# 4

# Linear Algebra Continued: Matrix Structure

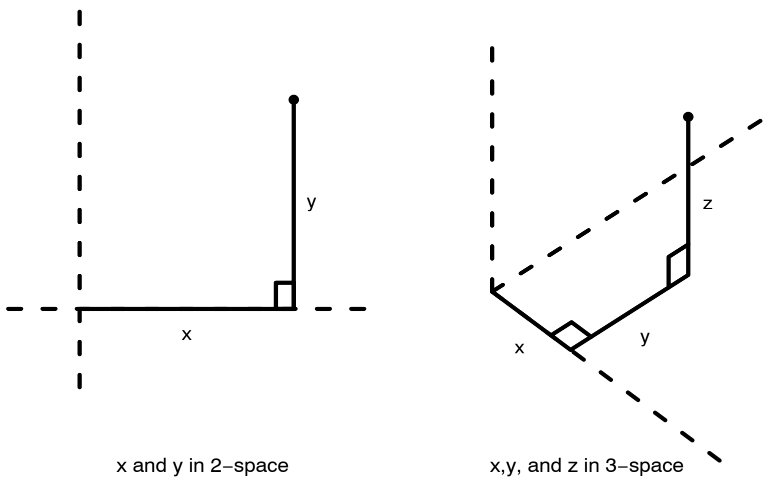
## 4.1 Objectives

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.

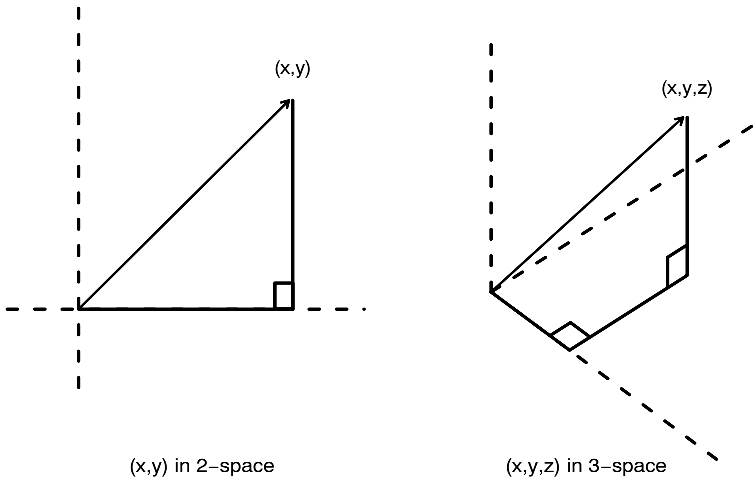
Fig. 4.1. Visualizing Space




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

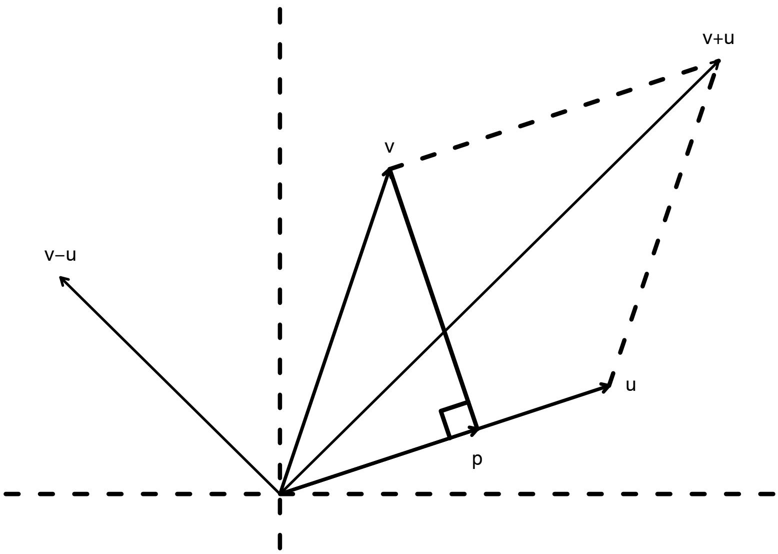



Thinking broadly about the two vectors in Figure 4.2, they take up an amount of "space" in the sense that they define a triangular planar region bounded by the vector itself and its two (left panel) or three (right panel) **projections** against the axes where the angle on the axis from this projection 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 $\theta$ 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 $\theta$ 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 $k$th column vector of $\mathbf{x}$ and $\mathbf{x}_{k.}$ denotes the $k$th 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 $\mathfrak{R}$. 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 $\omega$. 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 $\mathfrak{R}^i$ (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 $\mathfrak{R}^i$ 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 $\mathfrak{R}^i$ 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 $3 \times 3$ 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 $\left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{smallmatrix}\right]$ 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 $2 \times 2$ 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 $2 \times 2$. 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 $ij$th **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 $i$th row and $j$th 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 $3 \times 3$ 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 $2 \times 2$ 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 $7 \times 7$

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 $4 \times 4$ calculations after two steps. Fortunately this trick works again because we have the same structure remaining in the $3 \times 3$ case. Let us be a bit more strategic though and define the $2 \times 2$ 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.

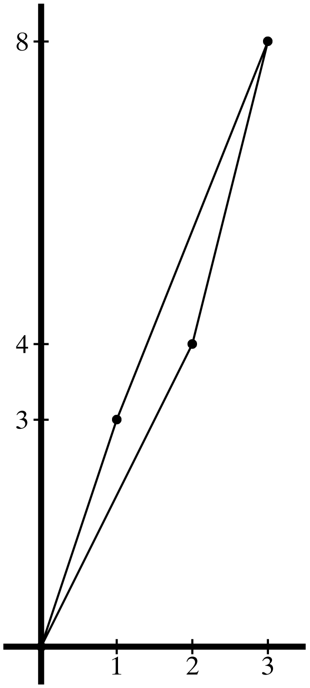
The trace and the determinant have interrelated uses and special properties as well. For instance, Kronecker products on square matrices have the properties $\text{tr}(\mathbf{X} \otimes \mathbf{Y}) = \text{tr}(\mathbf{X})\text{tr}(\mathbf{Y})$, and $|\mathbf{X} \otimes \mathbf{Y}| = |\mathbf{X}|^{\ell}|\mathbf{Y}|^{j}$ 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 $2\times2$ 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 $i$th minor (the matrix that remains when the $i$th row and column are removed) is multiplied by the $i$th 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 | $\text{rank}[\mathbf{X}\mathbf{Y}] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$<br>$\text{rank}\left[\begin{smallmatrix}\mathbf{X}\\ \mathbf{Y}\end{smallmatrix}\right] \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$ |
| $\rightarrowtail$ | 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 $\|\mathbf{X} \cdot \mathbf{Y}\| = \mathrm{tr}(\mathbf{X}'\mathbf{Y})$ |

★ **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

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

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 $2 \times 2$ 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 $3 \times 3$ 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.

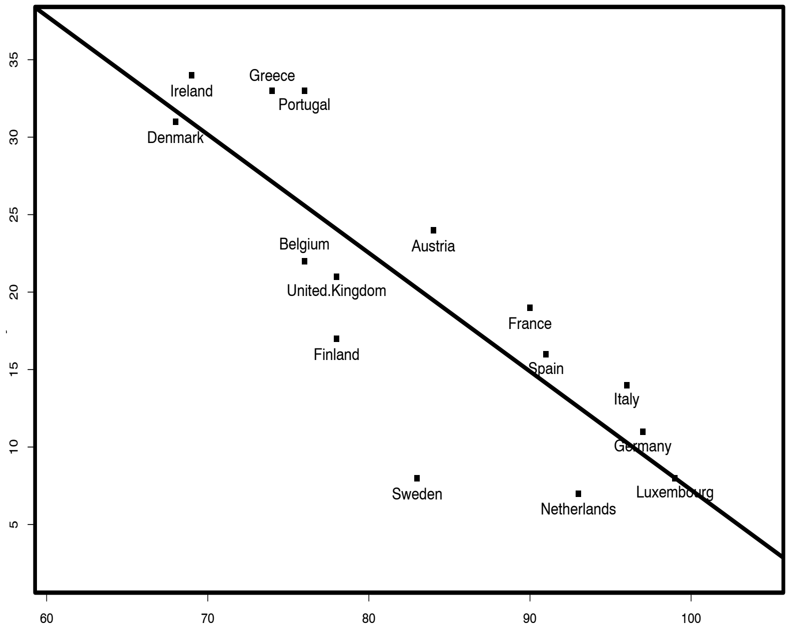
Once one understands linear regression in matrix notation, it is much easier to see what is happening. For instance, if there were a ***second*** explanatory variable (there are many more than one in social science models), then it would simply be an addition column of the $\mathbf{X}$ matrix and all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

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

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

$$\begin{aligned} 2x_1 - 3x_2 &= 4 \\ 5x_1 + 5x_2 &= 3, \end{aligned}$$

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 $\mathbf{y}$:

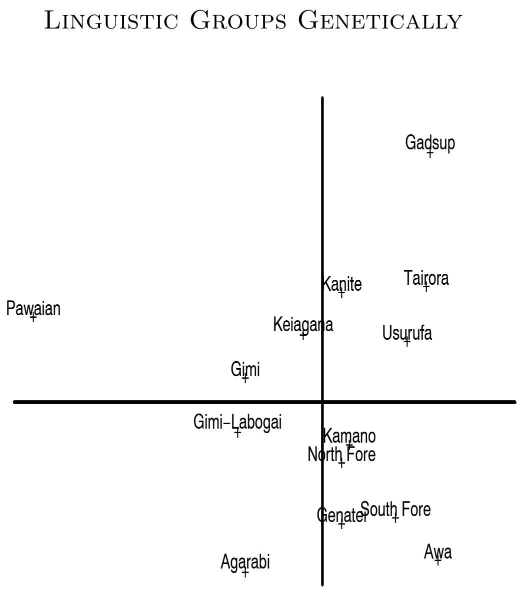
$$\mathbf{A}^{-1}\mathbf{y} = \begin{bmatrix} 0.2 & 0.12 \\ -0.2 & 0.08 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.16 \\ -0.56 \end{bmatrix},$$

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.




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, $\lambda_i$ 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 $3 \times 3$ 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 $\lambda$, 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 $\lambda_i$ 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 $\lambda_i$ is an eigenvalue of $\mathbf{X}$, then$\lambda_i^k$ 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 ($\mathfrak{R}^2$, $\mathfrak{R}^3$, 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 $i$th 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 $5 \times 5$ 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 $2 \times 3$ 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**

| | | | |
|---|---|---|---|
| Non-Negative Definite: | | | |
| | $\rightarrowtail$ | positive definite | $\mathbf{y}'\mathbf{X}\mathbf{y} > 0$ |
| | $\rightarrowtail$ | positive semidefinite | $\mathbf{y}'\mathbf{X}\mathbf{y} \geq 0$ |
| Non-Positive Definite: | | | |
| | $\rightarrowtail$ | 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 $3 \times 3$ 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 tr$(\mathbf{XY}) \neq$ tr$(\mathbf{X})$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 $|B|$, 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 $A = \left[\mathbf{X}'\mathbf{X} - \mathbf{XZ}(\mathbf{Z}'\mathbf{Z})^{-1}Z'X\right]^{-1}$. 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 $\omega$ 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

$4 \times 4$ 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 $8 \times 8$ 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 $2 \times 2$ 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 $j$th 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.