Essential Mathematics for Political and Social Research

Jeff Gill
University of California, Davis

Cambridge University Press

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:

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

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

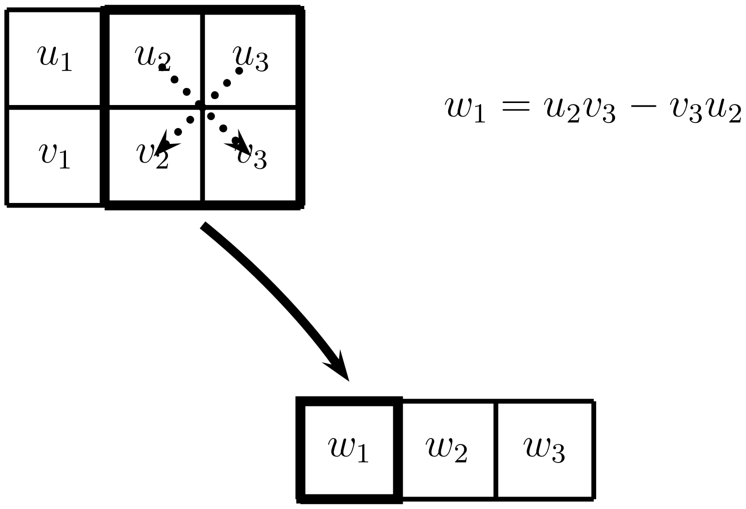
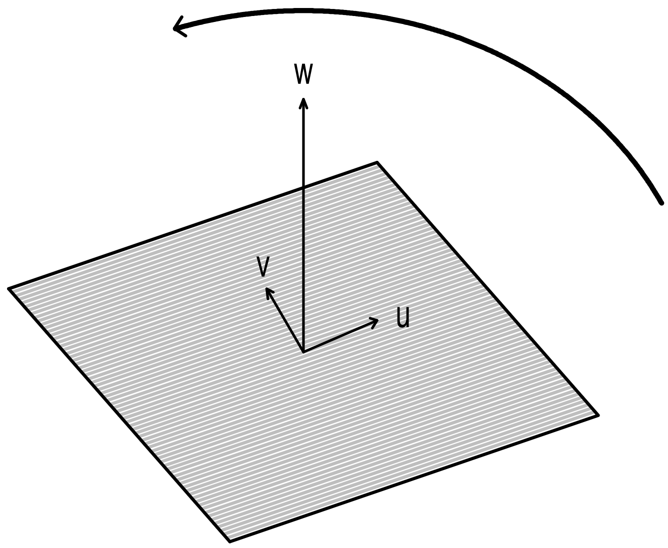


Figure 3.1 gives the intuition behind this product. First the $\mathbf{u}$ and $\mathbf{v}$ vectors are stacked on top of each other in the upper part of the illustration. The process of calculating the first vector value of the cross product, which we will call w_1, is done by "crossing" the elements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

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

★ **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 $\frac{1}{p}+\frac{1}{q}=1$, $\|\mathbf{v}\cdot\mathbf{w}\| \leq \|\mathbf{v}\|_p\|\mathbf{w}\|_q$
$\rightarrowtail$	Cauchy-Schwarz Ineq.	$\|\mathbf{v}\cdot\mathbf{w}\| \leq \|\mathbf{v}\|_2\|\mathbf{w}\|_2$
$\rightarrowtail$	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	$\|s\mathbf{v}\| = \|s\|\|\mathbf{v}\|$

★ **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):

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

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 $\boldsymbol{B}$: $\boldsymbol{A} = \boldsymbol{B} \Longleftrightarrow a_{ij} = b_{ij}\ \forall i, j$. 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:

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

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:

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

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

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

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

and the right-hand side is

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

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 vec$(\mathbf{X})$ and has some obvious properties, such as $s$vec$(\mathbf{X}) =$ vec$(s\mathbf{X})$ for some vector s and vec$(\mathbf{X} + \mathbf{Y}) =$ vec$(\mathbf{X}) +$ 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 vec$(\mathbf{X}) =$ 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 $\text{tr}()$ 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 $\cos(\theta) = \frac{\mathbf{vw}}{\|\mathbf{v}\| \, \|\mathbf{w}\|}$.

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:

$$\mathbf{P}^{\frac{1}{3}} = \left(\begin{array}{cc} \frac{1}{2}(1 - 1/\sqrt[3]{-\frac{1}{3}} & \frac{1}{2}(1 + 1/\sqrt[3]{-\frac{1}{3}} \\ \frac{1}{2}(1 + 1/\sqrt[3]{-\frac{1}{3}} & \frac{1}{2}(1 - 1/\sqrt[3]{-\frac{1}{3}}. \end{array} \right)$$

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: $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin(\theta)$.

4

Linear Algebra Continued: Matrix Structure

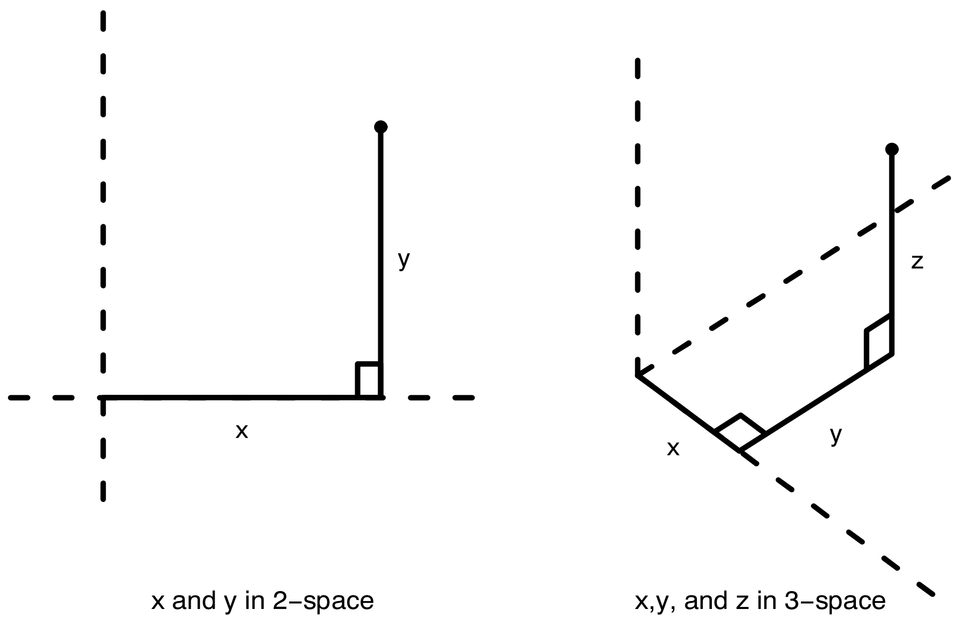
4.1 Objectives

This chapter introduces more theoretical and abstract properties of vectors and matrices. We already (by now!) know the mechanics of manipulating these forms, and it is important to carry on to a deeper understanding of the properties asserted by specific row and column formations. The last chapter gave some of the algebraic basics of matrix manipulation, but this is really insufficient for understanding the full scope of linear algebra. Importantly, there are characteristics of a matrix that are not immediately obvious from just looking at its elements and dimension. The structure of a given matrix depends not only on the arrangement of numbers within its rectangular arrangement, but also on the relationship between these elements and the "size" of the matrix. The idea of size is left vague for the moment, but we will shortly see that there are some very specific ways to claim size for matrices, and these have important theoretical properties that define how a matrix works with other structures. This chapter demonstrates some of these properties by providing information about the internal dynamics of matrix structure. Some of these topics are a bit more abstract than those in the last chapter.

4.2 Space and Time

We have already discussed basic Euclidean geometric systems in Chapter 1. Recall that Cartesian coordinate systems define real-measured axes whereby points are uniquely defined in the subsequent space. So in a Cartesian plane defined by $\Re^2$, points define an ordered pair designating a unique position on this 2-space. Similarly, an ordered triple defines a unique point in $\Re^3$ 3-space. Examples of these are given in Figure 4.1.

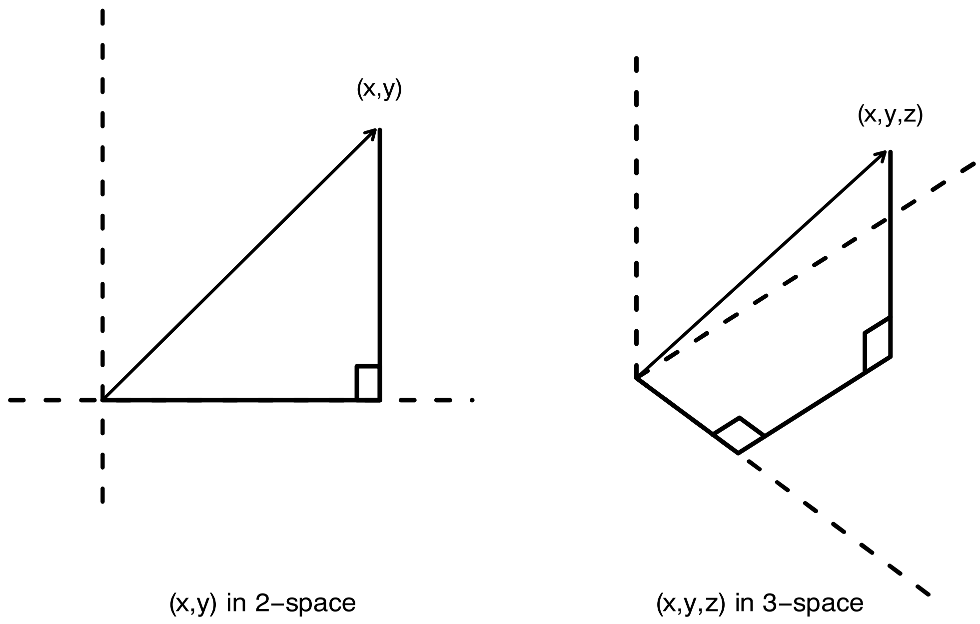
Fig. 4.1. VISUALIZING SPACE



What this figure shows with the lines is that the ordered pair or ordered triple defines a "path" in the associated space that uniquely arrives at a single point. Observe also that in both cases the path illustrated in the figure begins at the origin of the axes. So we are really defining a ***vector*** from the zero point to the arrival point, as shown in Figure 4.2.

Wait! This looks like a figure for illustrating the Pythagorean Theorem (the little squares are reminders that these angles are right angles). So if we wanted to get the length of the vectors, it would simply be $\sqrt{x^2 + y^2}$ in the first panel and $\sqrt{x^2 + y^2 + z^2}$ in the second panel. This is the intuition behind the basic vector norm in Section 3.2.1 of the last chapter.

Fig. 4.2. Visualizing Vectors in Spaces

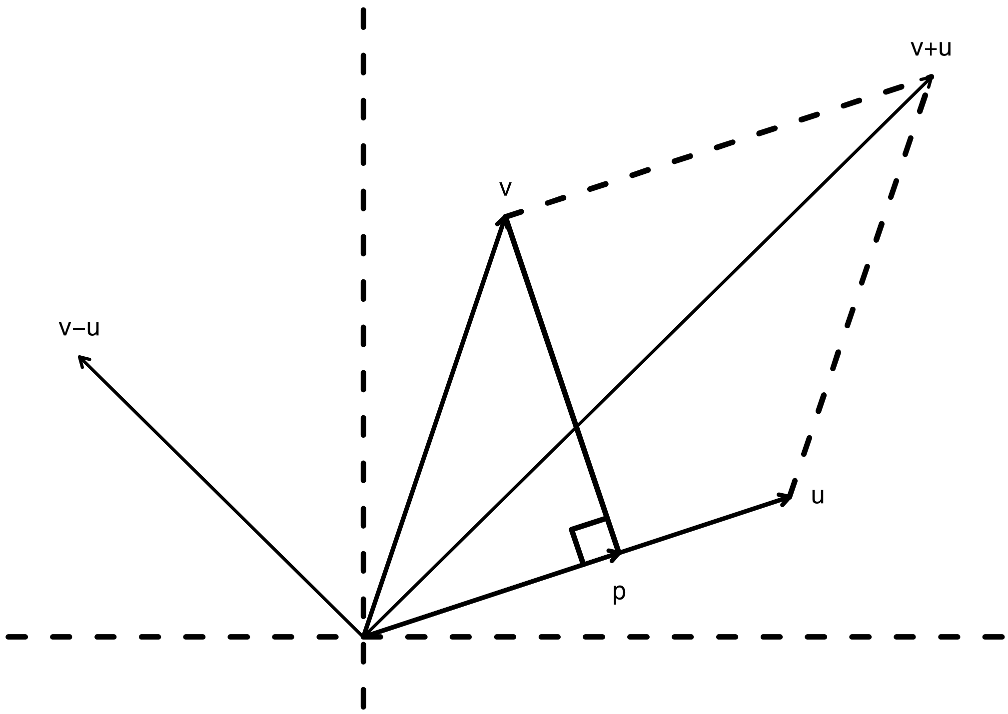


Thinking broadly about the two vectors in Figure 4.2, they take up an amount of "space" in the sense that they define a triangular planar region bounded by the vector itself and its two (left panel) or three (right panel) **projections** against the axes where the angle on the axis from this projection is necessarily a right angle (hence the reason that these are sometimes called **orthogonal projections**). Projections define how far along that axis the vector travels in total. Actually a projection does not have be just along the axes: We can project a vector $\mathbf{v}$ against another vector $\mathbf{u}$ with the following formula:

$$p = \text{ projection of } \mathbf{v} \text{ on to } \mathbf{u} \; = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|}\right)\left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right).$$

This is shown in Figure 4.3. We can think of the second fraction on the right-hand side above as the unit vector in the direction of $\mathbf{u}$, so the first fraction is a scalar multiplier giving length. Since the right angle is preserved, we can also think about rotating this arrangement until $\mathbf{v}$ is lying on the x-axis. Then it will be the same type of projection as before. Recall from before that two vectors at right angles, such as Cartesian axes, are called orthogonal. It should be reasonably easy to see now that orthogonal vectors produce zero-length projections.

Fig. 4.3. Vector Projection, Addition, and Subtraction



Another interesting case is when one vector is simply a multiple of another, say $(2, 4)$ and $(4, 8)$. The lines are then called **collinear** and the idea of a projection does not make sense. The plot of these vectors would be along the exact same line originating at zero, and we are thus adding no new geometric information. Therefore the vectors still consume the same space.

Also shown in Figure 4.3 are the vectors that result from $\mathbf{v}+\mathbf{u}$ and $\mathbf{v}-\mathbf{u}$ with angle θ between them. The area of the parallelogram defined by the vector $\mathbf{v}+\mathbf{u}$ shown in the figure is equal to the absolute value of the length of the orthogonal vector that results from the cross product: $\mathbf{u}\times\mathbf{v}$. This is related to the projection in the following manner: Call h the length of the line defining the projection in the figure (going from the point p to the point v). Then the parallelogram has size that is height times length: $h\|\mathbf{u}\|$ from basic geometry. Because the triangle created by the projection is a right triangle, from the trigonometry rules in Chapter 2 (page 55) we get $h = \|\mathbf{v}\| \sin\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Substituting we get $\mathbf{u} \times \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin\theta$ (from an exercise in the last chapter). Therefore the size of the parallelogram is $|\mathbf{v} + \mathbf{u}|$ since the order of the cross product could make this negative. Naturally all these principles apply in higher dimension as well.

These ideas get only slightly more complicated when discussing matrices because we can think of them as collections of vectors rather than as purely rectangular structures. The **column space** of an $i \times j$ matrix $\mathbf{X}$ consists of every possible linear combination of the j columns in $\mathbf{X}$, and the **row space** of the same matrix consists of every possible linear combination of the i rows in $\mathbf{X}$. This can be expressed more formally for the $i \times j$ matrix $\mathbf{X}$ as

- Column Space:
  all column vectors $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$,
  and scalars $s_1, s_2, \ldots, s_j$
  producing vectors $s_1\mathbf{x}_{.1} + s_2\mathbf{x}_{.2} + \cdots + s_j\mathbf{x}_{.j}$

- Row Space:
  all row vectors $\mathbf{x}_{1.}, \mathbf{x}_{2.}, \ldots, \mathbf{x}_{i.}$,
  and scalars $s_1, s_2, \ldots, s_i$
  producing vectors $s_1\mathbf{x}_{1.} + s_2\mathbf{x}_{2.} + \cdots + s_i\mathbf{x}_{i.}$,

where $\mathbf{x}_{.k}$ denotes the kth column vector of $\mathbf{x}$ and $\mathbf{x}_{k.}$ denotes the kth row vector of $\mathbf{x}$. It is now clear that the column space here consists of i-dimensional vectors and the row space consists of j-dimensional vectors. Note that the expression of space exactly fits the definition of a linear function given on page 24 in Chapter 1. This is why the field is called linear algebra. To make this process more practical, we return to our most basic example: The column space of the matrix $\left[\begin{smallmatrix}1 & 2\\3 & 4\end{smallmatrix}\right]$ includes (but is not limited to) the following resulting vectors:

$$3\begin{bmatrix}1\\3\end{bmatrix} + 1\begin{bmatrix}2\\4\end{bmatrix} = \begin{bmatrix}5\\13\end{bmatrix}, \qquad 5\begin{bmatrix}1\\3\end{bmatrix} + 0\begin{bmatrix}2\\4\end{bmatrix} = \begin{bmatrix}5\\15\end{bmatrix}.$$

★ **Example 4.1: Linear Transformation of Voter Assessments**. One difficult problem faced by analysts of survey data is that respondents often answer ordered questions based on their own interpretation of the scale. This means that an answer of "strongly agree" may have different meanings across a survey because individuals anchor against different response points, or they interpret the spacing between categories differently. Aldrich and McKelvey (1977) approached this problem by applying a linear transformation to data on the placement of presidents on a spatial issue dimension (recall the spatial representation in Figure 1.1). The key to their thinking was that while respondent i places candidate j at X_{ij} on an ordinal scale from the survey instrument, such as a 7-point "dove" to "hawk" measure, their real view was Y_{ij} along some smoother underlying metric with finer distinctions. Aldrich and McKelvey gave this hypothetical example for three voters:

Placement of Candidate Position on the Vietnam War, 1968

	Dove	1	2	3	4	5	6	7	Hawk
Voter 1		H,J,N			W			V	
Voter 2		H	J		N,V			W	
Voter 2		V		H	J,N			W	
Y									

H=Humphrey, J=Johnson, N=Nixon, W=Wallace, V=Voter

The graphic for Y above is done to suggest a noncategorical measure such as along $\Re$. To obtain a picture of this latent variable, Aldrich and McKelvey suggested a linear transformation for each voter to relate observed categorical scale to this underlying metric: $c_i + \omega_i X_{ij}$. Thus the perceived candidate positions for voter i are given by

$$Y_i = \begin{bmatrix} c_i + \omega_i X_{i1} \\ c_i + \omega_i X_{i2} \\ \vdots \\ c_i + \omega_i X_{iJ} \end{bmatrix},$$

which gives a better vector of estimates for the placement of all J candidates by respondent i because it accounts for individual-level "anchoring" by each respondent, c_i. Aldrich and McKelvey then estimated each of the values of c and ω. The value of this linear transformation is that it allows the researchers to see beyond the limitations of the categorical survey data.

Now let $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$ be a set of column vectors in $\Re^i$ (i.e., they are all length i). We say that the set of linear combinations of these vectors (in the sense above) is the **span** of that set. Furthermore, any additional vector in $\Re^i$ is spanned by these vectors if and only if it can be expressed as a linear combination of $\mathbf{x}_{.1}, \mathbf{x}_{.2}, \ldots, \mathbf{x}_{.j}$. It should be somewhat intuitive that to span $\Re^i$ here $j \geq i$ must be true. Obviously the minimal condition is $j = i$ for a set of linearly independent vectors, and in this case we then call the set a **basis**.

This brings us to a more general discussion focused on matrices rather than on vectors. A **linear space**, $\mathfrak{X}$, is the nonempty set of matrices such that remain **closed** under linear transformation:

- If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are in $\mathfrak{X}$,
- and $s_1, s_2, \ldots, s_n$ are any scalars,
- then $\mathbf{X}_{n+1} = s_1\mathbf{X}_1 + s_2\mathbf{X}_2 + \cdots + s_n\mathbf{X}_n$ is in $\mathfrak{X}$.

That is, linear combinations of matrices in the linear space have to remain in this linear space. In addition, we can define **linear subspaces** that represent some enclosed region of the full space. Obviously column and row spaces as discussed above also comprise linear spaces. Except for the pathological case where the linear space consists only of a null matrix, every linear space contains an infinite number of matrices.

Okay, so we still need some more terminology. The span of a finite set of matrices is the set of all matrices that can be achieved by a linear combination of the original matrices. This is confusing because a span is also a linear space. Where it is useful is in determining a minimal set of matrices that span a given linear space. In particular, the finite set of ***linearly independent*** matrices in a given linear space that span the linear space is called a basis for this linear space (note the word "a" here since it is not unique). That is, it cannot be made a smaller set because it would lose the ability to produce parts of the linear space, and it cannot be made a larger set because it would then no longer be linearly independent.

Let us make this more concrete with an example. A 3×3 identity matrix is clearly a basis for $\Re^3$ (the three-dimensional space of real numbers) because any three-dimensional coordinate, $[r_1, r_2, r_3]$ can be produced by multiplication of $\mathbf{I}$ by three chosen scalars. Yet, the matrices defined by $\begin{bmatrix} 1&0&0 \\ 0&0&1 \\ 0&0&1 \end{bmatrix}$ and $\begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&1 \end{bmatrix}$ do not qualify as a basis (although the second still ***spans*** $\Re^3$).

4.3 The Trace and Determinant of a Matrix

We have already noticed that the diagonals of a square matrix have special importance, particularly in the context of matrix multiplication. As mentioned in Chapter 3, a very simple way to summarize the overall magnitude of the diagonals is the **trace**. The trace of a square matrix is simply the sum of the diagonal values $\text{tr}(\mathbf{X}) = \sum_{i=1}^{k} x_{ii}$ and is usually denoted $\text{tr}(\mathbf{X})$ for the trace of square matrix $\mathbf{X}$. The trace can reveal structure in some surprising ways. For instance, an $i \times j$ matrix $\mathbf{X}$ is a zero matrix iff $\text{tr}(A'A) = 0$ (see the Exercises). In terms of calculation, the trace is probably the easiest matrix summary. For example,

$$\text{tr}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 + 4 = 5 \qquad \text{tr}\begin{pmatrix} 12 & \frac{1}{2} \\ 9 & \frac{1}{3} \end{pmatrix} = 12 + \frac{1}{3} = \frac{37}{3}.$$

One property of the trace has implications in statistics: $\text{tr}(\mathbf{X}'\mathbf{X})$ is the sum of the square of every value in the matrix $\mathbf{X}$. This is somewhat counterintuitive, so now we will do an illustrative example:

$$\text{tr}\left(\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}' \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\right) = \text{tr}\begin{pmatrix} 2 & 5 \\ 5 & 13 \end{pmatrix} = 15 = 1 + 1 + 4 + 9.$$

In general, though, the matrix trace has predictable properties:

Properties of (Conformable) Matrix Trace Operations

$\rightarrowtail$	Identity Matrix	$\text{tr}(\mathbf{I}_n) = n$
$\rightarrowtail$	Zero Matrix	$\text{tr}(\mathbf{0}) = 0$
$\rightarrowtail$	Square **J** Matrix	$\text{tr}(\mathbf{J}_n) = n$
$\rightarrowtail$	Scalar Multiplication	$\text{tr}(s\mathbf{X}) = s\text{tr}(\mathbf{X})$
$\rightarrowtail$	Matrix Addition	$\text{tr}(\mathbf{X} + \mathbf{Y}) = \text{tr}(\mathbf{X}) + \text{tr}(\mathbf{Y})$
$\rightarrowtail$	Matrix Multiplication	$\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX})$
$\rightarrowtail$	Transposition	$\text{tr}(\mathbf{X}') = \text{tr}(\mathbf{X})$

Another important, but more difficult to calculate, matrix summary is the **determinant**. The determinant uses all of the values of a square matrix to provide a summary of structure, not just the diagonal like the trace. First let us look at how to calculate the determinant for just 2×2 matrices, which is the difference in diagonal products:

$$\det(\mathbf{X}) = |\mathbf{X}| = \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} = x_{11}x_{22} - x_{12}x_{21}.$$

The notation for a determinant is expressed as $\det(\mathbf{X})$ or $|\mathbf{X}|$. Some simple numerical examples are

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (2)(3) = -2$$

$$\begin{vmatrix} 10 & \frac{1}{2} \\ 4 & 1 \end{vmatrix} = (10)(1) - \left(\frac{1}{2}\right)(4) = 8$$

$$\begin{vmatrix} 2 & 3 \\ 6 & 9 \end{vmatrix} = (2)(9) - (3)(6) = 0.$$

The last case, where the determinant is found to be zero, is an important case as we shall see shortly.

Unfortunately, calculating determinants gets much more involved with square matrices larger than 2×2. First we need to define a **submatrix**. The submatrix is simply a form achieved by deleting rows and/or columns of a matrix, leaving the remaining elements in their respective places. So for the matrix $\mathbf{X}$, notice the following submatrices whose deleted rows and columns are denoted by subscripting:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix},$$

$$\mathbf{X}_{[11]} = \begin{bmatrix} x_{22} & x_{23} & x_{24} \\ x_{32} & x_{33} & x_{34} \\ x_{42} & x_{43} & x_{44} \end{bmatrix}, \quad \mathbf{X}_{[24]} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}.$$

To generalize further for $n \times n$ matrices we first need to define the following: The ijth **minor** of $\mathbf{X}$ for x_{ij}, $|\mathbf{X}_{[ij]}|$ is the determinant of the $(n-1) \times (n-1)$ submatrix that results from taking the ith row and jth column out. Continuing, the **cofactor** of $\mathbf{X}$ for x_{ij} is the minor signed in this way: $(-1)^{i+j}|\mathbf{X}_{[ij]}|$. To exhaust the entire matrix we cycle recursively through the columns and take sums with a formula that multiplies the cofactor by the determining value:

$$\det(\mathbf{X}) = \sum_{j=1}^{n} (-1)^{i+j} x_{ij} |\mathbf{X}_{[ij]}|$$

for some constant i. This is not at all intuitive, and in fact there are some subtleties lurking in there (maybe I should have taken the ***blue*** pill). First, ***recursive*** means that the algorithm is applied iteratively through progressively smaller submatrices $\mathbf{X}_{[ij]}$. Second, this means that we lop off the top row and multiply the values across the resultant submatrices without the associated column. Actually we can pick any row or column to perform this operation, because the results will be equivalent. Rather than continue to pick apart this formula in detail, just look at the application to a 3×3 matrix:

$$\begin{vmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{vmatrix}$$

$$= (+1)x_{11} \begin{vmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{vmatrix} + (-1)x_{12} \begin{vmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{vmatrix} + (+1)x_{13} \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix}.$$

Now the problem is easy because the subsequent three determinant calculations are on 2×2 matrices. Here we picked the first row as the starting point as per the standard algorithm. In the bad old days before ubiquitous and powerful computers people who performed these calculations by hand first looked to start with rows or columns with lots of zeros because each one would mean that the subsequent contribution was automatically zero and did not need to be calculated. Using this more general process means that one has to be more careful about the alternating signs in the sum since picking the row or column to "pivot" on determines the order. For instance, here are the signs for a 7×7

matrix produced from the sign on the cofactor:

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

★ **Example 4.2: Structural Shortcuts**. There are a number of tricks for calculating the determinants of matrices of this magnitude and greater, but mostly these are relics from slide rule days. Sometimes the shortcuts are revealing about matrix structure. Ishizawa (1991), in looking at the return to scale of public inputs and its effect on the transformation curve of an economy, needed to solve a system of equations by taking the determinant of the matrix

$$\begin{bmatrix} \ell^1 & k^1 & 0 & 0 \\ \ell^2 & k^2 & 0 & 0 \\ L_w^D & L_r^D & \ell^1 & \ell^2 \\ K_w^D & K_r^D & k^1 & k^2 \end{bmatrix},$$

where these are all abbreviations for longer vectors or complex terms. We can start by being very mechanical about this:

$$\det = \ell^1 \begin{bmatrix} k^2 & 0 & 0 \\ L_r^D & \ell^1 & \ell^2 \\ K_r^D & k^1 & k^2 \end{bmatrix} - k^1 \begin{bmatrix} \ell^2 & 0 & 0 \\ L_w^D & \ell^1 & \ell^2 \\ K_w^D & k^1 & k^2 \end{bmatrix}.$$

The big help here was the two zeros on the top row that meant that we could stop our 4×4 calculations after two steps. Fortunately this trick works again because we have the same structure remaining in the 3×3 case. Let us be a bit more strategic though and define the 2×2 lower right matrix as

$\mathbf{D} = \begin{bmatrix} \ell^1 & \ell^2 \\ k^1 & k^2 \end{bmatrix}$, so that we get the neat simplification

$$\det = \ell^1 k^2 |\mathbf{D}| - k^1 \ell^2 |\mathbf{D}| = (\ell^1 k^2 - k^1 \ell^2)|\mathbf{D}| = |\mathbf{D}|^2.$$

Because of the squaring operations here this is guaranteed to be positive, which was substantively important to Ishizawa.

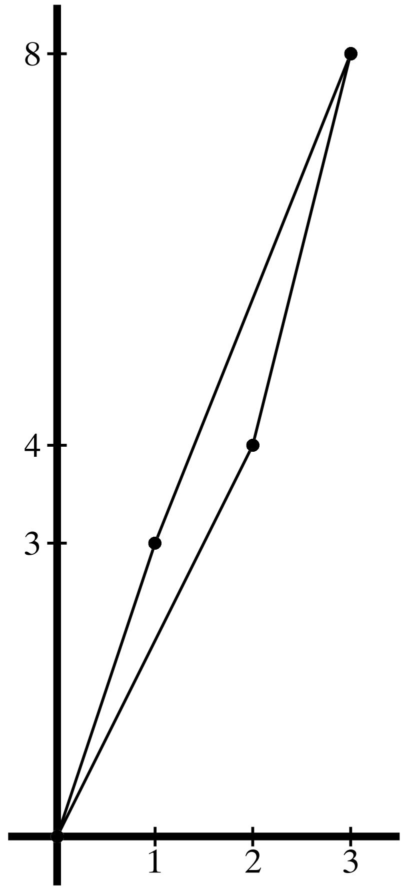
The trace and the determinant have interrelated uses and special properties as well. For instance, Kronecker products on square matrices have the properties $\text{tr}(\mathbf{X} \otimes \mathbf{Y}) = \text{tr}(\mathbf{X})\text{tr}(\mathbf{Y})$, and $|\mathbf{X} \otimes \mathbf{Y}| = |\mathbf{X}|^\ell |\mathbf{Y}|^j$ for the $j \times j$ matrix $\mathbf{X}$ and the $\ell \times \ell$ matrix $\mathbf{Y}$ (note the switching of exponents). There are some general properties of determinants to keep in mind:

Properties of $(n \times n)$ Matrix Determinants

$\rightarrowtail$	Diagonal Matrix	$\lvert\mathbf{D}\rvert = \prod_{i=1}^{n} \mathbf{D}_{ii}$
$\rightarrowtail$	(Therefore) Identity Matrix	$\lvert\mathbf{I}\rvert = 1$
$\rightarrowtail$	Triangular Matrix (upper or lower)	$\lvert\boldsymbol{\theta}\rvert = \prod_{i=1}^{n} \boldsymbol{\theta}_{ii}$
$\rightarrowtail$	Scalar Times Diagonal	$\lvert s\mathbf{D}\rvert = s^n \lvert\mathbf{D}\rvert$
$\rightarrowtail$	Transpose Property	$\lvert\mathbf{X}\rvert = \lvert\mathbf{X}'\rvert$
$\rightarrowtail$	**J** Matrix	$\lvert\mathbf{J}\rvert = 0$

It helps some people to think abstractly about the meaning of a determinant. If the columns of an $n \times n$ matrix $\mathbf{X}$ are treated as vectors, then the area of the parallelogram created by an n-dimensional space of these vectors is the absolute value of the determinant of $\mathbf{X}$, where the vectors originate at zero and the opposite point of the parallelogram is determined by the product of the columns (a cross product of these vectors, as in Section 4.2). Okay, maybe that is a bit ***too*** abstract! Now view the determinant of the 2×2 matrix $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$. The resulting parallelogram looks like the figure on the right. This figure indicates that the determinant is somehow a description of the size of a matrix in the geometric sense. Suppose that our example matrix were slightly different, say $\left[\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right]$.

Spatial Representation of a Determinant



This does not seem like a very drastic change, yet it is quite fundamentally different. It is not too hard to see that the size of the resulting parallelogram would be zero since the two column (or row) vectors would be right on top of each other in the figure, that is, collinear. We know this also almost immediately from looking at the calculation of the determinant ($ad - bc$). Here we see that two lines on top of each other produce no area. What does this mean? It means that the column dimension exceeds the offered "information" provided by this matrix form since the columns are simply scalar multiples of each other.

4.4 Matrix Rank

The ideas just described are actually more important than they might appear at first. An important characteristic of any matrix is its **rank**. Rank tells us the "space" in terms of columns or rows that a particular matrix occupies, in other words, how much unique information is held in the rows or columns of a matrix. For example, a matrix that has three columns but only two columns of unique information is given by $\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$. This is also true for the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}$, because the third column is just two times the second column and therefore has no new relational information to offer.

More specifically, when any one column of a matrix can be produced by nonzero scalar multiples of other columns added, then we say that the matrix is not **full rank** (sometimes called **short rank**). In this case at least one column is **linearly dependent**. This simply means that we can produce the relative relationships defined by this column from the other columns and it thus adds nothing to our understanding of the relationships defined by the matrix. One way to look at this is to say that the matrix in question does not "deserve" its number of columns.

Conversely, the collection of vectors determined by the columns is said to be **linearly independent** columns if the only set of scalars, $s_1, s_2, \ldots, s_j$, that satisfies $s_1\mathbf{x}_{.1} + s_2\mathbf{x}_{.2} + \cdots + s_j\mathbf{x}_{.j} = \mathbf{0}$ is a set of all zero values, $s_1 = s_2 = \ldots = s_j = 0$. This is just another way of looking at the same idea since such a condition means that we ***cannot*** reproduce one column vector from a linear combination of the others.

Actually this emphasis on columns is somewhat unwarranted because the rank of a matrix is equal to the rank of its transpose. Therefore, everything just said about columns can also be said about rows. To restate, ***the row rank of any matrix is also its column rank.*** This is a very important result and is proven in virtually every text on linear algebra. What makes this somewhat confusing is additional terminology. An $(i \times j)$ matrix is **full column rank** if its rank equals the number of columns, and it is **full row rank** if its rank equals its number of rows. Thus, if $i > j$, then the matrix can be full column rank but never full row rank. This does not necessarily mean that it ***has*** to be full column rank just because there are fewer columns than rows.

It should be clear from the example that a (square) matrix is full rank if and only if it has a nonzero determinant. This is the same thing as saying that a matrix is full rank if it is nonsingular or invertible (see Section 4.6 below). This is a handy way to calculate whether a matrix is full rank because the linear dependency within can be subtle (unlike our example above). In the next section we will explore matrix features of this type.

★ **Example 4.3: Structural Equation Models**. In their text Hanushek and Jackson (1977, Chapter 9) provided a technical overview of structural equation models where systems of equations are assumed to simultaneously affect each other to reflect endogenous social phenomena. Often these models are described in matrix terms, such as their example (p. 265)

$$\mathbf{A} = \begin{bmatrix} \gamma_{24} & 1 & \gamma_{26} & 0 & -1 \\ 0 & -1 & \gamma_{56} & 0 & 0 \\ 0 & \gamma_{65} & -1 & 0 & 0 \\ \beta_{34} & 0 & \beta_{36} & 0 & \beta_{32} \\ \beta_{44} & 0 & \beta_{46} & 0 & \beta_{42} \end{bmatrix}.$$

Without doing any calculations we can see that this matrix is of rank less than 5 because there is a column of all zeros. We can also produce this result by calculating the determinant, but that is too much trouble. Matrix determinants are not changed by multiplying the matrix by an identity in advance, multiplying by a permutation matrix in advance, or by taking transformations.

Therefore we can get a matrix

$$\mathbf{A}^* = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{24} & 1 & \gamma_{26} & 0 & -1 \\ 0 & -1 & \gamma_{56} & 0 & 0 \\ 0 & \gamma_{65} & -1 & 0 & 0 \\ \beta_{34} & 0 & \beta_{36} & 0 & \beta_{32} \\ \beta_{44} & 0 & \beta_{46} & 0 & \beta_{42} \end{bmatrix}'$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & \gamma_{65} & 0 & 0 \\ \gamma_{26} & \gamma_{56} & -1 & \beta_{36} & \beta_{46} \\ \gamma_{24} & 0 & 0 & \beta_{34} & \beta_{44} \\ -1 & 0 & 0 & \beta_{32} & \beta_{42} \end{bmatrix}$$

that is immediately identifiable as having a zero determinant by the general determinant form given on page 142 because each ith minor (the matrix that remains when the ith row and column are removed) is multiplied by the ith value on the first row.

Some rank properties are more specialized. An idempotent matrix has the property that

$$\text{rank}(\mathbf{X}) = \text{tr}(\mathbf{X}),$$

and more generally, for any square matrix with the property that $A^2 = sA$, for some scalar s

$$s\text{rank}(\mathbf{X}) = \text{tr}(\mathbf{X}).$$

To emphasize that matrix rank is a fundamental principle, we now give some standard properties related to other matrix characteristics.

Properties of Matrix Rank

$\rightarrowtail$	Transpose	$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}')$
$\rightarrowtail$	Scalar Multiplication (nonzero scalars)	$\text{rank}(s\mathbf{X}) = \text{rank}(\mathbf{X})$
$\rightarrowtail$	Matrix Addition	$\text{rank}(\mathbf{X}+\mathbf{Y}) \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$	Consecutive Blocks	$\text{rank}[\mathbf{X}\mathbf{Y}] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$ $\text{rank}\left[\begin{smallmatrix}\mathbf{X}\\ \mathbf{Y}\end{smallmatrix}\right] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$	Diagonal Blocks	$\text{rank}\left[\begin{smallmatrix}\mathbf{X} & 0\\ 0 & \mathbf{Y}\end{smallmatrix}\right] = \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$
$\rightarrowtail$	Kronecker Product	$\text{rank}(\mathbf{X}\otimes\mathbf{Y}) = \text{rank}(\mathbf{X})\text{rank}(\mathbf{Y})$

4.5 Matrix Norms

Recall that the vectors norm is a measure of length:

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

We have seen matrix "size" as described by the trace, determinant, and rank. Additionally, we can describe matrices by norming, but matrix norms are a little bit more involved than the vector norms we saw before. There are two general types, the **trace norm** (sometimes called the Euclidean norm or the Frobenius norm):

$$\|\mathbf{X}\|_F = \left[\sum_i\sum_j |x_{ij}|^2\right]^{\frac{1}{2}}$$

(the square root of the sum of each element squared), and the p-norm:

$$\|\mathbf{X}\|_p = \max_{\|\mathbf{v}\|_p} \|\mathbf{X}\mathbf{v}\|_p,$$

which is defined with regard to the unit vector $\mathbf{v}$ whose length is equal to the number of columns in $\mathbf{X}$. For $p = 1$ and an $I \times J$ matrix, this reduces to summing absolute values down columns and taking the maximum:

$$\|\mathbf{X}\|_1 = \max_J \sum_{i=1}^{I} |x_{ij}|.$$

Conversely, the infinity version of the matrix p-norm sums across rows before taking the maximum:

$$\|\mathbf{X}\|_\infty = \max_I \sum_{j=1}^{J} |x_{ij}|.$$

Like the infinity form of the vector norm, this is somewhat unintuitive because there is no apparent use of a limit. There are some interesting properties of matrix norms:

Properties of Matrix Norms, Size ($i \times j$)

$\rightarrowtail$	Constant Multiplication	$\|k\mathbf{X}\| = |k|\|\mathbf{X}\|$
$\rightarrowtail$	Addition	$\|\mathbf{X}+\mathbf{Y}\| \leq \|\mathbf{X}\| + \|\mathbf{Y}\|$
$\rightarrowtail$	Vector Multiplication	$\|\mathbf{X}\mathbf{v}\|_p \leq \|\mathbf{X}\|_p\|\mathbf{v}\|_p$
$\rightarrowtail$	Norm Relation	$\|\mathbf{X}\|_2 \leq \|\mathbf{X}\|_F \leq \sqrt{j}\|\mathbf{X}\|_2$
$\rightarrowtail$	Unit Vector Relation	$\mathbf{X}'\mathbf{X}\mathbf{v} = (\|\mathbf{X}\|_2)^2\mathbf{v}$
$\rightarrowtail$	P-norm Relation	$\|\mathbf{X}\|_2 \leq \sqrt{\|\mathbf{X}\|_1\ \|\mathbf{X}\|_\infty}$
$\rightarrowtail$	Schwarz Inequality	$|\mathbf{X}\cdot\mathbf{Y}| \leq \|\mathbf{X}\|\ \|\mathbf{Y}\|$, where $|\mathbf{X}\cdot\mathbf{Y}| = \text{tr}(\mathbf{X}'\mathbf{Y})$

★ **Example 4.4: Matrix Norm Sum Inequality**. Given matrices

$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 5 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} -1 & -2 \\ 3 & 4 \end{bmatrix},$$

observe that

$$\begin{array}{l|l} \|\mathbf{X}+\mathbf{Y}\|_\infty & \|\mathbf{X}\|_\infty + \|\mathbf{Y}\|_\infty \\ \left\| \begin{bmatrix} 2 & 0 \\ 8 & 5 \end{bmatrix} \right\|_\infty & \max(5,6) + \max(3,7) \\ \max(2,13) & 13, \end{array}$$

showing the second property above.

★ **Example 4.5: Schwarz Inequality for Matrices**. Using the same $\mathbf{X}$ and $\mathbf{Y}$ matrices and the $p=1$ norm, observe that

$$\begin{array}{c|l} |\mathbf{X}\cdot\mathbf{Y}| & \|\mathbf{X}\|_1 \, \|\mathbf{Y}\|_1 \\ (12)+(0) & \max(8,3)\cdot\max(4,6) \end{array}$$

showing that the inequality holds: $12 < 48$. This is a neat property because it shows a relationship between the trace and matrix norm.

4.6 Matrix Inversion

Just like scalars have inverses, some ***square*** matrices have a **matrix inverse**. The inverse of a matrix $\mathbf{X}$ is denoted $\mathbf{X}^{-1}$ and defined by the property

$$\mathbf{X}\mathbf{X}^{-1} = \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}.$$

That is, when a matrix is pre-multiplied or post-multiplied by its inverse the result is an identity matrix of the same size. For example, consider the following matrix and its inverse:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} -2.0 & 1.0 \\ 1.5 & -0.5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Not all square matrices are invertible. A singular matrix cannot be inverted, and often "singular" and "noninvertible" are used as synonyms. Usually matrix inverses are calculated by computer software because it is quite time-consuming with reasonably large matrices. However, there is a very nice trick for immediately inverting 2×2 matrices, which is given by

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

$$\mathbf{X}^{-1} = \det(\mathbf{X})^{-1} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix}.$$

A matrix inverse is unique: There is only one matrix that meets the multiplicative condition above for a nonsingular square matrix.

For inverting larger matrices there is a process based on **Gauss-Jordan elimination** that makes use of linear programming to invert the matrix. Although matrix inversion would normally be done courtesy of software for nearly all problems in the social sciences, the process of Gauss-Jordan elimination is a revealing insight into inversion because it highlights the "inverse" aspect with the role of the identity matrix as the linear algebra equivalent of 1. Start with the matrix of interest partitioned next to the identity matrix and allow the following operations:

- Any row may be multiplied or divided by a scalar.
- Any two rows may be switched.
- Any row may be multiplied or divided by a scalar and then added to another row. Note: This operation does not change the original row; its multiple is used but not saved.

Of course the goal of these operations has not yet been given. We want to iteratively apply these steps until the identity matrix on the right-hand side is on the left-hand side. So the operations are done with the intent of zeroing out the off-diagonals on the left matrix of the partition and then dividing to obtain 1's on the diagonal. During this process we do not care about what results on the right-hand side until the end, when this is known to be the inverse of the original matrix.

Let's perform this process on a 3×3 matrix:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 1 & 8 & 9 & 0 & 0 & 1 \end{array}\right].$$

Now multiply the first row by -4, adding it to the second row, and multiply the first row by -1, adding it to the third row:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 6 & 6 & -1 & 0 & 1 \end{array}\right].$$

Multiply the second row by $\frac{1}{2}$, adding it to the first row, and simply add this same row to the third row:

$$\left[\begin{array}{ccc|ccc} 1 & \frac{1}{2} & 0 & -1 & \frac{1}{2} & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 3 & 0 & -5 & 1 & 1 \end{array}\right].$$

Multiply the third row by $-\frac{1}{6}$, adding it to the first row, and add the third row (un)multiplied to the second row:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ 0 & 0 & -6 & -9 & 2 & 1 \\ 0 & 3 & 0 & -5 & 1 & 1 \end{array}\right].$$

Finally, just divide the second row by -6 and the third row by -3, and then switch their places:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ 0 & 1 & 0 & -\frac{5}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{3} & -\frac{1}{6} \end{array}\right],$$

thus completing the operation. This process also highlights the fact that matrices are representations of linear equations. The operations we performed are linear transformations, just like those discussed at the beginning of this chapter.

We already know that singular matrices cannot be inverted, but consider the described inversion process applied to an obvious case:

$$\left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right].$$

It is easy to see that there is nothing that can be done to put a nonzero value in the second column of the matrix to the left of the partition. In this way the Gauss-Jordan process helps to illustrate a theoretical concept.

Most of the properties of matrix inversion are predictable (the last property listed relies on the fact that the product of invertible matrices is always itself invertible):

Properties of $n \times n$ Nonsingular Matrix Inverse

$\rightarrowtail$	Diagonal Matrix	$\mathbf{D}^{-1}$ has diagonal values $1/d_{ii}$ and zeros elsewhere.
$\rightarrowtail$	(Therefore) Identity Matrix	$\mathbf{I}^{-1} = \mathbf{I}$
$\rightarrowtail$	(Non-zero) Scalar Multiplication	$(s\mathbf{X})^{-1} = \frac{1}{s}\mathbf{X}^{-1}$
$\rightarrowtail$	Iterative Inverse	$(\mathbf{X}^{-1})^{-1} = \mathbf{X}$
$\rightarrowtail$	Exponents	$\mathbf{X}^{-n} = (\mathbf{X}^{n})^{-1}$
$\rightarrowtail$	Multiplicative Property	$(\mathbf{XY})^{-1} = \mathbf{Y}^{-1}\mathbf{X}^{-1}$
$\rightarrowtail$	Transpose Property	$(\mathbf{X}')^{-1} = (\mathbf{X}^{-1})'$
$\rightarrowtail$	Orthogonal Property	If $\mathbf{X}$ is orthogonal, then $\mathbf{X}^{-1} = \mathbf{X}'$
$\rightarrowtail$	Determinant	$\lvert\mathbf{X}^{-1}\rvert = 1/\lvert\mathbf{X}\rvert$

★ **Example 4.6: Calculating Regression Parameters**. The classic "ordinary least squares" method for obtaining regression parameters proceeds as follows. Suppose that $\mathbf{y}$ is the outcome variable of interest and $\mathbf{X}$ is a matrix of explanatory variables with a leading column of 1's. What we would like is the vector $\hat{\mathbf{b}}$ that contains the intercept and the regression slope, which is calculated by the equation $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, which might have seemed hard before this point in the chapter. What we need to do then is just a series of multiplications, one inverse, and two transposes.

To make the example more informative, we can look at some actual data with two variables of interest (even though we could just do this in scalar algebra since it is just a bivariate problem). Governments often worry about the economic condition of senior citizens for political and social reasons. Typically in a large industrialized society, a substantial portion of these people obtain the bulk of their income from government pensions. One important question is whether there is enough support through these payments to provide subsistence above the poverty rate. To see if this is a concern, the European Union (EU) looked at this question in 1998 for the (then) 15 member countries with two variables: (1) the median (EU standardized) income of individuals age 65 and older as a percentage of the population age 0–64, and (2) the percentage with income below 60% of the median (EU standardized) income of the national population. The data from the European Household Community Panel Survey are

Nation	Relative Income	Poverty Rate
Netherlands	93.00	7.00
Luxembourg	99.00	8.00
Sweden	83.00	8.00
Germany	97.00	11.00
Italy	96.00	14.00
Spain	91.00	16.00
Finland	78.00	17.00
France	90.00	19.00
United.Kingdom	78.00	21.00
Belgium	76.00	22.00
Austria	84.00	24.00
Denmark	68.00	31.00
Portugal	76.00	33.00
Greece	74.00	33.00
Ireland	69.00	34.00

So the $\mathbf{y}$ vector is the second column of the table and the $\mathbf{X}$ matrix is the first column along with the leading column of 1's added to account for the intercept (also called the constant, which explains the 1's). The first quantity that we want to calculate is

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 15.00 & 1252 \\ 1252 & 105982 \end{bmatrix},$$

which has the inverse

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 4.76838 & -0.05633 \\ -0.05633 & 0.00067 \end{bmatrix}.$$

So the final calculation is

$$\begin{bmatrix} 4.76838 & -0.05633 \\ -0.05633 & 0.00067 \end{bmatrix} \begin{bmatrix} 1 & 93 \\ 1 & 99 \\ 1 & 83 \\ 1 & 97 \\ 1 & 96 \\ 1 & 91 \\ 1 & 78 \\ 1 & 90 \\ 1 & 78 \\ 1 & 76 \\ 1 & 84 \\ 1 & 68 \\ 1 & 76 \\ 1 & 74 \\ 1 & 69 \end{bmatrix}' \begin{bmatrix} 7 \\ 8 \\ 8 \\ 11 \\ 14 \\ 16 \\ 17 \\ 19 \\ 21 \\ 22 \\ 24 \\ 31 \\ 33 \\ 33 \\ 34 \end{bmatrix} = \begin{bmatrix} 83.69279 \\ -0.76469 \end{bmatrix}$$

These results are shown in Figure 4.4 for the 15 EU countries of the time, with a line for the estimated underlying trend that has a slope of $m = -0.77$ (rounded) and an intercept at $b = 84$ (also rounded). What does this mean? It means that for a one-unit positive change (say from 92 to 93) in over-65 relative income, there will be an ***expected*** change in over-65 poverty rate of -0.77 (i.e., a reduction). This is depicted in Figure 4.4.

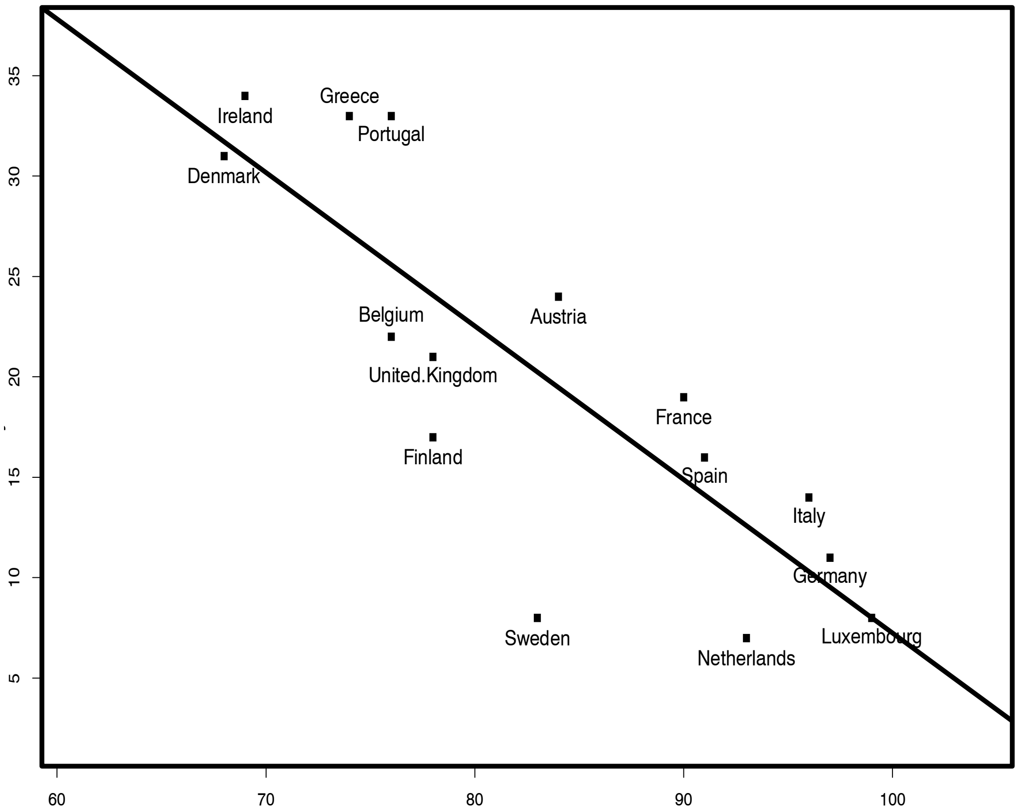
Once one understands linear regression in matrix notation, it is much easier to see what is happening. For instance, if there were a ***second*** explanatory variable (there are many more than one in social science models), then it would simply be an addition column of the $\mathbf{X}$ matrix and all the calculations would proceed exactly as we have done here.

4.7 Linear Systems of Equations

A basic and common problem in applied mathematics is the search for a solution, $\mathbf{x}$, to the system of simultaneous linear equations defined by

$$\mathbf{A}\mathbf{x} = \mathbf{y},$$

Fig. 4.4. Relative Income and Senior Poverty, EU Countries



where $\mathbf{A} \in \Re^{p \times q}$, $\mathbf{x} \in \Re^{q}$, and $\mathbf{y} \in \Re^{p}$. If the matrix $\mathbf{A}$ is invertible, then there exists a unique, easy-to-find, solution vector $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ satisfying $\mathbf{Ax} = \mathbf{y}$. Note that this shows the usefulness of a matrix inverse. However, if the system of linear equations in $\mathbf{Ax} = \mathbf{y}$ is not ***consistent***, then there exists no solution. Consistency simply means that if a linear relationship exists in the rows of $\mathbf{A}$, it must also exist in the corresponding rows of $\mathbf{y}$. For example, the following simple system of linear equations is consistent:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

because the second row is two times the first across $(\mathbf{x}|\mathbf{y})$. This implies that $\mathbf{y}$ is contained in the linear span of the columns (range) of $\mathbf{A}$, denoted as $\mathbf{y} \in R(\mathbf{A})$. Recall that a set of linearly independent vectors (i.e., the columns here) that span a vector subspace is called a basis of that subspace. Conversely, the following system of linear equations is not consistent:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 7 \end{bmatrix},$$

because there is no solution for $\mathbf{x}$ that satisfies both rows. In the notation above this is denoted $\mathbf{y} \notin R(\mathbf{A})$, and it provides no further use without modification of the original problem. It is worth noting, for purposes of the discussion below, that if $\mathbf{A}^{-1}$ exists, then $\mathbf{Ax} = \mathbf{y}$ is always consistent because there exist no linear relationships in the rows of $\mathbf{A}$ that must be satisfied in $\mathbf{y}$. The inconsistent case is the more common ***statistically*** in that a solution that minimizes the squared sum of the inconsistencies is typically applied (ordinary least squares).

In addition to the possibilities of the general system of equations $\mathbf{Ax} = \mathbf{y}$ having a unique solution and no solution, this arbitrary system of equations can also have an infinite number of solutions. In fact, the matrix $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ above is such a case. For example, we could solve to obtain $\mathbf{x} = (1, 1)'$, $\mathbf{x} = (-1, 2)'$, $\mathbf{x} = (5, -1)'$, and so on. This occurs when the $\mathbf{A}$ matrix is singular: rank($\mathbf{A}$) = dimension$(R(\mathbf{A})) < q$. When the $\mathbf{A}$ matrix is singular at least one column vector is a linear combination of the others, and the matrix therefore contains redundant information. In other words, there are $q' < q$ independent column vectors in $\mathbf{A}$.

★ **Example 4.7: Solving Systems of Equations by Inversion**. Consider the system of equations

$$2x_1 - 3x_2 = 4$$
$$5x_1 + 5x_2 = 3,$$

where $\mathbf{x} = [x_1, x_2]$, $\mathbf{y} = [4, 3]'$, and $\mathbf{A} = \left[\begin{smallmatrix} 2 & -3 \\ 5 & 5 \end{smallmatrix}\right]$. First invert $\mathbf{A}$:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2 & -3 \\ 5 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 & 0.12 \\ -0.2 & 0.08 \end{bmatrix}.$$

Then, to solve for $\mathbf{x}$ we simply need to multiply this inverse by $\mathbf{y}$:

$$\mathbf{A}^{-1}\mathbf{y} = \begin{bmatrix} 0.2 & 0.12 \\ -0.2 & 0.08 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.16 \\ -0.56 \end{bmatrix},$$

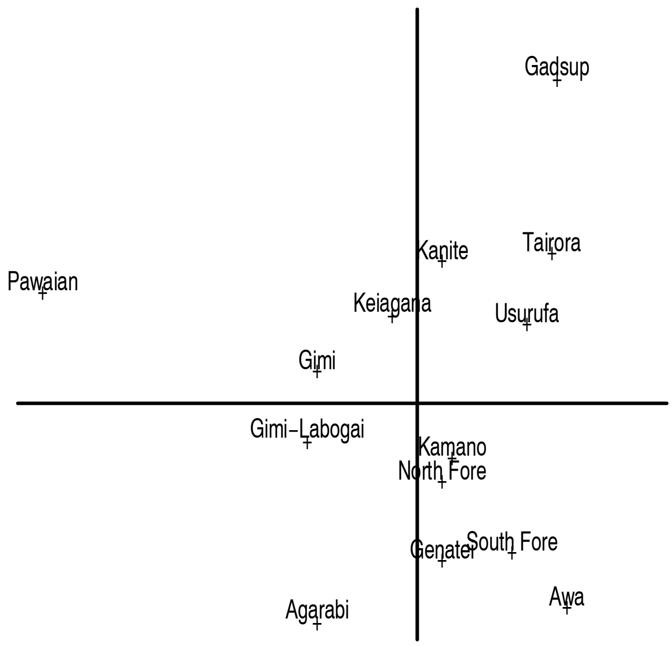
meaning that $\mathbf{x}_1 = 0.16$ and $\mathbf{x}_2 = -0.56$.

4.8 Eigen-Analysis of Matrices

We start this section with a brief motivation. Apparently a single original population undergoes genetic differentiation once it is dispersed into new geographic regions. Furthermore, it is interesting anthropologically to compare the rate of this genetic change with changes in nongenetic traits such as language, culture, and use of technology. Sorenson and Kenmore (1974) explored the genetic drift of proto-agricultural people in the Eastern Highlands of New Guinea with the idea that changes in horticulture and mountainous geography both determined patterns of dispersion. This is an interesting study because it uses biological evidence (nine alternative forms of a gene) to make claims about the relatedness of groups that are geographically distinct but similar ethnohistorically and linguistically. The raw genetic information can be summarized in a large matrix, but the information in this form is not really the primary interest. To see differences and similarities Sorenson and Kenmore transformed these variables into just two individual factors (new composite variables) that appear to explain the bulk of the genetic variation.

Once that is done it is easy to graph the groups in a single plot and then look at similarities geometrically. This useful result is shown in the figure at right, where we see the placement of these linguistic groups according to the similarity in blood-group genetics. The tool they used for turning the large multidimensional matrix of unwieldy data into an intuitive two-dimensional structure was eigenanalysis.

LINGUISTIC GROUPS GENETICALLY



A useful and theoretically important feature of a given square matrix is the set of **eigenvalues** associated with this matrix. Every $p \times p$ matrix $\mathbf{X}$ has p scalar values, $\lambda_i, i = 1, \ldots, p$, such that

$$\mathbf{X}\mathbf{h}_i = \lambda_i \mathbf{h}_i$$

for some corresponding vector $\mathbf{h}_i$. In this decomposition, λ_i is called an eigenvalue of $\mathbf{X}$ and h_i is called an **eigenvector** of $\mathbf{X}$. These eigenvalues show important structural features of the matrix. Confusingly, these are also called the **characteristic roots** and **characteristic vectors** of $\mathbf{X}$, and the process is also called **spectral decomposition**.

The expression above can also be rewritten to produce the **characteristic**

equation. Start with the algebraic rearrangement

$$(\mathbf{X} - \lambda_i \mathbf{I})\mathbf{h}_i = \mathbf{0}.$$

If the $p \times p$ matrix in the parentheses has a zero determinant, then there exist eigenvalues that are solutions to the equation:

$$|\mathbf{X} - \lambda_i \mathbf{I}| = \mathbf{0}.$$

★ **Example 4.8:** **Basic Eigenanalysis**. A symmetric matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \begin{bmatrix} 1.000 & 0.880 & 0.619 \\ 0.880 & 1.000 & 0.716 \\ 0.619 & 0.716 & 1.000 \end{bmatrix}.$$

The eigenvalues and eigenvectors are found by solving the characteristic equation $|\mathbf{X} - \lambda \mathbf{I}| = 0$. This produces the matrix

$$\lambda \mathbf{I} = \begin{bmatrix} 2.482 & 0.00 & 0.000 \\ 0.000 & 0.41 & 0.000 \\ 0.000 & 0.00 & 0.108 \end{bmatrix}$$

from which we take the eigenvalues from the diagonal. Note the descending order. To see the mechanics of this process more clearly, consider finding the eigenvalues of

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

To do this we expand and solve the determinant of the characteristic equation:

$$|\mathbf{Y} - \lambda \mathbf{I}| = (3 - \lambda)(0 - \lambda) - (-2) = \lambda^2 - 3\lambda + 2$$

and the only solutions to this quadratic expression are $\lambda_1 = 1, \lambda_2 = 2$. In fact, for a $p \times p$ matrix, the resulting characteristic equation will be a polynomial of order p. This is why we had a quadratic expression here.

Unfortunately, the eigenvalues that result from the characteristic equation can be zero, repeated (nonunique) values, or even complex numbers. However, all symmetric matrices like the 3×3 example above are guaranteed to have real-valued eigenvalues.

Eigenvalues and eigenvectors are associated. That is, for each eigenvector of a given matrix $\mathbf{X}$ there is exactly one corresponding eigenvalue such that

$$\lambda = \frac{\mathbf{h}'\mathbf{X}\mathbf{h}}{\mathbf{h}'\mathbf{h}}.$$

This uniqueness, however, is asymmetric. For each eigenvalue of the matrix there is an infinite number of eigenvectors, all determined by scalar multiplication: If $\mathbf{h}$ is an eigenvector corresponding to the eigenvalue λ, then $s\mathbf{h}$ is also an eigenvector corresponding to this same eigenvalue where s is any nonzero scalar.

There are many interesting matrix properties related to eigenvalues. For instance, the number of nonzero eigenvalues is the rank of the $\mathbf{X}$, the sum of the eigenvalues is the trace of $\mathbf{X}$, and the product of the eigenvalues is the determinant of $\mathbf{X}$. From these principles it follows immediately that a matrix is singular if and only if it has a zero eigenvalue, and the rank of the matrix is the number of nonzero eigenvalues.

Properties of Eigenvalues for a Nonsingular ($n \times n$) Matrix

$\rightarrowtail$	Inverse Property	If λ_i is an eigenvalue of $\mathbf{X}$, then $\frac{1}{\lambda_i}$ is an eigenvalue of $\mathbf{X}^{-1}$
$\rightarrowtail$	Transpose Property	$\mathbf{X}$ and $\mathbf{X}'$ have the same eigenvalues
$\rightarrowtail$	Identity Matrix	For $\mathbf{I}$, $\sum \lambda_i = n$
$\rightarrowtail$	Exponentiation	If λ_i is an eigenvalue of $\mathbf{X}$, thenλ_i^k is an eigenvalue of $\mathbf{X}^k$ and k a positive integer

It is also true that if there are no zero-value eigenvectors, then the eigenvalues determine a basis for the space determined by the size of the matrix ($\Re^2$, $\Re^3$, etc.). Even more interestingly, symmetric nonsingular matrices have eigenvectors that are perpendicular (see the Exercises).

A notion related to eigenvalues is matrix conditioning. For a symmetric definite matrix, the ratio of the largest eigenvalue to the smallest eigenvalue is the **condition number**. If this number is large, then we say that the matrix is "ill-conditioned," and it usually has poor properties. For example, if the matrix is nearly singular (but not quite), then the smallest eigenvalue will be close to zero and the ratio will be large for any reasonable value of the largest eigenvalue. As an example of this problem, in the use of matrix inversion to solve systems of linear equations, an ill-conditioned $\mathbf{A}$ matrix means that small changes in $\mathbf{A}$ will produce large changes in $\mathbf{A}^{-1}$ and therefore the calculation of $\mathbf{x}$ will differ dramatically.

★ **Example 4.9: Analyzing Social Mobility with Eigens**. Duncan (1966) analyzed social mobility between categories of employment (from the 1962 Current Population Survey) to produce probabilities for blacks and whites [also analyzed in McFarland (1981) from which this discussion is derived]. This well-known finding is summarized in two *transition* matrices, indicating probabilities for changing between *higher white collar*, *lower white collar*, *higher manual*, *lower manual*, and *farm*:

$$B = \begin{bmatrix} 0.112 & 0.105 & 0.210 & 0.573 & 0.000 \\ 0.156 & 0.098 & 0.000 & 0.745 & 0.000 \\ 0.094 & 0.073 & 0.120 & 0.684 & 0.030 \\ 0.087 & 0.077 & 0.126 & 0.691 & 0.020 \\ 0.035 & 0.034 & 0.072 & 0.676 & 0.183 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.576 & 0.162 & 0.122 & 0.126 & 0.014 \\ 0.485 & 0.197 & 0.145 & 0.157 & 0.016 \\ 0.303 & 0.127 & 0.301 & 0.259 & 0.011 \\ 0.229 & 0.124 & 0.242 & 0.387 & 0.018 \\ 0.178 & 0.076 & 0.214 & 0.311 & 0.221 \end{bmatrix},$$

where the rows and columns are in the order of employment categories given. So, for instance, 0.576 in the first row and first column of the W matrix means that we expect 57.6% of the children of white higher white collar workers will themselves become higher white collar workers. Contrastingly, 0.573 in the first row and fourth column of the B matrix means that we expect 57.4% of the children of black lower manual workers to become lower manual workers themselves.

A lot can be learned by staring at these matrices for some time, but what tools will let us understand long-run trends built into the data? Since these are transition probabilities, we could multiply one of these matrices to itself a large number of times as a simulation of future events (this is actually the topic of Chapter 9). It might be more convenient for answering simple questions to use eigenanalysis to pull structure out of the matrix instead.

It turns out that the eigenvector produced from $\mathbf{X}\mathbf{h}_i = \lambda_i \mathbf{h}_i$ is the **right eigenvector** because it sits on the right-hand side of $\mathbf{X}$ here. This is the default, so when an eigenvector is referenced without any qualifiers, the form derived above is the appropriate one. However, there is also the less-commonly used **left eigenvector** produced from $\mathbf{h}_i\mathbf{X} = \lambda_i \mathbf{h}_i$ and so-named for the obvious reason. If $\mathbf{X}$ is a symmetric matrix, then the two vectors are identical (the eigenvalues are the same in either case). If $\mathbf{X}$ is not symmetrical, they differ, but the left eigenvector can be produced from using the transpose: $\mathbf{X}'\mathbf{h}_i = \lambda_i \mathbf{h}_i$. The ***spectral component*** corresponding to the ith eigenvalue is the square matrix produced from the cross product of the right and left eigenvectors over the dot product of the right and left eigenvectors:

$$S_i = h_{i,\text{right}} \times h_{i,\text{left}}/h_{i,\text{right}} \cdot h_{i,\text{left}}.$$

This spectral decomposition into constituent components by eigenvalues is especially revealing for probability matrices like the two above, where the rows necessarily sum to 1.

Because of the probability structure of these matrices, the first eigenvalue is always 1. The associated spectral components are

$$B = \begin{bmatrix} 0.09448605 & 0.07980742 & 0.1218223 & 0.6819610 & 0.02114880 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.4293069 & 0.1509444 & 0.1862090 & 0.2148500 & 0.01840510. \end{bmatrix},$$

where only a single row of this 5×5 matrix is given here because all rows are identical (a result of $\lambda_1 = 1$). The spectral values corresponding to the first eigenvalue give the long-run (stable) proportions implied by the matrix probabilities. That is, if conditions do not change, these will be the eventual population proportions. So if the mobility trends persevere, eventually a little over two-thirds of the black population will be in lower manual occupations, and less than 10% will be in each of the white collar occupational categories (keep in mind that Duncan collected the data before the zenith of the civil rights movement). In contrast, for whites, about 40% will be in the higher white collar category with 15–20% in each of the other nonfarm occupational groups.

Subsequent spectral components from declining eigenvalues give weighted propensities for movement between individual matrix categories. The second eigenvalue produces the most important indicator, followed by the third, and so on. The second spectral components corresponding to the second eigenvalues $\lambda_{2,\text{black}} = 0.177676$, $\lambda_{2,\text{white}} = 0.348045$ are

$$B = \begin{bmatrix} 0.063066 & 0.043929 & 0.034644 & -0.019449 & -0.122154 \\ 0.103881 & 0.072359 & 0.057065 & -0.032037 & -0.201211 \\ -0.026499 & -0.018458 & -0.014557 & 0.008172 & 0.051327 \\ -0.002096 & -0.001460 & -0.001151 & 0.000646 & 0.004059 \\ -0.453545 & -0.315919 & -0.249145 & 0.139871 & 0.878486 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.409172 & 0.055125 & -0.187845 & -0.273221 & -0.002943 \\ 0.244645 & 0.032960 & -0.112313 & -0.163360 & -0.001759 \\ -0.3195779 & -0.043055 & 0.146714 & 0.213396 & 0.002298 \\ -0.6018242 & -0.081080 & 0.276289 & 0.401864 & 0.004328 \\ -1.2919141 & -0.174052 & 0.593099 & 0.862666 & 0.009292 \end{bmatrix}$$

Notice that the full matrix is given here because the rows now differ. McFarland noticed the structure highlighted here with the boxes containing positive values. For blacks there is a tendency for white collar status and higher manual to be self-reinforcing: Once landed in the upper left 2×3 submatrix, there is a tendency to remain and negative influences on leaving. The same phenomenon applies for blacks to manual/farm labor: Once there it is more difficult to leave. For whites the phenomenon is the same, except this barrier effect puts higher manual in the less desirable block. This suggests a racial differentiation with regard to higher manual occupations.

4.9 Quadratic Forms and Descriptions

This section describes a general attribute known as ***definiteness***, although this term means nothing on its own. The central question is what properties does an $n \times n$ matrix $\mathbf{X}$ possess when pre- and post-multiplied by a conformable nonzero vector $\mathbf{y} \in \Re^n$. The quadratic form of the matrix $\mathbf{X}$ is given by

$$\mathbf{y}'\mathbf{X}\mathbf{y} = s,$$

where the result is some scalar, s. If $s = 0$ for every possible vector $\mathbf{y}$, then $\mathbf{X}$ can only be the null matrix. But we are really interested in more nuanced properties. The following table gives the standard descriptions.

Properties of the Quadratic, $\mathbf{y}$ Non-Null

<u>Non-Negative Definite:</u>			
	$\rightarrowtail$	positive definite	$\mathbf{y}'\mathbf{X}\mathbf{y} > 0$
	$\rightarrowtail$	positive semidefinite	$\mathbf{y}'\mathbf{X}\mathbf{y} \geq 0$
<u>Non-Positive Definite:</u>			
	$\rightarrowtail$	negative definite	$\mathbf{y}'\mathbf{X}\mathbf{y} < 0$
	$\rightarrowtail$	negative semidefinite	$\mathbf{y}'\mathbf{X}\mathbf{y} \leq 0$

We can also say that $\mathbf{X}$ is **indefinite** if it is neither nonnegative definite nor nonpositive definite. The big result is worth stating with emphasis:

A positive definite matrix is always nonsingular.

Furthermore, a positive definite matrix is therefore invertible and the resulting inverse will also be positive definite. Positive semidefinite matrices are sometimes singular and sometimes not. If such a matrix is nonsingular, then its inverse is also nonsingular.

One theme that we keep returning to is the importance of the diagonal of a matrix. It turns out that every diagonal element of a positive definite matrix is positive, and every element of a negative definite matrix is negative. In addition, every element of a positive semidefinite matrix is nonnegative, and every element of a negative semidefinite matrix is nonpositive. This makes sense because we can switch properties between "negativeness" and "positiveness" by simply multiplying the matrix by -1.

★ **Example 4.10: LDU Decomposition**. In the last chapter we learned about LU decomposition as a way to triangularize matrices. The vague caveat at the time was that this could be done to "many" matrices. The condition, unstated at the time, is that the matrix must be nonsingular. We now know what that means, so it is now clear when LU decomposition is possible. More generally, though, ***any*** $p \times q$ matrix can be decomposed as follows:

$$\underset{(p\times q)}{\mathbf{A}} = \underset{(p\times p)}{\mathbf{L}}\underset{(p\times q)}{\mathbb{D}}\underset{(q\times q)}{\mathbf{U}}, \qquad \text{where} \qquad \mathbb{D} = \begin{bmatrix} \mathbf{D}_{r\times r} & 0 \\ 0 & 0 \end{bmatrix},$$

where $\mathbf{L}$ (lower triangular) and $\mathbf{U}$ (upper triangular) are nonsingular (even given a singular matrix $\mathbf{A}$). The diagonal matrix $\mathbf{D}_{r\times r}$ is unique and has dimension and rank r that corresponds to the rank of $\mathbf{A}$. If $\mathbf{A}$ is positive definite, and symmetric, then $\mathbf{D}_{r\times r} = \mathbb{D}$ (i.e., $r = q$) and $\mathbf{A} = \mathbf{L}\mathbb{D}\mathbf{L}'$ with unique $\mathbf{L}$.

For example, consider the LDU decomposition of the 3×3 unsymmetric, positive definite matrix A:

$$\mathbf{A} = \begin{bmatrix} 140 & 160 & 200 \\ 280 & 860 & 1060 \\ 420 & 1155 & 2145 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} 20 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 10 \end{bmatrix} \begin{bmatrix} 7 & 8 & 10 \\ 0 & 9 & 11 \\ 0 & 0 & 12 \end{bmatrix}.$$

Now look at the symmetric, positive definite matrix and its LDL' decomposition:

$$\mathbf{B} = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 21 & 21 \\ 5 & 21 & 30 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{bmatrix}',$$

which shows the symmetric principle above.

4.10 New Terminology

basis, 138

characteristic equation, 162

characteristic root, 161

characteristic vector, 161

closed, 410

cofactor, 141

collinear, 135

column space, 136

condition number, 164

Cramer's rule, 177

determinant, 140

eigenvalue, 161

eigenvector, 161

full column rank, 146

full rank, 146

full row rank, 146

Gauss-Jordan elimination, 152

indefinite, 168

left eigenvector, 165

linear space, 138

linear subspace, 138

linearly dependent, 146

linearly independent, 146

matrix inverse, 151

minor, 141

orthogonal projections, 134

projections, 134

right eigenvector, 165

row space, 136

short rank, 146

span, 138

spectral decomposition, 161

submatrix, 141

trace, 139

trace norm, 149

Exercises

4.1 For the matrix $\left[\begin{smallmatrix} 3 & 5 \\ 2 & 0 \end{smallmatrix}\right]$, show that the following vectors are or are not in the column space:

$$\begin{bmatrix} 11 \\ 4 \end{bmatrix}, \qquad \begin{bmatrix} 11 \\ 5 \end{bmatrix}.$$

4.2 Demonstrate that two orthogonal vectors have zero-length projections. Use unit vectors to make this easier.

4.3 Obtain the determinant and trace of the following matrix. Think about tricks to make the calculations easier.

$$\begin{bmatrix} 6 & 6 & 1 & 0 \\ 0 & 4 & 0 & 1 \\ 4 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

4.4 Prove that $\text{tr}(\mathbf{XY}) \neq \text{tr}(\mathbf{X})\text{tr}(\mathbf{Y})$, except for special cases.

4.5 In their formal study of models of group interaction, Bonacich and Bailey (1971) looked at linear and nonlinear systems of equations (their interest was in models that include factors such as free time, psychological compatibility, friendliness, and common interests). One of their conditions for a stable system was that the determinant of the matrix

$$\begin{pmatrix} -r & a & 0 \\ 0 & -r & a \\ 1 & 0 & -r \end{pmatrix}$$

must have a positive determinant for values of r and a. What is the arithmetic relationship that must exist for this to be true.

4.6 Find the eigenvalues of $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix}$.

4.7 Calculate $|B|$, $\text{tr}(B)$, and B^{-1} given $B = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

4.8 (Hanushek and Jackson 1977). Given the matrices

$$\mathbf{Y} = \begin{bmatrix} 10 \\ 13 \\ 7 \\ 5 \\ 2 \\ 6 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 1 & 4 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 5 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \\ 0 & 1 & 4 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 5 \end{bmatrix},$$

calculate $b_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$ and $b_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y}$. How are these vectors different? Similar?

4.9 Prove that the following matrix is or is not orthogonal:

$$\mathbf{B} = \begin{bmatrix} 1/3 & 2\sqrt{2}/3 & 0 \\ 2/3 & -\sqrt{2}/6 & \sqrt{2}/2 \\ -2/3 & \sqrt{2}/6 & \sqrt{2}/2 \end{bmatrix}.$$

4.10 Determine the rank of the following matrix:

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 0 \\ 2 & 4 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 1 & 2 & 1 & -1 & 1 \end{bmatrix}.$$

4.11 Clogg, Petkova, and Haritou (1995) give detailed guidance for deciding between different linear regression models using the same data. In this work they define the matrices $\mathbf{X}$, which is $n \times (p+1)$ rank $p+1$, and $\mathbf{Z}$, which is $n \times (q+1)$ rank $q+1$, with $p < q$. They calculate the matrix $A = \left[\mathbf{X}'\mathbf{X} - \mathbf{X}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}Z'X\right]^{-1}$. Find the dimension and rank of A.

4.12 For each of the following matrices, find the eigenvalues and eigenvectors:

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 \\ -3 & -4 \end{bmatrix} \qquad \begin{bmatrix} 11 & 3 \\ 9 & -4 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 7 & 4 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix} \qquad \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & 7 & \pi \\ 3 & e & 4 \\ 1 & 0 & \pi \end{bmatrix}.$$

4.13 Land (1980) develops a mathematical theory of social change based on a model of underlying demographic accounts. The corresponding population mathematical models are shown to help identify and track changing social indicators, although no data are used in the article. Label L_x as the number of people in a population that are between x and $x+1$ years old. Then the square matrix $\mathbf{P}'$ of order $(\omega+1)\times(\omega+1)$ is given by

$$\mathbf{P}' = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ L_1/L_0 & 0 & 0 & 0 & \dots \\ 0 & L_2/L_1 & 0 & 0 & \dots \\ 0 & 0 & L_2/L_1 & 0 & \dots \\ \vdots & \vdots & \ddots & \dots & \\ 0 & 0 & 0 & 0 & L_\omega/L_{\omega-1} \end{bmatrix},$$

where ω is the assumed maximum lifespan and each of the nonzero ratios gives the proportion of people living to the next age. The matrix $(\mathbf{I} - \mathbf{P}')$ is theoretically important. Calculate its trace and inverse. The inverse will be a lower triangular form with survivorship probabilities as the nonzero values, and the column sums are standard life expectations in the actuarial sense.

4.14 The Clement matrix is a symmetric, tridiagonal matrix with zero diagonal values. It is sometimes used to test algorithms for computing inverses and eigenvalues. Compute the eigenvalues of the following

4×4 Clement matrix:

$$\begin{bmatrix} 0 & 1.732051 & 0 & 0 \\ 1.732051 & 0 & 2.0 & 0 \\ 0 & 2.0 & 0 & 1.732051 \\ 0 & 0 & 1.732051 & 0 \end{bmatrix}.$$

4.15 Consider the two matrices

$$\mathbf{X_1} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.95 & 1 & 3 \end{bmatrix} \qquad \mathbf{X_2} = \begin{bmatrix} 5 & 2 & 5 \\ 2 & 1 & 2 \\ 3 & 2 & 3 \\ 2.99 & 1 & 3 \end{bmatrix}.$$

Given how similar these matrices are to each other, why is $(\mathbf{X_2}'\mathbf{X_2})^{-1}$ so different from $(\mathbf{X_1}'\mathbf{X_1})^{-1}$?

4.16 A Vandermonde matrix is a specialized matrix that occurs in a particular type of regression (polynomial). Find the determinant of the following general Vandermonde matrix:

$$\begin{bmatrix} 1 & v_1 & v_1^2 & v_1^3 & \dots & v_1^{n-1} & v_1^n \\ 1 & v_2 & v_2^2 & v_2^3 & \dots & v_2^{n-1} & v_2^n \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 1 & v_{n-1} & v_{n-1}^2 & v_{n-1}^3 & \dots & v_{n-1}^{n-1} & v_{n-1}^n \\ 1 & v_n & v_n^2 & v_n^3 & \dots & v_n^{n-1} & v_n^n \end{bmatrix}.$$

4.17 A Hilbert matrix has elements $x_{ij} = 1/(i + j - 1)$ for the entry in row i and column j. Is this always a symmetric matrix? Is it always positive definite?

4.18 Verify (replicate) the matrix calculations in the example with EU poverty data on page 155.

4.19 Solve the following systems of equations for x, y, and z:

$$\begin{aligned} x + y + 2z &= 2 \\ 3x - 2y + z &= 1 \\ y - z &= 3 \end{aligned}$$

$$\begin{aligned} 2x + 3y - z &= -8 \\ x + 2y - z &= 2 \\ -x - 4y + z &= -6 \end{aligned}$$

$$\begin{aligned} x - y + 2z &= 2 \\ 4x + y - 2z &= 10 \\ x + 3y + z &= 0. \end{aligned}$$

4.20 Show that the eigenvectors from the matrix $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$ are perpendicular.

4.21 A matrix is an M-matrix if $x_{ij} \leq 0, \; \forall i \neq j$, and all the elements of the inverse (X^{-1}) are nonnegative. Construct an example.

4.22 Williams and Griffin (1964) looked at executive compensation in the following way. An allowable bonus to managers, B, is computed as a percentage of net profit, P, before the bonus and before income taxes, T. But a reciprocal relationship exists because the size of the bonus affects net profit, and vice versa. They give the following example as a system of equations. Solve.

$$\begin{array}{rcrcrl} B & - & 0.10P & + & 0.10T & = 0 \\ 0.50B & - & 0.50P & + & T & = 0 \\ & & P & & & = 100,000. \end{array}$$

4.23 This question uses the following 8×8 matrix $\mathbf{X}$ of fiscal data by country:

$$
\begin{array}{l}
\\ \text{Australia} \\ \text{Britain} \\ \text{Canada} \\ \text{Denmark} \\ \text{Japan} \\ \text{Sweden} \\ \text{Switzerland} \\ \text{USA}
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
\mathbf{x}_{.1} & \mathbf{x}_{.2} & \mathbf{x}_{.3} & \mathbf{x}_{.4} & \mathbf{x}_{.5} & \mathbf{x}_{.6} & \mathbf{x}_{.7} & \mathbf{x}_{.8}
\end{array} \\
\begin{pmatrix}
3.3 & 9.9 & 5.41 & 5.57 & 5.15 & 5.35 & 5.72 & 6.24 \\
5.8 & 11.4 & 4.81 & 4.06 & 4.48 & 4.59 & 4.79 & 5.24 \\
12.1 & 9.9 & 2.43 & 2.24 & 2.82 & 4.29 & 4.63 & 5.65 \\
12.0 & 12.5 & 2.25 & 2.15 & 2.42 & 3.66 & 4.26 & 5.01 \\
4.1 & 2.0 & 0.02 & 0.03 & 0.10 & 1.34 & 1.32 & 1.43 \\
2.2 & 4.9 & 1.98 & 2.55 & 2.17 & 3.65 & 4.56 & 2.20 \\
-5.3 & 1.2 & 0.75 & 0.24 & 0.82 & 2.12 & 2.56 & 2.22 \\
5.4 & 6.2 & 2.56 & 1.00 & 3.26 & 4.19 & 4.19 & 5.44
\end{pmatrix}
\end{array}
$$

where $\mathbf{x}_{.1}$ is percent change in the money supply a year ago (narrow), $\mathbf{x}_{.2}$ is percent change in the money supply a year ago (broad), $\mathbf{x}_{.3}$ is the 3-month money market rate (latest), $\mathbf{x}_{.4}$ is the 3-month money market rate (1 year ago), $\mathbf{x}_{.5}$ is the 2-year government bond rate, $\mathbf{x}_{.6}$ is the 10-year government bond rate (latest), $\mathbf{x}_{.7}$ is the 10-year government bond rate (1 year ago), and $\mathbf{x}_{.8}$ is the corporate bond rate (source: ***The Economist***, January 29, 2005, page 97). We would expect a number of these figures to be stable over time or to relate across industrialized democracies. Test whether this makes the matrix $\mathbf{X}'\mathbf{X}$ ill-conditioned by obtaining the condition number. What is the rank of $\mathbf{X}'\mathbf{X}$. Calculate the determinant using eigenvalues. Do you expect near collinearity here?

4.24 Show that the inverse relation for the matrix $\mathbf{A}$ below is true:

$$
\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{e} & \frac{-b}{e} \\ \frac{-c}{e} & \frac{a}{e} \end{bmatrix}.
$$

Here e is the determinant of $\mathbf{A}$. Now apply this rule to invert the 2×2 matrix $\mathbf{X}'\mathbf{X}$ from the $n \times 2$ matrix $\mathbf{X}$, which has a leading column of 1's and a second column vector: $[x_{11}, x_{12}, \ldots, x_{1n}]$.

4.25 Another method for solving linear systems of equations of the form $\mathbf{A}^{-1}\mathbf{y} = \mathbf{x}$ is **Cramer's rule**. Define $\mathbf{A}_j$ as the matrix where $\mathbf{y}$ is plugged in for the jth column of $\mathbf{A}$. Perform this for every column $1, \ldots, q$ to produce q of these matrices, and the solution will be the vector $\left[\frac{|\mathbf{A}_1|}{\mathbf{A}}, \frac{|\mathbf{A}_2|}{\mathbf{A}}, \ldots \frac{|\mathbf{A}_q|}{\mathbf{A}}\right]$. Show that performing these steps on the matrix in the example on page 159 gives the same answer.