + 4$	$125 + 30 + 18$
173	173

★ **Example 3.11: Multiplication Norm Calculation**. The third property is also easy to demonstrate numerically. Suppose $\mathbf{u} = [-10, 5, 1]$ and $\mathbf{v} = [3, 3, 3]$. Then:

$\|\|\mathbf{u} \times \mathbf{v}\|\|$	$\|\|\mathbf{u}\|\|^2\|\|\mathbf{v}\|\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$
$\|\|[-10, 5, 1] \times [3, 3, 3]\|\|$	$\|\|[-10, 5, 1]\|\|^2\|\|[3, 3, 3]\|\|^2$
	$-([-10, 5, 1] \cdot [3, 3, 3])^2$
$\|\|[(15) - (3), (3) - (-30), (-30) - (15)]\|\|$	$((100 + 25 + 1)(9 + 9 + 9)$
	$-(-30 + 15 + 3)^2$
$(144) + (1089) + (2025)$	$(3402 - 144)$
3258	3258

Interestingly, norming can also be applied to find the n-dimensional distance between the endpoints of two vectors starting at the origin with a variant of the Pythagorean Theorem known as the **law of cosines**:

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\|\mathbf{w}\| \cos\theta,$$

where θ is the angle from the $\mathbf{w}$ vector to the $\mathbf{v}$ vector measured in radians. This is also called the cosine rule and leads to the property that $\cos(\theta) = \frac{\mathbf{vw}}{\|\mathbf{v}\| \|\mathbf{w}\|}$.

★ **Example 3.12: Votes in the House of Commons**. Casstevens (1970) looked at legislative cohesion in the British House of Commons. Prime Minister David Lloyd George claimed on April 9, 1918 that the French Army was stronger on January 1, 1918 than on January 1, 1917 (a statement that generated considerable controversy). Subsequently the leader of the Liberal Party moved that a select committee be appointed to investigate claims by the military that George was incorrect. The resulting motion was defeated by the following vote: Liberal Party 98 yes, 71 no; Labour Party 9 yes, 15 no; Conservative Party 1 yes, 206 no; others 0 yes, 3 no. The difficult in analyzing this vote is the fact that 267 Members of Parliament (MPs) did not vote. So do we include them in the denominator when making claims about voting patterns? Casstevens says no because large numbers of abstentions mean that such indicators are misleading. He alternatively looked at party cohesion for the two large parties as vector norms:

$$\|L\| = \|(98, 71)\| = 121.0165$$

$$\|C\| = \|(1, 206)\| = 206.0024.$$

From this we get the obvious conclusion that the Conservatives are more cohesive because their vector has greater magnitude. More interestingly, we can contrast the two parties by calculating the angle between these two vectors (in radians) using the cosine rule:

$$\theta = \arccos \left[\frac{(98, 71) \cdot ((1, 206)}{121.070 \times 206.002} \right] = 0.9389,$$

which is about 54 degrees. Recall that $\arccos$ is the inverse function to $\cos$. It is hard to say exactly how dramatic this angle is, but if we were analyzing a series of votes in a legislative body, this type of summary statistic would facilitate comparisons.

Actually, the norm used above is the most commonly used form of a **p-norm**:

$$\|\mathbf{v}\|_p = (|v_1|^p + |v_2|^p + \cdots + |v_n|^p)^{\frac{1}{p}},\ p \geq 0,$$

where $p = 2$ so far. Other important cases include $p = 1$ and $p = \infty$:

$$\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq n} |x_i|,$$

that is, just the maximum vector value. Whenever a vector has a p-norm of 1, it is called a **unit vector**. In general, if p is left off the norm, then one can safely assume that it is the $p = 2$ form discussed above. Vector p-norms have the following properties:

Properties of Vector Norms, Length-n

$\rightarrowtail$	Triangle Inequality	$\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$
$\rightarrowtail$	Hölder's Inequality	for $\frac{1}{p}+\frac{1}{q}=1$, $\lvert\mathbf{v}\cdot\mathbf{w}\rvert \leq \|\mathbf{v}\|_p\|\mathbf{w}\|_q$
$\rightarrowtail$	Cauchy-Schwarz Ineq.	$\lvert\mathbf{v}\cdot\mathbf{w}\rvert \leq \|\mathbf{v}\|_2\|\mathbf{w}\|_2$
$\rightarrowtail$	Cosine Rule	$\cos(\theta) = \frac{\mathbf{v}\mathbf{w}}{\|\mathbf{v}\|\,\|\mathbf{w}\|}$
$\rightarrowtail$	Vector Distance	$d(\mathbf{v},\mathbf{w}) = \|\mathbf{v}-\mathbf{w}\|$
$\rightarrowtail$	Scalar Property	$\|s\mathbf{v}\| = \lvert s\rvert\|\mathbf{v}\|$

★ **Example 3.13: Hölder's Inequality Calculation**. As a revealing mechanical demonstration that Hölders' Inequality holds, set $p = 3$ and $q = 3/2$ for the vectors $\mathbf{v} = [-1, 3]$ and $\mathbf{w} = [2, 2]$, respectively. Hölder's Inequality uses $|\mathbf{v}\cdot\mathbf{w}|$ to denote the absolute value of the dot product. Then:

$$\|\mathbf{v}\|_3 = (|-1|^3 + |3|^3)^{\frac{1}{3}} = 3.036589$$

$$\|\mathbf{w}\|_3 = (|2|^{\frac{3}{2}} + |2|^{\frac{3}{2}})^{\frac{2}{3}} = 3.174802$$

$$|\mathbf{v}\cdot\mathbf{w}| = |(-1)(2) + (3)(2)| = 4 < (3.036589)(3.174802) = 9.640569.$$

★ **Example 3.14: The Political Economy of Taxation**. While taxation is known to be an effective policy tool for democratic governments, it is also a very difficult political solution for many politicians because it can be unpopular and controversial. Swank and Steinmo (2002) looked at factors that lead to changes in tax policies in "advanced capitalist" democracies with the idea that factors like internationalization of economies, political pressure from budgets, and within-country economic factors are influential. They found that governments have a number of constraints on their ability to enact significant changes in tax rates, even when there is pressure to increase economic efficiency.

As part of this study the authors provided a total taxation from labor and consumption as a percentage of GDP in the form of two vectors: one for 1981 and another for 1995. These are reproduced as

Nation	**1981**	**1995**
Australia	30	31
Austria	44	42
Belgium	45	46
Canada	35	37
Denmark	45	51
Finland	38	46
France	42	44
Germany	38	39
Ireland	33	34
Italy	31	41
Japan	26	29
Netherlands	44	44
New Zealand	34	38
Norway	49	46
Sweden	50	50
Switzerland	31	34
United Kingdom	36	36
United States	29	28

A natural question to ask is, how much have taxation rates changed over the 14-year period for these countries collectively? The difference in mean averages, 38 versus 40, is not terribly revealing because it "washes out" important differences since some countries increased and other decreased. That is, what does a 5% difference in average change in total taxation over GDP say about how these countries changed as a group when some countries changed very little and some made considerable changes? Furthermore, when changes go in opposite directions it lowers the overall sense of an effect. In other words, summaries like averages are not good measures when we want some sense of net change.

One way of assessing total country change is employing the difference norm to compare aggregate vector difference.

$$||t_{1995} - t_{1981}||^2 = t'_{1995} \cdot t_{1995} - 2(t'_{1995} \cdot t_{1981}) + t'_{1981} \cdot t_{1981}$$

$$= \begin{bmatrix} 30 \\ 44 \\ 45 \\ 35 \\ 45 \\ 38 \\ 42 \\ 38 \\ 33 \\ 31 \\ 26 \\ 44 \\ 34 \\ 49 \\ 50 \\ 31 \\ 36 \\ 29 \end{bmatrix}' \cdot \begin{bmatrix} 30 \\ 44 \\ 45 \\ 35 \\ 45 \\ 38 \\ 42 \\ 38 \\ 33 \\ 31 \\ 26 \\ 44 \\ 34 \\ 49 \\ 50 \\ 31 \\ 36 \\ 29 \end{bmatrix} - 2 \begin{bmatrix} 30 \\ 44 \\ 45 \\ 35 \\ 45 \\ 38 \\ 42 \\ 38 \\ 33 \\ 31 \\ 26 \\ 44 \\ 34 \\ 49 \\ 50 \\ 31 \\ 36 \\ 29 \end{bmatrix}' \cdot \begin{bmatrix} 31 \\ 42 \\ 46 \\ 37 \\ 51 \\ 46 \\ 44 \\ 39 \\ 34 \\ 41 \\ 29 \\ 44 \\ 38 \\ 46 \\ 50 \\ 34 \\ 36 \\ 28 \end{bmatrix} + \begin{bmatrix} 31 \\ 42 \\ 46 \\ 37 \\ 51 \\ 46 \\ 44 \\ 39 \\ 34 \\ 41 \\ 29 \\ 44 \\ 38 \\ 46 \\ 50 \\ 34 \\ 36 \\ 28 \end{bmatrix}' \cdot \begin{bmatrix} 31 \\ 42 \\ 46 \\ 37 \\ 51 \\ 46 \\ 44 \\ 39 \\ 34 \\ 41 \\ 29 \\ 44 \\ 38 \\ 46 \\ 50 \\ 34 \\ 36 \\ 28 \end{bmatrix}$$

$$= 260$$

So what does this mean? For comparison, we can calculate the same vector norm except that instead of using t_{1995}, we will substitute a vector that increases the 1981 uniformly levels by 10% (a hypothetical increase of 10% for every country in the study):

$$\hat{t}_{1981} = 1.1 t_{1981} = [33.0, 48.4, 49.5, 38.5, 49.5, 41.8, 46.2, 41.8, 36.3$$
$$34.1, 28.6, 48.4, 37.4, 53.9, 55.0, 34.1, 39.6, 31.9].$$

This allows us to calculate the following benchmark difference:

$$||\hat{t}_{1981} - t_{1981}||^2 = 265.8.$$

So now it is clear that the observed vector difference for total country change from 1981 to 1995 is actually similar to a 10% across-the-board change rather than a 5% change implied by the vector means. In this sense we get a true multidimensional sense of change.

3.3 So What Is the Matrix?

Matrices are all around us: A **matrix** is nothing more than a rectangular arrangement of numbers. It is a way to individually assign numbers, now called **matrix elements** or **entries**, to specified positions in a single structure, referred to with a single name. Just as we saw that the order in which individual entries appear in the vector matters, the ordering of values within ***both*** rows and columns now matters. It turns out that this requirement adds a considerable amount of structure to the matrix, some of which is not immediately apparent (as we will see).

Matrices have two definable **dimensions**, the number of rows and the number of columns, whereas vectors only have one, and we denote matrix size by $row \times column$. Thus a matrix with i rows and j columns is said to be of dimension $i \times j$ (by convention rows comes before columns). For instance, a simple (and rather uncreative) 2×2 matrix named $\mathbf{X}$ (like vectors, matrix names are bolded) is given by:

$$\underset{2\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Note that matrices can also be explicitly notated with size.

Two things are important here. First, these four numbers are now treated together as a single unit: They are ***grouped*** together in the two-row by two-column matrix object. Second, the positioning of the numbers is specified.

That is, the matrix $\mathbf{X}$ is different than the following matrices:

$$\mathbf{W} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix},$$

as well as many others. Like vectors, the elements of a matrix can be integers, real numbers, or complex numbers. It is, however, rare to find applications that call for the use of matrices of complex numbers in the social sciences.

The matrix is a system. We can refer directly to the specific elements of a matrix by using ***subscripting*** of addresses. So, for instance, the elements of $\mathbf{X}$ are given by $x_{11} = 1$, $x_{12} = 2$, $x_{21} = 3$, and $x_{22} = 4$. Obviously this is much more powerful for larger matrix objects and we can even talk about arbitrary sizes. The element addresses of a $p \times n$ matrix can be described for large values of p and n by

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1(p-1)} & x_{1p} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2(p-1)} & x_{2p} \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ x_{(n-1)1} & x_{(n-1)2} & \cdots & \cdots & x_{(n-1)(p-1)} & x_{(n-1)p} \\ x_{n1} & x_{n2} & \cdots & \cdots & x_{n(p-1)} & x_{np} \end{bmatrix}.$$

Using this notation we can now define **matrix equality**. Matrix $\mathbf{A}$ is equal to matrix $\boldsymbol{B}$ if and only if every element of $\mathbf{A}$ is equal to the corresponding element of $\boldsymbol{B}$: $\boldsymbol{A} = \boldsymbol{B} \iff a_{ij} = b_{ij}\ \forall i, j$. Note that "subsumed" in this definition is the requirement that the two matrices be of the same dimension (same number of rows, i, and columns, j).

3.3.1 Some Special Matrices

There are some matrices that are quite routinely used in quantitative social science work. The most basic of these is the **square matrix**, which is, as the name implies, a matrix with the same number of rows and columns. Because one number identifies the complete size of the square matrix, we can say that a $k \times k$ matrix (for arbitrary size k) is a matrix of **order-k**. Square matrices can contain any values and remain square: The square property is independent of the contents. A very general square matrix form is the **symmetric matrix**. This is a matrix that is symmetric across the diagonal from the upper left-hand corner to the lower right-hand corner. More formally, $\mathbf{X}$ is a symmetric matrix iff $a_{ij} = a_{ji}\ \forall i, j$. Here is an unimaginative example of a symmetric matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 8 & 5 & 6 \\ 3 & 5 & 1 & 7 \\ 4 & 6 & 7 & 8 \end{bmatrix}.$$

A matrix can also be **skew-symmetric** if it has the property that the rows and column switching operation would provide the same matrix except for the sign. For example,

$$\mathbf{X} = \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & -3 \\ -2 & 3 & 0 \end{bmatrix}.$$

By the way, the symmetric property does not hold for the other diagonal, the one from the upper right-hand side to the lower left-hand side.

Just as the symmetric matrix is a special case of the square matrix, the **diagonal matrix** is a special case of the symmetric matrix (and therefore of the square matrix, too). A diagonal matrix is a symmetric matrix with all zeros on the off-diagonals (the values where $i \neq j$). If the (4×4) $\mathbf{X}$ matrix above were a diagonal matrix, it would look like

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

We can also define the diagonal matrix more generally with just a vector. A diagonal matrix with elements $[d_1, d_2, \ldots, d_{n-1}, d_n]$ is the matrix

$$\mathbf{X} = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & d_{n-1} & 0 \\ 0 & 0 & 0 & 0 & d_n \end{bmatrix}.$$

A diagonal matrix can have any values on the diagonal, but all of the other values must be zero. A very important special case of the diagonal matrix is the **identity matrix**, which has only the value 1 for each diagonal element: $d_i = 1,\ \forall i$. A 4×4 version is

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

This matrix form is always given the name **I**, and it is sometimes denoted to give size: $I_{4\times 4}$ or even just $\mathbf{I}(4)$. A seemingly similar, but actually very different, matrix is the **J** matrix, which consists of all 1's:

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

given here in a 4×4 version. As we shall soon see, the identity matrix is very commonly used because it is the matrix equivalent of the scalar number 1, whereas the **J** matrix is not (somewhat surprisingly). Analogously, the **zero**

matrix is a matrix of all zeros, the 4×4 case being

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

also given as a 4×4 matrix. Consider for a moment that the zero matrix and the **J** matrix here are also square, symmetric, diagonal, and particular named cases. Yet neither of these two ***must*** have these properties as both can be nonsquare as well: $i \neq j$.

This is a good time to also introduce a special nonsymmetric square matrix called the **triangular matrix**. This is a matrix with all zeros above the diagonal, **lower triangular**, or all zeros below the diagonal, **upper triangular**. Two versions based on the first square matrix given above are

$$\mathbf{X}_{LT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 8 & 0 & 0 \\ 3 & 5 & 1 & 0 \\ 4 & 6 & 7 & 8 \end{bmatrix}, \quad \mathbf{X}_{UT} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 8 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 8 \end{bmatrix},$$

where LT designates "lower triangular" and UT designates "upper triangular." This general form plays a special role in **matrix decomposition**: factoring matrices into multiplied components. This is also a common form in more pedestrian circumstances. Map books often tabulate distances between sets of cities in an upper triangular or lower triangular form because the distance from Miami to New York is also the distance from New York to Miami.

★ **Example 3.15: Marriage Satisfaction**. Sociologists who study marriage often focus on indicators of self-expressed satisfaction. Unfortunately marital satisfaction is sufficiently complex and sufficiently multidimensional that single measurements are often insufficient to get a full picture of underlying attitudes. Consequently, scholars such as Norton (1983) ask multiple questions designed to elicit varied expressions of marital satisfaction and therefore care a lot about the correlation between these. A correlation (described in detail in Chapter 8) shows how "tightly" two measures change with each other over a range from -1 to 1, with 0 being no evidence of moving together. His correlation matrix provides the correlational structure between answers to the following questions according to scales where higher numbers mean that the respondent agrees more (i.e., 1 is strong disagreement with the statement and 7 is strong agreement with the statement). The questions are

Question	**Measurement Scale**	**Valid Cases**
We have a good marriage	7-point	428
My relationship with my partner is very stable	7-point	429
Our marriage is strong	7-point	429
My relationship with my partner makes me happy	7-point	429
I really feel like *part of a team* with my partner	7-point	426
The degree of happiness, everything considered	10-point	407

Since the correlation between two variables is symmetric, it does not make sense to give a correlation matrix between these variables across a full matrix because the lower triangle will simply mirror the upper triangle and make the display more congested. Consequently, Norton only needs to show a triangular version of the matrix:

$$
\begin{array}{c}
\\
(1) \\
(2) \\
(3) \\
(4) \\
(5) \\
(6)
\end{array}
\begin{array}{c}
\begin{array}{cccccc}
(1) & (2) & (3) & (4) & (5) & (6)
\end{array} \\
\begin{pmatrix}
1.00 & 0.85 & 0.83 & 0.83 & 0.74 & 0.76 \\
 & 1.00 & 0.82 & 0.86 & 0.72 & 0.77 \\
 & & 1.00 & 0.78 & 0.68 & 0.70 \\
 & & & 1.00 & 0.71 & 0.76 \\
 & & & & 1.00 & 0.69 \\
 & & & & & 1.00
\end{pmatrix}
\end{array}.
$$

Interestingly, these analyzed questions all correlate highly (a 1 means a perfectly positive relationship). The question that seems to covary greatly with the others is the first (it is phrased somewhat as a summary, after all). Notice that strength of marriage and part of a team covary less than any others (a suggestive finding). This presentation is a bit different from an upper triangular matrix in the sense discussed above because we have just deliberately omitted redundant information, rather than the rest of matrix actually having zero values.

3.4 Controlling the Matrix

As with vectors we can perform arithmetic and algebraic operations on matrices. In particular addition, subtraction, and scalar operations are quite simple. Matrix addition and subtraction are performed only for two conformable matrices by performing the operation on an element-by-element basis for corresponding elements, so the number of rows and columns must match. Multiplication or division by a scalar proceeds exactly in the way that it did for vectors by affecting each element by the operation.

★ **Example 3.16: Matrix Addition.**

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} -2 & 2 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{X} + \mathbf{Y} = \begin{bmatrix} 1-2 & 2+2 \\ 3+0 & 4+1 \end{bmatrix} = \begin{bmatrix} -1 & 4 \\ 3 & 5 \end{bmatrix}.$$

★ **Example 3.17: Matrix Subtraction.**

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} -2 & 2 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{X} - \mathbf{Y} = \begin{bmatrix} 1-(-2) & 2-2 \\ 3-0 & 4-1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 3 & 3 \end{bmatrix}.$$

★ **Example 3.18: Scalar Multiplication.**

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad s = 5$$

$$s \times \mathbf{X} = \begin{bmatrix} 5\times 1 & 5\times 2 \\ 5\times 3 & 5\times 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix}.$$

★ **Example 3.19: Scalar Division.**

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad s = 5$$

$$\mathbf{X} \div s = \begin{bmatrix} 1\div 5 & 2\div 5 \\ 3\div 5 & 4\div 5 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}.$$

One special case is worth mentioning. A common implied scalar multiplication is the negative of a matrix, $-\mathbf{X}$. This is a shorthand means for saying that every matrix element in $\mathbf{X}$ is multiplied by -1.

These are the most basic matrix operations and obviously consist of nothing more than being careful about performing each individual elemental operation. As with vectors, we can summarize the arithmetic properties as follows.

Properties of (Conformable) Matrix Manipulation

$\rightarrowtail$	Commutative Property	$\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X}$
$\rightarrowtail$	Additive Associative Property	$(\mathbf{X} + \mathbf{Y}) + \mathbf{Z} = \mathbf{X} + (\mathbf{Y} + \mathbf{Z})$
$\rightarrowtail$	Matrix Distributive Property	$s(\mathbf{X} + \mathbf{Y}) = s\mathbf{X} + s\mathbf{Y}$
$\rightarrowtail$	Scalar Distributive Property	$(s + t)\mathbf{X} = s\mathbf{X} + t\mathbf{X}$
$\rightarrowtail$	Zero Property	$\mathbf{X} + \mathbf{0} = \mathbf{X}$ and $\mathbf{X} - \mathbf{X} = \mathbf{0}$

★ **Example 3.20: Matrix Calculations.** This example illustrates several of the properties above where $s = 7$, $t = 2$, $\mathbf{X} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$, and $\mathbf{Y} = \begin{bmatrix} 3 & 4 \\ 0 & -1 \end{bmatrix}$. The left-hand side is

$$(s + t)(\mathbf{X} + \mathbf{Y}) = (7 + 2)\left(\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 4 \\ 0 & -1 \end{bmatrix}\right)$$

$$= 9\begin{bmatrix} 5 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 45 & 36 \\ 9 & 0 \end{bmatrix},$$

and the right-hand side is

$$
\begin{aligned}
t\mathbf{Y}+s\mathbf{Y} + t\mathbf{X} + s\mathbf{X} \\
&= 2\begin{bmatrix} 3 & 4 \\ 0 & -1 \end{bmatrix} + 7\begin{bmatrix} 3 & 4 \\ 0 & -1 \end{bmatrix} + 2\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} + 7\begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 6 & 8 \\ 0 & -2 \end{bmatrix} + \begin{bmatrix} 21 & 28 \\ 0 & -7 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 14 & 0 \\ 7 & 7 \end{bmatrix} \\
&= \begin{bmatrix} 45 & 36 \\ 9 & 0 \end{bmatrix}.
\end{aligned}
$$

Matrix multiplication is necessarily more complicated than these simple operations. The first issue is conformability. Two matrices are conformable for multiplication if the number of columns in the first matrix match the number of rows in the second matrix. Note that this implies that ***the order of multiplication matters with matrices***. This is the first algebraic principle that deviates from the simple scalar world that we all learned early on in life. To be specific, suppose that $\mathbf{X}$ is size $k \times n$ and $\mathbf{Y}$ is size $n \times p$. Then the multiplication operation given by

$$\underset{(k\times n)}{\mathbf{X}}\underset{(n\times p)}{\mathbf{Y}}$$

is valid because the inner numbers match up, but the multiplication operation given by

$$\underset{(n\times p)}{\mathbf{Y}}\underset{(k\times n)}{\mathbf{X}}$$

is not unless $p = k$. Furthermore, the inner dimension numbers of the operation determine conformability and the outer dimension numbers determine the size of the resulting matrix. So in the example of $\mathbf{XY}$ above, the resulting matrix would be of size $k \times p$. To maintain awareness of this order of operation distinction, we say that $\mathbf{X}$ **pre-multiplies** $\mathbf{Y}$ or, equivalently, that $\mathbf{Y}$ **post-multiplies** $\mathbf{X}$.

So how is matrix multiplication done? In an attempt to be somewhat intuitive, we can think about the operation in ***vector terms***. For $\mathbf{X}_{k\times n}$ and $\mathbf{Y}_{n\times p}$, we take each of the n row vectors in $\mathbf{X}$ and perform a vector inner product with the n column vectors in $\mathbf{Y}$. This operation starts with performing the inner product of the first row in $\mathbf{X}$ with the first column in $\mathbf{Y}$ and the result will be the first element of the product matrix. Consider a simple case of two arbitrary 2×2 matrices:

$$
\begin{aligned}
\mathbf{XY} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \\
&= \begin{bmatrix} (x_{11}\ x_{12}) \cdot (y_{11}\ y_{21}) & (x_{11}\ x_{12}) \cdot (y_{12}\ y_{22}) \\ (x_{21}\ x_{22}) \cdot (y_{11}\ y_{21}) & (x_{21}\ x_{22}) \cdot (y_{12}\ y_{22}) \end{bmatrix} \\
&= \begin{bmatrix} x_{11}y_{11} + x_{12}y_{21} & x_{11}y_{12} + x_{12}y_{22} \\ x_{21}y_{11} + x_{22}y_{21} & x_{21}y_{12} + x_{22}y_{22} \end{bmatrix}.
\end{aligned}
$$

Perhaps we can make this more intuitive visually. Suppose that we notate the four values of the final matrix as $\mathbf{XY}[1,1], \mathbf{XY}[1,2], \mathbf{XY}[2,1], \mathbf{XY}[2,2]$ corresponding to their position in the 2×2 product. Then we can visualize how the rows of the first matrix operate against the columns of the second matrix to produce each value:

$$
\boxed{x_{11}\quad x_{12}}\ \ \boxed{\begin{matrix} y_{11} \\ y_{21} \end{matrix}} = \mathbf{XY}[1,1], \qquad \boxed{x_{11}\quad x_{12}}\ \ \boxed{\begin{matrix} y_{12} \\ y_{22} \end{matrix}} = \mathbf{XY}[1,2],
$$

$$\boxed{x_{21} \quad x_{22}} \; \boxed{\begin{matrix} y_{11} \\ y_{21} \end{matrix}} = \mathbf{XY}[2,1], \qquad \boxed{x_{21} \quad x_{22}} \; \boxed{\begin{matrix} y_{12} \\ y_{22} \end{matrix}} = \mathbf{XY}[2,2].$$

While it helps to visualize the process in this way, we can also express the product in a more general, but perhaps intimidating, scalar notation for an arbitrary-sized operation:

$$\underset{(k\times n)}{\mathbf{X}}\underset{(n\times p)}{\mathbf{Y}} = \begin{bmatrix} \sum_{i=1}^{n} x_{1i}y_{i1} & \sum_{i=1}^{n} x_{1i}y_{i2} & \cdots & \sum_{i=1}^{n} x_{1i}y_{ip} \\ \sum_{i=1}^{n} x_{2i}y_{i1} & \sum_{i=1}^{n} x_{2i}y_{i2} & \cdots & \sum_{i=1}^{n} x_{2i}y_{ip} \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ki}y_{i1} & \cdots & \cdots & \sum_{i=1}^{n} x_{ki}y_{ip} \end{bmatrix}.$$

To further clarify, now perform matrix multiplication with some actual values. Starting with the matrices

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} -2 & 2 \\ 0 & 1 \end{bmatrix},$$

calculate

$$\begin{aligned} \mathbf{XY} &= \begin{bmatrix} (1\;\;2)\cdot(-2\;\;0) & (1\;\;2)\cdot(2\;\;1) \\ (3\;\;4)\cdot(-2\;\;0) & (3\;\;4)\cdot(2\;\;1) \end{bmatrix} \\ &= \begin{bmatrix} (1)(-2)+(2)(0) & (1)(2)+(2)(1) \\ (3)(-2)+(4)(0) & (3)(2)+(4)(1) \end{bmatrix} \\ &= \begin{bmatrix} -2 & 4 \\ -6 & 10 \end{bmatrix}. \end{aligned}$$

As before with such topics, we consider the properties of matrix multiplication:

Properties of (Conformable) Matrix Multiplication

$\rightarrowtail$	Associative Property	$(\mathbf{XY})\mathbf{Z} = \mathbf{X}(\mathbf{YZ})$
$\rightarrowtail$	Additive Distributive Property	$(\mathbf{X} + \mathbf{Y})\mathbf{Z} = \mathbf{XZ} + \mathbf{YZ}$
$\rightarrowtail$	Scalar Distributive Property	$s\mathbf{XY} = (\mathbf{X}s)\mathbf{Y}$ $= \mathbf{X}(s\mathbf{Y}) = \mathbf{XY}s$
$\rightarrowtail$	Zero Property	$\mathbf{X0} = \mathbf{0}$

★ **Example 3.21: LU Matrix Decomposition**. Many square matrices can be decomposed as the product of lower and upper triangular matrices. This is a very general finding that we will return to and extend in the next chapter. The principle works like this for the matrix $\mathbf{A}$:

$$\underset{(p\times p)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{L}}\underset{(p\times p)}{\mathbf{U}},$$

where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{U}$ is an upper triangular matrix (sometimes a permutation matrix is also required; see the explanation of permutation matrices below).

Consider the following example matrix decomposition according to this scheme:

$$\begin{bmatrix} 2 & 3 & 3 \\ 1 & 2 & 9 \\ 1 & 1 & 12 \end{bmatrix} = \begin{bmatrix} 1.0 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3.0 & 3.0 \\ 0 & 0.5 & 7.5 \\ 0 & 0.0 & 18.0 \end{bmatrix}.$$

This decomposition is very useful for solving systems of equations because much of the mechanical work is already done by the triangularization.

Now that we have seen how matrix multiplication is performed, we can return to the principle that pre-multiplication is different than post-multiplication. In the case discussed we could perform one of these operations but not the other, so the difference was obvious. What about multiplying two square matrices? Both orders of multiplication are possible, but it turns out that except for special cases the result will differ. In fact, we need only provide one particular case to prove this point. Consider the matrices $\mathbf{X}$ and $\mathbf{Y}$:

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

$$\mathbf{YX} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

This is a very simple example, but the implications are obvious. Even in cases where pre-multiplication and post-multiplication are possible, these are different operations and matrix multiplication is not commutative.

Recall also the claim that the identity matrix $\mathbf{I}$ is operationally equivalent to 1 in matrix terms rather than the seemingly more obvious $\mathbf{J}$ matrix. Let us now test this claim on a simple matrix, first with $\mathbf{I}$:

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

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

and then with J:

$$\mathbf{XJ}_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(1) + (2)(1) & (1)(1) + (2)(1) \\ (3)(1) + (4)(1) & (3)(1) + (4)(1) \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 7 & 7 \end{bmatrix}.$$

The result here is interesting; post-multiplying by $\mathbf{I}$ returns the $\mathbf{X}$ matrix to its original form, but post-multiplying by $\mathbf{J}$ produces a matrix where values are the sum by row. What about pre-multiplication? Pre-multiplying by $\mathbf{I}$ also returns the original matrix (see the Exercises), but pre-multiplying by $\mathbf{J}$ gives

$$\mathbf{J}_2\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(1)+(1)(3) & (1)(2)+(1)(4) \\ (1)(1)+(1)(3) & (1)(2)+(1)(4) \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 4 & 6 \end{bmatrix},$$

which is now the sum down columns assigned as row values. This means that the $\mathbf{J}$ matrix can be very useful in calculations (including linear regression methods), but it does not work as a "one" in matrix terms. There is also a very interesting multiplicative property of the $\mathbf{J}$ matrix, particularly for nonsquare forms:

$$\underset{(p\times n)}{\mathbf{J}}\underset{(n\times k)}{\mathbf{J}} = n \underset{(p\times k)}{\mathbf{J}}.$$

Basic manipulations of the identity matrix can provide forms that are enormously useful in matrix multiplication calculations. Suppose we wish to switch two rows of a specific matrix. To accomplish this we can multiply by an identity matrix where the placement of the 1 values is switched:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{31} & x_{32} & x_{33} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}.$$

This pre-multiplying matrix is called a **permutation matrix** because it permutes the matrix that it operates on. Interestingly, a permutation matrix can be applied to a conformable vector with the obvious results.

The effect of changing a single 1 value to some other scalar is fairly obvious:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ sx_{31} & sx_{32} & sx_{33} \end{bmatrix},$$

but the effect of changing a single 0 value is not:

$$\begin{bmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} x_{11} + sx_{31} & x_{12} + sx_{32} & x_{13} + sx_{33} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}.$$

★ **Example 3.22: Matrix Permutation Calculation**. Consider the following example of permutation with an off-diagonal nonzero value:

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 2 & 3 \\ 7 & 0 & 1 \\ 3 & 3 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \cdot 3 + 0 \cdot 7 + 3 \cdot 3) & (1 \cdot 2 + 0 \cdot 0 + 3 \cdot 3) & (1 \cdot 3 + 0 \cdot 1 + 3 \cdot 3) \\ (0 \cdot 3 + 0 \cdot 7 + 1 \cdot 3) & (0 \cdot 2 + 0 \cdot 0 + 1 \cdot 3) & (0 \cdot 3 + 0 \cdot 1 + 1 \cdot 3) \\ (0 \cdot 3 + 1 \cdot 7 + 0 \cdot 3) & (0 \cdot 2 + 1 \cdot 0 + 0 \cdot 3) & (0 \cdot 3 + 1 \cdot 1 + 0 \cdot 3) \end{bmatrix}$$

$$= \begin{bmatrix} 12 & 11 & 12 \\ 3 & 3 & 3 \\ 7 & 0 & 1 \end{bmatrix},$$

which shows the switching of rows two and three as well as the confinement of multiplication by 3 to the first row.

3.5 Matrix Transposition

Another operation that is commonly performed on a single matrix is **transposition**. We saw this before in the context of vectors: switching between column and row forms. For matrices, this is slightly more involved but straightforward to understand: simply switch rows and columns. The transpose of an $i \times j$

matrix $\mathbf{X}$ is the $j \times i$ matrix $\mathbf{X}'$, usually called “X prime” (sometimes denoted $\mathbf{X}^T$ though). For example,

$$\mathbf{X}' = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

In this way the inner structure of the matrix is preserved but the shape of the matrix is changed. An interesting consequence is that transposition allows us to calculate the “square” of some arbitrary-sized $i \times j$ matrix: $\mathbf{X}'\mathbf{X}$ is always conformable, as is $\mathbf{X}\mathbf{X}'$, even if $i \neq j$. We can also be more precise about the definition of symmetric and skew-symmetric matrices. Consider now some basic properties of transposition.

Properties of Matrix Transposition

$\rightarrowtail$	Invertibility	$(\mathbf{X}')' = \mathbf{X}$
$\rightarrowtail$	Additive Property	$(\mathbf{X}+\mathbf{Y})' = \mathbf{X}' + \mathbf{Y}'$
$\rightarrowtail$	Multiplicative Property	$(\mathbf{X}\mathbf{Y})' = \mathbf{Y}'\mathbf{X}'$
$\rightarrowtail$	General Multiplicative Property	$(\mathbf{X}_1\mathbf{X}_2 \ldots \mathbf{X}_{n-1}\mathbf{X}_n)' = \mathbf{X}_n'\mathbf{X}_{n-1}' \ldots \mathbf{X}_2'\mathbf{X}_1'$
$\rightarrowtail$	Symmetric Matrix	$\mathbf{X}' = \mathbf{X}$
$\rightarrowtail$	Skew-Symmetric Matrix	$\mathbf{X} = -\mathbf{X}'$

Note, in particular, from this list that the multiplicative property of transposition reverses the order of the matrices.

★ **Example 3.23: Calculations with Matrix Transpositions**. Suppose we have the three matrices:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 3 & 7 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} -2 & -2 \\ 1 & 0 \end{bmatrix}.$$

Then the following calculation of $(\mathbf{XY}' + \mathbf{Z})' = \mathbf{Z}' + \mathbf{YX}'$ illustrates the invertibility, additive, and multiplicative properties of transposition. The left-hand side is

$$\begin{aligned}(\mathbf{XY}' + \mathbf{Z})' &= \left(\begin{bmatrix} 1 & 0 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}' + \begin{bmatrix} -2 & -2 \\ 1 & 0 \end{bmatrix} \right)' \\ &= \left(\begin{bmatrix} 2 & 2 \\ 27 & 20 \end{bmatrix} + \begin{bmatrix} -2 & -2 \\ 1 & 0 \end{bmatrix} \right)' \\ &= \left(\begin{bmatrix} 0 & 0 \\ 28 & 20 \end{bmatrix} \right)',\end{aligned}$$

and the right-hand side is

$$\begin{aligned}\mathbf{Z}' + \mathbf{YX}' &= \begin{bmatrix} -2 & -2 \\ 1 & 0 \end{bmatrix}' + \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 7 \end{bmatrix}' \\ &= \begin{bmatrix} -2 & 1 \\ -2 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 27 \\ 2 & 20 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 28 \\ 0 & 20 \end{bmatrix}.\end{aligned}$$

3.6 Advanced Topics

This section contains a set of topics that are less frequently used in the social sciences but may appear in some literatures. Readers may elect to skip this section or use it for reference only.

3.6.1 Special Matrix Forms

An interesting type of matrix that we did not discuss before is the **idempotent matrix**. This is a matrix that has the multiplication property

$$\mathbf{X}\mathbf{X} = \mathbf{X}^2 = \mathbf{X}$$

and therefore the property

$$\mathbf{X}^n = \mathbf{X}\mathbf{X}\cdots\mathbf{X} = \mathbf{X}, \quad n \in \mathcal{I}^+$$

(i.e., n is some positive integer). Obviously the identity matrix and the zero matrix are idempotent, but the somewhat weird truth is that there are lots of other idempotent matrices as well. This emphasizes how different matrix algebra can be from scalar algebra. For instance, the following matrix is idempotent, but you probably could not guess so by staring at it:

$$\begin{bmatrix} -1 & 1 & -1 \\ 2 & -2 & 2 \\ 4 & -4 & 4 \end{bmatrix}$$

(try multiplying it). Interestingly, if a matrix is idempotent, then the difference between this matrix and the identity matrix is also idempotent because

$$(\mathbf{I} - \mathbf{X})^2 = \mathbf{I}^2 - 2\mathbf{X} + \mathbf{X}^2 = \mathbf{I} - 2\mathbf{X} + \mathbf{X} = (\mathbf{I} - \mathbf{X}).$$

We can test this with the example matrix above:

$$(\mathbf{I} - \mathbf{X})^2 = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 & 1 & -1 \\ 2 & -2 & 2 \\ 4 & -4 & 4 \end{bmatrix} \right)^2$$

$$= \begin{bmatrix} 2 & -1 & 1 \\ -2 & 3 & -2 \\ -4 & 4 & -3 \end{bmatrix}^2 = \begin{bmatrix} 2 & -1 & 1 \\ -2 & 3 & -2 \\ -4 & 4 & -3 \end{bmatrix}.$$

Relatedly, a square **nilpotent** matrix is one with the property that $\mathbf{X}^n = \mathbf{0}$, for a positive integer n. Clearly the zero matrix is nilpotent, but others exist as well. A basic 2×2 example is the nilpotent matrix

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

Another particularistic matrix is a **involutory matrix**, which has the property that when squared it produces an identity matrix. For example,

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}^2 = \mathbf{I},$$

although more creative forms exist.

3.6.2 Vectorization of Matrices

Occasionally it is convenient to rearrange a matrix into vector form. The most common way to do this is to "stack" vectors from the matrix on top of each other, beginning with the first column vector of the matrix, to form one long column vector. Specifically, to **vectorize** an $i \times j$ matrix $\mathbf{X}$, we consecutively stack the j-length column vectors to obtain a single vector of length ij. This is denoted $\text{vec}(\mathbf{X})$ and has some obvious properties, such as $s\text{vec}(\mathbf{X}) = \text{vec}(s\mathbf{X})$ for some vector s and $\text{vec}(\mathbf{X} + \mathbf{Y}) = \text{vec}(\mathbf{X}) + \text{vec}(\mathbf{Y})$ for matrices conformable by addition. Returning to our simple example,

$$\text{vec}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}.$$

Interestingly, it is not true that $\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}')$ since the latter would stack rows instead of columns. And vectorization of products is considerably more involved (see the next section).

A final, and sometimes important, type of matrix multiplication is the **Kronecker product** (also called the ***tensor product***), which comes up naturally in the statistical analyses of time series data (data recorded on the same measures of interest at different points in time). This is a slightly more abstract process but has the advantage that there is no conformability requirement. For the $i \times j$ matrix $\mathbf{X}$ and $k \times \ell$ matrix $\mathbf{Y}$, a Kronecker product is the $(ik) \times (j\ell)$ matrix

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & x_{12}\mathbf{Y} & \cdots & \cdots & x_{1j}\mathbf{Y} \\ x_{21}\mathbf{Y} & x_{22}\mathbf{Y} & \cdots & \cdots & x_{2j}\mathbf{Y} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ x_{i1}\mathbf{Y} & x_{i2}\mathbf{Y} & \cdots & \cdots & x_{ij}\mathbf{Y} \end{bmatrix},$$

which is different than

$$\mathbf{Y} \otimes \mathbf{X} = \begin{bmatrix} y_{11}\mathbf{X} & y_{12}\mathbf{X} & \cdots & \cdots & y_{1j}\mathbf{X} \\ y_{21}\mathbf{X} & y_{22}\mathbf{X} & \cdots & \cdots & y_{2j}\mathbf{X} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ y_{i1}\mathbf{X} & y_{i2}\mathbf{X} & \cdots & \cdots & y_{1j}\mathbf{X} \end{bmatrix}.$$

As an example, consider the following numerical case.

★ **Example 3.24: Kronecker Product.** A numerical example of a Kronecker product follows for a (2×2) by (2×3) case:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

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

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

$$= \begin{bmatrix} -2 & 2 & 3 & -4 & 4 & 6 \\ 0 & 1 & 3 & 0 & 2 & 6 \\ -6 & 6 & 9 & -8 & 8 & 12 \\ 0 & 3 & 9 & 0 & 4 & 12 \end{bmatrix},$$

which is clearly different from the operation performed in reverse order:

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

$$= \begin{bmatrix} -2 & -4 & 2 & 4 & 3 & 6 \\ -6 & -8 & 6 & 8 & 9 & 12 \\ 0 & 0 & 1 & 2 & 3 & 6 \\ 0 & 0 & 3 & 4 & 9 & 12 \end{bmatrix},$$

even though the resulting matrices are of the same dimension.

The vectorize function above has a product that involves the Kronecker function. For $i \times j$ matrix $\mathbf{X}$ and $j \times k$ matrix $\mathbf{Y}$, we get $\text{vec}(\mathbf{XY}) = (\mathbf{I} \otimes \mathbf{X})\text{vec}(\mathbf{Y})$, where $\mathbf{I}$ is an identity matrix of order i. For three matrices this is only slightly more complex: $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}' \otimes \mathbf{X})\text{vec}(\mathbf{Y})$, for $k \times \ell$ matrix $\mathbf{Z}$. Kronecker products have some other interesting properties as well (matrix inversion is discussed in the next chapter):

Properties of Kronecker Products

$\rightarrowtail$	Trace	$\text{tr}(\mathbf{X} \otimes \mathbf{Y}) = tr\mathbf{X} \otimes tr\mathbf{Y}$
$\rightarrowtail$	Transpose	$(\mathbf{X} \otimes \mathbf{Y})' = \mathbf{X}' \otimes \mathbf{Y}'$
$\rightarrowtail$	Inversion	$(\mathbf{X} \otimes \mathbf{Y})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1}$
$\rightarrowtail$	Products	$(\mathbf{X} \otimes \mathbf{Y})(\mathbf{W} \otimes \mathbf{Z}) = \mathbf{XW} \otimes \mathbf{YZ}$
$\rightarrowtail$	Associative	$(\mathbf{X} \otimes \mathbf{Y}) \otimes \mathbf{W} = \mathbf{X} \otimes (\mathbf{Y} \otimes \mathbf{W})$
$\rightarrowtail$	Distributive	$(\mathbf{X} + \mathbf{Y}) \otimes \mathbf{W} = (\mathbf{X} \otimes \mathbf{W}) + (\mathbf{Y} \otimes \mathbf{W})$

Here the notation tr() denotes the "trace," which is just the sum of the diagonal values going from the uppermost left value to the lowermost right value, for square matrices. Thus the trace of an identity matrix would be just its order. This is where we will pick up next in Chapter 4.

★ **Example 3.25: Distributive Property of Kronecker Products Calculation**. Given the following matrices:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} -1 & -3 \\ 1 & 1 \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix},$$

we demonstrate that $(\mathbf{X}+\mathbf{Y})\otimes\mathbf{W} = (\mathbf{X}\otimes\mathbf{W})+(\mathbf{X}\otimes\mathbf{W})$. The left-hand side is

$$\begin{aligned}
(\mathbf{X}+\mathbf{Y})\otimes\mathbf{W} &= \left(\begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix} + \begin{bmatrix} -1 & -3 \\ 1 & 1 \end{bmatrix}\right) \otimes \begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -2 \\ 3 & 6 \end{bmatrix} \otimes \begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & -2\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \\ 3\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & 6\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & -4 & 4 \\ 0 & 0 & -6 & 0 \\ 6 & -6 & 12 & -12 \\ 9 & 0 & 18 & 0 \end{bmatrix},
\end{aligned}$$

and the right-hand side, $(\mathbf{X}\otimes\mathbf{W})+(\mathbf{X}\otimes\mathbf{W})$, is

$$= \left(\begin{bmatrix} 1 & 1 \\ 2 & 5 \end{bmatrix} \otimes \begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix}\right) + \left(\begin{bmatrix} -1 & -3 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & 1\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \\ 2\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & 5\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \end{bmatrix}$$

$$+ \begin{bmatrix} -1\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & -3\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \\ 1\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} & 1\begin{bmatrix} 2 & -2 \\ 3 & 0 \end{bmatrix} \end{bmatrix},$$

which simplifies down to

$$= \begin{bmatrix} 2 & -2 & 2 & -2 \\ 3 & 0 & 3 & 0 \\ 4 & -4 & 10 & -10 \\ 6 & 0 & 15 & 0 \end{bmatrix} + \begin{bmatrix} -2 & 2 & -6 & 6 \\ -3 & 0 & -9 & 0 \\ 2 & -2 & 2 & -2 \\ 3 & 0 & 3 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -4 & 4 \\ 0 & 0 & -6 & 0 \\ 6 & -6 & 12 & -12 \\ 9 & 0 & 18 & 0 \end{bmatrix}.$$

3.7 New Terminology

conformable, 85
diagonal matrix, 102
dimensions, 100
dot product, 87
entries, 100
equitable matrix, 127
field, 83
Hadamard product, 129
idempotent matrix, 118
identity matrix, 103
involutory matrix, 119
Jordan product, 130
Kronecker product, 119
law of cosines, 95
Lie product, 130
lower triangular, 104
matrix, 100
matrix decomposition, 104
matrix elements, 100
matrix equality, 101
matrix multiplication, 109
nilpotent matrix, 118
nonconformable, 85
order-k, 101
orthogonal, 87
outer product, 89
permutation matrix, 114
p-norm, 96
post-multiplies, 110
pre-multiplies, 110
scalar product, 87
skew-symmetric, 102
square matrix, 101
symmetric matrix, 102
transposition, 115
triangular matrix, 104
unit vector, 96
upper matrix, 104
vector, 83
vector cross product, 89
vector inner (dot) product, 87
vector norm, 93
vector transpose, 92
vectorize function, 119
zero matrix, 104

Exercises

3.1 Perform the following vector multiplication operations:

$$[1 \quad 1 \quad 1] \cdot [a \quad b \quad c]'$$

$$[1 \quad 1 \quad 1] \times [a \quad b \quad c]'$$

$$[-1 \quad 1 \quad -1] \cdot [4 \quad 3 \quad 12]'$$

$$[-1 \quad 1 \quad -1] \times [4 \quad 3 \quad 12]'$$

$$[0 \quad 9 \quad 0 \quad 11] \cdot [123.98211 \quad 6 \quad -6392.38743 \quad -5]'$$

$$[123.98211 \quad 6 \quad -6392.38743 \quad -5] \cdot [0 \quad 9 \quad 0 \quad 11]'.$$

3.2 Recalculate the two outer product operations in Example 3.2 only by using the vector $(-1) \times [3, 3, 3]$ instead of $[3, 3, 3]$. What is the interpretation of the result with regard to the direction of the resulting row and column vectors compared with those in the example?

3.3 Show that $\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$ implies $\cos(\theta) = \frac{\mathbf{v}\mathbf{w}}{\|\mathbf{v}\| \, \|\mathbf{w}\|}$.

3.4 What happens when you calculate the difference norm ($||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 - 2(\mathbf{u} \cdot \mathbf{v}) + ||\mathbf{v}||^2$) for two orthogonal vectors? How is this different from the multiplication norm for two such vectors?

3.5 Explain why the perpendicularity property is a special case of the triangle inequality for vector p-norms.

3.6 For p-norms, explain why the Cauchy-Schwarz inequality is a special case of Hölder's inequality.

3.7 Show that pre-multiplication and post-multiplication with the identity matrix are equivalent.

3.8 Recall that an involutory matrix is one that has the characteristic $X^2 = I$. Can an involutory matrix ever be idempotent?

3.9 For the following matrix, calculate $\mathbf{X}^n$ for $n = 2, 3, 4, 5$. Write a rule for calculating higher values of n.

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

3.10 Perform the following vector/matrix multiplications:

$$\begin{bmatrix} 1 & \frac{1}{2} & 2 \\ 1 & \frac{1}{3} & 5 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 7 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 7 & 5 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & 3 & 1 \\ 3 & 1 & 3 \\ 1 & 3 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

3.11 Perform the following matrix multiplications:

$$\begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & -2 \\ 6 & 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 7 \\ 3 & 0 \\ 1 & 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -9 \\ -1 & -4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -4 & -4 \\ -1 & 0 \\ -3 & -8 \end{bmatrix}' \qquad \begin{bmatrix} 0 & 0 \\ 0 & \infty \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

3.12 An **equitable matrix** is a square matrix of order n where all entries are positive and for any three values $i, j, k < n$, $x_{ij}x_{jk} = x_{ik}$. Show that for equitable matrices of order n, $X^2 = nX$. Give an example of an equitable matrix.

3.13 Communication within work groups can sometimes be studied by looking analytically at individual decision processes. Roby and Lanzetta (1956) studied at this process by constructing three matrices: OR, which maps six observations to six possible responses; PO, which indicates which type of person from three is a source of information for each observation; and PR, which maps who is responsible of the three for each of the six responses. They give these matrices (by example) as

$$OR = \begin{array}{c} \\ O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \\ O_6 \end{array} \begin{array}{c} \begin{array}{cccccc} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

$$PO = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \end{array} \begin{array}{c} \begin{array}{cccccc} O_1 & O_2 & O_3 & O_4 & O_5 & O_6 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}.$$

$$PR = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \end{array} \begin{array}{c} \begin{array}{ccc} P_1 & P_2 & P_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The claim is that multiplying these matrices in the order OR, PO, PR produces a personnel-only matrix (OPR) that reflects “the degree of operator interdependence entailed in a given task and personnel structure” where the total number of entries is proportional to the system complexity, the entries along the main diagonal show how autonomous the relevant agent is, and off-diagonals show sources of information in the organization. Perform matrix multiplication in this order to obtain the OPR matrix using transformations as needed where your final matrix has a zero in the last entry of the first row. Which matrix most affects the diagonal values of OPR when it is manipulated?

3.14 Singer and Spilerman (1973) used matrices to show social mobility between classes. These are stochastic matrices indicating different social class categories where the rows must sum to 1. In this construction a diagonal matrix means that there is no social mobility. Test their claim that the following matrix is the cube root of a stochastic matrix:

$$\mathbf{P}^{\frac{1}{3}} = \begin{pmatrix} \frac{1}{2}(1 - 1/\sqrt[3]{-\frac{1}{3}} & \frac{1}{2}(1 + 1/\sqrt[3]{-\frac{1}{3}} \\ \frac{1}{2}(1 + 1/\sqrt[3]{-\frac{1}{3}} & \frac{1}{2}(1 - 1/\sqrt[3]{-\frac{1}{3}}. \end{pmatrix}$$

3.15 Element-by-element matrix multiplication is a **Hadamard product** (and sometimes called a Schur product), and it is denoted with either “$*$” or “$\odot$” (and occasionally “$\circ$”) This element-wise process means that if $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary matrices of identical size, the Hadamard product is $\mathbf{X}\odot\mathbf{Y}$ whose ijth element ($\mathbf{XY}_{ij}$) is $\mathbf{X}_{ij}\mathbf{Y}_{ij}$. It is trivial to see that $\mathbf{X}\odot\mathbf{Y} = \mathbf{Y}\odot\mathbf{X}$ (an interesting exception to general matrix multiplication properties), but show that for two nonzero matrices $\text{tr}(\mathbf{X}\odot\mathbf{Y}) = \text{tr}(\mathbf{X})\cdot\text{tr}(\mathbf{Y})$. For some nonzero matrix $\mathbf{X}$ what does $\mathbf{I}\odot\mathbf{X}$ do? For an order k $\mathbf{J}$ matrix, is $\text{tr}(\mathbf{J}\odot\mathbf{J})$ different from $\text{tr}(\mathbf{JJ})$? Show why or why not.

3.16 For the following LU matrix decomposition, find the permutation matrix $\mathbf{P}$ that is necessary:

$$\begin{bmatrix} 1 & 3 & 7 \\ 1 & 1 & 12 \\ 4 & 2 & 9 \end{bmatrix} = \mathbf{P} \begin{bmatrix} 1.00 & 0.0 & 0 \\ 0.25 & 1.0 & 0 \\ 0.25 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2.0 & 9.00 \\ 0 & 2.5 & 4.75 \\ 0 & 0.0 & 8.80 \end{bmatrix}.$$

3.17 Prove that the product of an idempotent matrix is idempotent.

3.18 In the process of developing multilevel models of sociological data DiPrete and Grusky (1990) and others performed the matrix calculations $\mathbf{\Phi} = \mathbf{X}(\mathbf{I} \otimes \mathbf{\Delta}_\mu)\mathbf{X}' + \mathbf{\Sigma}_\epsilon$, where $\mathbf{\Sigma}_\epsilon$ is a $T \times T$ diagonal matrix with values $\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2$; $\mathbf{X}$ is an arbitrary (here) nonzero $n \times T$ matrix with $n > T$; and $\mathbf{\Delta}_\mu$ is a $T \times T$ diagonal matrix with values $\sigma_{\mu_1}^2, \sigma_{\mu_2}^2, \ldots, \sigma_{\mu_T}^2$. Perform this calculation to show that the result is a "block diagonal" matrix and explain this form. Use generic x_{ij} values or some other general form to denote elements of $\mathbf{X}$. Does this say anything about the Kronecker product using an identity matrix?

3.19 Calculate the LU decomposition of the matrix $\left[\begin{smallmatrix} 2 & 3 \\ 4 & 7 \end{smallmatrix}\right]$ using your preferred software such as with the `lu` function of the `Matrix` library in the `R` environment. Reassemble the matrix by doing the multiplication without using software.

3.20 The **Jordan product** for matrices is defined by

$$\mathbf{X} * \mathbf{Y} = \frac{1}{2}(\mathbf{XY} + \mathbf{YX}),$$

and the **Lie product** from group theory is

$$\mathbf{X}x\mathbf{Y} = \mathbf{XY} - \mathbf{YX}$$

(both assuming conformable $\mathbf{X}$ and $\mathbf{Y}$). The Lie product is also sometimes denoted with $[\mathbf{X}, \mathbf{Y}]$. Prove the identity relating standard matrix multiplication to the Jordan and Lie forms: $\mathbf{XY} = [\mathbf{X} * \mathbf{Y}] + [\mathbf{X}x\mathbf{Y}/2]$.

3.21 Demonstrate the inversion property for Kronecker products, $(\mathbf{X} \otimes \mathbf{Y})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1}$, with the following matrices:

$$\mathbf{X} = \begin{bmatrix} 9 & 1 \\ 2 & 8 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 2 & -5 & 1 \\ 2 & 1 & 7 \end{bmatrix}.$$

3.22 Vectorize the following matrix and find the vector norm. Can you think of any shortcuts that would make the calculations less repetitious?

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 3 \\ 3 & 1 & 2 \\ 4 & 3 & 6 \\ 5 & 5 & 5 \\ 6 & 7 & 6 \\ 7 & 9 & 9 \\ 8 & 8 & 8 \\ 9 & 8 & 3 \end{bmatrix}.$$

3.23 For two vectors in $\Re^3$ using $1 = \cos^2\theta + \sin^2\theta$ and $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \mathbf{u}^2 \cdot \mathbf{v}^2$, show that the norm of the cross product between two vectors, $\mathbf{u}$ and $\mathbf{v}$, is: $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin(\theta)$.