Section 5.1 Null Space and Range of a Matrix

The null space $\text{Null}(A)$ of a matrix $A$ is the set of vectors $x$ that are solutions to the system of equations $Ax = 0$. The range $\text{Range}(A)$ of $A$ is the set of vectors $b$ such that $Ax = b$ has a solution. Another name that is sometimes used for $\text{Null}(A)$ is the kernel of $A$. In this section we examine these two important sets of vectors associated with any matrix. We look at their role in linear models and learn how to determine these sets.

Both the $\text{Null}(A)$ and $\text{Range}(A)$ are vector spaces.

**Definition.** A vector space is any set $V$ of vectors such that if $x_1, x_2$ are in $V$, then any linear combination $rx_1 + sx_2$ is also in $V$.

If $Ax_1 = 0$ and $Ax_2 = 0$, then we have

$$A(rx_1 + sx_2) = r(Ax_1) + s(Ax_2) = r(0) + s(0) = 0$$

Thus $\text{Null}(A)$ is a vector space. A similarly simple proof, left as an exercise, shows that $\text{Range}(A)$ is a vector space.

Suppose that $A$ is an $n$-by-$n$ matrix for which the system $Ax = b$ has a unique solution for every $b$. Then $\text{Range}(A)$ is all possible $n$-vectors, and $\text{Null}(A)$ is just the zero vector $0$, since $0$ is always a solution to $Ax = 0$ and by assumption there can only be one solution to $Ax = 0$. 393
Ch. 5 Theory of Systems of Linear Equations and Eigenvalue/Eigenvector Problems

In this section we are concerned with matrices $A$ for which $Ax = b$ has nonunique solutions or no solutions. Then determining the range and null space of $A$ becomes important.

Let us briefly describe the geometry of systems of equations that do not have unique solutions. Consider the pair of equations

$$
\begin{align*}
2x - y &= 5 \\
3x + 2y &= 11
\end{align*}
$$

(1)

In Figure 5.1a we have plotted the graph of the two equations in (1) in $x - y$ space. The graph of a linear equation is a straight line. An $(x, y)$ pair that solves (1) must lie on the lines of both equations. That is, the $(x, y)$ pair must be the coordinates of the point where the two lines intersect. From Figure 5.1a we see that this intersection point has coordinates $(3, 1)$.

Suppose the two equations produce lines that are parallel:

$$
\begin{align*}
2x - y &= 5 \\
4x - 2y &= 4
\end{align*}
$$

(2)

(see Figure 5.1b). Then there is no common point—no solution to (2). Note that "parallel" means that the second equation's coefficients are multiples of the first's. We saw in the canoe-with-sail example in Section 1.1 that when two equations produce almost parallel lines, the equations can give strange results.

A system of three equations in three unknowns, such as

$$
\begin{align*}
2x + 5y + 4z &= 4 \\
x + 4y + 3z &= 1 \\
-x + 3y + 2z &= -5
\end{align*}
$$

(3)
Sec. 5.1 Null Space and Range of a Matrix

Figure 5.2 (a) Three planes intersect at a point. (b) The lines formed by intersections of two planes are parallel. (c) Three planes intersect along a line.

has a similar interpretation in three-dimensional space (see Figure 5.2a). Each equation now determines a plane of points. A solution of (3) will be the coordinates of a point where all three planes intersect. If two of the planes are parallel, no solution can exist.

Another possibility is that while each pair of equations intersects to form a line, there may be no point common to all three. This happens if the line formed by the intersection of two planes is parallel to the third plane. Figure 5.2b illustrates this situation; such a system of equations is

\[
\begin{align*}
    x - y + z &= 3 \\
    x + y + z &= 3 \\
    x + 3y + z &= 7
\end{align*}
\]

(4)

The points \((x, y, z)\) satisfying both of the first two equations in (4) form the line \(x + z = 3, y = 0\); the points satisfying the first and third equations form the line \(x + z = 4, y = 1\); and the points satisfying the second and third equations form the line \(x + z = 1, y = 2\).
Of course, when we have four or more equations in three unknowns, it is very likely that there will be no \((x, y, z)\) that lies on all four planes (satisfies all four equations). Conversely, if we had only the first two equations in (4), all the points on the line \(x + z = 3, y = 0\) mentioned above will be solutions.

Let us turn from systems with no solutions to 3-by-3 systems with multiple solutions. The points that lie on all three planes may happen to form a line, as in the case for the system

\[
\begin{align*}
4x - y + 2z &= 8 \\
2x + 5y + z &= 4 \\
-2x + 3y - z &= -4
\end{align*}
\]

which has the graph shown in Figure 5.2c. The line of points \((x, y, z)\) common to all three equations in (5) is \(2x + z = 4, y = 0\). So (5) has an infinite number of solutions. As we go to higher dimensions, the possibilities for multiple solutions or no solution increase.

We present two theorems that show the fundamental link between multiple solutions to the system \(Ax = b\) and the null space of \(A\).

**Theorem 1.** Let \(A\) be any \(m\)-by-\(n\) matrix.

(i) If \(\text{Null}(A)\) contains one nonzero vector \(x^0\), then \(\text{Null}(A)\) contains an infinite number of vectors; in particular, any multiple \(rx^0\) is in \(\text{Null}(A)\).

(ii) If \(x^0\) is in \(\text{Null}(A)\) and \(x^*\) is a solution to \(Ax = b\), then \(x^* + x^0\) is also a solution to \(Ax = b\).

(iii) If \(x_1, x_2\) are two different solutions to \(Ax = b\), for some given \(b\), then their difference \(x_1 - x_2\) is a vector in \(\text{Null}(A)\).

(iv) Given a solution \(x^*\) to \(Ax = b\), then any other solution \(x'\) to this matrix equation can be written as

\[x' = x^* + x^0\quad \text{for some } x^0 \text{ in } \text{Null}(A).\]

(v) If \(\text{Null}(A)\) consists of only the zero vector \(0\) (i.e., \(Ax = 0\) has only the solution \(x = 0\)), then \(Ax = b\) has at most one solution, for any given \(b\).

**Proof**

(i) \(A(rx^0) = r(Ax^0) = r0 = 0\), so \(rx^0\) is in \(\text{Null}(A)\) for any scalar \(r\). Thus \(\text{Null}(A)\) is infinite.

(ii) Let \(x^0\) be in \(\text{Null}(A)\), so \(Ax^0 = 0\). Since \(Ax^* = b\), then

\[
A(x^* + x^0) = Ax^* + A(x^0) = b + 0 = b
\]

Thus \(x^* + x^0\) is a solution to \(Ax = b\), as claimed.

(iii) Since \(A(x_1 - x_2) = Ax_1 - Ax_2 = b - b = 0\), then \(x_1 - x_2\) is in \(\text{Null}(A)\).
(iv) Given solutions \( x^* \) and \( x' \), choose \( x^0 = x' - x^* \). Then \( x' = x^* + x^0 \), and by part (iii), \( x^0 \) is in \( \text{Null}(A) \).

(v) Suppose that one solution \( x^* \) to \( Ax = b \) is known. By part (iv), all solutions \( x' \) of \( Ax = b \) can be expressed in the form \( x' = x^* + x^0 \), where \( x^0 \) is in \( \text{Null}(A) \). If \( \text{Null}(A) \) consists of just \( 0 \), then \( x' = x^* + 0 \)—there is only solution, \( x^* \), to \( Ax = b \).

**Theorem 2.** Let \( A \) be an \( m \)-by-\( n \) matrix. If \( Ax = b' \) has two solutions for some particular \( b' \):

(i) The null space of \( A \) has an infinite number of vectors.

(ii) For any \( b \), either \( Ax = b \) has no solution or an infinite number of solutions.

**Proof**

(i) By Theorem 1, part (iii), the difference of two solutions is a (nonzero) vector \( x^0 \) in \( \text{Null}(A) \), and then by Theorem 1, part (i), the multiples \( rx^0 \) yield an infinite number of vectors in \( \text{Null}(A) \).

(ii) Suppose that \( Ax = b \) has one solution \( x^* \). From Theorem 1, part (ii), \( x^* + x^0 \) is also a solution, for any \( x^0 \) in \( \text{Null}(A) \). By part (i) of this theorem, \( \text{Null}(A) \) is infinite.

Theorem 2’s result that two solutions lead to an infinite number of solutions corresponds to our geometric pictures in which multiple solutions always consisted of an (infinite) set of points along a line.

A system of equations with \( 0 \) on the right side—\( Ax = 0 \)—is called a **homogeneous system**. Solutions to the homogeneous system \( Ax = 0 \) form the null space of \( A \). One often speaks of the **null space of the system** \( Ax = 0 \), implicitly meaning the null space of \( A \).

Homogeneous systems \( Ax = 0 \) have arisen in several different settings in this book.

**Example 1. Multiple Solutions in an Oil Refinery Problem**

Let us suppose that in our familiar oil refinery problem, the three refineries produce only the first two products:

\[
\text{Heating oil: } 20x_1 + 4x_2 + 4x_3 = 500 \\
\text{Diesel oil: } 10x_1 + 14x_2 + 5x_3 = 850
\]  

Suppose that we are given one solution, \( x_1 = 15, x_2 = 50, x_3 = 0 \), using just the first two refineries. We want to find another solution with \( x_3 = 20 \). Let us find the null space of this coefficient matrix and then, using Theorem 1, part (iv), add an appropriate null space vector to the given solution \([15, 50, 0]\) to get a solution with \( x_3 = 20 \).

The null space for this system of equations is all solutions to the associated homogeneous system.
Solving with elimination by pivoting, we obtain
\[ 
\begin{align*}
20x_1 + 4x_2 + 4x_3 &= 0 \\
10x_1 + 14x_2 + 5x_3 &= 0
\end{align*}
\]
(8)

Thus \( x_1 = -\frac{3}{20}x_3 \) and \( x_2 = -\frac{1}{4}x_3 \). So vectors in the null space have the form
\[ 
[-\frac{3}{20}x_3, -\frac{1}{4}x_3, x_3] \quad \text{or} \quad x_3[-\frac{3}{20}, -\frac{1}{4}, 1]
\]
(9)

Theorem 1, part (iv), says that any solution \( x' \) to the system (7) can be expressed in the form \( x' = x + x^0 \), where \( x^0 \) is a null-space vector, in this case \( x^0 = r[-\frac{3}{20}, -\frac{1}{4}, 1] \) for some constant \( r \). We said above that we want \( x' \) to have \( x'_3 = 20 \). Then
\[ 
x' = x + x^0 \\
[x'_1, x'_2, 20] = [15, 50, 0] + r[-\frac{3}{20}, -\frac{1}{4}, 1]
\]
(10)

Matching the third entry on each side of (10), we have \( 20 = 0 + r \). So \( r = 20 \) and the desired solution is
\[ 
x' = x + x^0 = [15, 50, 0] + 20[-\frac{3}{20}, -\frac{1}{4}, 1]
\]
\[ 
= [12, 45, 20]
\]
(11)

**Example 2. Balancing Chemical Equations Revisited**

In Example 1 of Section 4.3 we obtained a system of equations for balancing the atomic equations for the chemical reaction in which permanganate (\( \text{MnO}_4 \)) and hydrogen (\( H \)) ions combine to form manganese (\( \text{Mn} \)) and water (\( \text{H}_2\text{O} \)):

\[ 
\text{MnO}_4 + H \rightarrow \text{Mn} + \text{H}_2\text{O}
\]
(12)

where \( H \) represents hydrogen and \( O \) oxygen. We let \( x_1 \) be the number of permanganate ions, \( x_2 \) the number of hydrogen ions, \( x_3 \) the number of manganese atoms, and \( x_4 \) the number of water molecules. To have the same number of atoms in the molecules on each side of the reaction, we obtained the system of equations

\[ 
\begin{align*}
\text{H:} & \quad x_2 = 2x_4 \\
\text{Mn:} & \quad x_1 = x_3 \\
\text{0:} & \quad 4x_1 = x_4
\end{align*}
\]
(13a)
or
\[\begin{align*}
x_2 - 2x_4 &= 0 \\
x_1 - x_3 &= 0 \\
4x_1 - x_4 &= 0
\end{align*}\]
(13b)

Notice that we have four unknowns but only three equations. When we solved system (13b) using elimination by pivoting (pivoting on entries (2, 1), (1, 2), (3, 3)), we obtained
\[\begin{align*}
x_2 &= 2x_4 \\
x_1 &= 4x_4 \\
x_3 &= 4x_4
\end{align*}\]
(14)

As vectors, the solutions in (14) have the form
\[\begin{bmatrix} 0, 4x_4, 4x_4, x_4 \end{bmatrix} \text{ or } x_4[1, 4, 4, 1] \]
(15)

These vectors form the null space for the system (13b).

For example, if \(x_4 = 4\), then \(x_2 = 8\), \(x_1 = x_3 = 1\), and the reaction equation becomes
\[\text{MnO}_4 + 8\text{H} \rightarrow \text{Mn} + 4\text{H}_2\text{O}\]

The solution we obtain makes the amounts of the first three types of molecules fixed ratios of the amount of the fourth type, which we are free to give any value (i.e., \(x_4\) is a free variable). In another series of pivots, the final free variable might end up being \(x_1\).

\[\begin{align*}
-8x_1 + x_2 &= 0 \\
-x_1 + x_3 &= 0 \\
-4x_1 + x_4 &= 0
\end{align*}\]
\[\begin{align*}
x_2 &= 8x_1 \\
x_3 &= x_1 \\
x_4 &= 4x_1
\end{align*}\]
(16)
yielding solution vectors
\[\begin{bmatrix} x_1, 8x_1, x_1, 4x_1 \end{bmatrix} \text{ or } x_1[1, 8, 1, 4] \]
(17)

However we formulate the solution of (13b), we always have the same set of vectors, namely, the null space of the coefficient matrix in (13b). For example, the vector [1, 8, 1, 4] in (17) is a multiple of the vector \([\frac{1}{4}, 2, \frac{1}{4}, 1]\) in (15).

Examples 1 and 2 show us how to determine the null space of any matrix. We simply apply elimination by pivoting to the homogeneous system \(Ax = 0\) and reduce it to the form (14), from which we obtain the null-space vectors as expressed in (15).
Example 3. The Null Space of the Frog Markov Chain

We have already performed elimination by pivoting on this transition matrix $A$ in Example 6 of Section 3.3 when we were trying to invert the matrix. We do the pivoting again, but now the right-side vector is $0$.

\[
\begin{bmatrix}
0.5 & 0.25 & 0 & 0 & 0 & 0 \\
0.5 & 0.5 & 0.25 & 0 & 0 & 0 \\
0 & 0.25 & 0.5 & 0.25 & 0 & 0 \\
0 & 0 & 0.25 & 0.5 & 0.25 & 0 \\
0 & 0 & 0 & 0.25 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0.25 & 0.5
\end{bmatrix}
\begin{bmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4 \\
p_5 \\
p_6
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

Pivoting on entries $(1, 1), (2, 2), (3, 3), (4, 4)$, and $(5, 5)$, we obtain

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & -2 \\
0 & 0 & 1 & 0 & 0 & 2 \\
0 & 0 & 0 & 1 & 0 & -2 \\
0 & 0 & 0 & 0 & 1 & 2 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
p_1 \\
p_2 \\
p_3 \\
p_4 \\
p_5 \\
p_6
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

(18)

Equations (18) express the first five $p_i$'s in terms of $p_6$. Rewriting (18), we get

\[
p_1 + p_6 = 0 \\
p_2 - 2p_6 = 0 \\
or \quad p_3 + 2p_6 = 0 \\
p_4 - 2p_6 = 0 \\
p_5 + 2p_6 = 0
\]

Equations (18) express the first five $p_i$'s in terms of $p_6$. Rewriting (18), we get

\[
p_1 = -p_6, \quad p_2 = 2p_6, \quad p_3 = -2p_6, \quad p_4 = 2p_6, \quad p_5 = -2p_6
\]

and thus the solutions to $Ap = 0$ have the form

\[
[-p_6, 2p_6, -2p_6, 2p_6, -2p_6, p_6] \quad \text{or} \quad p_6[-1, 2, -2, 2, -2, 1]
\]

(19)

The vectors in (19) are the null space of $A$.

We can add a null-space vector $p^0$ like (19) to a probability vector $p$, and if $Ap = p'$, then also $A(p + p^0) = p'$ [this fact is Theorem 1, part (ii)]. For example, we found earlier that $p^* =
[.1, .2, .2, .2, .2, .1] is the stable distribution for the frog Markov chain, so that $Ap^* = p^*$. So for any null-space vector $p^0$, we have $A(p^* + p^0) = Ap^* = p^*$. Suppose that we select for the null-space vector $p^0 = [-.1, .2, -.2, .2, -.2, .1]$ [with $p_6 = .1$ in (19)]. Then

$$p^* + p_0 = [.1, .2, .2, .2, .2, .1] + [-.1, .2, -.2, .2, -.2, .1] = [0, .4, 0, .4, .2]$$

Thus if we start with distribution $[0, .4, 0, .4, 0, .2]$, we will reach the stable distribution $p^*$ after just one period. (The reader should verify this result numerically.)

**Example 4. A Two-Variable Null Space**

The following system of equations formed constraints in the transportation problem presented in Section 4.6; the names of the variables have been changed for simplicity. *(Background: The first equation represents the fact that the amount $x_1$ of food shipped from the first warehouse to college $A$ plus the amount $x_2$ shipped from the first warehouse to college $B$ equals 20, the amount of food in the first warehouse. The other equations represent the amounts available at the second and third warehouses and the amounts needed at colleges $A$ and $B$. There are many solutions—ways to ship the food between the three warehouses and the two colleges. In Section 4.6 each $x_i$ had an associated cost and we sought to minimize the total cost.)*

\[
\begin{align*}
    x_1 + x_2 &= 20 \\
    x_3 + x_4 &= 30 \\
    x_5 + x_6 &= 15 \\
    x_1 + x_3 + x_5 &= 25 \\
    x_2 + x_4 + x_6 &= 40
\end{align*}
\]

To change from one solution $x^*$ of (20) to another (cheaper) solution $x^{**}$, we would add some null-space vector to $x^*$ according to Theorem 1. To find the null space, we solve the associated homogeneous system

\[
\begin{align*}
    x_1 + x_2 &= 0 \\
    x_3 + x_4 &= 0 \\
    x_5 + x_6 &= 0 \\
    x_1 + x_3 + x_5 &= 0 \\
    x_2 + x_4 + x_6 &= 0
\end{align*}
\]

Pivoting on entries (1, 2), (2, 4), (3, 6), (4, 3), we obtain
The solutions in (22) to the homogeneous system (21) produce the set of null-space vectors

\[ \begin{bmatrix} x_1, -x_1, -x_1 - x_5, x_1 + x_5, x_3, -x_3 \end{bmatrix} \]  

(23)

Breaking the \( x_1 \) and \( x_5 \) components apart, we can rewrite (23) as

\[ x_1[1, -1, -1, 1, 0, 0] + x_5[0, 0, -1, 1, 1, -1] \]  

(24)

So the null space is all linear combinations of the two vectors

\[ x_1^* = [1, -1, -1, 1, 0, 0] \quad \text{and} \quad x_5^* = [0, 0, -1, 1, 1, -1]. \]

Suppose that we are given the solution to (20) \( \mathbf{x} \): \( x_1 = 20, x_3 = 5, x_4 = 25, x_6 = 15, \) and \( x_2 = x_5 = 0. \) Thus

\[ \mathbf{x} = [20, 0, 5, 25, 0, 15] \]

Also suppose that we want a solution in which \( x_2 = 10 \) and \( x_5 = 5. \) We can achieve this by adding the right linear combination of null-space vectors \( x_1^* = [1, -1, -1, 1, 0, 0] \) and \( x_5^* = [0, 0, -1, 1, 1, -1]. \) Remember that adding any null-space vector to a solution vector yields another solution vector.

To make \( x_2 = 10 \) (it is now 0), we can add to \( \mathbf{x} \) the vector \( -10x_1^* = [-10, 10, 10, -10, 0, 0]. \) To make \( x_5 = 5 \) (also now 0), we can add to \( \mathbf{x} \) the vector \( 5x_5^* = [0, 0, -5, 5, 5, -5] \) to \( \mathbf{x}. \) So our desired solution is

\[ \mathbf{x} - 10x_1^* + 5x_5^* = [20, 0, 5, 25, 0, 15] \]
\[ + [-10, 10, 10, -10, 0, 0] \]
\[ + [0, 0, -5, 5, 5, -5] \]
\[ = [10, 10, 10, 20, 5, 10] \]

The null space in Example 4 is a little more complicated. If we had performed a different pivot sequence in (22), the two vectors that generated the null space would be different from the two vectors in (24). It is not even obvious that we would end up with two vectors. Maybe this same null space could be expressed as combinations of three vectors. Could it be expressed as multiples of a single vector? Probably not.
In Section 5.2 we use the theory of vector spaces to prove that any pivot sequence yields the same number of vectors for building the null space.

Optional Example on the Null Space of the Differentiation Transformation

In the Interlude before this chapter, we noted that the operation $D(f(x))$ of taking the derivative of a function $f(x)$ was a linear transformation of functions. $D$ acts on the abstract vector space of all differentiable functions.

Let us determine the null space of the differentiation transformation and see how the results in Theorem 1 apply to $D$. $\text{Null}(D)$ is the set of functions $f(x)$ for which $D(f(x)) = 0$, that is, $f'(x) = 0$. These are just the constant functions, $f(x) = c$ for some constant $c$.

$$\text{Null}(D) = \{ \text{all constant functions} \}$$

Suppose that we want to solve the problem

$$D(f(x)) = 9x^2 + x \quad \text{[find } f(x) \text{ such that } f'(x) = 9x^2 + x]\)$$

(25)

By Theorem 1, part (ii), if $f^*(x)$ is a solution to (25), then $f^*(x)$ plus any constant function [i.e., plus any member of $\text{Null}(D)$] is also a solution, say, $f^*(x) = 3x^3 + x^2/2$. Then $3x^3 + x^2/2 + c$, for any constant $c$, is also a solution.

Conversely, by Theorem 1, part (iv), every solution to (25) can be written as the sum of a specific solution $f^*(x)$ to (25) plus some constant function [a member of $\text{Null}(D)$]. In this case, every solution has the form $3x^3 + x^2/2 + c$. This result is the basic formula about the form of the integral (the antiderivative) of a function.

We close this discussion of null spaces by noting a close relationship between null spaces and eigenvectors. An eigenvector $\mathbf{u}$, with associated eigenvalue $\lambda$, satisfies the $n$-by-$n$ homogeneous system

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \quad \text{(or more familiarly, } \mathbf{A}\mathbf{u} = \lambda\mathbf{u})$$

Thus an eigenvector associated with $\lambda$ is a member of the null space of $(\mathbf{A} - \lambda\mathbf{I})$. For example, if $\mathbf{A}$ is the transition matrix of the frog Markov chain, then the stable distribution $\mathbf{p}^*$ satisfies $\mathbf{A}\mathbf{p}^* = \mathbf{p}^*$ or $(\mathbf{A} - \mathbf{I})\mathbf{p}^* = \mathbf{0}$, so $\mathbf{p}^*$ is in the null space of $\mathbf{A} - \mathbf{I}$.

Conversely, $\mathbf{A}\mathbf{u} = \mathbf{0}$ is equivalent to $(\mathbf{A} - \mathbf{I})\mathbf{u} = \mathbf{0}$. So any nonzero vector $\mathbf{u}$ in the null space of $\mathbf{A}$ is an eigenvector of $\mathbf{A}$ associated with eigenvalue $\lambda = 0$.

So far we have considered systems with multiple solutions. Next we discuss systems with no solution—where $\mathbf{A}\mathbf{x} = \mathbf{b}$ cannot be solved for some vectors $\mathbf{b}$. Recall that the set of $\mathbf{b}$'s for which $\mathbf{A}\mathbf{x} = \mathbf{b}$ can be solved is
called the range of $A$. The following simple growth model presents a system with no solution.

**Example 5. An Inconsistent System of Equations**

Let $c_1$ be the value of currency on isle 1 and $c_2$ the currency on isle 2. Suppose that the growth equations for currency in the next period are

$$c_1' = .4c_1 + .5c_2$$
$$c_2' = .6c_1 + .5c_2$$

(26)

Consider the problem of picking $c_1$ and $c_2$ so that in the next period, isle 1 has $100$ more and isle 2 has $200$ less: $c_1' = c_1 + 100$ and $c_2' = c_2 - 200$. Then $c_1$, $c_2$ must satisfy the equations

$$c_1 + 100 = c_1' = .4c_1 + .5c_2$$
$$c_2 - 200 = c_2' = .6c_1 + .5c_2$$

or

$$0.6c_1 - 0.5c_2 = -100$$
$$-0.6c_1 + 0.5c_2 = 200$$

(27)

When we pivot on entry $(1, 1)$ in (27), we obtain

$$c_1 - \frac{5}{6}c_2 = \frac{100}{6}$$
$$0 = -100$$

(28)

The second equation in (28) is an impossibility.

A system of equations is called **inconsistent** if, when reduced, it yields an impossible equation. If a system of equations is inconsistent, no solution is possible. Conversely, if no solution is possible, elimination must reveal an inconsistency—otherwise, elimination will produce a solution.

An extreme case of inconsistency arises in regression.

**Example 6. The Regression Model $\hat{y} = qx + r$ as System of Equations to Be Solved**

Suppose that we have the set of $(x, y)$ points $(1, 3)$, $(2, 5)$, $(3, 4)$, $(3, 6)$, $(4, 7)$, and $(4, 6)$. The $x$-value might represent the number of years of college and the $y$-value the score on some graduate admissions test. We want to fit a line of the form $y = qx + r$ to these data. That is, we want the best possible estimates $\hat{y}_i$ for each $y_i$ when $\hat{y}$ is the...
linear function \( qx + r \). Trying to draw a line that actually passes through these six points is equivalent to trying to solve the system of equations

\[
\begin{align*}
3 &= 1q + r \\
5 &= 2q + r \\
4 &= 3q + r \quad \text{or} \quad y = Xq, \quad \text{with} \quad X = \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \\
6 &= 3q + r \\
7 &= 4q + r \\
6 &= 4q + r
\end{align*}
\]

Clearly, no solution is possible here in the regular sense—system (29) is inconsistent. However, an approximate solution to \( y = Xq \) is provided by the least-squares fit of regression theory.

Let us turn our attention to the task of determining the range of a matrix, that is, to finding those \( b \)'s for which the system \( Ax = b \) has a solution.

**Example 7. Range of a Projection Transformation**

In Section 4.1 we observed that the linear transformation

\[
\begin{align*}
x' &= .5x + .5y \\
y' &= .5x + .5y
\end{align*}
\]

projects all \((x, y)\) points onto the line \( x' = y' \): It maps any \([x, y]\) to the point \([(x + y)/2, (x + y)/2]\) (see Figure 4.3).

Let \( A \) be the coefficient matrix in (30). The range of \( A \) is the set of vectors \([x', y']\) such that \( x' = y' \). Let us derive this defining equation for the range directly from \( A \). Letting \( w = [x, y] \) and \( w' = [x', y'] \), we rewrite the transformation \( Aw = w' \) in (30) as \( Aw = Iw' \):

\[
\begin{align*}
.5x + .5y &= 1x' + 0y' \\
.5x + .5y &= 0x' + 1y'
\end{align*}
\]

Now we try to perform elimination by pivoting to convert \( Aw = Iw' \) into the form \( Iw = A^{-1}w' \), as we do when computing the inverse of \( A \). We subtract the first equation from the second (to eliminate \( x \) from the second equation) and divide the first equation by .5 to obtain

\[
\begin{align*}
x + y &= 2x' \\
0 &= -x' + y'
\end{align*}
\]
The second equation in (32), which can be rewritten \( x' = y' \) gives a condition that must be true for any \( w' = [x', y'] \) satisfying \( Aw = Iw' \). We have verified that vectors \([x', y']\) with \( x' = y' \) form the range of \( A \).

**Example 8. The Range of a Refinery Production Problem**

Consider the following variation on our refinery production problem. Now there are just two refineries and their collective output vector \([b_1, b_2, b_3]\) is given by

\[
\begin{align*}
\text{Heating oil:} & \quad 20x_1 + 4x_2 = b_1 \\
\text{Diesel oil:} & \quad 10x_1 + 14x_2 = b_2 \\
\text{Gasoline:} & \quad 5x_1 + 5x_2 = b_3
\end{align*}
\]

(33)

We seek an equation describing possible output vectors. That is, if \( A \) is the coefficient matrix of (33), we seek a defining constraint on the range vectors of \( A \).

We use the technique introduced in Example 7. We write the system \( Ax = b \) in (33) as \( Ax = Ib \) and perform elimination by pivoting at entry (1, 1) and then entry (2, 2) in the augmented matrix \([A | I]\):

\[
\begin{bmatrix}
20 & 4 & 1 & 0 & 0 \\
10 & 14 & 0 & 1 & 0 \\
5 & 5 & 0 & 0 & 1
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0.2 & \frac{1}{20} & 0 & 0 \\
0 & 12 & -\frac{1}{2} & 1 & 0 \\
0 & 4 & -\frac{1}{4} & 0 & 1
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0 & \frac{7}{120} & -\frac{1}{60} & 0 \\
0 & 1 & -\frac{1}{24} & \frac{1}{12} & 0 \\
0 & 0 & -\frac{1}{12} & -\frac{1}{3} & 1
\end{bmatrix}
\]

(34)

The reduced augmented matrix in (34) corresponds to the system of equations

\[
\begin{align*}
x_1 & = \frac{7}{120}b_1 + \frac{1}{60}b_2 \\
x_2 & = -\frac{1}{24}b_1 + \frac{1}{12}b_2 \\
0 & = -\frac{1}{12}b_1 - \frac{1}{3}b_2 + b_3
\end{align*}
\]

(35)

The last equation in (35) can be rewritten as

\[
b_3 = \frac{1}{12}b_1 + \frac{1}{4}b_3 = \frac{1}{12}(b_1 + 3b_2)
\]

(36)

This is the range constraint we were looking for. In terms of refinery production, it means that we can achieve any production vector \( b \) in
which the gasoline production \( b_2 \) equals \( \frac{1}{2} \) the sum of the heating oil production \( b_1 \) plus three times the diesel oil production \( b_2 \).

Suppose that heating oil \( b_1 \) and diesel oil \( b_2 \) are the outputs of primary interest and we want \( b_1 = 300 \) and \( b_2 = 300 \). Then we pick \( b_3 \) using (36) to get a vector in the range. We set

\[
b_3 = \frac{1}{12}(b_1 + 3b_2) = \frac{1}{12}(300 + 3 \cdot 300) = 100
\]

Now we can determine the appropriate production levels \( x_1, x_2 \) from the first two equation in (35). With \( b = [300, 300, 100] \), then

\[
x_1 = \frac{1}{12}b_1 - \frac{1}{60}b_2 = \frac{1}{12} \cdot 300 - \frac{1}{60} \cdot 300
\]
\[
= 17.5 - 5 = 12.5
\]
\[
x_2 = -\frac{1}{24}b_1 + \frac{1}{12}b_2 = -\frac{1}{24} \cdot 300 + \frac{1}{12} \cdot 300
\]
\[
= -12.5 + 25 = 12.5
\]

(Note that some vectors in the range may be infeasible because they would make \( x_1 \) or \( x_2 \) negative.)

Let us try out this method for finding the range of the transition matrix \( A \) in our frog Markov chain.

**Example 9. Range of Frog Markov Chain**

As in Example 8 we try to convert \( Ap = Ip' \) into \( Ip = A^{-1}p' \) using elimination by pivoting. We already attempted this inversion in Example 6 of Section 3.3. We started with the augmented matrix of \( Ap = Ip' \).

\[
\begin{bmatrix}
.50 & .25 & 0 & 0 & 0 & 0 \\
.50 & .50 & .25 & 0 & 0 & 0 \\
0 & .25 & .50 & .25 & 0 & 0 \\
0 & 0 & .25 & .50 & .25 & 0 \\
0 & 0 & 0 & .25 & .50 & .50 \\
0 & 0 & 0 & 0 & .25 & .50 \\
\end{bmatrix}
\]

and ended after pivoting on entries (1, 1), (2, 2), (3, 3), (4, 4), and (5, 5) with

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -2 \\
0 & 0 & 1 & 0 & 0 & 2 \\
0 & 0 & 0 & 1 & 0 & -2 \\
0 & 0 & 0 & 0 & 1 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(37)
Again the last equation in (37) gives the constraint required for a vector $p'$ to be in the range of $A$, namely,

$$0 = -p'_1 + p'_2 - p'_3 + p'_4 - p'_5 + p'_6$$

or

$$p'_2 + p'_4 + p'_6 = p'_1 + p'_3 + p'_5$$

(38)

This is a nice simple formula that allows us to test whether a probability vector $p'$ can be a next-state distribution.

It is left as an exercise for the reader to explain in terms of the frog Markov chain model why even-state probabilities must equal odd-state probabilities.

Before leaving this example, note that we can apply this technique for determining the range not only to $A$ but to powers of $A$. The range of $A^k$ tells us possible distributions for the Markov chain after $k$ periods. What happens as $k$ goes to infinity? The range should contract to multiples of the stable distribution $[.1, .2, .2, .2, .2, .1]$. This behavior is explored in Exercise 25.

To summarize what we now know about the range of an $m$-by-$n$ matrix $A$, elimination by pivoting when applied to $Ax = lb$ results in one of three possibilities [cases 2 and 3 may both apply]:

1. The reduced form of $A$ is $I$. Then for each $b$, $Ax = b$ has a unique solution, and $\text{Null}(A) = 0$.
2. The reduced form of $A$ contains one or more rows of zeros. Then the right sides of these zero rows yield defining constraints that a vector $b$ in the range of $A$ must satisfy.
3. The reduced form of $A$ contains an $m$-by-$m$ $I$ plus additional columns. Then, for each $b$, $Ax = b$ has an infinite number of solutions (the additional columns give rise to the null-space vectors).

Section 5.1 Exercises

Summary of Exercises
Exercises 1–4 involve plotting and graphically solving systems of equations. Exercises 5–17 require finding null spaces and related particular solutions to various matrix systems. Exercises 18–25 require finding a constraint equa-
tion for the range and finding particular range vectors. Exercises 26–28 are some simple theory questions.

1. Plot the lines of solutions to the following systems of equations.
   (a) \(2x + 3y = 10\)  
   (b) \(x - 2y = -4\)  
   \(6x + 9y = 30\)  
   \(-3x + 6y = 12\)

2. Describe and sketch the solution sets, if any, of the following systems of equations.
   (a) \(x + 2y + 3z = 0\)  
   \(4x + 3y + 5z = 0\)  
   \(-4x + 2y + 2z = 0\)
   (b) \(2x + y + z = 3\)  
   \(x - 2y - 3z = 2\)  
   \(3x + 4y + 5z = 3\)
   (c) \(x + 3y - z = 5\)  
   \(2x + y + 2z = 8\)  
   \(3x - y + 5z = 11\)
   (d) \(x + 2y + 3z = 3\)  
   \(x - y + z = 4\)  
   \(2x + 10y + 10z = 8\)
   (e) \(2x + 2y + 2z = 6\)  
   \(x + y - z = 3\)  
   \(-x + y + 3z = -7\)

3. For each of the following systems of equations, plot each line and from your drawing determine whether there is no solution, one solution, or an infinite set of solutions.
   (a) \(3x - 2y = 5\)  
   \(-6x + 4y = 10\)
   (b) \(3x - 2y = 5\)  
   \(2x - 3y = 5\)  
   \(9x - 6y = 15\)

4. For each of the systems of equations in Exercise 2, sketch as best you can the plane determined by each equation. From your sketch, guess whether there is no solution, one solution, or an infinite set of solutions. Verify your guess by solving the system.

5. Give a vector, if one exists, that generates the null space of the following systems of equations or matrices. Which of these seven systems/matrices are invertible? [Consider the coefficient matrix and ignore the particular right-side values in parts (e) and (f).]
   (a) \[
   \begin{bmatrix}
   1 & -2 \\
   -2 & 4
   \end{bmatrix}
   \]
   (b) \[
   \begin{bmatrix}
   4 & -1 & 2 \\
   2 & 5 & 1 \\
   -2 & 3 & -1
   \end{bmatrix}
   \]
   (c) \[
   \begin{bmatrix}
   -1 & 3 & 1 \\
   5 & -1 & 3 \\
   2 & 1 & 2
   \end{bmatrix}
   \]
   (d) \[
   \begin{bmatrix}
   2 & 1 & 1 \\
   1 & -2 & -3 \\
   3 & 4 & 5
   \end{bmatrix}
   \]
   (e) \(x - 2y + z = 6\)  
   \(-x + y - 2z = 4\)
   (f) \(x - y + z = 3\)  
   \(x + y + z = 3\)  
   \(2x - 3y + z = 7\)
(g) The coefficient matrix $X$ in the six regression equations in Example 6.

6. For the following matrices, describe the set of vectors in the null space.

(a) \[
\begin{bmatrix}
-3 & 4 & 1 \\
2 & -1 & 1 \\
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
2 & 1 & 5 & 0 \\
1 & 2 & 4 & -3 \\
1 & 1 & 3 & -1 \\
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 \\
\end{bmatrix}
\]

7. (a) For $A$ in Exercise 6, part (a) and $b = [10, 10]$, if $Ax = b$ has the given solution $x' = [0, 0, 10]$, find the family of all solutions to $Ax = b$.

(b) Find a solution to $Ax = b$ in part (a) with $x_1 = 3$.

8. (a) For $A$ in Exercise 6, part (b) and $b = [30, 30, 20]$, if $Ax = b$ has the given solution $x' = [10, 10, 0, 0]$, find the family of all solutions to $Ax = b$.

(b) Find a solution to $Ax = b$ in part (a) with $x_1 = 5$.

9. (a) For $A$ in Exercise 6, part (c) and $b = [10, 15, 5, 0, 15]$, if $Ax = b$ has the given solution $x' = [5, 0, 5, 0, 5]$, find the family of all solutions to $Ax = b$.

(b) Find a solution to $Ax = b$ in part (a) with $x_4 = 10$ and $x_5 = 10$.

(c) Find a solution to $Ax = b$ in part (a) with $x_1 = 10$ and $x_2 = 5$.

10. Consider the modified refinery system from Example 1:

\[
\begin{align*}
20x_1 & + 4x_2 + 4x_3 = 700 \\
10x_1 & + 14x_2 + 5x_3 = 500
\end{align*}
\]

Given the solution $x_1 = 31$, $x_2 = 10$, $x_3 = 10$, use the appropriate null-space vector to obtain a second solution in which the following is true.

(a) $x_3 = 22$  
(b) $x_2 = 25$  
(c) $x_1 = 28$

11. Consider the following refinery-type problem:

\[
\begin{align*}
10x_1 & + 5x_2 + 5x_3 = 300 \\
5x_1 & + 10x_2 + 8x_3 = 300
\end{align*}
\]
(a) Find the null space of the system.
(b) Given the solution \( x_1 = x_2 = 20, x_3 = 0 \), find a second solution with \( x_3 = 10 \).
(c) Repeat part (b) to find a second solution with \( x_1 = 15 \).

12. The rabbit–fox growth model from Section 1.3,

\[
\begin{align*}
R' &= R + .1R - .15F \\
F' &= F + .1R - .15F
\end{align*}
\]

had stable values for which \( R' = R \) and \( F' = F \) when the monthly change was zero:

\[
\begin{align*}
R &= R' \rightarrow .1R - .15F = 0 \\
F &= F' \rightarrow .1R - .15F = 0
\end{align*}
\]

Solutions for these homogeneous equations were of the form \([R, F] = r[3, 2] \). Suppose we want a vector \([R, F]\) that remains stable when 30 rabbits and 30 foxes are killed each month by hunters.

\[
\begin{align*}
R &= R + .1R - .15F - 30 \rightarrow .1R - .15F = 30 \\
F &= F + .1R - .15F - 30 \rightarrow .1R - .15F = 30
\end{align*}
\]

Find the family of stable population vectors in this case by adding one particular solution to the set of homogeneous solutions. Find a stable vector with \( F = 400 \).

13. Write out a system of equations required to balance the following chemical reaction and solve.

\[
\text{SO}_2 + \text{NO}_3 + \text{H}_2\text{O} \rightarrow \text{H} + \text{SO}_4 + \text{NO}
\]

where S represents sulfur, N nitrogen, H hydrogen, and O oxygen.

14. Write out a system of equations required to balance the following chemical reaction and solve.

\[
\text{PbN}_6 + \text{CrMn}_2\text{O}_8 \rightarrow \text{Cr}_2\text{O}_3 + \text{MnO}_2 + \text{Pb}_3\text{O}_4 + \text{NO}
\]

where Pb represents lead, N nitrogen, Cr chromium, Mn manganese, and O oxygen.

15. Find the null space for these Markov transition matrices.

(a) \[
\begin{bmatrix}
1 & .5 & 0 \\
0 & 0 & 0 \\
0 & .5 & 1
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
.4 & 0 & .2 \\
.3 & .5 & .4 \\
.3 & .5 & .4
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
.5 & 0 & .5 \\
0 & 1 & 0 \\
.5 & 0 & .5
\end{bmatrix}
\]
For each transition matrix $A$ in Exercise 15, find the set of probability vectors $p$ such that the next-period vector $p'$ ($= Ap$) is a stable probability vector (as was done in Example 3).

Prove that if $A$ is the 3-by-3 transition matrix of some Markov chain and the null space of $A$ is infinite, then either two columns of $A$ are equal or else one column is the weighted average of the other two (one-half the sum of the other two).

Give a constraint equation, if one exists, on the vectors in the range of the matrices in Exercise 5.

For the matrix $A$ in Exercise 5, part (a), find a range vector $b$ in which $b_1 = 5$.

For the $b$ in part (a), solve $Ax = b$ (give the family of solutions using Theorem 2).

Give a constraint equation, if one exists, on the vectors in the range of the matrices in Exercise 6.

For matrix $A$ in Exercise 6, part (b), find a range vector $b$ in which $b_1 = 5$ and $b_2 = 2$.

For the $b$ in part (a), solve $Ax = b$ [give the family of solutions using Theorem 1, part (iv)].

For matrix $A$ in Exercise 6, part (c), find a range vector $b$ in which $b_1 = b_2 = b_3 = 15$.

For the $b$ in part (a), solve $Ax = b$ [give the family of solutions using Theorem 1, part (iv)].

Repeat part (a) to find a range vector with $b_1 = 20$, $b_3 = 10$, $b_4 = 20$. 
23. Give a constraint equation, if one exists, on the probability vectors in the range of the transition matrices in Exercise 15.

24. By looking at the transition matrix for the frog Markov chain, explain why the even-state probabilities in the next period must equal the odd-state probabilities in the next period.

*Hint:* Show that this equality is true if we start (this period) from a specific state.

25. Compute the constraint on the range of the following powers of the frog transition matrix. You will need a matrix software package.
   (a) $A^2$  (b) $A^3$  (c) $A^{10}$

26. Show that the range $R(A)$ of a matrix is a vector space. That is, if $b$ and $b'$ are in $R(A)$ (for some $x$ and $x'$, $Ax = b$ and $Ax' = b'$), show that $rb + sb'$ is in $R(A)$, for any scalars $r, s$.

27. Using matrix algebra, show that if $x_1$ and $x_2$ are solutions to the matrix equation $Ax = b$, then any linear combination $x' = cx_1 + dx_2$, with $c + d = 1$, is also a solution.

28. Show that the intersection $V_1 \cap V_2$ of two vector spaces $V_1, V_2$ is again a vector space.

---

**Section 5.2 Theory of Vector Spaces Associated with Systems of Equations**

In this section we introduce basic concepts about vector spaces and use them to obtain important information about the range and null space of a matrix. Recall that a **vector space** $V$ is a collection of vectors such that if $u, v \in V$, then any linear combination $ru + sv$ is in $V$. In Section 5.1 we introduced the range and null space of a matrix $A$:

$$
\text{Range}(A) = \{b : Ax = b \text{ for some } x\} \\
\text{Null}(A) = \{x : Ax = 0\}
$$

We noted that $\text{Range}(A)$ and $\text{Null}(A)$ are both vector spaces.

In Examples 2, 3, and 4 of Section 5.1 we used the elimination process to find a vector or pair of vectors that generated the null spaces of certain matrices. For example, multiples of $[-1, 2, -2, 2, -2, 1]$ formed the null space of the frog Markov transition matrix. In Examples 7, 8, and 9 of Section 5.1, we used elimination to find constraint equations that vectors in the range must satisfy. For the frog Markov matrix, the constraint for range vectors $p$ was $p_1 + p_3 + p_5 = p_2 + p_4 + p_6$. 
The number of vectors generating the null space and number of constraint equations for the range were dependent on how many pivots we made during elimination. Our goal in this section is to show that the sizes of $\text{Null}(A)$ and $\text{Range}(A)$ are independent of how elimination is performed.

The vector space $V$ generated by a set $Q = \{q_1, q_2, \ldots, q_r\}$ of vectors is the collection of all vectors that can be expressed as a linear combination of the $q_i$'s. That is,

$$V = \{v: v = r_1q_1 + r_2q_2 + \cdots + r_rq_r, r_i \text{ scalars}\}$$

For example, if $Q$ consists of the unit $n$-vectors $e_j$ (with all 0's except for a 1 in position $j$), then $V$ is the vector space of all $n$-vectors, that is, euclidean $n$-space. Another name for a generating set is a spanning set.

The column space of $A$, denoted $\text{Col}(A)$, is the vector space generated by the column vectors $a_1^C, \ldots, a_n^C$ of $A$. When we write $Ax = b$ as

$$a_1^Cx_1 + a_2^Cx_2 + \cdots + a_n^Cx_n = b$$

we see that the system $Ax = b$ has a solution if and only if $b$ can be expressed as a linear combination of the column vectors of $A$, or

**Lemma 1.** The system $Ax = b$ has a solution if and only if $b$ is in $\text{Col}(A)$. Equivalently, $\text{Col}(A) = \text{Range}(A)$.

The components $x_i$ of the solution $x$ give the weights in the linear combination of columns that yield $b$. Note that Lemma 1 is true for any $m$-by-$n$ matrix $A$ and any $m$-vector $b$.

**Example 1. Refinery Problem as a Column Space Problem**

The refinery problem introduced in Section 1.2 involved three refineries each producing different amounts of heating oil, diesel oil, and gasoline from a barrel of crude oil. Production levels of each refinery were sought to satisfy a vector of demands. The resulting system of equations was

\[
\begin{align*}
\text{Heating oil:} & \quad \quad 20x_1 + 4x_2 + 4x_3 = 500 \\
\text{Diesel oil:} & \quad \quad 10x_1 + 14x_2 + 5x_3 = 850 \\
\text{Gasoline:} & \quad \quad 5x_1 + 5x_2 + 12x_3 = 1000
\end{align*}
\]

But this system is just seeking to express the demand vector $[500, 850, 1000]$ as a linear combination of the production vectors of the three refineries. That is, we seek $x_1, x_2, x_3$ such that

$$x_1 \begin{bmatrix} 20 \\ 10 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 14 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ 5 \\ 12 \end{bmatrix} = \begin{bmatrix} 500 \\ 850 \\ 1000 \end{bmatrix}$$
Sec. 5.2 Theory of Vector Spaces Associated with Systems of Equations

Up to this point, solving a system \( Ax = b \) was viewed as a problem about the rows of \( A \), that is, about the equations specified by the rows. Gaussian elimination involves forming linear combinations of the equations (rows of \( A \)) to obtain a new reduced system that can be solved by back substitution. Lemma 1 says that solving \( Ax = b \) can equally be viewed as a problem about a linear combination of the columns of \( A \). This vector approach to solving \( Ax = b \) has an associated geometric picture.

**Example 2. Geometric Picture of Solution to a System of Equations**

Consider the system of equations

\[
\begin{align*}
x_1 + x_2 &= 4 \\
x_1 - 2x_2 &= 1
\end{align*}
\]

or

\[
\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 1 \\ -2 \end{bmatrix} x_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}
\]

Solving by elimination, we find that \( x_1 = 3 \) and \( x_2 = 1 \). Figure 5.3 graphs this solution in vector-space terms, showing the right-side vector \([4, 1]\) as a linear combination of the column vectors \([1, 1]\) and \([1, -2]\). Note that the picture gives no insight into why \( x_1 = 3 \), \( x_2 = 1 \) is the solution.

To determine the size of \( \text{Range}(A) \), we analyze the structure of \( \text{Col}(A) \), the column space of \( A \), which by Lemma 1 equals \( \text{Range}(A) \). The key question is: How many of the columns of \( A \) are actually needed to generate \( \text{Col}(A) \). Some columns in \( A \) may be redundant.

A set of vectors \( a_1, a_2, \ldots, a_i \) is called **linearly dependent** if one of them can be expressed as a linear combination of the others. Another way to say this is that there is a nonzero solution \( x \) to

\[
x_1a_1 + x_2a_2 + \cdots + x_ia_i = 0 \quad \text{or, equivalently,} \quad Ax = 0 \quad (1)
\]

Figure 5.3
where \( A \) is the matrix whose columns are the vectors \( \mathbf{a}_i \). Linear dependence is equivalent to (1) because if \( x_i \neq 0 \), we can rewrite (1) as

\[
-x_i \mathbf{a}_i = x_1 \mathbf{a}_1 + \cdots + x_n \mathbf{a}_n
\]
or

\[
\mathbf{a}_i = - \frac{x_1}{x_i} \mathbf{a}_1 - \frac{x_2}{x_i} \mathbf{a}_2 - \cdots - \frac{x_n}{x_i} \mathbf{a}_n
\]

So any \( \mathbf{a}_i \), for which \( x_i \neq 0 \), can be written as a linear combination of the other \( \mathbf{a}'s \).

A set of vectors are linearly independent if they are not linearly dependent. For example, the columns of an identity matrix

\[
I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]

are linearly independent. If vectors \( \mathbf{a}_i \) are linearly independent, then the only solution to \( x_1 \mathbf{a}_1 + \cdots + x_n \mathbf{a}_n = \mathbf{0} \) (i.e., \( Ax = \mathbf{0} \)) can be \( \mathbf{x} = \mathbf{0} \).

**Example 3. Example of a Linearly Dependent Set of Columns**

Consider the matrix \( A = \begin{bmatrix} 1 & 1 & 4 \\ 1 & -2 & 1 \end{bmatrix} \). By inspection we see that

\[
\mathbf{a}_3 = 3 \mathbf{a}_1 + \mathbf{a}_2: \begin{bmatrix} 4 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \end{bmatrix}
\]

So the columns of \( A \) are linearly dependent.

The following method illustrates a systematic way to find this linear dependence. We perform elimination by pivoting on \( A \). We pivot on entry \((1, 1)\) and then on \((2, 2)\):

\[
\begin{bmatrix} 1 & 1 & 4 \\ 1 & -2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 4 \\ 0 & -3 & -3 \end{bmatrix} \rightarrow A^* = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix}
\]

where \( A^* \) represents the reduced form of \( A \).

Remember that \( \mathbf{x} \) is a solution to \( Ax = \mathbf{0} \) if and only if \( \mathbf{x} \) is a solution to \( A^* \mathbf{x} = \mathbf{0} \). Equivalently, there is a linear dependence among the columns of \( A \) if and only if there are is linear dependence among the columns of \( A^* \). But the first two columns of \( A^* \) are unit vectors (they form the 2-by-2 identity matrix). So trivially, the third column of \( A^* \) is dependent on the first two:
This is exactly the relationship that we found in (2).

If we were to apply the method in Example 3 to the coefficient matrix in the refinery problem of Example 1, elimination by pivoting would reduce the matrix to a 3-by-3 identity matrix. Since the columns of the identity matrix are trivially linearly independent, this means that the original columns in the refinery problem were linearly independent.

We state the method used to find linear dependence in Example 3 and its consequences as a theorem.

**Theorem 1**

(i) Let $A$ be any $m$-by-$n$ matrix and let $A^*$ be the reduced matrix obtained from $A$ using elimination by pivoting. Then a set of columns of $A$ is linearly dependent (linearly independent) if and only if the corresponding columns in $A^*$ are linearly dependent (linearly independent).

(ii) Any unpivoted column of $A^*$ (a column that was not reduced to a unit vector) is linearly dependent on the set of columns containing pivots.

(iii) The columns of $A^*$ with pivots are linearly independent. The corresponding columns of $A$ generate the column space of $A$.

The following example illustrates this method further.

**Example 4. Redundant Columns in Transportation Problem Constraints**

In Example 4 of Section 5.1 we examined the following system of equations (that were transportation problem constraints seen in Section 4.6).

$$
\begin{align*}
    x_1 + x_2 &= 20 \\
    x_3 + x_4 &= 30 \\
    x_5 + x_6 &= 15 \\
    x_1 + x_3 + x_5 &= 25 \\
    x_2 + x_4 + x_6 &= 40
\end{align*}
$$

with coefficient matrix $A$

$$
A = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
$$

(4)
In Section 5.1 we performed elimination by pivoting on \( A \) using entries (1, 2), (2, 4), (3, 6) and (4, 3). The reduced matrix \( A^* \) was

\[
A^* = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]  

(5)

First note that the last row of zeros can be ignored. Columns 2, 4, 6, and 3 of \( A^* \) (in that order) are the unit vectors of the 4-by-4 identity matrix. Columns 1 and 5 of \( A^* \) are each linearly dependent on the four unit-vector columns. For example,

\[
a_1^* = a_2^* - a_4^* + a_3^*:
\]

\[
\begin{bmatrix}
1 \\
-1 \\
0 \\
1 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
- 
\begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}
+ 
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
1
\end{bmatrix}
\]  

(6)

The relation (6) among columns in \( A^* \) is mirrored in \( A \), where

\[
a_1 = a_2 - a_4 + a_3: 
\]

\[
\begin{bmatrix}
1 \\
0 \\
1 \\
0 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
- 
\begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}
+ 
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
1
\end{bmatrix}
\]

Thus the columns where pivots were performed, columns \( a_2, a_3, a_4, \) and \( a_6 \), generate the range of \( A \).

From Theorem 1, part (iii), it follows that the number of pivots performed equals the number of linearly independent columns that generate the column space of \( A \).

A basis of a vector space \( V \) is a minimal-sized set of vectors that generate \( V \). Implicit in this definition is the fact that a basis is a set of linearly independent vectors. As an example, the \( n \) coordinate vectors \( e_j \) form a basis for the space of all \( n \)-dimensional vectors. Since a basis is a minimal-sized generating set, every generating set contains a basis. For example, while the column space of a matrix \( A \) is defined to be generated by the columns of \( A \), only the pivot columns are needed to generate the column space, as shown in Example 4.

The following result, which we prove in two ways, shows the theo-
Proposition. If \( \{v_1, v_2, \ldots, v_n\} \) is a basis for a vector space \( V \), every vector in \( V \) has a unique representation as a linear combination of the \( v_i \)'s.

Proof 1 (Using the Definition of Linear Independence): Suppose that \( w \) is a vector in \( V \) that has two representations as a linear combination of the \( v_i \). So

\[
\begin{align*}
w &= a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \\
w &= b_1 v_1 + b_2 v_2 + \cdots + b_n v_n
\end{align*}
\]

Then

\[
0 = w - w = (a_1 - b_1) v_1 + (a_2 - b_2) v_2 + \cdots + (a_n - b_n) v_n
\]

Since the \( v_i \)'s form a basis and hence are linearly independent, the linear combination of \( v_i \)'s in (7) can only equal \( 0 \) if the terms \( (a_i - b_i) \) are all zero. So the two representations must be the same.

Proof 2 (Using Elimination). To find the representation of \( w \) in terms of the \( v_i \)'s, we solve the system of equations for the \( x_i \):

\[
x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = w \quad \text{or} \quad A x = w
\]

where \( A \) has the \( v_i \)'s as its columns. Since the \( v_i \)'s are a basis and hence linearly independent, we can pivot in every column [otherwise, by Theorem 1, part (ii), each unpivoted column is linearly dependent on the pivot columns]. Then \( A x = w \) has a unique solution (see the summary at the end of Section 5.1).

Proof 2 shows us how to compute the unique representation of a vector \( w \) in terms of the \( v_i \)'s, simply solve (8). For example, if \( v_1 = [1, 2, 3] \) and \( v_2 = [0, -1, 2] \) are a basis for vector space \( V \) and \( w = [3, 8, 5] \) is in \( V \), then to determine the right linear combination of the \( v_i \)'s to get \( w \), we solve the system

\[
\begin{bmatrix}
1 & 0 \\
2 & -1 \\
3 & 2
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix}
3 \\
8 \\
5
\end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 0
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix}
3 \\
-2 \\
0
\end{bmatrix}
\]

or \( x_1 = 3, x_2 = -2 \).
Now we prove a critical vector-space lemma that resolves the question of whether all pivot sequences have the same length.

**Lemma 2.** All linearly independent sets of vectors that generate a given vector space $V$ have the same size. Any such set is a basis for $V$.

**Proof.** Let $S = \{s\}$ and $T = \{t\}$ be two sets of linearly independent vectors that generate $V$. Suppose that $S$ and $T$ have different sizes; for concreteness, let $S$ have four vectors and $T$ have five vectors. Then we can use linear combinations of the vectors of $S$ to represent the vectors in $T$. If $t_1 = c_1 s_1 + c_2 s_2 + c_3 s_3 + c_4 s_4$, define the $S$-coordinate vector of $t_1$ to be $[c_1, c_2, c_3, c_4]$. Consider the equation, defined in (1), for dependence of the $t_i$ (with the $t_i$ represented in $S$-coordinates):

$$x_1 t_1 + x_2 t_2 + x_3 t_3 + x_4 t_4 + x_5 t_5 = 0 \quad (10)$$

Since the $t_i$ are four-dimensional vectors (in $S$-coordinates), (10) is a system with four equations in five variables. Solving (10) by elimination by pivoting leaves at least one unpivoted column and hence by Theorem 1, part (ii), there is linear dependence among the $5 t_i$. Contradiction.

The **dimension** of a vector space $V$, written $\dim(V)$, is the number of vectors in a basis for $V$. For example, the set of $n$ unit vectors $e_j$ is a basis for $n$-dimensional space”; thus this space does indeed have dimension $n$.

Combining Lemma 2 with Theorem 1, part (iii), we have

**Theorem 2**

(i) The columns of $A$ used in a pivot sequence are a basis for the range of $A$.

(ii) All pivot sequences have the same size; the size is the dimension of the range of $A$.

By Theorem 2, it now makes sense to talk about the number of pivots in a pivot sequence. The **rank** of a matrix $A$, written $\text{rank}(A)$, is the number of pivots in any pivot sequence.

**Corollary**

$$\text{Rank}(A) = \dim(\text{Range}(A))$$

We have been concerned about which sets of columns of $A$ are linearly dependent, that is, when there is a nonzero $x$ so that

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0 \quad \text{or} \quad Ax = 0 \quad (11)$$

Such an $x$ in (11) is a vector in $\text{Null}(A)$, the null space of $A$.

If $A^*$ is the reduced matrix, then we know that $x$ is a solution of
Sec. 5.2 Theory of Vector Spaces Associated with Systems of Equations

Ax = 0 if and only if it is a solution to $A^*x = 0$. Thus Null($A^*$) equals Null($A$).

**Example 5. Relation Between Null Space and Column Space**

Consider the matrix

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

and perform elimination by pivoting.

Pivoting on entry (1, 1) yields

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

Pivoting on entry (2, 2) yields

$$A^* = \begin{bmatrix}
1 & 0 & 1 & -1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix}$$

Clearly, the first two columns of $A^*$, the pivot columns, are a basis for the column space of $A^*$—they are linearly independent and generate Col($A^*$). Then by Theorem 1, part (iii), the first two columns in $A$ generate Col($A$).

Since Null($A^*$) = Null($A$), a basis for Null($A^*$) will be a basis for Null($A$). Looking at $A^*$, we see that

$$a_3^* = a_1^* + a_2^* \quad \text{and} \quad a_4^* = -a_1^* + a_2^* \quad (12)$$

(Where $a_i^*$ denotes the $i$th column of $A^*$). The vector equations in (12) can be rewritten as

$$-a_1^* - a_2^* + a_3^* = 0 \quad \text{or} \quad A^* \begin{bmatrix}
-1 \\
-1 \\
1 \\
0
\end{bmatrix} = 0 \quad (13)$$

$$a_1^* - a_2^* + a_4^* = 0 \quad \text{or} \quad A^* \begin{bmatrix}
1 \\
-1 \\
0 \\
1
\end{bmatrix} = 0$$
Let \( x^*_3 = [-1, -1, 1, 0] \) and \( x^*_4 = [1, -1, 0, 1] \). Since \( x^*_3 \) and \( x^*_4 \) are linearly independent (look at their last two entries) and generate \( \text{Null}(A^*) \), they form a basis for \( \text{Null}(A^*) \) and hence for \( \text{Null}(A) \).

In general, given a reduced matrix \( A^* \), each unpivoted column can be expressed as a linear combination of the pivoted columns, which are unit vectors in \( A^* \) [see (12)]. This linear combination yields a solution to \( A^*x = 0 \), as shown in (13). If unpivoted column \( h \) of \( A^* \) has an entry \( a^*_{ih} \) in the \( i \)th row, the solution \( x^*_h \) we obtain is

\[
x^*_h = [-a^*_{1h}, -a^*_{2h}, \ldots, -a^*_{mh}, 0, 0, \ldots, 1, \ldots, 0]
\]

where the entries for unpivoted columns are all 0 except for entry \( h \), which is 1 (assuming pivots were performed in the first \( m \) rows and \( m \) columns).

For example, in Example 5, the fourth column of \( A^* \) begins

\[
\begin{bmatrix}
-1 \\
1 \\
\vdots \\
\end{bmatrix}
\]

(in the two pivot rows), so \( x^*_4 = [1, -1, 0, 1] \). The entries in (14) from the unpivoted columns form a unit vector, so the set of \( x^*_h \)'s are linearly independent and form a basis of the null space of \( A \).

Observe that every column in \( A^* \) is now either (i) a unit vector that is in the basis of the column space; or else (ii) gives rise to a vector in the basis of the null space. That is, every column contributes to the size of the range of \( A \) or to the diversity of different solutions possible to \( Ax = b \), for a given \( b \).

**Theorem 3.** Let \( A \) be a matrix with \( n \) columns. The vectors \( x^*_h \) in (14) corresponding to unpivoted columns form a basis for \( \text{Null}(A) \). Furthermore,

\[
\dim(\text{Range}(A)) + \dim(\text{Null}(A)) = n
\]

**Corollary A**

\[
\text{Dim}(\text{Null}(A)) = n - \dim(\text{Range}(A)) = n - \text{rank}(A)
\]

**Corollary B.** Any solution \( x' \) to \( Ax = b \) can be written in the form

\[
x' = x^* + r_1x^*_1 + r_2x^*_2 + \cdots + r_kx^*_k
\]

where \( x^* \) is a given particular solution to \( Ax = b \) and the \( x^*_h \)'s are as given in (14).
Proof of Corollary B. By Theorem 1, part (iv) of Section 5.1, any solution \( x' \) can be written \( x' = x^* + x^0 \), the sum of a particular solution \( x^* \) and some null space vector \( x^0 \). Since the \( x_h^* \) generate \( \text{Null}(A) \), any such \( x^0 \) is some linear combination of the \( x_h^* \)'s.

We complete our brief survey of vector spaces of \( A \) with \( \text{Row}(A) \), the vector space generated by the rows of \( A \). As noted when elimination was introduced in Section 3.2, the elimination process repeatedly replaces a row with a linear combination of rows. When rows are zeroed out in the elimination process, they are linearly dependent on the preceding rows *in which pivots were performed*. Conversely, every nonzero row in \( A^* \) is a pivot row (where a pivot was performed).

Because the submatrix of \( A^* \) formed by the pivot rows and pivot columns is an identity matrix, these pivot rows are linearly independent (see Exercises for details) and will be shown shortly to form a basis for \( \text{Row}(A) \). Hence the dimension of the row space equals \( \text{rank}(A) \) (= number of pivots).

**Theorem 4.** Let \( A \) be any \( m \)-by-\( n \) matrix. The maximum number of linearly independent rows in \( A \) and the maximum number of linearly independent columns in \( A \) are equal. Both are \( \text{rank}(A) \). That is,

\[
\dim(\text{Row}(A)) = \text{rank}(A) = \dim(\text{Col}(A))
\]

The results in Theorems 2, 3, and 4 yield several more equivalent conditions for when a system of equations has a unique solution.

**Theorem 5.** Let \( A \) be an \( n \)-by-\( n \) matrix. The system \( Ax = b \) has a unique solution, for any \( b \), if and only if any of the following equivalent conditions are satisfied.

(i) The dimension of \( \text{Range}(A) \) is \( n \).
(ii) The column vectors of \( A \) are linearly independent.
(iii) The dimension of \( \text{Row}(A) \) is \( n \).
(iv) The row vectors of \( A \) are linearly independent.
(v) The null space of \( A \) has dimension 0 (consists of only the 0 vector).

The following example illustrates the uses of Theorem 5.

**Example 6. Row Space Test for Unique Solution**

Let us consider the following variation of our refinery model introduced in Section 1.2. Suppose that we change the numbers in gasoline production so that the third row is the sum of the first two rows.

| Heating oil: | 20x_1 + 4x_2 + 4x_3 = 500 |
| Diesel oil:   | 10x_1 + 14x_2 + 5x_3 = 850 |
| Gasoline:     | 30x_1 + 18x_2 + 9x_3 = 1000 |
Once we observe that the last row in the coefficient matrix is the sum (a linear combination) of the first two rows, so that the rows are not linearly independent, then we know by Theorem 5, part (iv), that there will not be a unique solution to this production problem: either no solution or multiple solutions. If we tried to perform elimination, we would only be able to pivot twice (if we could pivot in all three rows, they would have to be linearly independent).

Theorem 4 tells us that if the rows are dependent, the columns also are. However, that column dependence is far from obvious.

We now give a theoretical application of Theorems 3 and 4. Suppose that \( A \) is \( m \times n \), where \( m < n \). All the columns cannot be linearly independent, since \( \dim(\text{Col}(A)) = \dim(\text{Row}(A)) \) and there are only \( m \) rows, \( m < n \). Therefore, \( \text{rank}(A) \leq m < n \). We conclude from Theorem 3, \( \dim(\text{Null}(A)) = n - \text{rank}(A) = n - \dim(\text{Row}(A)) > 0 \). Then \( \text{Null}(A) \) is infinite and \( Ax = b \) cannot have a unique solution:

**Theorem 6.** If \( A \) is an \( m \times n \) matrix, where \( m < n \), the system \( Ax = b \) can never have a unique solution (either multiple solutions or no solution).

We close this section with a discussion of another way to interpret the elimination process and the rank of a matrix. To do this, we must introduce the concept of a simple matrix.

A simple matrix \( K \) is formed by the product \( c \times d \) of two vectors \( c \) and \( d \) in which \( c \) is treated as an \( m \times 1 \) matrix and \( d \) as a \( 1 \times n \) matrix. Thus entry \( k_{ij} \) of \( K \) equals \( a_{ij}b_j \). We refer to this product \( c \times d \) as a matrix product of vectors. For example, the following matrix product of vectors yields a simple matrix.

\[
[3, \ -1] \times [1, 2, 3] = \begin{bmatrix} 3 \\ -1 \end{bmatrix} [1, 2, 3] = \begin{bmatrix} 3 \cdot 1 & 3 \cdot 2 & 3 \cdot 3 \\ -1 \cdot 1 & -1 \cdot 2 & -1 \cdot 3 \end{bmatrix}
\]

All rows in a simple matrix are multiples of each other, and similarly for columns. If we pivot on an entry \( (i, j) \) in a simple matrix, the elimination computation will convert all other rows to 0’s (verification is left as an exercise). This means that simple matrices have rank 1.

Simple matrices will be used extensively in Section 5.5. For now, the property of simple matrices of interest is

**Theorem 7.** Let \( C \) be a \( m \times r \) matrix with columns \( e_i \) and \( D \) be an \( r \times n \) matrix with rows \( d_j \). Then the matrix multiplication \( CD \) can be
decomposed into a sum of the simple matrices $c_j^C \ast d_i^R$ of the column vectors of $C$ times the row vectors of $D$.

$$CD = c_1^C \ast d_1^R + c_2^C \ast d_2^R + \cdots + c_r^C \ast d_r^R$$  \hspace{1cm} (17)

One way to verify (17) is using the rules for partitioned matrices, that is, we partition $C$ into $r$ $m$-by-$1$ matrices (the $c_i^C$) and partition $D$ into $r$ $1$-by-$n$ matrices (the $d_i^R$); see Exercise 15 of Section 2.6 for details.

We illustrate this theorem with the following product of two matrices.

**Example 7. Decomposition of Matrix Multiplication into a Sum of Simple Matrices**

Let

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 11 & 12 & 13 \\ 14 & 15 & 16 \\ 17 & 18 & 19 \end{bmatrix}$$

Then

$$CD = \begin{bmatrix} 1 \times 11 & + & 2 \times 14 & + & 3 \times 17 \\ 4 \times 11 & + & 5 \times 14 & + & 6 \times 17 \end{bmatrix} \begin{bmatrix} 1 \times 12 & + & 2 \times 15 & + & 3 \times 18 \\ 4 \times 12 & + & 5 \times 15 & + & 6 \times 18 \end{bmatrix} \begin{bmatrix} 1 \times 13 & + & 2 \times 16 & + & 3 \times 19 \end{bmatrix}$$  \hspace{1cm} (18a)

$$= \begin{bmatrix} 1 \times 11 & 1 \times 12 & 1 \times 13 \\ 4 \times 11 & 4 \times 12 & 4 \times 13 \end{bmatrix} + \begin{bmatrix} 2 \times 14 & 2 \times 15 & 2 \times 16 \\ 5 \times 14 & 5 \times 15 & 5 \times 16 \end{bmatrix} + \begin{bmatrix} 3 \times 17 & 3 \times 18 & 3 \times 19 \\ 6 \times 17 & 6 \times 18 & 6 \times 19 \end{bmatrix}$$  \hspace{1cm} (18b)

$$= \begin{bmatrix} 1 \\ 4 \end{bmatrix} \ast \begin{bmatrix} 11 & 12 & 13 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} \ast \begin{bmatrix} 14 & 15 & 16 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \end{bmatrix} \ast \begin{bmatrix} 17 & 18 & 19 \end{bmatrix}$$  \hspace{1cm} (18c)

$$= c_1^C \ast d_1^R + c_2^C \ast d_2^R + c_3^C \ast d_3^R$$

The first simple matrix $c_1^C \ast d_1^R$ in (18c) is a matrix containing the first term of each scalar product in the entries of $CD$ in (18a). Similarly for the second and third simple matrices.

We now show how an $m$-by-$n$ matrix $A$ can be decomposed into a sum of $k$ simple matrices, where $k = \text{rank}(A)$. Another way to say this is that we subtract a set of simple matrices from $A$ to eliminate all entries in $A$ (to reduce $A$ to the $O$ matrix).

Our strategy will be to form a simple matrix $K_1 = I_1 \ast u_1$ whose first row equals $a_1^R$ (the first row of $A$) and whose first column equals $a_1^C$ (the first column of $A$). Then $A - K_1$ will have $0$'s in its first row and column. We form $K_2$ to remove the second row and column of $A$; possibly we zero out additional rows and columns in the process. We continue similarly with $K_3$, and so on.

Let $u_1 = a_1^R$ and let
\[ I_1 = \left( \frac{1}{a_{11}} \right) a_1^C = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{11} & a_{11} & a_{11} \end{bmatrix} \]

(actually the first entry of \( I_1 \) is \( 1 = a_{11}/a_{11} \)). Then the entries of first row of \( I_1 \cdot u_1 \) are 1 (first entry of \( I_1 \)) times \( u_1 \), which equals \( u_1 = a_1^R \), as required. And the entries in the first column of \( I_1 \cdot u_1 \) are \( I_1 \) times the first entry of \( u_1 \), \( a_{11} \). We have

\[
\begin{bmatrix}
1 \\
a_{21} \\
a_{11} \\
a_{31} \\
a_{11}
\end{bmatrix}
* \begin{bmatrix}
a_{11} \\
a_{12} \\
a_{13} \\
a_{21} \\
a_{31}
\end{bmatrix}
= \begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & \cdots & \cdots \\
a_{31} & \cdots & \cdots
\end{bmatrix}
\]

So \( A - K_1 (= A - I_1 \cdot u_1) \) has zeros in the first row and column. Observe that outside the first row and column, the new entry \((i, j)\) in \( A - K_1 \) equals

\[ a_{ij} - \left( \frac{a_{1j}}{a_{11}} \right) a_{ij} \]

Surprise! This is our old friend the elimination operation [when we pivot on entry \((1, 1)\)]. Vector \( I_1 \) is just the first column in the matrix \( L \) of elimination multipliers from the \( A = LU \) decomposition, and \( u_1 \) is the first row of \( U \) (which equals the first row of \( A \)). So we have shown that when we subtract from \( A \), the simple matrix \( I_1^C \cdot u_1^R \) formed by \( I_1^C \) (the first column of \( L \)) and \( u_1^R \) (the first row of \( U \)), we obtain a matrix with 0’s in the first row and first column. This new matrix is just the coefficient matrix (ignoring the first row of 0’s) for the remaining \( n - 1 \) equations in \( Ax = b \) when we pivot on entry \((1, 1)\).

Repeating this argument, we let \( K_2 = I_2^C \cdot u_2^R \) and subtracting \( K_2 \) from \( A - K_1 \) will have the effect of next pivoting on entry \((2, 2)\), and zeroing out the second row and column. The other \( K_i \) are defined and perform similarly, so ultimately we see that

\[ A = I_1^C \cdot u_1^R + I_2^C \cdot u_2^R + \cdots + I_m^C \cdot u_m^R \quad (19) \]

**Example 8. Refinery Matrix Expressed as a Sum of Simple Matrices**

In Section 3.2 we gave the \( LU \) decomposition of our refinery matrix

\[
A = \begin{bmatrix}
20 & 4 & 4 \\
10 & 14 & 5 \\
5 & 5 & 12
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
\frac{1}{2} & 1 & 0 \\
\frac{1}{1} & \frac{1}{3} & 1
\end{bmatrix}\begin{bmatrix}
20 & 4 & 4 \\
0 & 12 & 3 \\
0 & 0 & 10
\end{bmatrix}
\]
By (19), we can write $A$ as

$$A = I_1^c * u_1^R + I_2^c * u_2^R + I_3^c * u_3^R$$

$$= \begin{bmatrix} 1 \\ \frac{1}{4} \end{bmatrix} * \begin{bmatrix} 20 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{3} \end{bmatrix} * \begin{bmatrix} 4 \\ 12 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} * \begin{bmatrix} 4 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 4 & 4 \\ 10 & 2 & 2 \\ 5 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 12 & 3 \\ 0 & 4 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The reader should check that this set of three simple matrices adds up to $A$. 


**Theorem 8.** Gaussian elimination can be viewed as a decomposition of $A$ into a sum of rank($A$) simple matrices:

$$A = I_1^c * u_1^R + I_2^c * u_2^R + \cdots + I_k^c * u_k^R$$

(21)

where $k = \text{rank}(A)$, $I_i^c$ is the $i$th column of $L$ (the matrix of elimination multipliers), and $u_i^R$ is the $i$th row of $U$ (the reduced matrix in Gaussian elimination).

The minimum number of simple matrices whose sum equals matrix $A$ is rank($A$).

The last sentence of Theorem 8 is proved in Exercise 32. The symmetric role of columns and rows in Theorem 8 explains why the dimensions of the row and column spaces of a matrix are equal.

It is not hard to show (see Exercise 31) that if $A$ has the simple-matrix decomposition $A = c_1 * d_1 + \cdots + c_k * d_k$, then the $c_i$ are a basis for the column space of $A$ and the $d_i$ are a basis for the row space of $A$. It follows that

**Corollary**

(i) The nonzero rows of $U$ generate the row space of $A$ and the nonzero (below main diagonal) columns of $L$ generate the column space of $A$.

(ii) Theorem 8 reproves the fact that the dimension of the column space of $A$ equals the dimension of the row space of $A$, equals rank($A$).

From Theorem 7 it follows if $A$ equals the sum of simple matrices $I_i^c * u_i^R$, then $A = LU$—we have proved the $LU$ decomposition.

The decomposition (19) of a matrix $A$ into a sum of rank($A$) simple matrices is of more theoretical than practical interest.
Optional (Based on Section 4.6)

Let us reinterpret the simplex algorithm of linear programming using the concept of a basis for the column space. When slack variables were added, as say to the table–chair production problem in Section 4.6, the form of the constraint equations was

$$
\begin{bmatrix}
40 & 200 & 0 & 0 & 0 & 0 \\
1 & 4 & 1 & 0 & 0 & 0 | 1400 \\
2 & 3 & 0 & 1 & 0 & 0 | 2000 \\
1 & 12 & 0 & 0 & 1 & 0 | 3600 \\
2 & 0 & 0 & 0 & 0 & 1 | 1800 \\
\end{bmatrix}
$$

(22)

Observe that the columns associated with the slack variables \(x_3, x_4, x_5, x_6\) form an identity matrix and hence are the basis for the column space. For this reason, \(x_3, x_4, x_5, x_6\) are called basic variables and variables \(x_1, x_2\) nonbasic, for the linear program (22): Clearly, the columns of nonbasic variables \(x_1, x_2\) in (22) are linearly dependent on the basic variables’ columns. Recall that the simplex algorithm sets the nonbasic variables equal to 0 so that the basic variables then have nonnegative values equal to the corresponding right-side entry.

The pivot step in the simplex algorithm can be viewed as picking some nonbasic variable to enter the basis while a basic variable leaves in the basis. For (22), we chose \(x_2\) (whose coefficient 200 in the objective function is largest) to enter and \(x_5\) to leave the basis. After pivoting, we have

$$
\begin{bmatrix}
\frac{7}{3} & 0 & 0 & 0 & \frac{50}{3} & 0 & -60,000 \\
\frac{1}{3} & 1 & 0 & -\frac{1}{3} & 0 & | 200 \\
\frac{1}{4} & 0 & 0 & 1 & -\frac{1}{4} & 0 & | 1,100 \\
\frac{1}{12} & 1 & 0 & 0 & \frac{1}{12} & 0 & | 300 \\
2 & 0 & 0 & 0 & 1 & 0 & | 1,800 \\
\end{bmatrix}
$$

(23)

Note that now the columns of \(x_2, x_3, x_4, x_6\) form the basis for the column space.

Whereas our discussion in Section 4.6 focused on which were the independent (nonbasic) variables, the traditional approach is to concentrate on which are the basic variables.

Section 5.2 Exercises

Summary of Exercises

Exercises 1–10 involve the column space, linear dependence, and generators of the column space. Exercises 11–25 involve associated theory. Exercises 26–32 involve simple matrices and the representation of a matrix as a sum of simple matrices.
1. Three soup factories $F_1$, $F_2$, and $F_3$ generate production vectors, $\mathbf{f}_1 = [20, 100, 20]$, $\mathbf{f}_2 = [200, 0, 50]$, and $\mathbf{f}_3 = [0, 100, 200]$, of the amounts (in gallons) of tomato, chicken, and split-pea soup produced each hour. If the demand is $\mathbf{d} = [5000, 3000, 3000]$, write a system of equations for determining the right linear combination (how long each factory should work) of production vectors to meet the demand. Determine the weights.

2. There are three refineries producing heating oil, diesel oil, and gasoline. The production vector of refinery $A$ (per barrel of crude oil) is $[10, 5, 10]$ and of refinery $B$ is $[4, 11, 8]$. The production vector for refinery $C$ is the average of the vector for refineries $A$ and $B$. If the demand vector is $[380, 370, 460]$, write a system of equations for finding the right linear combination of refinery production vectors to equal the demand vector. Find the set of such linear combinations.

3. For each of the following sets of vectors, express the first vector as a linear combination of the remaining vectors if possible.
   (a) $[1, 1]$: $[2, 1], [2, -1]$  (b) $[3, 2]$: $[2, -3], [-3, 6]$
   (c) $[3, -1]$: $[1, 3], [-2, 3]$
   (d) $[1, 1, 1]$: $[2, 1, 0], [0, 1, 2], [3, 2, 1]$

4. For each of the following pairs of a matrix and a vector, express the vector as a linear combination of the columns of the matrix. Plot this linear combination as was done in Figure 5.3.
   (a) $\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$  (b) $\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$  (c) $\begin{bmatrix} 2 & 3 \\ 5 & 8 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

5. The first column in the inverse $A^{-1}$ of a 2-by-2 matrix $A$ gives the weights in a linear combination of $A$'s columns that equals $\mathbf{e}_1 = [1, 0]$. The second column in $A^{-1}$ gives the weights in a linear combination of $A$'s columns that equals $\mathbf{e}_2 = [0, 1]$. Find these weights for expressing $\mathbf{e}_1$ and $\mathbf{e}_2$ as linear combinations of the columns and plot the linear combinations as in Figure 5.3 for the following matrices.
   (a) $\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$  (b) $\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$  (c) $\begin{bmatrix} 2 & 3 \\ 5 & 8 \end{bmatrix}$

6. Tell which of the following sets of vectors are linearly independent. If linearly dependent, express one vector as a linear combination of the others.
   (a) $[1, 2], [-2, 4]$  (b) $[1, 3], [3, -1]$
   (c) $[2, 1], [2, 3], [2, 8]$  (d) $[1, 1, 1], [-2, 0, -2], [2, 1, 2]$
   (e) $[2, 1, 0], [1, 1, 3], [0, 2, 1]$

7. Find a set of columns that form a basis for the column space of each of the following matrices (use the reduced matrix $A^*$ as in Examples 3 and 4). Give the rank of each matrix.
Ch. 5 Theory of Systems of Linear Equations and Eigenvalue/Eigenvector Problems

\[
\begin{bmatrix}
1 & -2 \\
-2 & 4 \\
\end{bmatrix}
\quad \begin{bmatrix}
4 & -1 & 2 \\
2 & 5 & 1 \\
-2 & 3 & -1 \\
\end{bmatrix}
\begin{bmatrix}
-1 & 3 & 1 \\
5 & -1 & 3 \\
2 & 1 & 2 \\
\end{bmatrix}
\quad \begin{bmatrix}
2 & 1 & -1 \\
1 & -2 & 2 \\
3 & 4 & -4 \\
\end{bmatrix}
\]

(a) 
(b)  
(c) 
(d)  

8. For each matrix in Exercise 7, find all sets of columns that form a basis for the column space.

9. Find a set of columns that form a basis for the column space of each of the following matrices. Give the rank of each matrix. Also find a basis for the null space of each matrix.

(a) \[
\begin{bmatrix}
-3 & 5 \\
6 & -10 \\
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
2 & 1 & 7 \\
1 & 2 & 5 \\
1 & 1 & 4 \\
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 \\
\end{bmatrix}
\]

(d) \[
\begin{bmatrix}
1 & 0 & 2 & -1 & 1 \\
1 & 1 & 1 & 1 & 2 \\
1 & 0 & 2 & -1 & 1 \\
0 & 1 & -1 & 2 & 1 \\
1 & 0 & 2 & -1 & 1 \\
\end{bmatrix}
\]

(e) \[
\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 \\
\end{bmatrix}
\]

10. Let A be a coefficient matrix in a refinery problem, as in Example 1, with each column representing the production vector of a refinery. Explain the practical significance of having one column be a linear combination of the others. What constraints and what freedom does this permit the manager of the refineries?

11. For a m-by-n matrix A, the reduced matrix A* can be written in the partitioned form \( A^* = \begin{bmatrix} I & R \\ O & O \end{bmatrix} \), where I is an r-by-r identity matrix \((r = \text{rank}(A))\) and R is r-by-(n - r). Using the submatrix R and an appropriate size identity matrix I, give a matrix N in partitioned form whose columns are the basis of Null(A).

Hint: See expression (14).
12. Determine the rank of matrix $A$, if possible, from the given information.
   (a) $A$ is an $n$-by-$n$ matrix with linearly independent columns.
   (b) $A$ is a 6-by-4 matrix and $\text{Null}(A) = \{0\}$.
   (c) $A$ is a 5-by-6 matrix and $\dim(\text{Null}(A)) = 3$.
   (d) $A$ is a 3-by-3 matrix and $\det(A) = 17$.
   (e) $A$ is a 5-by-5 and $\dim(\text{Row}(A)) = 3$.
   (f) $A$ is an invertible 4-by-4 matrix.
   (g) $A$ is a 4-by-3 matrix and $Ax = b$ has either a unique solution or else no solution.
   (h) $A$ is a 8-by-8 matrix and $\dim(\text{Row}(A^T)) = 6$.
   (i) $A$ is a 7-by-5 matrix in which $\dim(\text{Null}(A^T)) = 3$.

13. In this exercise, the reader should try to find by inspection a linear dependence among the rows of each matrix. If dependence is found, use elimination by pivoting to find a linear dependence among the columns (as in Examples 3 and 4).

   (a) $\begin{bmatrix} 5 & 2 & 3 \\ 0 & 1 & 2 \\ 5 & 3 & 5 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 3 & 4 & 2 \end{bmatrix}$
   (c) $\begin{bmatrix} 5 & 7 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}$

14. Show that the number of pivots performed in Gaussian elimination will be the same as the number of pivots in elimination by (full) pivoting. (Thus the rank of a matrix can be defined in terms of either type of elimination.)

15. Let $A^*$ be the reduced-form matrix of $A$.
   (a) Show that nonzero rows of $A^*$ generate $\text{Row}(A)$ (i.e., linear combinations of the rows of $A^*$ generate the same vectors as linear combinations of rows of $A$).
   (b) Show that the nonzero rows of $A^*$ must be linearly independent. 
      *Hint:* Look at the form of $A^*$.
   (c) Conclude that the nonzero rows of $A^*$ are a basis of $\text{Row}(A)$ and hence that $\dim(\text{Row}(A)) = \text{rank}(A)$.

16. Let $U$ be the upper triangular matrix produced at the end of Gaussian elimination on the matrix $A$.
   (a) Show that nonzero rows of $U$ generate $\text{Row}(A)$ (i.e., linear combinations of the rows of $U$ generate the same vectors as linear combinations of rows of $A$).
   (b) Show that the nonzero rows of $U$ must be linearly independent.
      *Hint:* Look at the form of $U$.
   (c) Conclude that the nonzero rows of $U$ are a basis of $\text{Row}(A)$ and hence that $\dim(\text{Row}(A)) = \text{rank}(A)$.

17. (a) Suppose that the rows of $A$ are linearly dependent. Show that at the end of Gaussian elimination, the resulting upper triangular matrix $U$ will have at least one row of zeros.
   (b) Suppose that $A$ is a square matrix with linearly dependent columns.
Show that at the end of Gaussian elimination, the resulting matrix \( U \) will have at least one row of zeros.

*Hint:* Use part (a) and Theorem 5.

18. Show that \( \text{Row}(A^T) \) (\( A^T \) is the transpose of \( A \)) equals the \( \text{Col}(A) \) and that \( \text{Col}(A^T) \) equals \( \text{Row}(A) \). Show that \( \text{rank}(A) = \text{rank}(A^T) \).

19. (a) Use the results of Exercise 17 and 18 to show that the nonzero rows in the reduced form \( A^T^* \) of \( A^T \) are a basis for the \( \text{Col}(A) \).
   (b) Use part (a) to compute a basis for \( \text{Col}(A) \) for the matrix \( A \) in Example 3.
   (c) Repeat part (b) for the matrix in Example 4.

20. (a) Show that the rank of a matrix does not change when a multiple of one row is subtracted from another row.
   (b) Show that the rank of a matrix does not change when a multiple of one column is subtracted from another column.

*Hint:* Use \( A^T \).

21. (a) Show that if \( A \) is an \( m \times n \) matrix and \( b \) an \( m \)-vector, \( b \) is in \( \text{Range}(A) \) if and only if \( \text{rank}([A\ b]) = \text{rank}(A) \), where \([A\ b]\) denotes the augmented \( m \times (n + 1) \) matrix with \( b \) added as an extra column to \( A \).
   (b) If \( Ax = b \) has no solution, show that \( \text{rank}([A\ b]) \) must be \( \text{rank}(A) + 1 \).

22. Let \( A \) be an \( n \times n \) matrix. Show that \( \det(A) = 0 \) if and only if the rows of \( A \) are linearly dependent or if the columns of \( A \) are linearly dependent.

*Hint:* Use Theorem 5 and Theorem 4 of Section 3.3.

23. This exercise examines the vectors in the column space of two matrices \( A \) and \( B \), that is, vectors in \( \text{Col}(A) \cap \text{Col}(B) \). If \( d \) is such a vector, then \( Ax' = d \) and \( Bx'' = d \), for some \( x' \), \( x'' \). Show that if \( C = [A \ -B] \) and \( x^* = [x' \ x''] \), then \( d \) is in \( \text{Col}(A) \cap \text{Col}(B) \) if and only if \( x^* \) is in \( \text{Null}(C) \).

24. Show that any set \( H \) of \( k \) linearly independent \( n \)-vectors, \( k < n \), can be extended to a basis for all \( n \)-vectors.

*Hint:* Form an \( n \times (k + n) \) matrix \( A \) whose first \( k \) columns come from \( H \) and whose last \( n \) columns are the identity matrix—thus \( \text{dim}(\text{Col}(A)) = n \); show that a basis for \( \text{Col}(A) \) using the elimination by pivoting approach in Example 5 will include the columns of \( H \).

25. Show that \( \lambda \) is an eigenvalue of \( A \) if and only if \( \det(A - \lambda I) = 0 \).

*Hint:* "If" part is immediate; for the "only if" part, use Theorem 3 and Theorem 4 of Section 3.3.
26. Let \( \mathbf{a} = [1, 2, 3], \mathbf{b} = [2, 0], \mathbf{c} = [-1, 2, 1] \). Compute the following simple matrices. 
(a) \( \mathbf{a} \ast \mathbf{b} \)  
(b) \( \mathbf{a} \ast \mathbf{c} \)  
(c) \( \mathbf{c} \ast \mathbf{c} \)

27. Verify that the simple matrices in Exercise 26 have rank 1.

28. (a) Show that \( \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \) is a simple matrix by giving the two vectors whose matrix product is \( \mathbf{A} \).
(b) Repeat part (a) for \( \mathbf{B} = \begin{bmatrix} 12 & -6 & 9 \\ 8 & -4 & 6 \\ 4 & -2 & 3 \end{bmatrix} \)

29. Write each of the following matrices as the sum of two simple matrices.
(a) \( \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \)  
(b) \( \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 7 & 8 & 9 \end{bmatrix} \)

30. (a) Find the LU decomposition of matrix in Exercise 9, part (b) and use the decomposition to write the matrix as the sum of two simple matrices as in Example 8.
(b) Repeat part (a) for the matrix in Example 5.
(c) Repeat part (a) for the matrix in Exercise 9, part (d).

31. Describe the column space and row space of a simple matrix \( \mathbf{a} \ast \mathbf{b} \) and give a basis for each.

32. (a) Show that if a matrix \( \mathbf{A} \) of rank \( k \) is expressed as the sum of \( k \) simple matrices \( \mathbf{c}_i \ast \mathbf{d}_i, i = 1, 2, \ldots, k \), then the \( \mathbf{c}_i \) are a basis of \( \text{Col}(\mathbf{A}) \) and the \( \mathbf{d}_i \) are a basis of \( \text{Row}(\mathbf{A}) \).
(b) Prove the last sentence in Theorem 8, the minimum number of simple matrices whose sum is matrix \( \mathbf{A} \) is \( \text{rank}(\mathbf{A}) \), as follows: The LU decomposition yields a sum of \( k \) simple matrices equaling \( \mathbf{A} \), where \( k = \text{rank}(\mathbf{A}) \), by the first part of Theorem 8; if fewer than \( k \) simple matrices could sum to \( \mathbf{A} \), use part (a) to show that then \( \dim(\text{Col}(\mathbf{A})) < \text{rank}(\mathbf{A}) \) — impossible.

Section 5.3 Approximate Solutions and Pseudoinverses

This section presents a method for obtaining an approximate solution that can be used to "solve" an \( m \)-by-\( n \) system \( \mathbf{A}\mathbf{x} = \mathbf{b} \) that has no solution, that is, when \( \mathbf{b} \) is not in the range of \( \mathbf{A} \). We seek a "solution" \( \mathbf{w} \) that gives a vector \( \mathbf{p} = \mathbf{A}\mathbf{w} \) which is as close as possible to \( \mathbf{b} \). In the following discussion, we use the euclidean norm \( |\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \), because
we will be treating the vectors $p$, $b$ as points in euclidean space and minimizing the distance of the error $|b - p|$.

**Example 1. Refinery Problem Revisited**

Recall the refinery model first presented in Section 1.2 with three refineries producing three petroleum-based products, heating oil, diesel oil, and gasoline.

- **Heating oil:** $20x_1 + 4x_2 + 4x_3 = 500$
- **Diesel oil:** $10x_1 + 14x_2 + 5x_3 = 850$ \hspace{1cm} (1)
- **Gasoline:** $5x_1 + 5x_2 + 12x_3 = 1000$

Suppose that the third refinery is out of service. We still want to attempt to produce the same amounts of these products. That is, we want to satisfy the system (as best we can)

- **Heating oil:** $20x_1 + 4x_2 = 500$
- **Diesel oil:** $10x_1 + 14x_2 = 850$ \hspace{1cm} or (2)
- **Gasoline:** $5x_1 + 5x_2 = 1000$

Let $A$ be the matrix of coefficients in (2) and let $b$ be the right-side vector. We seek to minimize $|b - Aw|$ (recall that we are using the euclidean distance as our norm). The approximation we want requires a vector $w$ so that

$$Aw = w_1 \begin{bmatrix} 20 \\ 10 \\ 5 \end{bmatrix} + w_2 \begin{bmatrix} 4 \\ 14 \\ 5 \end{bmatrix}$$

is as close to

$$\begin{bmatrix} 500 \\ 850 \\ 1000 \end{bmatrix}$$

as possible.

This type of approximate solution is called a **least-squares solution**, because the euclidean distance $|b - Aw|$ to be minimized involves a sum of squares. For such approximate solutions to be meaningful, we require
that $A$ have more rows than columns; otherwise, a more sophisticated theory is needed.

Recall that we encountered least-squares solutions in regression. We review the regression problem presented in Section 4.2.

**Example 2. Simple Linear Regression Reviewed**

We wanted to fit the three $(x, y)$ points $(0, 1), (2, 1), \text{ and } (4, 4)$ to a line of the form $y = qx$ (where the $x$-value might be the number of college mathematics courses taken and the $y$-value a score on some test). The requirement $qx = y$ for these points yields a system of equations

\[
\begin{align*}
0q &= 1 \\
2q &= 1 \quad \text{or} \quad qx = y, \quad \text{where } x = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \\
4q &= 4
\end{align*}
\]

Figure 5.4a shows the points to be estimated by this line, and Figure 5.4b shows the vectors $y$ and $qx$ in 3-space. Our goal is to find $q$ so that the estimates $\hat{y}_i = qx_i$ in (3) are as close as possible to the true $y_i$. A way to view $qx$ in Figure 5.4b is as the projection of $y$ onto the line through $x$ from the origin.

To obtain the value for $q$, we minimized the sum of the squares of the errors (SSE).

\[
\text{SSE} = \sum (qx_i - y_i)^2 = (0q - 1)^2 + (2q - 1)^2 + (4q - 4)^2 = 20q^2 - 36q + 18
\]

In Section 4.2 we found the optimal $q$ by differentiating (4), setting the derivative equal to 0, and obtaining $q = .9$ (see line $\hat{y} = .9x$ in Figure 5.4a). We also calculated how to minimize SSE for an arbitrary number of $x - y$ pairs and obtained the formula

\[
q = \frac{\sum x_iy_i}{\sum x_i^2} = \frac{x \cdot y}{x \cdot x}
\]

where $x$ and $y$ are the vectors of $x$- and $y$-values. Note that for this example

\[
q = \frac{0 \times 1 + 2 \times 1 + 4 \times 4}{0 \times 0 + 2 \times 2 + 4 \times 4} = \frac{18}{20} = .9
\]

When the model $y = qx + r$ is used, (3) is changed to

\[
\begin{align*}
0q + r &= 1 \\
2q + r &= 1 \quad \text{or} \quad Aw = y, \quad \text{where } A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}, \quad w = \begin{bmatrix} q \\ r \end{bmatrix}
\end{align*}
\]
Figure 5.4 (a) Regression estimates for points (0, 1), (2, 1), and (4, 4).
(b) \( \hat{y} = 0.9x \) is regression solution when \( x = [0, 2, 4], y = [1, 1, 4] \). (c) \( p \) is closest vector to \( b \) in range of \( A \).

Now the least-squares solution is a pair \( q, r \) such that the vector \( q[0, 2, 4] + r[1, 1, 1] \) (which is in the range of \( A \)) is as close as possible to \( y \) (see Figure 5.4c, where \( p = Aw, b = y \)). Again, we can view \( p \) as the projection of \( b \) onto the range of \( A \).

We have now motivated the importance of finding \( w \) so that the vector \( p = Aw \) is as close as possible to a given vector \( b \), that is, so that \( p \) is the projection of \( b \) onto the range of \( A \). To determine \( w \) and \( p \), we shall use the following geometric property of the projection \( p \) of \( b \) onto the range of \( A \): The error vector \( b - p \) is at right angles to—perpendicular to—vectors in the range of \( A \) (see Figure 5.4b and c).
The term **orthogonal** is used in linear algebra instead of "perpendicular" to describe two vectors at right angles. The following theorem provides a simple numerical test for orthogonality.

**Theorem 1.** Two $n$-vectors $a, b$ are orthogonal if and only if their scalar product is zero, $a \cdot b = 0$.

This theorem is a consequence of a more general fact about angles between vectors that is proved later in this section.

We use Theorem 1 to obtain a matrix equation satisfied by $p$ (where $p = Aw$) when $b - p$ is orthogonal to vectors in the range of $A$ (implying that $p$ is the closest vector to $b$ in the range of $A$).

First we illustrate the procedure by obtaining a formula for $q$ in the simple regression model $\hat{y} = qx$ in Example 1, that is, to find $q$ so that $qx$ is as close as possible to $y$. The error vector in this case is $y - qx$, and the range is simply all multiples of $x$. The error vector $y - qx$ should be orthogonal to the range, that is, orthogonal to $x$. By Theorem 1, this yields

$$x \cdot (y - qx) = 0 \quad \text{or} \quad x \cdot y - qx \cdot x = 0$$

Solving for $q$, we obtain the same regression formula as in (5):

$$q = \frac{x \cdot y}{x \cdot x} \quad (7)$$

As noted above, $qx$ is the projection of vector $y$ onto vector $x$. Thus

**Theorem 2.** The projection of $y$ onto $x$ (i.e., onto the line from the origin through $x$) is $qx$, where $q = x \cdot y / x \cdot x$.

Next consider the general case where we want an approximate solution to $Ax = b$ for any $m$-by-$n$ matrix $A$. The error vector $b - p = b - Aw$ should be orthogonal to every vector in the range of $A$. Recall that the range of $A$ is formed by linear combinations of the column vectors of $A$, $r_1a_1^C + r_2a_2^C + \cdots + r_na_n^C$. If $b - Aw$ is orthogonal to any linear combination of the column vectors, then it certainly must be orthogonal to these column vectors $a_i^C$ themselves. By Theorem 1 we have

$$a_i^C \cdot (b - Aw) = 0 \quad \text{for } i = 1, 2, \ldots, n \quad (8)$$

If we make a matrix $A^*$ whose rows are the columns of $A$, then (8) gives

$$A^*(b - Aw) = 0 \quad (9)$$

But this matrix $A^*$ is simply $A^T$, the transpose of $A$ (whose rows are the columns of $A$). So (9) is
Assuming that the matrix $A^T A$ is invertible, we can solve (10) for $w$ to obtain

$$w = (A^T A)^{-1} A^T b$$  \hspace{1cm} (11)

The right side of (11) is a pretty messy expression. Since the rows of $A^T$ are the columns of $A$, entry $(i, j)$ in the matrix product $A^T A$ is just the scalar product $a_i^T \cdot a_j^T$ of the $i$th column of $A$ times the $j$th column of $A$. When $A$ consists of a single column $x$, as in the regression model $qx = y$, (11) reduces to $q = (x \cdot x)^{-1} x \cdot y$ or $q = x \cdot y / x \cdot x$—the formula we obtained above.

We call the product of matrices on the right in (11)

$$A^+ = (A^T A)^{-1} A^T$$  \hspace{1cm} (12)

the pseudoinverse of $A$ (the term generalized inverse is also used). If $A$ is an $m$-by-$n$ matrix, $A^+$ will be an $n$-by-$m$ matrix.

**Theorem 3.** The least-squares solution $w$ to the system of equations $Ax = b$ is $w = A^+ b$, where $A^+ = (A^T A)^{-1} A^T$. Further, $A^+$ is the left inverse of $A$: $A^+ A = I$.

The second sentence of the theorem is easily verified: $A^+ A = (A^T A)^{-1} (A^T A) = I$, since we are multiplying $A^T A$ times its inverse. The identity, $A^+ A = I$, can be used to check that you have computed $A^+$ correctly.

If $A$ is an invertible $n$-by-$n$ matrix, the pseudoinverse $A^+$ equals the regular inverse $A^{-1}$ (see Exercise 22). If $b$ happens to lie in the range of $A$, then $Aw$ will the exact solution, that is, $Aw$ equals $b$.

Although (12) is complex, the fact that such a matrix $A^+$ exists at all is impressive. Applying Theorem 2 to the general regression model, we obtain

**Corollary.** Consider the regression model $\hat{y} = q_1 x_1 + q_2 x_2 + \cdots + q_n x_n + r$ with associated matrix equation

$$y = Xq$$

where $y$ is the set of $y$-value observations, $X$ is the matrix whose $j$th column is the set of $x_j$-value observations and whose last column is the 1's vector, and $q = [q_1, q_2, \ldots, q_n, r]$. Then the regression model parameters $q$ are given by

$$q = (X^T X)^{-1} X^T y$$  \hspace{1cm} (13)
Example 3. Least-Squares Solution to Refinery Problem

Let us find the least-squares solution to the system of equations we had in Example 1, where the first two refineries alone had to try to satisfy the demand vector.

\[
\begin{align*}
20x_1 + 4x_2 &= 500 \\
10x_1 + 14x_2 &= 850 \\
5x_1 + 5x_2 &= 1000
\end{align*}
\] (14)

If \( A \) is the coefficient matrix in (14), then we compute \( A^T A \) and \( (A^T A)^{-1} \) to be [recall that entry \((i, j)\) in \( A^T A \) is the scalar product of columns \( i \) and \( j \) of \( A \)].

\[
A^T A = \begin{bmatrix} 525 & 245 \\ 245 & 237 \end{bmatrix} \quad \text{and} \quad (A^T A)^{-1} = \begin{bmatrix} .00368 & -.00380 \\ -.00380 & .00815 \end{bmatrix}
\]

The pseudoinverse \( A^+ \) of \( A \) is

\[
A^+ = (A^T A)^{-1}A^T = \begin{bmatrix} .00368 & -.00380 \\ -.00380 & .00815 \end{bmatrix} \begin{bmatrix} 20 & 10 & 5 \\ 4 & 14 & 5 \end{bmatrix} = \begin{bmatrix} .0584 & -.0164 & -.0006 \\ -.0435 & .0761 & .0217 \end{bmatrix}
\] (15)

With (15), we can now find the least-squares solution \( w \) to (14):

\[
w = A^+ b = \begin{bmatrix} .0584 & -.0164 & -.0006 \\ -.0435 & .0761 & .0217 \end{bmatrix} \begin{bmatrix} 500 \\ 850 \\ 1000 \end{bmatrix} = \begin{bmatrix} 14.6 \\ 64.7 \end{bmatrix}
\] (16)

This solution produces the following approximating output vector:

\[
Aw = [551, 1051, 394]
\]

with an error vector of

\[
b - Aw = [500 - 551, 850 - 1051, 1000 - 394] = [-51, -201, 606]
\]

This is a terrible approximation. We vastly underproduce the third product (gasoline). With the third refinery shut down, we shall always get much more of the second product than the third product [see (14)].
Example 4. Solution to Regression Model
\[
\hat{y} = qx + r
\]
Let us use the pseudoinverse to solve our regression problem with points \((0, 1), (2, 1), (4, 4)\) and the model \(\hat{y} = qx + r\). The system of equations is

\[
\begin{align*}
0q + r &= 1 \\
2q + r &= 1 \\
4q + r &= 4
\end{align*}
\]

which can be expressed as \(Xq = y\), where

\[
X = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}, \quad q = \begin{bmatrix} q \\ r \end{bmatrix}
\]

Then

\[
X^T X = \begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix}, \quad (X^T X)^{-1} = \begin{bmatrix} \frac{1}{1} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{5}{6} \end{bmatrix}
\]

Using \((X^T X)^{-1}\), we obtain the pseudoinverse

\[
X^+ = (X^T X)^{-1} X^T = \begin{bmatrix} \frac{1}{8} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{5}{8} \end{bmatrix} \begin{bmatrix} 0 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{5}{8} & \frac{1}{3} & -\frac{1}{8} \end{bmatrix}
\]

Then

\[
q = \begin{bmatrix} q \\ r \end{bmatrix} = X^+ y = \begin{bmatrix} -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{5}{8} & \frac{1}{3} & -\frac{1}{8} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{2} \end{bmatrix}
\]

So \(q = .75, r = .5\); this is the same answer that we obtained for this problem in Section 4.2 (see Figure 5.4a). Our regression estimates for the \(y\)-values are given by \(Xq = [.5, 2, 3.5]\). 

Example 5. Least-Squares Polynomial Fitting
Suppose that we want to try to fit a quadratic curve through the set of points \((0, 7), (1, 5), (2, 4), (3, 4), (4, 8), (5, 12)\) using a least-squares approximation. Our model is

\[
\hat{y} = ax^2 + bx + c
\]
We shall treat $x^2$ as a separate variable, say let $z = x^2$, so that a linear multivariate regression model can be used

$$\hat{y} = az + bx + c$$

(20)

For the given set of points our $X$ matrix is

$$X = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \\ 25 & 5 & 1 \end{bmatrix} \quad \text{with} \quad y = \begin{bmatrix} 7 \\ 5 \\ 4 \\ 4 \\ 8 \\ 12 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

(21)

Using a computer program, we obtain

$$X^+ = (X^TX)^{-1}X^T = \frac{1}{56} \begin{bmatrix} 5 & -1 & -4 & -1 & 5 \\ -33 & .2 & 18.4 & 21.6 & 9.8 & -17 \end{bmatrix}$$

and

$$q = X^+y = \begin{bmatrix} .893 \\ -3.493 \\ 7.214 \end{bmatrix} \quad \text{with estimates} \quad \hat{y} = Xq = \begin{bmatrix} 7.2 \\ 4.6 \\ 3.8 \\ 4.8 \\ 5.3 \\ 12.1 \end{bmatrix}$$

(22)

Our quadratic estimate is thus $\hat{y} = .893x^2 - 3.493x + 7.214$.

Although the estimated $y$-values work out closely to the observed $y$-values, a word of warning is important. This is a very poorly conditioned problem—the columns of $X$ are all fairly similar. In fact, the condition number of the matrix $(X^TX)$ is 2000 (in the sum norm)! A small change in the data could produce a large change in our answer.

To compute the pseudoinverse $(A^TA)^{-1}A^T$, we need to know that the matrix $A^TA$ is invertible. The following result gives us the information we need.

Theorem 4. Let $A$ be an $m$-by-$n$ matrix. Then the $n$-by-$n$ matrix $A^TA$ is invertible (and the pseudoinverse $A^+$ exists) if the $n$ columns of $A$ are linearly independent, or equivalently, if $\text{rank}(A) = n$. 
Proof (Optional). We shall prove the stronger result that \( \text{rank}(A^T A) = \text{rank}(A) \). Since \( \text{rank}(A) = n \), then \( \text{rank}(A^T A) = n \) and any \( n \times n \) matrix of rank \( n \) is invertible.

We work with null spaces. By Corollary A to Theorem 3 of Section 5.2, for a matrix \( B \) with \( n \) columns, \( \dim(\text{Null}(B)) = n - \text{rank}(B) \). Then \( \text{rank}(A^T A) = \text{rank}(A) \) is a consequence of showing that \( \text{Null}(A^T A) = \text{Null}(A) \).

Let \( x \) be any vector in \( \text{Null}(A) \), that is, \( Ax = 0 \). Then

\[
(A^T A)x = A^T(Ax) = A^T(0) = 0
\]

so \( x \) is in \( \text{Null}(A^T A) \).

Conversely, let \( y \) be any vector in \( \text{Null}(A^T A) \). We want to show that \( y \) is in \( \text{Null}(A) \). To do this, we show that \( (Ay) \cdot (Ay) = 0 \), which implies that \( Ay = 0 \) (since \( c \cdot c = 0 \) means \( \sum c_i^2 = 0 \)). We need the fact that \( (Ay) \cdot (Ay) \) is the same as \( (Ay)^T(Ay) \) (the latter is the product of the \( 1 \times m \) matrix \( (Ay)^T \) times the \( m \times 1 \) matrix \( Ay \)). Then we have

\[
(Ay) \cdot (Ay) = (Ay)^T(Ay) = (y^T A^T)(Ay)
\]

\[
= y^T(A^T A)y = y^T \cdot 0 = 0
\]

We note that in regression problems, practical considerations dictate that the matrix \( X \) is virtually certain to have linearly independent columns.

There is an important special case in which the computation of the pseudoinverse becomes very easy. This is when the columns of the matrix \( A \) are orthogonal. Now by Theorem 1, the scalar product of columns \( a_i^c \cdot a_j^c \) equals 0. Since entry \( (i, j) \) in \( A^T A \) is exactly this scalar product, \( A^T A \) will be all 0's except on the main diagonal. This simple form of \( A^T A \) leads to a simple form for \( (A^T A)^{-1} \) and for \( A^+ \).

Example 6. Regression with Orthogonal Columns

We shall repeat the analysis of Example 5 with points \((0, 1), (2, 1),\) and \((4, 4)\) and regression model \( \hat{y} = qx + r \), but we shall shift the \( x \)-values so that the average \( x \)-value is 0 (in Section 4.2 we noted that such a shift simplified our regression formulas). The average \( x \)-value is \((0 + 2 + 4)/3 = 2\). If we subtract 2 from each \( x \)-value, obtaining points \((-2, 1), (0, 1), \) and \( (2, 4) \), then the new average \( x \)-value is 0 (subtracting off the average value always makes the new average be 0).

Let us repeat the pseudoinverse computations of Example 5 for these new points.

\[
\begin{align*}
-2q + r &= 1 \\
0q + r &= 1 \quad \text{or} \quad Xq = y, \\
2q + r &= 4
\end{align*}
\]
where \( X = \begin{bmatrix} -2 & 1 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \), \( y = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \) (23)

Observe that the two columns of \( X \) are now orthogonal. The scalar product of the two columns \( x \cdot 1 = \Sigma x_i \) is 0, since the average \( x \)-value \((= \Sigma (x_i)/m)\) is 0. Then

\[
X^T X = \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad (X^T X)^{-1} = \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}
\] (24)

The inverse \((X^T X)^{-1}\) in (24) can be computed by the determinant formula, but we note that the inverse of a diagonal matrix \( D \) (with 0's everywhere off the main diagonal) is obtained by replacing each diagonal entry with its inverse, as in (24).

The two diagonal entries in \( X^T X \) are, in symbolic terms, \( x \cdot x \) and \( 1 \cdot 1 = m \) (number of points in regression problem). Thus, when the average \( x \)-value is 0, \((X^T X)^{-1}\) has the form

\[
(X^T X)^{-1} = \begin{bmatrix} \frac{1}{x \cdot x} & 0 \\ 0 & 1/m \end{bmatrix}
\] (25)

The pseudoinverse \( X^+ \) is now

\[
X^+ = (X^T X)^{-1} X^T = \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} -2 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{8} & 0 & \frac{2}{8} \\ \frac{1}{8} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}
\] (26)

When we premultiply any matrix \( B \) by a diagonal matrix \( D \), then \( D \) has the effect of multiplying the \( i \)th row of \( B \) by the \( i \)th diagonal entry of \( D \), as in (26).

Looking at the values of the diagonal entries in \((X^T X)^{-1}\) [see (25)], we see that \( X^+ \) is simply the transpose of \( X \) with the first column of \( X \) divided by its sum of squares \((x \cdot x)\) and the second column divided by \( m \) (the number of points).

Finally,

\[
q = \begin{bmatrix} q \\ r \end{bmatrix} = X^+ y = \begin{bmatrix} -\frac{2}{8} & 0 & \frac{2}{8} \\ \frac{1}{8} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{2}{2} \end{bmatrix}
\]

Observe that \( q \) is the scalar product of the first row of \( X^+ \) with \( y \). But we just noted that the first row of \( X^+ \) is simply \( x \) divided by the number \( x \cdot x \). Similarly, \( r \) equals the scalar product of the second row of \( X^+ \) times \( y \), and this second row is just \((1/m)1\). Thus we have the simple formulas
By Theorem 2, \( q \) and \( r \) are simply the projections of \( y \) onto \( x \) and \( 1 \), respectively. The nice results obtained in Example 6 will be true for the pseudoinverse of any matrix with orthogonal columns.

**Theorem 5.** If the \( m \)-by-\( n \) matrix \( A \) \((m > n)\) has orthogonal columns, the pseudoinverse \( A^+ \) is obtained by dividing each column of \( A \) by the sum of the squares of the column's entries and then taking the transpose of the resulting matrix; the \( i \)th row of \( A^+ \) is \( a_i^C/(a_i^C \cdot a_i^C) \). Further, the least-squares solution \( w = A^+ b \) is just the projection of \( b \) onto the columns of \( A \): \( w_i = a_i^C \cdot b / (a_i^C \cdot a_i^C) \).

Suppose that we have a regression model with several input variables, such as

\[
\hat{y} = q_1 u + q_2 v + q_3 x + r
\]

and suppose that the vectors \( u, v, x, \) and \( 1 \) (of the \( u \)-values, \( v \)-values, \( x \)-values, and \( 1 \)'s vector) are orthogonal. Then Theorem 5 tells us, generalizing (27), that the regression parameters are the projections of \( y \) onto \( u, v, x, \) and \( 1 \):

\[
q_1 = \frac{u \cdot y}{u \cdot u}, \quad q_2 = \frac{v \cdot y}{v \cdot v}, \quad q_3 = \frac{x \cdot y}{x \cdot x}, \quad r = \frac{1}{m} \sum y_i
\]

But what chance is there that the \( u, v, x, \) and \( 1 \) vectors will be orthogonal? The answer is often up to the person who collects the data. If the \( u-, v-, \) and \( x \)-values measure settings of control knobs on a complex machine and the \( y \)-value measures the task performed by the machine, then a researcher who knows about Theorem 5 could pick settings to make the vectors orthogonal. This is a problem in a statistical subject called design of experiments.

We asserted in Theorem 1 that if \( a \) and \( b \) are orthogonal, that is, they form an angle of 90°, then \( a \cdot b = 0 \). This result was central to the derivation of the pseudoinverse. Now we prove the following theorem about the angle between two vectors, and Theorem 1 follows directly from this result. We measure the angle between two vectors \( a, b \) by treating the vectors as line segments from the origin to the points with coordinates given by \( a \) and \( b \) (see Figure 5.5).

**Theorem 6.** The cosine of the angle \( \theta \) between any vectors \( a, b \) is

\[
\cos \theta = \frac{a \cdot b}{|a| |b|}
\]

If \( a \) and \( b \) are unit-length vectors, (30) becomes \( \cos \theta = a \cdot b \).
Figure 5.5 \( \theta \) is the angle between \( \mathbf{a} = [3, 4] \) and \( \mathbf{b} = [12, 5] \).

Example 7. Examples of Angles Between Vectors

(i) If \( \mathbf{a} = [1, 0, 0] \) and \( \mathbf{b} = [0, 1, 0] \), then \( \cos \theta = \mathbf{a} \cdot \mathbf{b} = 0 \) and we conclude that \( \mathbf{a}, \mathbf{b} \) form a 90° angle, that is, they are orthogonal.

(ii) If \( \mathbf{a} = [3, 4] \) and \( \mathbf{b} = [12, 5] \), then \( |\mathbf{a}| = 5 \) (\( = \sqrt{9 + 16} \)) and \( |\mathbf{b}| = 13 \) (\( = \sqrt{144 + 25} \)). So \( \cos \theta = \mathbf{a} \cdot \mathbf{b} / |\mathbf{a}| |\mathbf{b}| = (3 \cdot 12 + 4 \cdot 5) / 5 \cdot 13 = \frac{44}{65} = .86 \). The angle with a cosine of .86 is 36° (see Figure 5.5).

(iii) If \( \mathbf{a} = [.6, .8] \), with \( |\mathbf{a}| = 1 \) and \( \mathbf{b} = [1, 0] \), then \( \cos \theta = \mathbf{a} \cdot \mathbf{b} = .6 \cdot 1 + .8 \cdot 0 = .6 \)—just the first coordinate.

The proof of Theorem 6 uses the law of cosines:

\[
|\mathbf{a} - \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}| \cdot |\mathbf{b}| \cos \theta \tag{31}
\]

The square of the euclidean norm \( |c|^2 \) is simply \( \mathbf{c} \cdot \mathbf{c} \), and we can write \( |\mathbf{a} - \mathbf{b}|^2 \) as \( (\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b}) \). Expanding with matrix algebra, we have

\[
|\mathbf{a} - \mathbf{b}|^2 = (\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b}) = \mathbf{a} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} - 2\mathbf{a} \cdot \mathbf{b} \tag{32}
\]

The right side of (32) is the same as the right side of (31) except for the last terms. So these last terms must be equal:

\[-2|\mathbf{a}| \cdot |\mathbf{b}| \cos \theta = -2\mathbf{a} \cdot \mathbf{b}\]

Solving for \( \cos \theta \) yields Theorem 6.

Theorem 6 has a very important application in statistics. The cosine of the angle \( \theta(x, y) \) between two vectors \( x \) and \( y \) tells us if the vectors are close together [when \( \cos \theta(x, y) \) is near 1] or opposites of one another [when \( \cos \theta(x, y) \) is near -1], or are unrelated, that is, close to orthogonal [when \( \cos \theta(x, y) \) is near 0].

Suppose that \( x \) and \( y \) are vectors of data from an experiment, say \( x \) is the scores of 10 students on a math test and \( y \) is the scores of the 10 students on a language test. Then \( \cos \theta(x, y) \) tells us how closely related these two
sets of data are and helps us predict future relations between math and language scores. If cos θ(x, y) is .8, performance on these two tests is closely related and we can view a student’s score on one test as a reasonably good predictor of how he or she will do on the other test. If cos θ(x, y) is −.7, the score vectors point in almost opposite directions and a high score on one test is very likely to produce a below-average score on the other test. If cos θ(x, y) = 0 (the vectors x, y are orthogonal), then performance on one test tells us nothing about the likely performance on the other test (in statistics, one says that the two sets of data are independent).

Definition. Let \( \mathbf{x} = [x_1, x_2, \ldots, x_n] \) and \( \mathbf{y} = [y_1, y_2, \ldots, y_n] \) be two sets of observations with the property that the average x-value and the average y-value are each 0. Then the correlation coefficient \( \text{Cor}(x, y) \) of x and y is defined to be \( \cos \theta(x, y) \).

\[
\text{Cor}(x, y) = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} \tag{33}
\]

Recall that the average x-value is \( \bar{x} = (1/n) \sum x_i \). If \( \bar{x} \neq 0 \), we can subtract \( \bar{x} \) from each \( x_i \) to get a revised vector that does have an average value of 0. Similarly for y-values. We need an average value of 0 so that the opposite of a high score (a positive value) will be a low score (a negative value). This way the terms \( x_i y_i \) in (33) for pairs of oppositely correlated entries \( x_i, y_i \) will be negative (when \( x_i y_i \) is the product of a positive and a negative number), leading to a negative correlation.

**Example 8. Correlation Coefficient**

Suppose that we ask the eight faculty members of the Podunk University Alchemy Department to rate the quality of their graduate students and we poll the students to get a rating of the quality of each of the eight. The results of our experiment are presented in Table 5.1 (where we have processed the data to make the average value 0 in each category).

**Table 5.1**

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Quality of Students ((x_i))</th>
<th>Student Rating ((y_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aristotle</td>
<td>+5</td>
<td>+2</td>
</tr>
<tr>
<td>2. Galileo</td>
<td>−5</td>
<td>−7</td>
</tr>
<tr>
<td>3. Goldbrick</td>
<td>−2</td>
<td>0</td>
</tr>
<tr>
<td>4. Hasbeen</td>
<td>+3</td>
<td>−1</td>
</tr>
<tr>
<td>5. Leadbottom</td>
<td>−4</td>
<td>−5</td>
</tr>
<tr>
<td>6. Merlin</td>
<td>+5</td>
<td>+3</td>
</tr>
<tr>
<td>7. Midas</td>
<td>+5</td>
<td>−0</td>
</tr>
<tr>
<td>8. Santa Claus</td>
<td>−7</td>
<td>+8</td>
</tr>
</tbody>
</table>
Applying formula (33) to these data, we obtain

\[ \text{Cor}(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} = \frac{5 \cdot 2 + (-5)(-7) + \cdots + (-7)8}{\sqrt{178} \cdot \sqrt{152}} \]

\[ = \frac{21}{13.3 \cdot 12.3} = .1 \]

Looking back at the data in Table 5.1, we are a little surprised to see such a low correlation, since the numbers in the two columns correspond fairly well for most faculty with the glaring exception of Santa Claus. Statisticians would call Santa Claus’s data pair \((-7, 8)\) an outlier, an observation that fits poorly with the rest of the data. We warned in the regression section (Section 4.2) that one or two outliers can distort a statistical analysis. (A little investigating reveals that Santa Claus is a terrible teacher but is still well liked because he gives the students lots of candy every December.)

Let us throw out Santa Claus’s numbers and recompute the correlation coefficient. This requires us to adjust the data so that the averages in each column are again 0. The new numbers are shown in Table 5.2.

**Table 5.2**

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Quality of Students ((x_i))</th>
<th>Students’ Rating ((y_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aristotle</td>
<td>+4</td>
<td>+3</td>
</tr>
<tr>
<td>2. Galileo</td>
<td>-6</td>
<td>-6</td>
</tr>
<tr>
<td>3. Goldbrick</td>
<td>-3</td>
<td>1</td>
</tr>
<tr>
<td>4. Hasbeen</td>
<td>+2</td>
<td>0</td>
</tr>
<tr>
<td>5. Leadbottom</td>
<td>-5</td>
<td>-4</td>
</tr>
<tr>
<td>6. Merlin</td>
<td>+4</td>
<td>+4</td>
</tr>
<tr>
<td>7. Midas</td>
<td>+4</td>
<td>+1</td>
</tr>
</tbody>
</table>

\[ \text{Cor}(x, y) = \frac{85}{\sqrt{122} \cdot \sqrt{79}} \approx \frac{85}{11 \cdot 8.9} = .9 \]

a high degree of correlation.

We finish this section with a discussion of orthogonal vector spaces associated with least-squares solutions. A least-squares solution \(w\) to \(Ax = b\) involves breaking \(b\) into two parts,

\[ b = p + b - p \quad (34) \]

where \(p = Aw\) is the least-squares solution and \(b - p\) is the error vector.
Recall that \( p \) and \( b - p \) must be orthogonal. The decomposition of \( b \) in (34) into a range and an error vector is unique (\( p \) is the unique vector \( A w \)). We assume here that \( A \) has linearly independent columns. The vector \( p \) is in the range of \( A \) (= column space of \( A \)). Now we shall identify the vector space \( V \) containing \( b - p \). \( V \) is the error space of \( A \) consisting of all error vectors; these are vectors \( e \) orthogonal to the columns of \( A \). This means that \( a_T^e \cdot e = 0 \) for all columns \( a_T^e \) of \( A \).

Earlier in this section we expressed the fact that the error vector \( b - Aw \) was orthogonal to all the columns of \( A \) as \( A^T(b - Aw) = 0 \). That is, \( A^T \) has rows that are the columns of \( A \). Then the error space \( V \) can be defined:

\[
V = \{v : A^Tv = 0\} \tag{35}
\]

But from (35), we see that \( V \) is simply the null space \( \text{Null}(A^T) \) of \( A^T \). Thus \( \text{Range}(A) \) is orthogonal to \( \text{Null}(A^T) \). Here we are calling two vector spaces orthogonal if all pairs of vectors, one from each, are orthogonal.

Let us next determine the dimension of the error space \([= \text{Null}(A^T)]\). By Theorem 3 of Section 5.2, the dimension of the \( \text{Null}(A^T) \) is \( m - \text{rank}(A^T) \) (where \( m \) is the number of columns in \( A^T \)). So

\[
\text{dim}(\text{Error space}(A)) = m - \text{rank}(A^T) \tag{36}
\]

But \( \text{rank}(A^T) = \text{rank}(A) \) (this simple consequence of Theorem 4 of Section 5.2 is proved in Exercise 18 of that section). So (36) is the same as

\[
\text{dim}(\text{Error space}(A)) = m - \text{rank}(A) \tag{37}
\]

Using the fact that the dimension of \( \text{Range}(A) \) is \( \text{rank}(A) \), we have the expected result [in light of (34)]:

\[
\text{dim}(\text{Range}(A)) + \text{dim}(\text{Error space}(A)) = m \tag{38}
\]

Summarizing, we have

**Theorem 7**

(i) Let \( A \) be a \( m \)-by-\( n \) matrix and \( b \) be any \( m \)-vector. Then \( b \) can be written as a unique sum

\[
b = b_1 + b_2 \tag{39}
\]

where \( b_1 \) is in \( \text{Range}(A) \), \( b_2 \) is in \( \text{Error space}(A) \), and \( b_1, b_2 \) are orthogonal. Further,

\[
\text{dim}(\text{Range}(A)) + \text{dim}(\text{Error space}(A)) = m
\]

The vector \( b_1 \) in (39) equals \( Aw \) and is the projection of \( b \) onto the column space of \( A \); \( w \) is the least-squares solution \( w = A^+b \).
The error space of \( A \) equals Null(\( A^T \)), so Null(\( A^T \)) is orthogonal to Range(\( A \)).

Theorem 7 is valid even if \( A \) does not have linearly independent columns.

Recall that the Range(\( A \)) equals Col(\( A \)), the column space of \( A \), which equals Row(\( A^r \)). So Theorem 7, part (ii) implies that Null(\( A^T \)) is orthogonal to Row(\( A^r \)), or, interchanging the names of \( A \) and \( A^T \), Null(\( A \)) is orthogonal to Row(\( A \)). Reinterpreting (39) in terms of these vector spaces, we have

**Theorem 8.** Let \( A \) be an \( m \)-by-\( n \) matrix. The row space of \( A \) and the null space of \( A \) are orthogonal. Further, any \( n \)-vector \( x \) can be written as a unique sum

\[
x = x_1 + x_2
\]

where \( x_1 \) is in the Row(\( A \)) and \( x_2 \) is in Null(\( A \)).

Note that Null(\( A \)) being orthogonal to Row(\( A \)) follows directly from the definition of Null(\( A \)): \( x \) is in Null(\( A \)) when \( ax = 0 \), but \( Ax = 0 \) just says that \( x \) is orthogonal to the rows of \( A \). However, the unique sum (40) is not so obvious.

The reader should recall Theorem 1 of Section 5.1, which asserted that any solution \( x' \) to \( Ax = b \) could be expressed as \( x' = x* + x^0 \), where \( x* \) is some particular solution to \( Ax = b \) and \( x^0 \) is some solution to the homogeneous system \( Ax = 0 \) [i.e., \( x^0 \) is in Null(\( A \))]. Theorem 8 tells us that the decomposition of \( x' \) can be chosen so that \( x* \) is in the row space of \( A \).

**Corollary A.** Let \( x' \) be a solution to the system \( Ax = b \). Then \( x' \) can be uniquely decomposed \( x' = x* + x^0 \), where \( x* \) is in the row space of \( A \) and \( x^0 \) is in the null space of \( A \). Further, \( x* \) and \( x^0 \) are orthogonal.

Note in Corollary A that if \( Ax' = b \), then \( Ax* + Ax^0 = b \) (since \( x' = x* + x^0 \)). But \( Ax^0 = 0 \) since \( x^0 \) is in Null(\( A \)), so \( Ax* = b \). We have thus proved the surprising result:

**Corollary B.** If the system \( Ax = b \) has a solution, it has a solution \( x* \) that lies in the row space of \( A \).

**Section 5.3 Exercises**

**Summary of Exercises**

Exercises 1–16 involve regression and least-squares solutions; when asked if a 2-by-2 matrix is poorly conditioned, say yes if the condition number is \( \geq 10 \). Exercises 17–21 involve angles between vectors and the correlation coefficient. Exercises 22–30 involve examples and extensions of vector-space theory.
1. Determine the condition number of the matrix \((A^TA)\) in Example 3. Is this matrix poorly conditioned?

2. Seven students earned the following scores on a test after studying the subject matter different numbers of weeks.

<table>
<thead>
<tr>
<th>Student</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Study ((x_i))</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Test Score ((y_i))</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

Fit these data with a regression model of the form \(\hat{y} = qx + r\). Determine \(q\) and \(r\) by computing the pseudoinverse of \(X\), the matrix whose first column is the vector of \(x_i\)'s and whose second column is a 1's vector. Plot the observed scores and the predicted scores.

3. The following data indicate the numbers of accidents bus drivers had in one year as a function of the numbers of years on the job.

<table>
<thead>
<tr>
<th>Years on Job ((x_i))</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents ((y_i))</td>
<td>10</td>
<td>8</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

(a) Fit these data with a regression model of the form \(\hat{y} = qx + r\). Determine \(q\) and \(r\) by computing the pseudoinverse of \(X\), the matrix whose first column is the vector of \(x_i\)'s and whose second column is a 1's vector.

(b) What is the condition number of the matrix \((X^TX)\)? Is the problem poorly conditioned?

(c) Repeat the calculations in part (a) by first shifting the \(x\)-values to make the average \(x\)-value be 0 (see Example 6).

4. The following data shows the GPA and the job salary (five years after graduation) of six mathematics majors from Podunk U.

<table>
<thead>
<tr>
<th>GPA</th>
<th>2.3</th>
<th>3.1</th>
<th>2.7</th>
<th>3.4</th>
<th>3.7</th>
<th>2.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td>25,000</td>
<td>38,000</td>
<td>28,000</td>
<td>35,000</td>
<td>30,000</td>
<td>32,000</td>
</tr>
</tbody>
</table>

(a) Fit these data with a regression model of the form \(\hat{y} = qx + r\) using pseudoinverses.

(b) What is the condition number of the matrix \((X^TX)\)? Is the problem poorly conditioned?

(c) Repeat the calculations in part (a) by first shifting the \(x\)-values to make the average \(x\)-value be 0 (see Example 6).
5. Compute the pseudoinverse, and then solve the refinery problem in Example 1 when refinery 1 is shut down (the other two refineries operate).

6. (a) Compute the pseudoinverse, and then solve, the refinery problem in Example 1 when refinery 2 is shut down (the other two refineries operate).
   (b) Compute the error vector \( e = b - Aw \) for part (a). Compute the angle between the error vector \( e \) in part (a) and the solution vector \( Aw \). It should be about 90°. Is it?
   (c) Which refinery closing, of the three refineries, has the smallest error vector (in the sum norm)—this assumes that you have done Exercise 5.

7. Compute the pseudoinverse of the following matrices.
   (a) \[ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \]
   (b) \[ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \]
   (c) \[ \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \]
   (d) \[ \begin{bmatrix} 2 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -2 & 1 \\ -1 & 1 & 2 \end{bmatrix} \]
   (e) \[ \begin{bmatrix} 0 & 2 & -1 \\ 1 & -1 & 3 \\ 2 & 1 & 0 \\ -1 & 4 & 1 \end{bmatrix} \]

8. In each case, find the linear combination of the first two vectors that is as close as possible to the third vector.
   (a) \([1, 2, 1], [2, 0, -1]; [3, -1, 0]\)
   (b) \([1, 0, 1], [0, 1, 1]; [0, 0, 5]\)
   (c) \([0, -2, 3], [1, 1, 1]; [1, -5, 10]\)
   (d) \([2, 0, 1], [-1, 0, 1]; [4, 3, 2]\)
   (e) \([0, 1, 1, 0], [1, -1, -1, 1]; [2, 0, 2, 0]\)

9. (a) Factory A produces 30 cars, 40 light trucks, and 20 heavy trucks per day, while factory B produces 60 cars, 20 light trucks, and 20 heavy trucks a day. If the monthly demand is 1000 cars, 500 light trucks, and 400 heavy trucks, what is the least-squares solution (days of production for each factory)?
   (b) If the monthly demand increased by 10 cars, how much longer would factory A have to work each month?
   Hint: See Example 4 of Section 3.3.
   (c) If the monthly demand increased by 10 light trucks and 5 heavy trucks, how much longer would factory B have to work each month?

10. (a) Bureaucratic office A produces 40 new regulations, inspects 90 defective appliances, and approves 300 applications a week. Bureaucratic office B produces 80 new regulations, inspects 40 defec-
tive appliances, and approves 200 applications a week. How many weeks would each office have to work in order to best approximate (in the least-squares sense) a demand of producing 1000 new regulations, inspecting 700 defective appliances, and approving 2000 applications.

(b) What is the condition number of the matrix \( (A^T A) \) in the pseudoinverse computations? Is this problem poorly conditioned?

(c) If the demand for new regulations increased by 10, how much longer would office A have to work?

11. Consider the regression problem in which high school GPA and total SAT score (verbal plus math) are used to predict a person’s college GPA:

\[
\text{GPA college} = q_1 (\text{GPA high school}) + q_2 \left( \frac{\text{total SAT}}{1000} \right) + r
\]

Suppose that our data for five people are as follows:

<table>
<thead>
<tr>
<th>GPA_{col}</th>
<th>GPA_{hi}</th>
<th>SAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.8</td>
<td>3.0</td>
</tr>
<tr>
<td>B</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>C</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td>D</td>
<td>3.2</td>
<td>3.6</td>
</tr>
<tr>
<td>E</td>
<td>3.8</td>
<td>3.4</td>
</tr>
</tbody>
</table>

(a) Compute the pseudoinverse \( (X^T X)^{-1} X^T \). In the process, determine the condition number of \( (X^T X) \). Is this problem poorly conditioned?

(b) Determine \( q_1 \), \( q_2 \), and \( r \).

(c) Determine the error vector \( e \) (differences between true GPA-college and estimated GPA-college). Is it orthogonal to the estimated GPA-college vector?

12. In Example 5, re-solve the quadratic least-squares approximation problem for the following data points. Note that for parts (a) and (b) the \( x \)-values are the same, so the pseudoinverse will be the same (just use the \( X^+ \) in the text).

(a) Same as in Example 5 but the fourth point is \((3, 5)\).

(b) Same as in Example 5 but the first point is \((0, 9)\).

(c) \((0, 7), (1, 5), (2, 7), (3, 9), (4, 13)\)

(d) \((-2, 7), (-1, 5), (0, 4), (1, 4), (2, 8), (3, 12); how is this problem related to the original problem in Example 5?\)

(e) \((0, 2), (1, 4), (2, 10)\)

13. Fit a cubic polynomial to the following data points using the same idea as in the quadratic fit in Example 5: \((-1, -2), (0, 3), (1, 2), (2, 8), (3, 12), (4, 100)\). What is the condition number of \( (X^T X) \)?
14. Consider the regression model \( \tilde{z} = qx + ry + s \) for the following data, where the \( x \)-value is a scaled score (to have average value of 0) of high school grades, the \( y \)-value is a scaled score of SAT scores, and the \( z \)-value is a scaled score of college grades.

\[
\begin{array}{c|cccc}
  x & -4 & -2 & 0 & 2 & 4 \\
  y & 2 & -1 & -2 & -1 & 2 \\
  z & 3 & 6 & 7 & 7 & 6 \\
\end{array}
\]

Determine \( q \), \( r \), and \( s \). Note that the \( x \), \( y \), and \( 1 \) vectors (in the regression matrix equation \( z = qx + ry + s1 \)) are orthogonal.

15. Consider the regression model \( y_i = qx_i \), \( i = 1, 2, \ldots, n \). Compute the pseudoinverse for this regression problem and solve for \( q \) (in terms of the \( x_i \), \( y_i \) values). As a matrix system \( Xq = y \), the matrix \( X \) is the \( n \)-by-1 column vector of \( x \)-values and \( y \) is the vector of \( y \)-values. Your answer should agree with the formula for \( q \) in equation (7).

16. Use a geometric picture to explain why if \( Aw \) is very close to \( b \), then the error vector \( b - Aw \) may not be exactly orthogonal to the projection vector \( Aw \) in a least-squares solution to \( Ax = b \).

17. Compute the cosine of angle, and determine the angle, made by the following pairs of vectors.

   (a) \([1, 0], [1, 1]\)  
   (b) \([3, 4], [-3, 4]\)  
   (c) \([1, 2], [3, 1]\)  
   (d) \([1, 0, 1], [0, 1, 0]\)  
   (e) \([1, 1, 1], [1, -1, 2]\)  
   (f) \([1, 1, 1], [2, -1, 3]\)  

18. Compute the correlation coefficient between the vectors of \( x \)- and \( y \)-values in Exercise 14, and between the vectors of \( x \)- and \( z \)-values in Exercise 14.

19. The following data show scores that three students received on a battery of six different tests.

<table>
<thead>
<tr>
<th></th>
<th>Gerry</th>
<th>Jimmie</th>
<th>Ronnie</th>
</tr>
</thead>
<tbody>
<tr>
<td>General IQ</td>
<td>12</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Mathematics</td>
<td>8</td>
<td>.22</td>
<td>4</td>
</tr>
<tr>
<td>Reading</td>
<td>16</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Running</td>
<td>24</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Speaking</td>
<td>12</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>Watching</td>
<td>12</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>
Compute the correlation coefficient between
(a) Gerry and Jimmie  (b) Gerry and Ronnie
(c) Jimmie and Ronnie

*Hint*: Remember first to subtract the average value from each number.

20. (a) Compute the correlation coefficient of the following readings from eight students of their IQs and scores at Zaxxon.

<table>
<thead>
<tr>
<th>Student</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>120</td>
<td>130</td>
<td>105</td>
<td>90</td>
<td>125</td>
<td>120</td>
<td>110</td>
<td>160</td>
</tr>
<tr>
<td>Zaxxon</td>
<td>11,000</td>
<td>7,000</td>
<td>10,000</td>
<td>12,000</td>
<td>8,000</td>
<td>100,000</td>
<td>8,000</td>
<td>6,000</td>
</tr>
</tbody>
</table>

(b) Delete student F and recompute the correlation coefficient.

*Hint*: Remember first to subtract the average value from each number.

21. Suppose that there are two dials A and B on a machine that produces steel. We want to find out how settings $a_i$, $b_i$ of the two dials affect the quality $c_i$ of the steel. We use a regression model $\hat{\mathbf{c}} = p\mathbf{a} + q\mathbf{b} + r$. For each of the following vectors a of settings for dial A, find a vector b of settings of dial B that is orthogonal to the dial A vector and also orthogonal to the 1’s vector.

(a) $\mathbf{a} = [2, 1, 0, -1, -2]$
(b) $\mathbf{a} = [-4, -1, 0, 2, 3]$
(c) $\mathbf{a} = [2, 6, 1, -4, -3, -2]$

22. Prove that the pseudoinverse $A^+$ equals the true inverse $A^{-1}$ if the $n$-by-$n$ matrix A is invertible (and hence has rank $n$).

23. (a) Show that in the euclidean norm $|\mathbf{a} - \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$ by squaring both sides of this inequality and using the law of cosines (on the left side).
(b) Show that in the euclidean norm $|\mathbf{a} + \mathbf{c}| \leq |\mathbf{a}| + |\mathbf{c}|$ by letting $\mathbf{b} = -\mathbf{c}$ (and hence $-\mathbf{b} = \mathbf{c}$) and using part (a).

24. Show that if $\mathbf{a}$ is orthogonal to $\mathbf{b}$ and $\mathbf{c}$, then $\mathbf{a}$ cannot be linearly dependent on $\mathbf{b}$ and $\mathbf{c}$.

25. If the vectors in a basis of vector space $V$ are mutually orthogonal to the vectors in a basis of vector space $W$, show that every vector in $V$ is orthogonal to every vector in $W$.

26. For each of the following matrices $A$, express the 1’s vector $\mathbf{1}$ (of the appropriate length) as the unique sum of two vectors, $\mathbf{1} = \mathbf{b}_1 + \mathbf{b}_2$, such that $\mathbf{b}_1$ is in Range($A$) and $\mathbf{b}_2$ is in the error space of $A$. This unique sum exists by Theorem 7, part (i).
Sec. 5.4 Orthogonal Systems

27. For each matrix $A$ in Exercise 26, find a basis $\{v_i\}$ for the Range($A$) [= Col($A$)] and a basis $\{w_i\}$ for Null($A^T$). Then verify that the $v_i$ are orthogonal to the $w_i$, as required by Theorem 7, part (ii).

28. For each of the following matrices $A$, express the 1’s vector $\mathbf{1}$ as a unique sum, $\mathbf{1} = x_1 + x_2$ of a vector $x_1$ in Row($A$) and a vector $x_2$ in Null($A$).

\[
(a) \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \quad (b) \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad (c) \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix}
\]

29. Find a solution to $Ax = \mathbf{1}$, for $A$ the matrix in Exercise 28, part (c), in which $x$ is in Row($A$).

30. Use Theorem 8 to prove that if $v_1, v_2, \ldots, v_k$ are a linearly independent set of vectors in the row space of a matrix $A$, then $w_i = Av_i$ are a linearly independent set of vectors in the range of $A$. Thus, if $\{v_i\}$ are a basis for Row($A$), then $\{Av_i\}$ are a basis for Col($A$).
The inverse $A^{-1}$ of a matrix $A$ with orthogonal columns $a_i^C$ is easy to describe. It is essentially the same as the pseudoinverse: $A^{-1}$ is formed by dividing each column $a_i^C$ by $a_i^C \cdot a_i^C$, the sum of the squares of its entries, and forming the transpose of the resulting matrix. Thus, if $s_i = a_i^C \cdot a_i^C$, then

$$A^{-1} = \begin{bmatrix}
\frac{1}{s_1} a_1^C \\
\frac{1}{s_2} a_2^C \\
\vdots \\
\frac{1}{s_n} a_n^C
\end{bmatrix} \quad (1)$$

We verify (1) by noting that entry $(i, j)$ in $A^{-1}A$ will be 0 if $i \neq j$ because $a_i^C \cdot a_j^C = 0$ (the columns are orthogonal). Entry $(i, i)$ equals $(a_i^C/s_i) \cdot a_i^C = a_i^C \cdot a_i^C/(a_i^C \cdot a_i^C) = 1$.

**Example 1. Inverse of Matrix with Orthogonal Columns**

(i) Consider the matrix $A = \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$, whose columns are orthogonal. The sum of the squares of the entries in each column of $A$ is $3^2 + 4^2 = 25$. If we divide each column by 25 and take the transpose, we obtain

$$A^{-1} = \begin{bmatrix}
\frac{3}{25} & \frac{4}{25} \\
-\frac{4}{25} & \frac{3}{25}
\end{bmatrix}$$

The reader should check that this matrix is exactly what one would get by computing this 2-by-2 inverse using elimination.

(ii) Consider the orthogonal-column matrix

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

Its inverse, by (1), is

$$A^{-1} = \begin{bmatrix}
\frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & \frac{1}{3} & -\frac{1}{3}
\end{bmatrix}$$

Again the reader should check that $A^{-1}A = I$. \(\blacksquare\)
Sec. 5.4 Orthogonal Systems

Let us use (1) to obtain a formula for the \( i \)th component \( x_i \) in the solution \( \mathbf{x} \) to \( \mathbf{A}\mathbf{x} = \mathbf{b} \). Given the inverse \( \mathbf{A}^{-1} \), we can find \( \mathbf{x} \) as \( \mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \). The \( i \)th component in \( \mathbf{A}^{-1}\mathbf{b} \) is the scalar product of \( i \)th row of \( \mathbf{A}^{-1} \) with \( \mathbf{b} \). By (1), the \( i \)th row of \( \mathbf{A}^{-1} \) is \( \mathbf{a}_i^C/(a_i^C \cdot a_i^C) \) and thus

\[
    x_i = \frac{\mathbf{a}_i^C \cdot \mathbf{b}}{\mathbf{a}_i^C \cdot \mathbf{a}_i^C} \tag{2}
\]

Our old friend, the length of the projection of \( \mathbf{b} \) onto column \( \mathbf{a}_i^C \) (see Theorem 2 of Section 5.3).

A set of orthogonal vectors of unit length (whose norm is 1) are called orthonormal. The preceding formulas for \( x_i \) and \( \mathbf{A}^{-1} \) become even nicer if the columns of \( \mathbf{A} \) are orthonormal. In this case, \( \mathbf{a}_i^C \cdot \mathbf{a}_i^C = 1 \). Then the denominator in (2) is 1, so now the projection formula is \( x_i = \mathbf{a}_i^C \cdot \mathbf{b} \). To obtain \( \mathbf{A}^{-1} \), we divide each column of \( \mathbf{A} \) by 1 and form the transpose: that is, \( \mathbf{A}^{-1} = \mathbf{A}^T \). Summarizing this discussion, we have

**Theorem 1**

(i) If \( \mathbf{A} \) is an \( n \)-by-\( n \) matrix whose columns are orthogonal, then \( \mathbf{A}^{-1} \) is obtained by dividing the \( i \)th column of \( \mathbf{A} \) by the sum of the squares of its entries and transposing the resulting matrix [see (1)].

The \( i \)th component \( x_i \) in the solution of \( \mathbf{A}\mathbf{x} = \mathbf{b} \) is the length of the projection of \( \mathbf{b} \) on \( \mathbf{a}_i^C \): \( x_i = \mathbf{a}_i^C \cdot \mathbf{b}/\mathbf{a}_i^C \cdot \mathbf{a}_i^C \).

(ii) If the columns of \( \mathbf{A} \) are orthonormal, then the inverse \( \mathbf{A}^{-1} \) is \( \mathbf{A}^T \) and the length of the projection is just \( x_i = \mathbf{a}_i^C \cdot \mathbf{b} \).

Suppose that we have a basis of \( n \) orthogonal vectors \( \mathbf{q}_i \) for \( n \)-space. If \( \mathbf{Q} \) has the \( \mathbf{q}_i \) as its columns, the solution \( \mathbf{x} = \mathbf{b}^* \) of \( \mathbf{Q}\mathbf{x} = \mathbf{b} \) will be a vector \( \mathbf{b}^* \) of lengths of the projections of \( \mathbf{b} \) onto each \( \mathbf{q}_i \):

\[
    \mathbf{Q}\mathbf{b}^* = b_1^*\mathbf{q}_1 + b_2^*\mathbf{q}_2 + \cdots + b_n^*\mathbf{q}_n = \mathbf{b} \tag{3}
\]

Here the term \( b_i^*\mathbf{q}_1 \) is just the projection of \( \mathbf{b} \) onto \( \mathbf{q}_1 \). So (3) simply says

**Corollary.** Any \( n \)-vector \( \mathbf{b} \) can be expressed as the sum of the projections of \( \mathbf{b} \) onto a set of \( n \) orthogonal vectors \( \mathbf{q}_i \).

**Example 2. Conversion of Coordinates from One Basis to Another**

Consider the orthonormal basis \( \mathbf{q}_1 = [0.8, 0.6], \mathbf{q}_2 = [-0.6, 0.8] \) for 2-space. To express the vector \( \mathbf{b} = [1, 2] \) in terms of \( \mathbf{q}_1, \mathbf{q}_2 \) coordinates, we need to solve the system

\[
    b_1^* \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix} + b_2^* \begin{bmatrix} -0.6 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\]
Figure 5.6

By Theorem 1,

\[ b_1^* = q_1 \cdot b = .8 \times 1 + .6 \times 2 = 2, \]
\[ b_2^* = q_2 \cdot b = -.6 \times 1 + .8 \times 2 = 1 \]

where \( b_1^* q_1 = 2[.8, .6] \) is the projection of \( b \) on \( q_1 \), and \( b_2^* q_2 = [-.6, .8] \) is the projection of \( b \) on \( q_2 \). Thus \( b = [1, 2] \) is expressed as an \( e_1 - e_2 \) coordinate vector, while \( b^* = [2, 1] \) is the same vector expressed in \( q_1 - q_2 \) coordinates. A geometric picture of this conversion is given in Figure 5.6, where the vector \([2, 1]\) is depicted as the sum of its projection onto \( q_1 \) and onto \( q_2 \).

Theorem 1 is a carbon copy of Theorem 5 of Section 5.3 about pseudoinverses when columns are orthogonal. As with the inverse, if \( A \)'s columns are orthonormal, the pseudoinverse \( A^+ \) of \( A \) will simply be \( A^T \). The following example gives a familiar illustration of this result and shows why orthogonal columns make inverses and pseudoinverse so similar.

**Example 3. Pseudoinverse of Matrix with Orthonormal Columns**

Let \( I_2 \) be the first two columns of the 3-by-3 identity matrix.

\[ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \]

Then

\[ I_2^+ = I_2^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \]

For any vector \( b = [b_1, b_2, b_3] \), the least-squares solution \( x = b^* \) to \( I_2 x = b \) is
Sec. 5.4 Orthogonal Systems

**Optional**

There is another interesting geometric fact about orthonormal columns (see the Exercises for the two-dimensional case).

**Theorem 2.** When $Q$ has orthonormal columns, then solving $Qx = b$ for $b^* = Q^Tb$ is equivalent to performing the orthonormal change of basis $b \rightarrow b^* = Q^Tb$. Such a basis change is simply a rotation of the coordinate axes, a reflection through a plane, or a combination of both. The entries in $Q$ can be expressed in terms of the sines and cosines of the angles of this rotation.

For example, the rotation of axis in the plane by $\theta^\circ$ is a linear transformation $R$ of 2-space:

$$R: \begin{align*}
x' &= x \cos \theta^\circ + y \sin \theta^\circ \\
y' &= -x \sin \theta^\circ + y \cos \theta^\circ
\end{align*}$$

where

$$A = \begin{bmatrix}
\cos \theta^\circ & \sin \theta^\circ \\
-\sin \theta^\circ & \cos \theta^\circ
\end{bmatrix}$$

It is easy to check that $A$ has orthonormal columns.

It follows that the distance between a pair of vectors and the angle that they form do not change with an orthonormal change of basis.

(Note: End of optional material.)

Orthogonal columns have another important advantage besides easy formulas. A highly nonorthogonal set of columns—that is, columns that are almost parallel—can result in unstable computations.

**Example 4. Nonorthogonal Columns**

Consider the following system of equations:

\begin{align*}
1x_1 + .75x_2 &= 5 \\
1x_1 + 1x_2 &= 7
\end{align*}
Let us call the two column vectors in the coefficient matrix of (4): $u = [1, 1]$ and $v = [.75, 1]$. The cosine of their angle is, by Theorem 6 of Section 5.3,

$$\cos \theta(u, v) = \frac{u \cdot v}{|u||v|} = \frac{1.75}{\sqrt{2} \cdot 1.25} = .99$$

The angle with cosine of .99 is $8^\circ$. Thus $u$ and $v$ are almost parallel (almost the same vector). Representing any 2-vector $b$ as a linear combination of two vectors that are almost the same is tricky, that is, unstable. For example, to solve (4) we must we find weights $x_1$, $x_2$ such that

$$x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} .75 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

(6)

The system (4) is the canoe-with-sail system from Section 1.1. We already know that calculations with $A$, the coefficient matrix in (4), are very unstable. In Section 3.5 we computed the condition number of $A$ to be $c(A) = 16$. Recall that the condition number $c(A) = \|A\| \cdot \|A^{-1}\|$ measures how much a relative error in the entries of $A$ (or in $b$) could affect the relative error in $x = [x_1, x_2]$; in this case, a 5% error in $b$ could cause an error $16 [= c(A)]$ times greater in $x$, a $16 \times 5% = 80\%$ error.

We solved (4) in Section 1.1 and obtained $x_1 = -1$, $x_2 = 7$. If we had solved for $b' = [7, 5]$, we would have obtained the answer $x_1 = 13$, $x_2 = -8$ (see Figure 5.7 for a picture of this result). Or for $b'' = [6, 6]$, $x_1 = 6$, $x_2 = 0$.  

Figure 5.7
Reading the results of Example 4 in reverse, we see that when errors arise in solving an ill-conditioned system of equations $Ax = b$ (in which $A$ has a large condition number), the problem should be that some column vector (or a linear combination of them) forms a small angle with another column vector—this means that the columns are almost linearly dependent. If the columns were close to mutually orthogonal, the system $Ax = b$ would be well-conditioned.

**Principle.** Let $A$ be an $n$-by-$n$ matrix with rank($A$) = $n$ so that the system of equations $Ax = b$ has a unique solution. The solution to $Ax = b$ will be more or less stable according to how close or far from orthogonal the column vectors of $A$ are.

Suppose that the columns of the $n$-by-$n$ matrix $A$ are linearly independent but not orthogonal. We shall show how to find a new $n$-by-$n$ matrix $A^*$ of orthonormal columns (orthogonal and unit length) that are linear combinations of the columns of $A$.

Our procedure can be applied to any basis $a_1, a_2, \ldots, a_m$ of an $m$-dimensional space $V$ and will yield a new basis of $m$ orthonormal vectors $q_i$ for $V$ (unit-length vectors make calculations especially simple). The procedure is inductive in the sense that the first $k$ $q_i$ will be an orthonormal basis for the space $V_k$ generated by the first $k$ $a_i$. The method is called **Gram–Schmidt orthogonalization**.

For $k = 1$, $q_1$ should be a multiple of $a_1$. To make $q_1$ have norm 1, we set $q_1 = a_1/|a_1|$. Next we must construct from $a_2$ a second unit vector $q_2$ orthogonal to $q_1$. We divide $a_2$ into two "parts": the part of $a_2$ parallel to $q_1$ and the part of $a_2$ orthogonal (perpendicular) to $q_1$ (see Figure 5.8). The component of $a_2$ in $q_1$'s direction is simply the projection of $a_2$ onto $q_1$. This projection is $sq_1$, where the length $s$ of the projection is

$$s = \frac{a_2 \cdot q_1}{q_1 \cdot q_1} = \frac{a_2 \cdot q_1}{|a_2|}$$

since $q_1 \cdot q_1 = 1$. The rest of $a_2$, the vector $a_2 - sq_1$, is orthogonal to the projection $sq_1$, and hence orthogonal to $q_1$. So $a_2 - sq_1$ is the orthogonal vector we want for $q_2$. To have unit norm, we set $q_2 = (a_2 - sq_1)/|a_2 - sq_1|$.

Let us show how the procedure works thus far.

**Figure 5.8** Gram–Schmidt orthogonalization.
Example 5. Gram–Schmidt Orthogonalization in Two Dimensions

Suppose that $a_1 = [3, 4]$ and $a_2 = [2, 1]$ (see Figure 5.8). We set

$$q_1 = \frac{a_1}{|a_1|} = \frac{[3, 4]}{5} = \left[\frac{3}{5}, \frac{4}{5}\right]$$

We project $a_2$ onto $q_1$ to get the part of $a_2$ parallel to $q_1$. From (7), the length of the projection is

$$s = a_2 \cdot q_1 = [2, 1] \cdot \left[\frac{3}{5}, \frac{4}{5}\right] = \frac{2 \cdot 3 + 1 \cdot 4}{5} = \frac{10}{5} = 2$$

and the projection is $s q_1 = 2 \left[\frac{3}{5}, \frac{4}{5}\right] = \left[\frac{6}{5}, \frac{8}{5}\right]$. Next we determine the other part of $a_2$, the part orthogonal to $s q_1$:

$$a_2 - s q_1 = [2, 1] - \left[\frac{6}{5}, \frac{8}{5}\right] = \left[\frac{4}{5}, -\frac{3}{5}\right]$$

Since $\left|\left[\frac{4}{5}, -\frac{3}{5}\right]\right| = 1$, then

$$q_2 = \frac{a_2 - s q_1}{|a_2 - s q_1|} = \frac{\left[\frac{4}{5}, -\frac{3}{5}\right]}{1} = \left[\frac{4}{5}, -\frac{3}{5}\right]$$

We extend the previous construction by finding the projections of $a_3$ onto $q_1$ and $q_2$. Then the vector $a_3 - s_1 q_1 - s_2 q_2$, which is orthogonal to $q_1$ and $q_2$ should be $q_3$; as before, we divide $a_3 - s_1 q_1 - s_2 q_2$ by its norm to make $q_3$ unit length. We continue this process to find $q_4$, $q_5$, and so on.

Example 6. Gram–Schmidt Orthogonalization of 3-by-3 Matrix

Let us perform orthogonalization on the matrix $A$ whose $i$th column we denote by $a_i$.

$$A = \begin{bmatrix} 0 & 3 & 2 \\ 3 & 5 & 5 \\ 4 & 0 & 5 \end{bmatrix}$$

(8)

First $q_1 = a_1/|a_1| = [0, 3, 4]/5 = \left[0, \frac{3}{5}, \frac{4}{5}\right]$. The length of the projection $a_2$ onto $q_1$ is

$$s = a_2 \cdot q_1 = 3 \cdot 0 + 5 \cdot \frac{3}{5} + 0 \cdot \frac{4}{5} = 3$$

(9a)

So the projection of $a_2$ onto $q_1$ is

$$s q_1 = 3 \left[0, \frac{3}{5}, \frac{4}{5}\right] = \left[0, \frac{9}{5}, \frac{12}{5}\right]$$
Next we compute
\[
\mathbf{a}_2 - s\mathbf{q}_1 = [3, 5, 0] - \left[0, \frac{9}{5}, \frac{12}{5}\right] = [3, \frac{16}{5}, -\frac{12}{5}]
\]
where \(|\mathbf{a}_2 - s\mathbf{q}_1| = \sqrt{9 + 256/25 + 144/25} = 5\). Then
\[
\mathbf{q}_2 = \frac{\mathbf{a}_2 - s\mathbf{q}_1}{|\mathbf{a}_2 - s\mathbf{q}_1|} = \left[\frac{3}{5}, \frac{16}{25}, -\frac{12}{25}\right]
\]
We compute the length of the projections of \(\mathbf{a}_3\) onto \(\mathbf{q}_1\) and \(\mathbf{q}_2\):
\[
s_1 = \mathbf{a}_3 \cdot \mathbf{q}_1 = 2 \cdot 0 + 5 \cdot \frac{3}{5} + 5 \cdot \frac{4}{5} = 3 + 4 = 7
\]
\[
s_2 = \mathbf{a}_3 \cdot \mathbf{q}_2 = 2 \cdot \frac{3}{5} + 5 \cdot \frac{16}{25} + 5 \cdot \left(-\frac{12}{25}\right)
= \frac{6}{5} + \frac{16}{5} - \frac{12}{5} = 2
\]
(9b)
Then
\[
s_1\mathbf{q}_1 = 7\left[0, \frac{3}{5}, \frac{4}{5}\right] = \left[0, \frac{21}{5}, \frac{28}{5}\right]
\]
\[
s_2\mathbf{q}_2 = 2\left[\frac{2}{5}, \frac{12}{25}, -\frac{12}{25}\right] = \left[\frac{4}{5}, \frac{24}{25}, -\frac{24}{25}\right]
\]
and
\[
\mathbf{a}_3 - s_1\mathbf{q}_1 - s_2\mathbf{q}_2 = [2, 5, 5] - \left[0, \frac{21}{5}, \frac{28}{5}\right] - \left[\frac{4}{5}, \frac{24}{25}, -\frac{24}{25}\right]
= \left[\frac{3}{5}, -\frac{12}{25}, \frac{9}{25}\right]
\]
Since computation reveals that \(|\mathbf{a}_3 - s_1\mathbf{q}_1 - s_2\mathbf{q}_2| = 1\), then
\[
\mathbf{q}_3 = (\mathbf{a}_3 - s_1\mathbf{q}_1 - s_2\mathbf{q}_2) = \left[\frac{3}{5}, -\frac{12}{25}, \frac{9}{25}\right]
\]
The matrix of these new orthogonal column vectors is
\[
\mathbf{Q} = \begin{bmatrix}
0 & \frac{3}{5} & \frac{3}{5} \\
\frac{3}{5} & \frac{12}{25} & -\frac{12}{25} \\
\frac{4}{5} & -\frac{12}{25} & \frac{9}{25}
\end{bmatrix}
\]
(10)

In keeping with the principle above, the accuracy of this procedure depends on how close to and far from orthogonality the columns \(\mathbf{a}_i\) are. If a linear combination of some \(\mathbf{a}_i\) forms a small angle with another vector \(\mathbf{a}_k\) (this means the matrix \(\mathbf{A}\) has a large condition number), then the resulting \(\mathbf{q}_i\) will have errors, making them not exactly orthogonal. However, more stable methods are available using advanced techniques, such as Householder transformations.
Suppose that the columns of $A$ are not linearly independent. If, say, $a_3$ is a linear combination of $a_1$ and $a_2$, then in the Gram–Schmidt procedure the error vector $a_3 - s_1q_1 - s_2q_2$ with respect to $q_1$ and $q_2$ will be $0$. In this case we skip $a_3$ and use $a_4 - s_1q_1 - s_2q_2$ to define $q_3$. The number of vectors $q_i$ formed will be the dimension of the column space of $A$, that is, $\text{rank}(A)$.

The effect of the orthogonalization process can be represented by an upper triangular matrix $R$ so that one obtains the matrix factorization

**Theorem 3.** Any $m$-by-$n$ matrix $A$ can be factored in the form

$$A = QR$$

where $Q$ is the $m$-by-$\text{rank}(A)$ matrix with orthonormal columns $q_i$ obtained by Gram–Schmidt orthogonalization, and $R$ is an upper triangular matrix of size $\text{rank}(A)$-by-$n$ (described below).

For $i < j$, entry $r_{ij}$ of $R$ is $a_j \cdot q_i$, the projection of $a_j$ onto $q_i$. The diagonal entries in $R$ are the sizes, before normalization, of the new columns: $r_{11} = |a_1|$, $r_{22} = |a_2 - s_1q_1|$, $r_{33} = |a_3 - s_1q_1 - s_2q_2|$, and so on.

**Example 7. QR Decomposition**

Give the $QR$ decomposition for the matrix $A$ in Example 6.

$$A = \begin{bmatrix} 0 & 3 & 2 \\ 3 & 5 & 5 \\ 4 & 0 & 5 \end{bmatrix}$$

The orthonormal matrix $Q$ is given in (10). We form $R$ from the information about the sizes of new columns and the projections as described in the preceding paragraph. Here $r_{12} = s = 3$ in (9a), and $r_{13} = s_1 = 7, r_{23} = s_2 = 2$ in (9b). Then

$$QR = \begin{bmatrix} 0 & \frac{3}{5} & \frac{4}{5} \\ \frac{3}{5} & \frac{16}{25} & -\frac{12}{25} \\ \frac{4}{5} & -\frac{12}{25} & \frac{9}{25} \end{bmatrix} \begin{bmatrix} 5 & 3 & 7 \\ 0 & 5 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Let us compute the second column of $QR$—multiplying $Q$ by $r^C_2$, the second column of $R$—and show that the result is $a_2$, the second column of $A$.

$$Qr^C_2 = \begin{bmatrix} 0 & \frac{3}{5} & \frac{4}{5} \\ \frac{3}{5} & \frac{16}{25} & -\frac{12}{25} \\ \frac{4}{5} & -\frac{12}{25} & \frac{9}{25} \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}$$

$$= 3 \begin{bmatrix} \frac{3}{5} \\ \frac{16}{25} \\ -\frac{12}{25} \end{bmatrix} + 5 \begin{bmatrix} \frac{4}{5} \\ -\frac{12}{25} \\ \frac{9}{25} \end{bmatrix} + 0 \begin{bmatrix} \frac{4}{5} \\ \frac{16}{25} \\ \frac{9}{25} \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} = a_2$$

$\blacksquare$
Columns of $Q$ are obtained from linear combinations of the columns of $A$. Reversing this procedure yields the columns of $A$ as linear combinations of the columns of $Q$. This reversal is what is accomplished by the matrix product $QR$. Consider the computation in (12). In terms of the columns $q_i$ of $Q$, (12) is

$$3q_1 + 5q_2 + 0q_3 = a_2$$

or, in terms of $R$,

$$r_{12}q_1 + r_{22}q_2 = a_2 \tag{13}$$

($a_2$ equals its projection onto $q_1$ plus its projection onto $q_2$).

Next consider the formula for $q_2$:

$$q_2 = \frac{a_2 - sq_1}{|a_2 - sq_1|} = \frac{a_2 - r_{12}q_1}{r_{22}} \tag{14}$$

since $r_{22} = |a_2 - sq_1|$ and $r_{12} = s$. Solving for $a_2$ in (14), we obtain (13)

$$q_2 = \frac{a_2 - r_{12}q_1}{r_{22}} \rightarrow r_{12}q_1 + r_{22}q_2 = a_2$$

The same analysis shows that the $j$th column in the product $QR$ is just a reversal of the orthogonalization steps for finding $q_j$.

The matrix $R$ is upper triangular because column $a_j$ is only involved in building columns $q_i$, $q_{i+1}$, $q_n$ of $Q$. The $QR$ decomposition is the column counterpart to the $LU$ decomposition, given in Section 3.2, in which the row combinations of Gaussian elimination are reversed to obtain the matrix $A$ from its row-reduced matrix $U$.

The $QR$ decomposition is used frequently in numerical procedures. We use it to find eigenvalues in the appendix to Section 5.5.

We will sketch one of its most frequent uses, finding the inverse or pseudoinverse of an ill-conditioned matrix. If $A$ is an $n$-by-$n$ matrix with linearly independent columns, the decomposition $A = QR$ yields

$$A^{-1} = (QR)^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T \tag{15}$$

The fact that $Q^{-1} = Q^T$ when $Q$ has orthonormal columns was part of Theorem 1. Given the QR decomposition of $A$, (15) says that to get $A^{-1}$, we only need to determine $R^{-1}$. Since $R$ is an upper triangular matrix, its inverse is obtained quickly by back substitution (see Exercise 12 of Section 3.5). When $A$ is very ill-conditioned, one should compute $A^{-1}$ via (15): first, determining the QR decomposition of $A$, using advanced (more stable) variations of the Gram–Schmidt procedure; then determining $R^{-1}$; and thus obtaining $A^{-1} = R^{-1}Q^T$.

Equation (15) extends to pseudoinverses. That is, if $A$ is an $m$-by-$n$
matrix with linearly independent columns and \( m > n \), then its pseudoinverse \( A^+ \) can be computed as

\[
A^+ = R^{-1}Q^T
\]  

(16)

See the Exercises for instructions on how to verify (16) and examples of its use. *This formula for the pseudoinverse is the standard way pseudoinverses are computed in practice.* Even if one determines \( Q \) and \( R \) using the basic Gram–Schmidt procedure given above, the resulting \( A^+ \) from (16) will be substantially more accurate than computing \( A^+ \) using the standard formula \( A^+ = (A^T A)^{-1} A^T \), because the matrix \( A^T A \) tends to be ill-conditioned. For example, in the least-squares polynomial-fitting problem in Example 5 of Section 5.4, the condition number of the 3-by-3 matrix \( X^T X \) was around 2000!

**Principle.** Because of conditioning problems, the pseudoinverse \( A^+ \) of a matrix \( A \) should be computed by the formula \( A^+ = R^{-1}Q^T \), where \( Q \) and \( R \) are the matrices in the QR decomposition of \( A \).

We now introduce a very different use of orthogonality. Our goal is to make a vector space for the set of all continuous functions. To make matters a little easier, let us focus on functions that can be expressed as a polynomial or infinite series in powers of \( x \), such as \( x^3 + 3x^2 - 4x + 1 \) or \( e^x \) or \( \sin x \).

Recall that the defining property of a vector space \( V \) is that if \( u \) and \( v \) are in \( V \), then \( ru + sv \) is also in \( V \), for any scalars \( r, s \). Clearly, linear combinations of polynomials (or infinite series) are again polynomials (or infinite series), so these functions form a vector space.

*For a vector space of functions to be useful, we need a coordinate system, that is, a basis of independent functions \( u_1(x) \) (functions that are not linearly dependent on each other) so that any function \( f(x) \) can be expressed as a linear combination of these basis functions.*

\[
f(x) = f_1u_1(x) + f_2u_2(x) + \cdots
\]

(17)

This basis will need to be infinite and the linear combinations of basis functions may also be infinite. The best basis would use orthogonal, or even better, orthonormal functions.

To make an orthogonal basis, we first need to extend the definition of a scalar, or inner, product \( \mathbf{c} \cdot \mathbf{d} \) of vectors to an inner product of functions. The *inner product of two functions* \( f(x) \) and \( g(x) \) on the interval \([a, b]\) is defined as

\[
f(x) \cdot g(x) = \int_a^b f(x)g(x) \, dx
\]

(18)

This definition is a natural generalization of the standard inner product \( \mathbf{c} \cdot \mathbf{d} \) in that both \( \mathbf{c} \cdot \mathbf{d} \) and \( f(x) \cdot g(x) \) form sums of term-by-term products of the respective entities, but in (18) we have a continuous sum, an integral.
With an inner product defined, most of the theory and formulas defined for vector spaces can be applied to our space of functions. The inner product tells us when two vectors $c$, $d$ are orthogonal (if $c \cdot d = 0$), and allows us to compute coordinates $c_i^*$ of $c$ in an orthonormal basis $u_i$: $c_i^* = c \cdot u_i$ (these coordinates are just the projections of $c$ onto the $u_i$). We can now do the same calculations for functions with (18).

The functional equivalent of the euclidean norm is defined by

$$|f(x)|^2 = f(x) \cdot f(x) = \int_{a}^{b} f(x)^2 \, dx \tag{19}$$

The counterpart of the sum norm $|c|_s = \sum |c_i|$ for vectors is $|f(x)|_s = \int |f(x)| \, dx$.

An orthonormal basis for our functions on the interval $[a, b]$ will be a set of functions $\{u_i(x)\}$ which are orthogonal—by (18), $\int u_i(x)u_j(x) \, dx = 0$, for all $i \neq j$—and whose norms are 1—by (19), $\int u_i(x)^2 \, dx = 1$. Given such an orthonormal basis $\{u_i(x)\}$, the coordinates $f_i$ of a function $f(x)$ in terms of the $u_i(x)$ are computed by the projection formula $f_i = f(x) \cdot u_i(x)$ used for $n$-dimensional orthonormal bases:

$$f(x) = [f(x) \cdot u_1(x)]u_1(x) + [f(x) \cdot u_2(x)]u_2(x) + \cdots \tag{20}$$

How do we find such an orthonormal basis? The first obvious choice is the set of powers of $x$: $1, x, x^2, x^3, \ldots$. These are linearly independent; that is, $x^k$ cannot be expressed as a linear combination of smaller powers of $x$. Unfortunately, there is no interval on which $1, x, x^2$ are mutually orthogonal. On $[-1, 1]$, $1 \cdot x = \int x \, dx = 0$ and $x \cdot x^2 = \int x^3 \, dx = 0$, but $1 \cdot x^2 = \int x^2 \, dx = \frac{2}{3}$.

There are many sets of orthogonal functions that have been developed over the years. We shall mention two, Legendre polynomials and Fourier trigonometric functions.

The Gram-Schmidt orthogonalization procedure provides a way to build an orthonormal basis out of a basis of linearly independent vectors. The calculations in this procedure use inner products, and hence this procedure can be applied to the powers of $x$ (which are linearly independent but, as we just said, far from orthogonal) to find an orthonormal set of polynomials.

When the interval is $[-1, 1]$, the polynomials obtained by orthogonalization are called Legendre polynomials $L_k(x)$. Actually, we shall not worry about making their norms equal to 1. As noted above, the functions $x^0 = 1$ and $x$ are orthogonal on $[-1, 1]$. So $L_0(x) = 1$ and $L_1(x) = x$. Also, $x^2$ is orthogonal to $x$ but not to 1 on $[-1, 1]$. We must subtract off the projection of $x^2$ onto 1:

$$L_2(x) = x^2 - \left( \frac{1 \cdot x^2}{1 \cdot 1} \right) 1 = x^2 - \frac{\int x^2 \, dx}{\int 1 \, dx} = x^2 - \frac{\frac{2}{3}}{2} = x^2 - \frac{1}{3} \tag{21}$$

A similar orthogonalization computation shows that $L_3(x) = x^3 - \frac{3}{8}x$. 


Example 8. Approximating $e^x$ by Legendre Polynomials

Let us use the first four Legendre polynomials $L_0(x) = 1$, $L_1(x) = x$, $L_2(x) = x^2 - \frac{1}{3}$, $L_3(x) = x^3 - 3x/5$ to approximate $e^x$ on the interval $[-1, 1]$. We want the first four terms in (20):

$$e^x = w_0 L_0(x) + w_1 L_1(x) + w_2 L_2(x) + w_3 L_3(x)$$

$$= w_0 + w_1 x + w_2 \left( x^2 - \frac{1}{3} \right) + w_3 \left( x^3 - \frac{3x}{5} \right)$$

where $w_i = e^x \cdot L_i(x)/L_i(x) \cdot L_i(x) = \int e^x L_i(x) \, dx/\int L_i(x)^2 \, dx$. For example,

$$w_2 = \frac{\int_{-1}^{1} e^x (x^2 - \frac{1}{3}) \, dx}{\int_{-1}^{1} (x^2 - \frac{1}{3})^2 \, dx}$$

With a little calculus, we compute the $w_i$ to be (approximately)

$$w_0 = \frac{2.35}{2} = 1.18, \quad w_1 = \frac{.736}{.667} = 1.10,$$

$$w_2 = \frac{.096}{.178} = .53, \quad w_3 = \frac{.008}{.046} = .18$$

Then (22) becomes

$$e^x = 1.18 + 1.10x + .53 \left( x^2 - \frac{1}{3} \right) + .18 \left( x^3 - \frac{3x}{5} \right)$$

If we collect like powers of $x$ together on the right side, (23) simplifies to

$$e^x = 1 + x + .53x^2 + .18x^3$$

Comparing our approximation against the real values of $e^x$ at the points $-1$, $-0.5$, $0$, $0.5$, $1$, we find

<table>
<thead>
<tr>
<th>$x$</th>
<th>-1</th>
<th>-0.5</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^x$</td>
<td>.37</td>
<td>.61</td>
<td>1</td>
<td>1.64</td>
<td>2.72</td>
</tr>
<tr>
<td>Legendre approximation</td>
<td>.37</td>
<td>.61</td>
<td>1</td>
<td>1.65</td>
<td>2.71</td>
</tr>
</tbody>
</table>
Sec. 5.4 Orthogonal Systems

A pretty good fit. In particular, it is a better fit on \([-1, 1]\) than simply using the first terms of the power series for \(e^x\), namely, \(1 + x + x^2/2 + x^3/6\). The approximation gets more accurate as more Legendre polynomials are used.

Over the interval \([0, 2\pi]\) the trigonometric functions \((1/\sqrt{\pi}) \sin kx\) and \((1/\sqrt{\pi}) \cos kx\), for \(k = 1, 2, \ldots\), plus the constant function \(1/\sqrt{2\pi}\) are an orthonormal basis. To verify that they are orthogonal requires showing that

\[
\frac{1}{\sqrt{\pi}} \sin jx \cdot \frac{1}{\sqrt{\pi}} \cos kx = \frac{1}{\pi} \int_0^{2\pi} \sin jx \cos kx \, dx = 0
\]

for all \(j, k\)

\[
\frac{1}{\sqrt{\pi}} \sin jx \cdot \frac{1}{\sqrt{\pi}} \sin kx = \frac{1}{\pi} \int_0^{2\pi} \sin jx \sin kx \, dx = 0
\]

for all \(j \neq k\)

\[
\frac{1}{\sqrt{\pi}} \cos jx \cdot \frac{1}{\sqrt{\pi}} \cos kx = \frac{1}{\pi} \int_0^{2\pi} \cos jx \cos kx \, dx = 0
\]

for all \(j \neq k\)

plus showing these trigonometric functions are orthogonal to a constant function. To verify that these trigonometric functions have unit length requires showing

\[
\frac{1}{\sqrt{\pi}} \sin kx \cdot \frac{1}{\sqrt{\pi}} \sin kx = \frac{1}{\pi} \int_0^{2\pi} \sin^2 kx \, dx = 1 \quad \text{for all} \ k
\]

\[
\frac{1}{\sqrt{\pi}} \cos kx \cdot \frac{1}{\sqrt{\pi}} \cos kx = \frac{1}{\pi} \int_0^{2\pi} \cos^2 kx \, dx = 1 \quad \text{for all} \ k
\]

When \(u_{2k-1}(x) = (1/\sqrt{\pi}) \sin kx\) and \(u_{2k}(x) = (1/\sqrt{\pi}) \cos kx, k = 1, 2, \ldots\) and \(u_0(x) = 1/\sqrt{2\pi}\) in (20), this representation of \(f(x)\) is called a Fourier series, and the coefficients \(f(x) \cdot u_i(x)\) in (20) are called Fourier coefficients. Using Fourier series, we see that any piecewise continuous function can be expressed as a linear combination of sine and cosine waves. One important physical interpretation of this fact is that any complex electrical signal can be expressed as a sum of simple sinusoidal signals.

**Example 9. Fourier Series Representation of a Jump Function**

Let us determine the Fourier series representation of the discontinuous function: \(f(x) = 1\) for \(0 < x \leq \pi\) and \(= 0\) for \(\pi < x \leq 2\pi\). The Fourier coefficients \(f(x) \cdot u_i(x)\) in (20) are
\[
f(x) \cdot u_{2k-1}(x) = f(x) \cdot \frac{1}{\sqrt{\pi}} \sin kx = \frac{1}{\sqrt{\pi}} \int_0^\pi \sin kx \, dx
\]
\[
= \frac{1}{k\sqrt{\pi}} \left[ -\cos kx \right]_0^\pi = \begin{cases} \frac{2}{k\sqrt{\pi}} & k \text{ odd} \\ 0 & k \text{ even} \end{cases}
\]
\[
f(x) \cdot u_{2k}(x) = f(x) \cdot \frac{1}{\sqrt{\pi}} \cos kx = \frac{1}{\sqrt{\pi}} \int_0^\pi \cos kx \, dx
\]
\[
= \frac{1}{k\sqrt{\pi}} [\sin kx]_0^\pi = 0
\]
Representing a function in terms of an orthonormal set of functions as in (20) has a virtually unlimited number of applications in the physical sciences and elsewhere. If one can solve a physical problem for the orthonormal basis functions, then one can typically obtain a solution for any function as a linear combination of the solutions for the basis functions. This is true for most differential equations associated with electrical circuits, vibrating bodies, and so on. Statisticians use Fourier series to analyze time-series patterns (see Example 3 of Section 1.5). The study of Fourier series is one of the major fields of mathematics.

We complete our discussion of vector spaces of functions by showing how badly conditioned the powers of $x$ are as a basis for representing functions. Remember that the powers of $x, x^i, i = 0, 1, \ldots$, are linearly independent. The problem is that they are far from orthogonal.

Let us consider how we might approximate an arbitrary function $f(x)$ as a linear combination of, say, the powers of $x$ up to $x^5$:

$$f(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4 + w_5x^5 \quad (27)$$

using the continuous version of least-squares theory. If $f(x)$ and the powers of $x$ were vectors, not functions, then (27) would have the familiar matrix form $f = Aw$ and the approximate solution $w$ would be given by $w = A^+f$, where $A^+ = (A^TA)^{-1}A^T$.

Let us generalize $f = Aw$ to functions by letting the columns of a matrix be functions. We define the functional "matrix" $A(x)$:

$$A(x) = [1, x, x^2, x^3, x^4, x^5]$$

Now (27) becomes

$$f(x) = A(x)w \quad (28)$$

To find the approximate solution to (28), we need to compute the functional version of the pseudoinverse $A(x)^+: A(x)^+ = (A(x)^TA(x))^{-1}A(x)^T$ and then find the vector $w$ of coefficients in (27):

$$w = A(x)^+f(x) = (A(x)^TA(x))^{-1}(A(x)^Tf(x) \quad (29)$$

The matrix $A(x)^T$ has $x^i$ as its $i$th "row", so the matrix product $A(x)^TA(x)$ involves computing the inner product of each "row" of $A(x)^T$ with each "column" of $A(x)$:

$$\text{entry } (i, j) \text{ in } A(x)^TA(x) \text{ is } x^i \cdot x^j (= \int x^i x^j \, dx)$$

Similarly, the matrix-"'vector"' product $A(x)^Tf(x)$ is the vector of inner prod-
The computations are simplest if we use the interval $[0, 1]$. Then entry $(i, j)$ of $A(x)^T A(x)$ is

$$x^i \cdot x^j = \int_0^1 x^{i+j} \, dx = \left[ \frac{x^{i+j+1}}{i+j+1} \right]_0^1 = \frac{1}{i+j+1}$$  \hspace{1cm} (30)

For example, entry $(1, 2)$ is $\int x^2 \, dx = \frac{1}{3} x^3$. Note that we consider the constant function 1 ($= x^0$) to be the zeroth row of $A(x)^T$.

Computing all the inner products for $A(x)^T A(x)$ yields

$$A(x)^T A(x) = \begin{bmatrix}
1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\
\frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\
\frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \\
\frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} \\
\frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} & \frac{1}{11}
\end{bmatrix}$$  \hspace{1cm} (31)

This matrix is very ill-conditioned since the columns are all similar to each other. When the fractions in (31) are expressed to six decimal places, such as $\frac{1}{3} = .333333$, the inverse given by the author’s microcomputer was (with entries rounded to integer values)

Fractions expressed to six decimal places

$$(A(x)^T A(x))^{-1} = \begin{bmatrix}
17 & -116 & -47 & 1,180 & -1,986 & 958 \\
-116 & 342 & 7,584 & -34,881 & 49,482 & -22,548 \\
-47 & 7,584 & -76,499 & 242,494 & -301,846 & 129,004 \\
1,180 & -34,881 & 242,494 & 644,439 & 723,636 & -289,134 \\
-1,986 & 49,482 & -301,846 & 723,636 & -747,725 & 278,975 \\
958 & -22,548 & 129,004 & -289,134 & 278,975 & -97,180
\end{bmatrix}$$  \hspace{1cm} (32)

The (absolute) sum of the fifth column in (32) is about 2,000,000. The first column in (31) sums to about 2.5. So the condition number of $A(x)^T A(x)$, in the sum norm, is about $2,000,000 \times 2.5 = 5,000,000$. Now that is an ill-conditioned matrix!

We rounded fractions to six significant digits, but our condition number tells us that without a seventh significant digit, our numbers in (32) could be off by 500% error [a relative error of .000001 in $A(x)^T A(x)$ could yield answers off by a factor of 5 in pseudoinverse calculations]. Thus the numbers in (32) are worthless.

Suppose that we enter the matrix in (31) again, now expressing fractions to seven decimal places. The new inverse computation yields
Fractions expressed to seven decimal places

\[(A(x)TA(x))^{-1} = \begin{bmatrix}
51 & -1,051 & 6,160 & -1,475 & 15,419 & -5,845 \\
-1,051 & 26,385 & -165,765 & 410,749 & -438,029 & 168,208 \\
6,160 & -165,765 & 1,079,198 & -2,731,939 & 2,955,103 & -1,146,281 \\
-1,475 & 410,749 & -2,731,939 & 7,017,359 & -7,671,190 & 2,999,546 \\
15,419 & -438,029 & 2,955,103 & -7,671,190 & 8,454,598 & -3,327,362 \\
-5,845 & 168,208 & -1,146,281 & 2,999,546 & -3,327,362 & 1,316,523 \\
\end{bmatrix}\]

We have a totally different matrix. Most of the entries in (33) are about 10 times larger than corresponding entries in (32). The sum of the fifth column in (33) is about 23,000,000. If we use (33), the condition number of \(A(x)TA(x)\) is around 56,000,000. Our entries in (33) were rounded to seven significant digits, but the condition number says eight significant digits were needed. Again our numbers are worthless. To compute the inverse accurately would require double-precision computation.

It is only fair to note that the ill-conditioned matrix (31) is famously bad. It is called a 6-by-6 Hilbert matrix [a Hilbert matrix has \(1/((i+j)+1)\) in entry \((i, j)\)].

Suppose that we used the numbers in (32) for \((A(x)^TA(x))^{-1}\) in computing the pseudoinverse. Let us proceed to calculate \(A(x)^+\) and then compute the coefficients in an approximation for a function by a fifth-degree polynomial. Let us choose \(f(x) = e^x\). Then \((A(x)^Te^x)\) is the vector of inner products \(x^i \cdot e^i = \int x^i e^x \, dx\), \(i = 0, 1, \ldots, 5\). Some calculus yields \(A(x)^Te^x = [2.718, 1, .718, .563, .465, .396]\) (expressed to three significant digits).

Now inserting our values for \((A(x)^TA(x))^{-1}\) and \(A^Te^x\) into (27), we obtain

\[
w = (A(x)^TA(x))^{-1}(A(x)^Te^x) = \begin{bmatrix}
17 & -116 & -47 & 1,180 & -1,986 & 958 \\
-116 & 342 & 7,584 & -34,881 & 49,482 & -22,548 \\
-47 & 7,584 & -76,499 & 242,494 & -301,846 & 129,004 \\
1,180 & -34,881 & 242,494 & 644,439 & 723,636 & -289,134 \\
-1,986 & 49,482 & -301,846 & 723,636 & -747,725 & 278,975 \\
958 & -22,548 & 129,004 & -289,134 & 278,975 & -97,180 \\
\end{bmatrix} \begin{bmatrix}
2.718 \\
1 \\
.718 \\
.563 \\
.465 \\
.396 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
17 \\
-87 \\
-219 \\
1,611 \\
2,449 \\
1,135 \\
\end{bmatrix}
\]
Thus our fifth-degree polynomial approximation of $e^x$ on the interval $[0, 1]$ is

$$e^x = 17 - 87x - 219x^2 + 1611x^3 + 2449x^4 + 1135x^5 \quad (35)$$

Setting $x = 1$ in (35), we have $e^1 = 17 - 86 - 219 + 1611 + 2449 + 1135 = 4907$, pretty bad. Since our computed values in $(A(x)^TA(x))^{-1}$ are meaningless, such a bad approximation of $e^x$ was to be expected.

Compare (35) with the Legendre polynomial approximation in Example 8.

**Section 5.4 Exercises**

**Summary of Exercises**

1. Compute the inverses of these matrices with orthogonal columns. Solve

$$Ax = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

where $A$ is the matrix in part (b).

(a) $\begin{bmatrix} .6 & .8 \\ -.8 & .6 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{bmatrix}$

2. Compute the inverses of these matrices with orthogonal columns.

(a) $\begin{bmatrix} -1 & 4 & -1 \\ 2 & 1 & -2 \\ 1 & 2 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & -3 & 6 \\ -6 & 2 & 3 \\ 3 & 6 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} .5 & - .5 & 1 \\ - .5 & .5 & 1 \\ 1 & .5 & 0 \end{bmatrix}$

Solve $Ax = 1$, where $A$ is the matrix in part (a).

3. Show that if $A$ is an $n$-by-$n$ upper triangular matrix with orthonormal columns, $A$ is the identity matrix $I$.

4. Compute the length $k$ of the projection of $b$ onto $a$ and give the projection vector $ka$.  


(a) \( \mathbf{a} = [0, 1, 0], \mathbf{b} = [3, 2, 4] \)
(b) \( \mathbf{a} = [1, -1, 2], \mathbf{b} = [2, 3, 1] \)
(c) \( \mathbf{a} = [\frac{1}{3}, \frac{2}{3}, \frac{1}{3}], \mathbf{b} = [4, 1, 3] \)
(d) \( \mathbf{a} = [2, -1, 3], \mathbf{b} = [-2, 5, 3] \)

5. Express the vector \([2, 1, 2]\) as a linear combination of the following orthogonal bases for three-dimensional space.
(a) \([1, -1, 2], [2, 2, 0], [-1, 1, 1]\)
(b) \([\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}], [\frac{2}{3}, \frac{1}{3}, \frac{1}{3}], [\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}]\)
(c) \([3, 1.5, 1], [1, -3, 1.5], [-1.5, 1, 3]\)

6. Compute the pseudoinverse of

\[
\mathbf{A} = \begin{bmatrix}
\frac{1}{3} & -\frac{2}{3} \\
-\frac{2}{3} & \frac{1}{3} \\
\frac{2}{3} & \frac{2}{3}
\end{bmatrix}
\]

Find the least-squares solution to \(\mathbf{A}\mathbf{x} = \mathbf{1}\).

7. Compute the pseudoinverse of

\[
\mathbf{A} = \begin{bmatrix}
3 & 4 \\
1 & -2 \\
2 & -5
\end{bmatrix}
\]

Find the least-squares solution to \(\mathbf{A}\mathbf{x} = \mathbf{1}\).

8. Consider the regression model \(\hat{z} = qx + ry + s\) for the following data, where the \(x\)-value is a scaled score (to have average value of 0) of high school grades, the \(y\)-value is a scaled score of SAT scores, and the \(z\)-value is a score of college grades.

<table>
<thead>
<tr>
<th>(x)</th>
<th>-4</th>
<th>-2</th>
<th>0</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y)</td>
<td>2</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>(z)</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

Determine \(q\), \(r\), and \(s\). Note that the \(x\), \(y\), and \(1\) vectors are mutually orthogonal.

9. Verify that Theorem 2 is true in two dimensions, namely, that a change from the standard \(\{\mathbf{e}_1, \mathbf{e}_2\}\) basis to some other orthonormal basis \(\{\mathbf{q}_1, \mathbf{q}_2\}\) corresponds to a rotation (around the origin) and possibly a reflection. Note that since \(\mathbf{q}_1, \mathbf{q}_2\) have unit length, they are completely determined by knowing the (counterclockwise) angles \(\theta_1, \theta_2\) they make with the positive \(\mathbf{e}_1\) axis; also since \(\mathbf{q}_1, \mathbf{q}_2\) are orthogonal, \(|\theta_1 - \theta_2| = 90^\circ|\).
10. (a) Show that an orthonormal change of basis preserves lengths (in euclidean norm).
   \( \text{Hint: } \) Verify that \((Qv) \cdot (Qv) = v \cdot v\) (where \(Q\) has orthonormal columns) by using the identity \((Ab) \cdot (Cd) = b^T(A^TC)d\).

(b) Show that an orthonormal change of basis preserves angles.
   \( \text{Hint: } \) Show that the cosine formula for the angle is unchanged by the method in part (a).

11. Compute the angle between the following pairs of nonorthogonal vectors. Which are close to orthogonal?
   (a) \([3, 2], [-3, 4]\)      (b) \([1, 2, 5], [2, 5, 3]\)
   (c) \([1, -3, 2], [-2, 4, -3]\)

12. Find the QR decomposition of the following matrices.
   (a) \[\begin{bmatrix} 3 & -1 \\ 4 & 1 \end{bmatrix}\]  (b) \[\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\]  (c) \[\begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 2 \end{bmatrix}\]  (d) \[\begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}\]

13. Use the Gram–Schmidt orthogonalization to find an orthonormal basis that generates the same vector space as the following bases:
   (a) \([1, 1], [2, -1]\)      (b) \([2, 1, 2], [4, 1, 1]\),
   (c) \([3, 1, 1], [1, 2, 1], [1, 1, 2]\)

14. (a) Compute the inverse of the matrix in Exercise 12, part (c) by first finding the QR decomposition of the matrix and then using (15) to get the inverse. (See Exercise 12 of Section 3.5 for instructions on computing \(R^{-1}\).) What is its condition number?
   (b) Check your answer by computing the inverse by the regular elimination by pivoting method.

15. (a) Find the pseudoinverse \(A^+\) of the matrix \(A\) in Exercise 12, part (b) by using the QR decomposition of \(A\) and computing \(A^+ = R^{-1}Q^T\).
   (b) Check your answer by finding the pseudoinverse from the formula \(A^+ = (A^TA)^{-1}A^T\). Note that this is a very poorly conditioned matrix; compute the condition number of \((A^TA)\).

16. Use (16) to find the pseudoinverse in solving the refinery problem in Example 3 of Section 5.3.

17. Use (16) to find the pseudoinverse in the following regression problems using the model \(\hat{y} = qx + r\).
   (a) \((x, y)\) points: \((0, 1), (2, 1), (4, 4)\)
   (b) \((x, y)\) points: \((3, 2), (4, 5), (5, 5), (6, 5)\)
   (c) \((x, y)\) points: \((-2, 1), (0, 1), (2, 4)\)

18. Use (16) to find the pseudoinverse in the least-squares polynomial-fitting problem in Example 5 of Section 5.3.
19. Verify (16): \( A^+ = R^{-1}Q^T \), by substituting \( QR \) for \( A \) and \( R^TQ^T \) for \( A^T \) in the pseudoinverse formula \( A^+ = (A^TA)^{-1}A^T \) and simplifying (remember that \( R \) is invertible; we assume that the columns of \( A \) are linearly independent).

20. Show that if the columns of the \( m \)-by-\( n \) matrix \( A \) are linearly independent, the \( m \)-by-\( m \) matrix \( R \) of the QR decomposition must be invertible.

**Hint:** Show main diagonal entries of \( R \) are nonzero and then see Exercise 12 of Section 3.5 for instructions on computing inverse of \( R \).

21. Show that any set \( H \) of \( k \) orthonormal \( n \)-vectors can be extended to an orthonormal basis for \( n \)-dimensional space.

**Hint:** Form an \( n \)-by-(\( k + n \)) matrix whose first \( k \) columns come from \( H \) and whose remaining \( n \) columns form the identity matrix; now apply the Gram–Schmidt orthogonalization to this matrix.

22. Over the interval \([0, 1]\), compute the following inner products: \( x \cdot x \), \( x \cdot x^3 \), \( x^3 \cdot x^3 \).

23. Verify that the fourth Legendre polynomial is \( x^3 - \frac{3}{8}x \).

24. Verify the values found for the weights \( w_1 \), \( w_2 \), \( w_3 \), and \( w_4 \) in Example 8.

**Note:** You must use integration by parts—or a table of integrals.

25. Approximate the following functions \( f(x) \) as a linear combination of the first four Legendre polynomials over the interval \([-1, 1]\): \( L_0(x) = 1 \), \( L_1(x) = x \), \( L_2(x) = x^2 - \frac{1}{3} \), \( L_3(x) = x^3 - 3x/5 \).

(a) \( f(x) = x^4 \)  
(b) \( f(x) = |x| \)  
(c) \( f(x) = -1: x < 0, = 1: x \geq 0 \)

26. Approximate \( x^3 + 2x - 1 \) as a linear combination of the first four Legendre polynomials over the interval \([-1, 1]\): \( L_0(x) = 1, L_1(x) = x, L_2(x) = x^2 - \frac{1}{3}, L_3(x) = x^3 - 3x/5 \). Your "approximation" should equal \( x^3 + 2x - 1 \), since this polynomial is a linear combination of the functions \( 1, x, x^2, \) and \( x^3 \), from which the Legendre polynomials were derived by orthogonalization.

27. (a) Find the Legendre polynomial of degree 4. 
(b) Find the Legendre polynomial of degree 5.

28. (a) Using the interval \([0, 1]\), instead of \([-1, 1]\), find three orthogonal polynomials of the form \( K_0(x) = a, K_1(x) = bx + c, \) and \( K_2(x) = dx^2 + ex + f \).

(b) Find a least-squares approximation of \( x^4 \) on the interval \([0, 1]\) using your three polynomials in part (a).
(c) Find a least-squares approximation of $x^{1/2}$ on the interval $[0, 1]$ using your three polynomials in part (a).

(d) Find a least-squares approximation of $x^2 - 2x + 1$ on the interval $[0, 1]$ using your three polynomials in part (a). Hopefully, your approximation will equal $x^2 - 2x + 1$, since this polynomial is a linear combination of $1, x,$ and $x^2,$ the functions used to build your set of orthogonal polynomials.

29. (a) Find a fourth polynomial $K_i(x)$ of order 3 orthogonal on $[0, 1]$ to the three polynomials in Exercise 28, part (a).

(b) Find a least-squares approximation to $x^4$ on the interval $[0, 1]$ using your four orthogonal polynomials.

30. Compute the inverse and find the condition number (in sum norm) of the following Hilbert-like matrices.

(a) \[
\begin{bmatrix}
\frac{1}{8} & \frac{1}{9} \\
\frac{1}{9} & \frac{1}{10}
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
\frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\
\frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\
\frac{1}{6} & \frac{1}{7} & \frac{1}{8}
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
\frac{1}{8} & \frac{1}{9} & \frac{1}{10} \\
\frac{1}{9} & \frac{1}{10} & \frac{1}{11} \\
\frac{1}{10} & \frac{1}{11} & \frac{1}{12}
\end{bmatrix}
\]

Section 5.5 Eigenvector Bases and the Eigenvalue Decomposition

In this section we use eigenvectors to gain insight into the structure of a matrix. We review how an eigenvector basis simplifies the computation of powers of $A$. Then we present a way to decompose a matrix into simple matrices formed by the eigenvectors. This decomposition yields a way to compute all eigenvectors and eigenvalues of a symmetric matrix.

Recall that a vector $u$ is an eigenvector of the $n$-by-$n$ matrix $A$ if for some $\lambda$ (an eigenvalue), $Au = \lambda u$. In words, multiplying $u$ by a matrix $A$ has the same effect as multiplying $u$ by the scalar $\lambda$. It follows that $A^k u = \lambda^k u$. A stable distribution $p$ of a Markov chain is an eigenvector of the transition matrix $A$ (associated with eigenvalue 1: $Ap = p$). The dominant eigenvector (associated with the largest eigenvalue) gives the long-term distribution of a growth model (see Sections 2.5 and 4.5).

As noted in Section 2.5, if a vector $x$ can be expressed as a linear combination of the eigenvectors $u_i$ of $A$;

$$x = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n$$  \hspace{1cm} (1)

then

$$Ax = a_1 Au_1 + a_2 Au_2 + \cdots + a_n Au_n$$

$$= a_1 \lambda_1 u_1 + a_2 \lambda_2 u_2 + \cdots + a_n \lambda_n u_n$$  \hspace{1cm} (2)
In words, matrix multiplication becomes scalar multiplication of eigenvector coordinates.

Further,

\[ A^k x = a_1 A^k u_1 + a_2 A^k u_2 + \cdots + a_n A^k u_n \]

(3)

When we express \( x \) in terms of a basis of \( A \)'s eigenvectors, \( A^k x \) can quickly be computed by (3).

Assume that we index the eigenvalues so that \( \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \), \( \lambda_1 \) is going to be much larger than the other \( \lambda \)'s. Thus the first term on the right in (3) dominates the other terms, and we have

\[ A^k x = a_1 \lambda_1^k u_1 \]

(4)

We review the example from Section 2.5 that illustrated these results.

\textbf{Example 1. Computing Powers of a Matrix with Eigenvectors}

The computer \((C)\) and dog \((D)\) growth model from Section 2.5 is

\[ x' = Ax, \quad \text{where} \quad A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \quad x = [C, D] \quad x' = [C', D'] \]

The two eigenvalues and associated eigenvectors of \( A \) are \( \lambda_1 = 4 \) with \( u_1 = [1, 1] \) and \( \lambda_2 = 1 \) with \( u_2 = [1, -2] \). Note that since \( u_1 \) and \( u_2 \) are linearly independent, they form a basis for 2-space.

Suppose that we want to determine the effects of this growth model over 20 periods with the starting vector \( x = [1, 7] \). We want to express \( x \) as a linear combination of \( u_1 \) and \( u_2 \): \( x = w_1 u_1 + w_2 u_2 \). Determining the set of weights \( w = [w_1, w_2] \) requires solving

\[ x = w_1 u_1 + w_2 u_2 \quad \text{or} \quad x = Uw, \]

where \( U = [u_1, u_2] = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \)

(5)

The solution to (5) is \( w = U^{-1} x \), which yields \( w_1 = 3 \), \( w_2 = -2 \), so that \( x = 3u_1 - 2u_2 \). (Here \( w_1, w_2 \) are simply the coordinates of \( x \) in the eigenvector basis for 2-space.) Then

\[ Ax = A(3u_1 - 2u_2) = 3Au_1 - 2Au_2 \]

\[ = 3(4u_1) - 2(1u_2) \]

\[ = 12u_1 - 2u_2 \]

(6)

For 20 periods, we have
In Example 7 of Section 3.3 we observed that these steps were represented by a matrix equation

\[ Ax = U D_\lambda U^{-1} x, \quad \text{where } D_\lambda = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \]  

(8a)

and

\[ A = U D_\lambda U^{-1} \]  

(8b)

where \( U \), as in (5), has the eigenvectors as its columns.

In words, we explain (8a) and (8b) as follows. The vector \( w \) in (5) can be viewed as the vector \( x \) expressed in terms of eigenvector coordinates, and as noted above, \( w = U^{-1} x \). Looking at \( U D_\lambda U^{-1} \) times \( x \) from right to left, the product \( U^{-1} x \) converts \( x \) to the eigenvector-coordinate vector \( w \). Next the matrix \( D_\lambda \) multiplies each eigenvector coordinate by the appropriate eigenvalue \( (D_x w = [4w_1, w_2]) \) to get the eigenvector-based coordinates after the matrix multiplication. Finally, multiplying \( U \) converts back to the original coordinate system.

The matrix-vector product \( x' = Ax \) is transformed, in eigenvector coordinates into \( w' = D_\lambda w \). Further, \( A^k = U D_\lambda^k U^{-1} \), so

\[ x^{(20)} = A^{20} x \quad \text{becomes} \quad w^{(20)} = D_\lambda^{20} w \]  

(9)

For our particular \( A \),

\[ w^{(20)} = D_\lambda^{20} w = \begin{bmatrix} 4^{20} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 4^{20} w_1 \\ w_2 \end{bmatrix} \]  

(10)

Summarizing (part of this is simply Theorem 5 of Section 3.3), we obtain

**Theorem 1.** Let \( A \) be an \( n \)-by-\( n \) matrix and let \( U \) be an \( n \)-by-\( n \) matrix whose columns \( u_i \) are \( n \) linearly independent eigenvectors of \( A \). If \( c = Ab \), then in \( u_i \)-coordinates \( c^*, b^* \), we have

\[ c^* = D_\lambda b^* \quad \text{or} \quad c^*_i = \lambda_i b^*_i \]  

(11)

where \( b^* = U^{-1} b \) and \( c^* = U^{-1} c \) are the \( u_i \)-coordinate vectors for \( b \) and \( c \) and \( D_\lambda \) is the diagonal matrix whose diagonal entries are the eigenvalues of \( A \), \( \lambda_1, \lambda_2, \ldots, \lambda_n \). Similarly, if \( c = A^k b \), then \( c^*_i = \lambda_i^k b^*_i \).
Section 5.5 Eigenvector Bases and the Eigenvalue Decomposition

Theorem 2. For $A$, $D_{\lambda}$, and $U$ as in Theorem 1, $A$ can be written

$$A = UD_{\lambda}U^{-1}$$

(12a)

and

$$A^k = UD_{\lambda}^kU^{-1}$$

(12b)

where $D_{\lambda}^k$ has the eigenvalues raised to the $k$th power. Further,

$$D_{\lambda} = U^{-1}AU$$

(13)

Example 2. Conversion of $A$ to $D_{\lambda}$

Let us use the matrix $A$ in Example 1, $A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$. We had found $U$ to be: $U = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$. By the determinant formula for 2-by-2 inverses, $U^{-1} = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$. Then, by (13),

$$D_{\lambda} = U^{-1}AU = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{8}{3} & \frac{4}{3} \\ \frac{4}{3} & -\frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

(14)

Let us next use (12b) to compute $A^k$. This formula is most useful for large $k$, but for illustrative purpose we use $k = 2$. First we compute $A^2$ directly:

$$A^2 = AA = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 5 \\ 10 & 6 \end{bmatrix}$$

Next we compute $A^2$ as

$$A^2 = UD_{\lambda}^2U^{-1} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

$$= \begin{bmatrix} 16 & 1 \\ 16 & -2 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

$$= \begin{bmatrix} 11 & 5 \\ 10 & 6 \end{bmatrix}$$
The formula $A = UD\lambda U^{-1}$ leads to an important way to decompose a matrix into simple matrices. Just as Gaussian elimination yields the $LU$ decomposition and Gram–Schmidt orthogonalization yields the $QR$ decomposition, so our eigenvector coordinates formula also yields a matrix decomposition.

Recall that if $c = [1, 2]$ and $d = [3, 4, -1]$, the simple matrix $c \ast d$ equals

$$c \ast d = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \ast [3, 4, -1] = \begin{bmatrix} 3 & 4 & -1 \\ 6 & 8 & -2 \end{bmatrix}$$

More generally, entry $(i, j)$ in $c \ast d$ equals $c_i d_j$.

Theorem 7 of Section 5.2 says that the matrix product $CD$ can be decomposed into a sum of simple matrices formed by columns $c_i$ of $C$ and the rows $d_j$ of $D$:

$$CD = c_1^* d_1^T + c_2^* d_2^T + \cdots + c_n^* d_n^T$$

Letting $C = U$ and $D = D\lambda U^{-1}$ (the $i$th row of $D$ is the $i$th row of $U^{-1}$ multiplied by $\lambda_i$), we obtain the following decomposition.

**Theorem 3. Eigenvalue Decomposition.** Let $A$ be an $n$-by-$n$ matrix with $n$ linearly independent eigenvectors $u_i$ associated with eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Let $u_i$ denote the $i$th row of $U^{-1}$. Then $A$ is the weighted sum of simple matrices:

$$A = UD\lambda U^{-1} = \lambda_1 u_1 \ast u_1' + \lambda_2 u_2 \ast u_2' + \cdots + \lambda_n u_n \ast u_n'$$

In (16) the $\lambda_i$'s are factored out in front of the simple matrices.

For typical large matrices, the eigenvalues tend to decline in size quickly. For example, if $n = 20$, perhaps $\lambda_1 = 5$, $\lambda_2 = 2$, $\lambda_3 = .6$, $\lambda_4 = .02$; so the sum of the first three simple matrices in (16) would yield a very good approximation of the matrix.

**Example 3. Eigenvalue Decomposition of 2-by-2 Matrix**

We illustrate Theorem 3 with the 2-by-2 matrix of Examples 1 and 2. From those examples we have $\lambda_1 = 4$, $\lambda_2 = 1$ and

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}, \quad U^{-1} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

Then (16) says
Sec. 5.5 Eigenvector Bases and the Eigenvalue Decomposition

\[
A = 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix} * \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 1 \end{bmatrix} * \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}
\]

\[
= 4 \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{bmatrix} + \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 2 \end{bmatrix}
\]

We next state a very useful fact about symmetric matrices.

**Theorem 4**

(i) Any symmetric \( n \)-by-\( n \) matrix \( A \) has a set of \( n \) orthogonal eigenvectors \( u_i \).

(ii) Two eigenvectors associated with distinct eigenvalues are always orthogonal.

**Corollary.** When \( A \) is symmetric, the matrix \( U \) in Theorem 1 has orthogonal columns, and \( U^{-1} \) is obtained from \( U^T \) by dividing each row by \( |u_i|^2 = u_i \cdot u_i \). When additionally the columns have length 1 (orthonormal), the eigenvalue decomposition in Theorem 3 becomes

\[
A = \lambda_1 u_1 * u_1 + \lambda_2 u_2 * u_2 + \cdots + \lambda_n u_n * u_n
\]

(17)

The corollary’s claim about how to obtain \( U^{-1} \) comes from Theorem 1 of Section 5.4.

The eigenvalue decomposition (17) sheds new light on what happens when we multiply a symmetric matrix \( A \) times some vector \( x \). If in \( u_i \)-coordinates (\( u_i \) are orthonormal), \( x \) is

\[
x = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n
\]

(18)

and we compute \( Ax \) using (17), we have

\[
Ax = (\lambda_1 u_1 * u_1 + \lambda_2 u_2 * u_2 + \cdots + \lambda_n u_n * u_n)x
\]

(19)

and (19) is equivalent to

\[
Ax = \lambda_1 (u_1 \cdot x)u_1 + \lambda_2 (u_2 \cdot x)u_2 + \cdots + \lambda_n (u_n \cdot x)u_n
\]

(20)

since

\[
(u_1 \cdot u_1)x = u_1(u_1 \cdot x) \quad \text{or} \quad (u_1 \cdot x)u_1
\]

(21)

and similarly for the other \( u_i \).
We verify (21) as follows. The first $u_1$ in $u_1 \cdot x$ is to be treated as an $n$-by-$1$ matrix, the second $u_1$ as a $1$-by-$n$ matrix, and $x$ as a column vector (an $n$-by-$1$ matrix). Now use the associative law of matrix multiplication in (21).

So multiplying $A$ by $x$, when $A$ is represented as a sum of simple matrices, has the effect of first projecting $x$ onto the eigenvectors $u_i$ [$\langle u_i, x \rangle u_i$ is this projection] and then multiplying each such projection by the eigenvalue $\lambda_i$.

In Section 3.4 we used equation (4)—that $A^kx$ is approximately a multiple of $u_1$ to determine $u_1$ and $\lambda_1$. But we had no way to compute other eigenvalues or eigenvectors. The eigenvalue decomposition (17) gives us a way.

Suppose that $A$ is a symmetric matrix and we have determined the dominant (largest) eigenvalue $\lambda_1$ and an associated eigenvector $u_1$ (of length 1) by an iterative method. Consider then the matrix

$$A_2 = A - \lambda_1 u_1 \cdot x_1$$

(22)

The matrix $A_2$ is $A$ minus the first simple matrix in the eigenvalue decomposition in (17). It follows that

$$A_2 = \lambda_2 u_2 \cdot x_2 + \lambda_3 u_3 \cdot x_3 + \cdots + \lambda_n u_n \cdot x_n$$

(23)

Since the eigenvalue decomposition of a square matrix is unique, it follows that the (nonzero) eigenvalues of $A_2$ are $\lambda_2$, $\lambda_3$, $\ldots$, $\lambda_n$ with associated eigenvectors $u_2$, $u_3$, $\ldots$, $u_n$. In particular, $\lambda_2$ is now the dominant eigenvalue of $A_2$ and applying an iterative method to $A_2$ will yield $\lambda_2$ and $u_2$.

If we subtract the second simple matrix in (17) from $A_2$, we will get a matrix $A_3$ whose dominant eigenvalue is $\lambda_3$, and so on. This method of getting the eigenvalues and eigenvectors of $A$ is called deflation. We note that if we have a small error in $\lambda_1$ or $u_1$, the resulting $A_2$ still has the same dominant eigenvalue and eigenvector (the consequences of such errors are explored in the Exercises).

**Deflation Method to Compute Eigenvalues and Eigenvectors of a Symmetric Matrix $A$**

**Step 0.** Set $A_1 = A$; set $i = 1$.

**Step 1.** Use the iterative method to determine $A_i$‘s dominant eigenvalue $\lambda_i$ and an associated eigenvector $u_i$ (of length 1).

**Step 2.** Set $A_{i+1} = A_i - \lambda_i u_i \cdot x_i$. Increase $i$ by 1. If $i \leq n$, go to step 1.

A faster method for determining all eigenvalues and eigenvectors of any $n$-by-$n$ matrix, symmetric or not, is presented in the appendix to this section.
Example 4. Finding Eigenvalues and Eigenvectors of a 3-by-3 Symmetric Matrix

Let us use the deflation method to find all eigenvalues and eigenvectors of the symmetric matrix

\[
A = \begin{bmatrix}
3 & 2 & 1 \\
2 & 3 & 1 \\
1 & 1 & 0
\end{bmatrix}
\]

Setting \( \mathbf{x} = [1, 0, 0] \) and computing \( A^2 \mathbf{x} \), we get a multiple of the unit vector

\[
\mathbf{u}_1 = [.684, .684, .254] \quad \text{with } \lambda_1 = 5.37
\]

Next we compute the deflated matrix \( A_2 \):

\[
A_2 = A - \lambda_1 \mathbf{u}_1 \times \mathbf{u}_1
\]

\[
= \begin{bmatrix}
3 & 2 & 1 \\
2 & 3 & 1 \\
1 & 1 & 0
\end{bmatrix} - \begin{bmatrix}
2.51 & 2.51 & .93 \\
2.51 & 2.51 & .93 \\
.93 & .93 & .35
\end{bmatrix}
\]

\[
= \begin{bmatrix}
.49 & -.51 & .07 \\
-.51 & .49 & .07 \\
.07 & .07 & -.35
\end{bmatrix}
\]

Again with \( \mathbf{x} = [1, 0, 0] \), we compute \( A_1^2 \mathbf{x} \) and get \([.5, -.5, 0]\). So

\[
\mathbf{u}_2 = [.707, -.707, 0] \quad \text{with } \lambda_2 = 1
\]

(The reader should verify that \([.5, -.5, 0]\) was an eigenvector of the original matrix \( A \).) Deflating again, we have

\[
A_3 = A_2 - \lambda_2 \mathbf{u}_2 \times \mathbf{u}_2
\]

\[
= \begin{bmatrix}
.49 & -.51 & .07 \\
-.51 & .49 & .07 \\
.07 & .07 & -.35
\end{bmatrix} - \begin{bmatrix}
.5 & -.5 & 0 \\
-.5 & .5 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

\[
= \begin{bmatrix}
-.01 & -.01 & .07 \\
-.01 & -.01 & .07 \\
.07 & .07 & -.35
\end{bmatrix}
\]

Next we find that \( \mathbf{u}_3 = [.181, .181, .967] \) and \( \lambda_3 = -.37 \). Computing the last simple matrix, we obtain
This simple matrix equals $A_3$, as required. Thus we have confirmed that $A$ is the sum of three simple matrices.

Substantial savings in computer storage can be realized by representing a large matrix as a sum of simple matrices. If a symmetric 20-by-20 matrix is well approximated as the sum of two simple matrices, then only 10% as much storage is needed: 2 columns instead of all 20 columns.

Example 5. Approximating a Digital Picture with Simple Matrices

Approximate the 8-by-8 digital "picture" $A$ whose entries represent varying levels of darkness between 0 and 1.

$$A = \begin{bmatrix}
0 & 0 & 0 & .2 & .2 & 0 & 0 & 0 \\
0 & .1 & .2 & .3 & .3 & .2 & .1 & 0 \\
0 & .2 & .5 & .6 & .6 & .5 & .2 & 0 \\
.. & .. & .. & .. & .. & .. & .. & .. \\
0 & .2 & .5 & .6 & .6 & .5 & .2 & 0 \\
0 & .1 & .2 & .3 & .3 & .2 & .1 & 0 \\
0 & 0 & 0 & .2 & .2 & 0 & 0 & 0 
\end{bmatrix} \quad (25)$$

We first approximate $A$ with the simple matrix $c * c$, where $c = [.1, .3, .7, .9, .9, .7, .3, .1]$ has $c_i$ equal to the (approximate) square root of the diagonal entry $(i, i)$ of $A$.

$$c * c = \begin{bmatrix}
.01 & .03 & .07 & .09 & .09 & .07 & .03 & .01 \\
.03 & .09 & .14 & .27 & .27 & .14 & .09 & .03 \\
.07 & .14 & .49 & .63 & .63 & .49 & .14 & .07 \\
.. & .. & .. & .. & .. & .. & .. & .. \\
.09 & .27 & .63 & .81 & .81 & .63 & .27 & .09 \\
.09 & .27 & .63 & .81 & .81 & .63 & .27 & .09 \\
.07 & .14 & .49 & .63 & .63 & .49 & .14 & .07 \\
.03 & .09 & .14 & .27 & .27 & .14 & .09 & .03 \\
.01 & .03 & .07 & .09 & .09 & .07 & .03 & .01 
\end{bmatrix} \quad (26)$$

Now we shall use the eigenvalue decomposition of $A$ into simple matrices to approximate $A$ more accurately. We use the first three terms involving the three largest eigenvalues,
Actually, for this $A$, the first simple matrix $\lambda_1 u_1 * u_1$ alone will be a good approximation.

For large matrices, the iterative method of computing $A^k x$ for a large $k$ to approximate $u_i$ is unnecessarily time consuming. One of the faster methods mentioned in the appendix to this section should be used. For $u_1$, we shall give the results using simple iteration because it converges quickly. Computing $A^{10} x$ with $x = [1, 1, 1, 1, 1, 1, 1, 1]$ yields a multiple of

\[
\begin{pmatrix}
0.078 & 0.189 & 0.408 & 0.540 & 0.540 & 0.408 & 0.189 & 0.078 \\
-0.017 & -0.041 & -0.088 & 0.083 & 0.083 & -0.088 & -0.041 & -0.017 \\
0.041 & 0.099 & 0.214 & 0.283 & 0.283 & 0.214 & 0.099 & 0.041 \\
0.088 & 0.214 & 0.461 & 0.610 & 0.610 & 0.461 & 0.214 & 0.088 \\
0.117 & 0.283 & 0.610 & 0.808 & 0.808 & 0.610 & 0.283 & 0.117 \\
0.088 & 0.214 & 0.461 & 0.610 & 0.610 & 0.461 & 0.214 & 0.088 \\
0.041 & 0.099 & 0.214 & 0.283 & 0.283 & 0.214 & 0.099 & 0.041 \\
0.017 & 0.041 & 0.088 & 0.117 & 0.117 & 0.088 & 0.041 & 0.017 \\
\end{pmatrix}
\]

This first simple matrix is very close to $A$. Except for entries $(1, 3)$ and $(1, 4)$ (and symmetrically equivalent entries), every entry in (28) is within about .04 of the corresponding entry in $A$. Since the numbers in $A$ were probably rounded off to one decimal digit, one could argue that (28) is as good an approximation to $A$ as we should seek.

Next we compute the deflated matrix $A_2$:

\[
A_2 = A - \lambda_1 u_1 * u_1
\]

\[
= \begin{pmatrix}
-0.017 & -0.041 & -0.088 & 0.083 & 0.083 & -0.088 & -0.041 & -0.017 \\
-0.041 & 0.001 & -0.014 & 0.017 & 0.017 & -0.014 & 0.001 & -0.041 \\
-0.088 & -0.014 & 0.039 & -0.010 & -0.010 & 0.039 & -0.014 & -0.088 \\
0.083 & 0.017 & -0.010 & -0.008 & -0.008 & -0.010 & 0.017 & 0.083 \\
-0.088 & -0.014 & 0.039 & -0.010 & -0.010 & 0.039 & -0.014 & -0.088 \\
-0.041 & 0.001 & -0.014 & 0.017 & 0.017 & -0.014 & 0.001 & -0.041 \\
-0.017 & -0.041 & -0.088 & 0.083 & 0.083 & -0.088 & -0.041 & -0.017 \\
\end{pmatrix}
\]
Using the iterative method on $A_2$ takes a long time to converge but finally gives us a multiple of

$$u_2 = [.514, .224, .258, -.345, -.345, .258, .224, .514]$$

with $\lambda_2 = -.27$

Instead of computing the second simple matrix, we give the sum of the first two simple matrices,

$$\lambda_1 u_1 * u_1 + \lambda_2 u_2 * u_2$$

$$= \begin{bmatrix}
-0.054 & 0.10 & 0.052 & 0.165 & 0.165 & 0.052 & 0.10 & -0.054 \\
0.10 & 0.084 & 0.198 & 0.304 & 0.304 & 0.198 & 0.084 & 0.10 \\
0.052 & 0.198 & 0.443 & 0.634 & 0.634 & 0.443 & 0.198 & 0.052 \\
0.165 & 0.304 & 0.634 & 0.776 & 0.776 & 0.634 & 0.304 & 0.165 \\
0.165 & 0.304 & 0.634 & 0.776 & 0.776 & 0.634 & 0.304 & 0.165 \\
0.052 & 0.198 & 0.443 & 0.634 & 0.634 & 0.443 & 0.198 & 0.052 \\
0.10 & 0.084 & 0.198 & 0.304 & 0.304 & 0.198 & 0.084 & 0.10 \\
-0.054 & 0.10 & 0.052 & 0.165 & 0.165 & 0.052 & 0.10 & -0.054 \\
\end{bmatrix}$$

(30)

Next we form $A_3$ and find that

$$u_3 = [.451, -.054, -.454, .295, .295, -.454, -.054, .451]$$

with $\lambda_3 = .263$

Note how close in absolute value $\lambda_2$ and $\lambda_3$ are; this is why the iterative method converged so slowly for $A_2$.

Now we can give the desired approximation of $A$ by the first three simple matrices associated with the eigenvalue decomposition of $A$.

$$A = \lambda_1 u_1 * u_1 + \lambda_2 u_2 * u_2 + \lambda_3 u_3 * u_3$$

$$= \begin{bmatrix}
-0.001 & 0.004 & -0.002 & 0.200 & 0.200 & -0.002 & 0.004 & -0.001 \\
0.004 & 0.089 & 0.204 & 0.300 & 0.300 & 0.204 & 0.089 & 0.004 \\
-0.002 & 0.204 & 0.497 & 0.599 & 0.599 & 0.497 & 0.204 & -0.002 \\
0.200 & 0.300 & 0.599 & 0.799 & 0.799 & 0.599 & 0.300 & 0.200 \\
0.200 & 0.300 & 0.599 & 0.799 & 0.799 & 0.599 & 0.300 & 0.200 \\
-0.002 & 0.204 & 0.497 & 0.599 & 0.599 & 0.497 & 0.204 & -0.002 \\
0.004 & 0.089 & 0.204 & 0.300 & 0.300 & 0.204 & 0.089 & 0.004 \\
-0.001 & 0.004 & -0.002 & 0.200 & 0.200 & -0.002 & 0.004 & -0.001 \\
\end{bmatrix}$$

(31)

The average deviation of an entry of (31) from $A$ is .002, and only one entry has an error exceeding .004.
We next give a statistical application of the eigenvalue decomposition of a symmetric matrix.

**Example 6. Principal Components in Statistical Analysis**

Suppose that an anthropologist has collected data on 25 physical characteristics, variables \( x_1, x_2, \ldots, x_{25} \), for 100 prehuman fossil remains. The researcher computes a measure of the variability \( V_i \) of each variable \( x_i \) called the **variance** of \( x_i \). A large variance means that the \( x_i \)-variable varies substantially from fossil to fossil. The anthropologist also computes a measure of the joint variability \( \text{Cov}_{ij} \) of each pair of variables \( x_i, x_j \) called the **covariance**. The covariance is proportional to the correlation coefficient (which was discussed in Section 5.3). A positive \( \text{Cov}_{ij} \) means variables \( x_i \) and \( x_j \) have similar values; a negative \( \text{Cov}_{ij} \) means the variables are opposites [if the \( k \)th fossil has a large \( x_i \)-value, the \( k \)th fossil probably has a small or negative \( x_j \)-value; and \( \text{Cov}_{ij} \) near 0 means values of \( x_i \) and \( x_j \) are unrelated (uncorrelated)].

The anthropologist would like to find good linear combinations of the characteristics that "explain" the variability of the data. For example, one might define the **length index** \( L \) to be

\[
L = 0.2x_1 + 0.4x_3 - 0.3x_{11} + 0.4x_{17} + 0.5x_{18}
\]  

where \( x_1 \) might be length of forearm, \( x_3 \) length of thigh, and so on. The idea is that although we may find a certain amount of variability in individual variables from fossil to fossil, such as varying length of forearm, the "right" measure that gives the best way to distinguish one fossil from another is some composite index, such as the length index.

Among all possible indices formed by a linear combination of variables, the index \( L_1 \) that shows the greatest variability (i.e., largest variance) is called the **first principal component**. Among those other indices that are **uncorrelated to** \( L_1 \) (covariance is 0), the index \( L_2 \) with the largest variance is called the **second principal component**, and so on. We want index \( L_2 \) uncorrelated so that it gives us new (additional) information about variability that was not contained in \( L_1 \).

In summary, the first principal component gives an index that explains the maximum variability of the data from one fossil to another. The first four principal components will typically account for over 90% of the variability in a set of 25 variables. Clearly, there are great advantages in describing each fossil with three or four numbers rather than 25 numbers. The same is true for studies in psychology, finance, quality control, and any other field where people collect large amounts of data.

So how do we find these principal components? That is, how do we determine the weights (coefficients) of the \( x_i \), such as the weights in (32)? The answer is that we form a covariance matrix \( C \) of all the covariances of the fossil data, where entry \( (i, j) \) is \( \text{Cov}_{ij} \) (and \( \text{Cov}_{ii} = \))
Then one can show that the vector of weights used in the first principal component is the unit-length dominant eigenvector \( \mathbf{u}_1 \) of \( \mathbf{C} \). The first simple matrix \( \lambda_1 (\mathbf{u}_1 \ast \mathbf{u}_1) \) in the eigenvalue decomposition of \( \mathbf{C} \) shows how much of the variability in \( \mathbf{C} \) is explained by the first principal component.

As an example, consider the following 4-by-4 covariance matrix (representing 4 of the 25 characteristics mentioned above)

\[
\mathbf{C} = \begin{bmatrix}
0.86 & 1.19 & 2.02 & 1.45 \\
1.19 & 1.68 & 2.86 & 2.06 \\
2.02 & 2.86 & 5.05 & 3.50 \\
1.45 & 2.06 & 3.50 & 2.53 \\
\end{bmatrix}
\]  

(33)

All the covariances here happen to be positive (this is often the case), although they can be negative. We can determine the eigenvalues and associated eigenvectors by deflation, as in the previous examples, or by using some computer package. We find that

\[
\lambda_1 = 10, \quad \lambda_2 \approx 0.098, \quad \lambda_3 \approx 0.022, \quad \lambda_4 = 0.003
\]  

(34)

The dominant (unit-length) eigenvector is

\[
\mathbf{u}_1 = [0.289, 0.408, 0.707, 0.5]
\]

The simple matrix \( \lambda_1 \mathbf{u}_1 \ast \mathbf{u}_1 \) should approximate \( \mathbf{C} \) well, since \( \lambda_1 \) is much larger than the other eigenvalues.

\[
\lambda_1 \mathbf{u}_1 \ast \mathbf{u}_1 = \begin{bmatrix}
0.84 & 1.18 & 2.04 & 1.44 \\
1.18 & 1.67 & 2.88 & 2.04 \\
2.04 & 2.88 & 5.00 & 3.54 \\
1.44 & 2.04 & 3.54 & 2.50 \\
\end{bmatrix}
\]  

(35)

Upon comparing (35) with \( \mathbf{C} \), it is clear that \( I_1 \) accounts for almost all the variability in the covariance matrix \( \mathbf{C} \).

The first principal component index \( I_1 \) is the linear combination of variables \( x_1, x_2, x_3, x_4 \) with weights given by \( \mathbf{u}_1 \):

\[
I_1 = 0.289x_1 + 0.408x_2 + 0.707x_3 + 0.5x_4
\]  

(36)

The eigenvalue decomposition of a matrix \( \mathbf{A} \) and the deflation method for finding successive eigenvalues and eigenvectors depended on special properties of \( \mathbf{A} \). There had to be a set of \( n \) linearly independent eigenvectors for the eigenvalue decomposition and \( \mathbf{A} \) had to be symmetric for the deflation method to work. Symmetry is easy to recognize. What about linearly independent eigenvectors? The following theorem answers this question. It is a companion to Theorem 4 (which stated that an \( n \)-by-\( n \) symmetric matrix has \( n \) orthogonal eigenvectors).
Theorem 5

(i) Eigenvectors associated with different eigenvalues are linearly independent.

(ii) If the \( n \)-by-\( n \) matrix \( A \) has \( n \) distinct eigenvalues, any set of \( n \) eigenvectors, each associated with a different eigenvalue, will form a basis for \( n \)-space, and the results in Theorems 1, 2, and 3 apply.

Proof. For explicitness, assume that \( n = 3 \) and let \( \mathbf{u}_1 \), \( \mathbf{u}_2 \), and \( \mathbf{u}_3 \) be three eigenvectors of \( A \) associated with different eigenvalues \( \lambda_1 \), \( \lambda_2 \), and \( \lambda_3 \), respectively. It is easy to show that \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) are linearly independent (Exercise 14). Suppose that \( \mathbf{u}_1 \), \( \mathbf{u}_2 \), and \( \mathbf{u}_3 \) are not linearly independent, so that \( \mathbf{u}_3 \) can be expressed as a unique linear combination of \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \): \( \mathbf{u}_3 = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 \). Now we compute \( A \mathbf{u}_3 \) in two ways.

\[
A \mathbf{u}_3 = \lambda_3 \mathbf{u}_3 = \lambda_3 (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2) \tag{37}
\]

and

\[
A \mathbf{u}_3 = A (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2) = \lambda_1 c_1 \mathbf{u}_1 + \lambda_2 c_2 \mathbf{u}_2 \tag{38}
\]

The representation of \( A \mathbf{u}_3 \) as a linear combination of \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) is unique. That is, the weights of \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) on the right sides of (37) and (38) must be equal: \( \lambda_3 c_1 = \lambda_1 c_1 \) and \( \lambda_3 c_2 = \lambda_2 c_2 \). Thus \( \lambda_3 = \lambda_1 \) and \( \lambda_3 = \lambda_2 \). This contradiction proves that \( \mathbf{u}_1 \), \( \mathbf{u}_2 \), and \( \mathbf{u}_3 \) must be linearly independent.

The following is an example of a "defective" matrix to which Theorem 5 does not apply.

Example 7. A Matrix Without an Eigenvector Basis

The matrix

\[
A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}
\]

has the characteristic polynomial which equals \( \det(A - \lambda I) = \lambda^2 \) (check this), so its two eigenvalues are both 0: \( \lambda_1 = \lambda_2 = 0 \). \( A \) is not symmetric and does not have two different eigenvalues. Thus Theorem 5 does not apply to \( A \).

Any eigenvector \( \mathbf{u} \) of \( A \) must satisfy \( (A - 0I)\mathbf{u} = 0 \). That is,

\[
A\mathbf{u} = \mathbf{0} \quad \text{or} \quad 0u_1 + 1u_2 = 0 \tag{39}
\]

The first equation reduces to \( u_2 = 0 \), and the second equation is vacuous. Thus a solution \( \mathbf{u} \) to (39) can have any value for \( u_1 \) while \( u_2 \)
must be 0. However, all such vectors $u = [u_1, 0]$ are multiples of one another, so the eigenvectors of $A$ do not form a basis for 2-space. ■

Another issue that we have skirted is complex-valued eigenvalues and eigenvectors. Complex eigenvalues were encountered in our discussion of population models in Section 4.5. In many other applications complex numbers arise naturally. (Incidentally, one can show that the eigenvalues of a symmetric matrix are always real.)

What happens when a matrix is not square (so that it has no eigenvalues or eigenvectors)? Can we find something like an eigenvalue decomposition for these matrices just as a pseudoinverse substitutes for the inverse in a nonsquare matrix? The answer is yes.

In the spirit of the development of the pseudoinverse of a nonsquare matrix, we again turn to the $n$-by-$n$ matrix $A^T A$ which is square and symmetric [since entry $(i, j)$ is just the scalar product of the $i$th and $j$th columns of $A$]. From the eigenvalue decomposition of $A^T A$, one can obtain a decomposition for $A$.

**Theorem 6. Singular-Value Decomposition.** For any $m$-by-$n$ matrix $A$ with linearly independent columns, let $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ be the eigenvalues of $A^T A$ and $U$ be an $n$-by-$n$ matrix whose columns $u_i$ are the associated (orthonormal) eigenvectors of $A^T A$. The $u_i$ form a basis for the row space of $A$.

Define the $i$th **singular value** $s_i$ to be $s_i = \sqrt{\lambda_i}$. Let $U'$ be an $m$-by-$n$ matrix with columns $u_i' = (1/s_i)Au_i$. The $u_i'$ form an orthonormal basis for the range of $A$.

Then $A$ can be decomposed in the form

$$A = U'D_sU'^T$$  \hspace{1cm} (40)

where $D_s$ is the diagonal matrix whose $i$th diagonal entry is $s_i$.

The proof of Theorem 6 is given in the Exercises. Recall from Theorem 4 of Section 5.3 that if $A$ has linearly independent columns, then $A^T A$ has linearly independent columns. There is a generalized form of Theorem 6 for linearly dependent columns.

The definition of the columns of $U'$ means that $U'$ has the matrix formula

$$U' = AUD_s^{-1}$$

The factorization (40) leads to the following simple matrix decomposition (by the same argument that led up to Theorem 3).

**Corollary A.** Let $A$ be as in Theorem 6, let $u_i$ be the $i$th column of $U$, and let $u_i'$ be the $i$th column of $U'$. Then
Eigenvector Bases and the Eigenvalue Decomposition

We also get from (40) a result like Theorem 1.

**Corollary B.** The computation \( c = Ab \) becomes \( c^* = Ds, b^* \) (or \( c_i^* = s_i b_i^* \)) when \( b^* = U^Tb \) and \( c^* = U^Tc \). That is, multiplying a vector by \( A \) reduces to scalar multiplication of the coordinates when we express \( b \) in the proper row space coordinates and \( c \) in the proper column space (range) coordinates.

The singular-value decomposition of \( A \) can be "inverted" to obtain the pseudoinverse of \( A \).

**Corollary C.** Let \( A \) be an \( m \)-by-\( n \) matrix. Then the pseudoinverse of \( A \) equals

\[
A^+ = UD_s^{-1}U^T
\]

Recall that \( D_s^{-1} \) is a diagonal matrix with entry \( (i, i) = 1/s_i \).

We now give the singular-value decomposition for the two-refinery matrix discussed in Section 5.3.

**Example 8. Singular-Value Decomposition of Two-Refinery Model**

The two-refinery variant was

\[
\begin{align*}
20x_1 + 4x_2 &= b_1 \\
10x_1 + 14x_2 &= b_2 \\
5x_1 + 5x_2 &= b_3
\end{align*}
\]

Let us compute the singular-value decomposition of \( A \). First we form \( A^T A \):

\[
A^T A = \begin{bmatrix} 525 & 245 \\ 245 & 237 \end{bmatrix}
\]

We need to find the eigenvalues \( \lambda_1, \lambda_2 \) of \( A^T A \), since the singular values \( s_1, s_2 \) are their square roots. Further, the eigenvectors of \( A^T A \) are the columns of the matrix \( U \). By some method, discussed previously or in the appendix to this section, we find that

\[
\begin{align*}
\lambda_1 &= 665 & u_1 &= [0.87, 0.49] \\
\lambda_2 &= 97 & u_2 &= [-0.49, 0.87]
\end{align*}
\]

So the singular values of \( A \) are \( s_1 = \sqrt{665} = 25.6, s_2 = \sqrt{97} = 9.8 \), and
As an aside, we note that $U$ performs a rotation of $20^\circ$ (see Theorem 3 of Section 5.4).

Next we compute $U' = [u'_1, u'_2]$, where $u'_i = (1/s_i)Au_i$. We obtain

$$U' = \begin{bmatrix} .76 & -.65 \\ .60 & .74 \\ .26 & .19 \end{bmatrix} \quad (47)$$
Then the singular-value decomposition of $A = U'DSU'$ is

$$
\begin{bmatrix}
20 & 4 \\
10 & 14 \\
5 & 5 \\
\end{bmatrix} = \begin{bmatrix}
.76 & -.65 \\
.60 & .74 \\
.26 & .19 \\
\end{bmatrix} \begin{bmatrix}
25.6 & 0 \\
0 & 9.8 \\
\end{bmatrix} \begin{bmatrix}
.87 & .49 \\
-.49 & .87 \\
\end{bmatrix}
$$

(48)

As a sum of simple matrices [see (41)], (48) becomes

$$
A = \begin{bmatrix}
16.9 & 9.5 \\
13.4 & 7.5 \\
5.8 & 3.3 \\
\end{bmatrix} + \begin{bmatrix}
3.1 & -5.5 \\
-3.4 & 6.5 \\
-0.8 & 1.7 \\
\end{bmatrix}
$$

(49)

The decomposition in (48) can be interpreted as follows. There are two basic "input" and "output" units for the refinery model. The first input unit consists of .87 barrel of petroleum for refinery 1 and .49 barrel for refinery 2, and one such input unit yields 25.6 output units, each consisting of .76 gallon of diesel oil, .60 heating oil, and .26 gasoline. The second input unit consists of -.49 barrel (production is "reversed") for refinery 1 and .87 barrel for refinery 2, and it yields 9.8 output units, each consisting of -.65 diesel, .74 heating, and .19 gasoline.

A more impressive use of the singular-value decomposition is given in Figure 5.10. Figure 5.10 (top) shows a 49-by-36 digitized image of a bust of Abe Lincoln [entry $(i,j)$ is the height of the bust in that position]. The remaining figures in this set show the digitized image produced by the matrix $A_k$, the sum of the first $k$ simple matrices in the singular-value decomposition of $A$.

Section 5.5 Exercises

Summary of Exercises
Exercises 1–5 involve the diagonalization of matrices presented in Theorem 2. Exercises 6–13 involve the eigenvalue decomposition of matrices into a sum of simple matrices (many require deflation to find the eigenvalues). Exercises 14 and 15 are about independence of eigenvectors. Exercise 16 involves defective matrices. Exercises 17–22 involve computing the singular-value decomposition. Exercises 23–25 prove the results in Theorem 6.

1. Compute the representation $UDSU^{-1}$ of Theorem 2 for the following matrices whose eigenvalues and largest eigenvector you were asked to determine in Exercise 23 of Section 3.1.

(a) $\begin{bmatrix} 4 & 0 \\ 2 & 2 \end{bmatrix}$  
(b) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  
(c) $\begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$  
(d) $\begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}$

2. For a starting vector of $p = [10, 10]$, compute $p^{(10)} = A^{10}p$ for each matrix $A$ in Exercise 1.
3. Compute $A^5$ for each matrix in Exercise 1 using the formula $A^5 = UD^2U^{-1}$.

4. (a) Given that $A = UD_U U^{-1}$, prove that $A^2 = UD^2 U^{-1}$.
(b) Use induction to prove that $A^k = UD^k U^{-1}$.

5. (a) Obtain a formula for $A^{-1}$ similar to $A = UD_U U^{-1}$.
Hint: Only the matrix $D$ will be different.
(b) Verify your formula in part (a) for $A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$ (the matrix in Example 1).

6. Give the eigenvalue decomposition into a sum of two simple matrices (Theorem 3) for each matrix in Exercise 1.

7. Give the eigenvalue decomposition into a sum of two simple matrices for the following symmetric matrices.
   (a) $\begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & -4 \\ -4 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} -1 & -5 \\ -5 & -1 \end{bmatrix}$

8. Use the deflation method to compute all eigenvalues and eigenvectors of the following symmetric matrices.
   (a) $\begin{bmatrix} 1 & 6 \\ 6 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
   (d) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -1 \\ 3 & -1 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

9. Using a software package for finding eigenvectors and eigenvalues, determine the eigenvalue decomposition into simple matrices for the following symmetric matrices.
   (a) $\begin{bmatrix} 1 & 3 & 0 \\ 3 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

10. Approximate the following symmetric digital pictures by:
(i) The first simple matrix in the eigenvalue decomposition.
(ii) The sum of the first two simple matrices in the eigenvalue decomposition.
Use deflation (or a software package) to determine the two first eigenvectors and eigenvalues.
11. Explain how the symmetry in the 8-by-8 digital picture (25) in Example 5 would allow one to find the eigenvalue decomposition for the 4-by-4 upper right corner submatrix $A'$ and use this to get the eigenvalue decomposition for the whole matrix.

Find the dominant eigenvalue and associated (normalized) eigenvector for $A'$; approximate $A'$ by the first simple matrix in the eigenvalue decomposition.

12. Verify that the eigenvalues (34) and dominant eigenvector in Example 6 are correct. Determine the second principal component in Example 6 (the normalized eigenvector for $\lambda_2$).

13. Determine the first principal component for each of the following covariance matrices. In each case, tell how well it accounts for the variability (how well does $\lambda_1 u_1 \ast u_1$ approximate

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>1.1</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.1</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>0.6</td>
<td>3.1</td>
<td>1.5</td>
</tr>
<tr>
<td>0.6</td>
<td>4.1</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>3.1</td>
<td>0.8</td>
<td>2.7</td>
<td>5.2</td>
</tr>
<tr>
<td>1.5</td>
<td>1.2</td>
<td>5.2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

14. In the proof Theorem 4, show that $u_1$ and $u_2$ are linearly independent by supposing the opposite, that $u_2 = ru_1$, and obtaining a contradiction when $Au_2 \neq A(ru_1)$.

15. Two $n$-by-$n$ matrices $A$ and $B$ are called similar if there exists an invertible $n$-by-$n$ matrix $U$ such that $A = UBU^{-1}$ (or equivalently, $B = U^{-1}AU$).
(a) Show that similar matrices have the same set of eigenvalues.
(b) Show that if $A$ has a set of $n$ linearly independent eigenvectors, then $A$ is similar to a diagonal matrix.

16. Find the eigenvalues and as many eigenvectors as you can for the following defective matrices.

(a) \[
\begin{bmatrix}
1 & 0 \\
2 & 1
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
1 & -1 \\
1 & 3
\end{bmatrix}
\]
(c) \[
\begin{bmatrix}
2 & 1 & 0 \\
0 & 2 & 1 \\
0 & 0 & 2
\end{bmatrix}
\]

17. Find the singular-value decomposition [equation (40)] for the following matrices and use the decomposition to write the matrices as a sum of simple matrices.

(a) \[
\begin{bmatrix}
2 \\
1 \\
0
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
1 & 0 \\
3 & 3 \\
2 & 5
\end{bmatrix}
\]
(c) \[
\begin{bmatrix}
3 & 0 \\
2 & 5 \\
1 & 3 \\
4 & 0
\end{bmatrix}
\]
(d) \[
\begin{bmatrix}
1 & 2 \\
2 & 3 \\
3 & 4
\end{bmatrix}
\]

18. For a refinery problem with $p$ refineries and $q$ products as in Example 8, suppose that the coefficient matrix is the matrix in each part of Exercise 17. In each case, interpret the singular-value decomposition in terms of units of "input" and "output," as was done at the end of Example 8.

19. Use the formula for the pseudoinverse in Corollary C to compute the pseudoinverse of each matrix in Exercise 17.

20. With the help of deflation or a software package to find eigenvectors of $A^TA$, find the singular-value decomposition for the following matrices and use the decomposition to approximate these matrices as a sum of two simple matrices.

(a) \[
\begin{bmatrix}
2 & 0 & -1 \\
1 & 2 & 0 \\
4 & -1 & 2 \\
5 & 0 & -1
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
.1 & .2 & .3 & .4 \\
.4 & 0 & .1 & .2 \\
.3 & .7 & .1 & 0 \\
.1 & .6 & .1 & 0 \\
.4 & 0 & .1 & .3
\end{bmatrix}
\]

21. If $A$ is a symmetric matrix, show that the singular-value decomposition reduces to the standard eigenvalue decomposition in Theorem 2.

22. Verify the formula for the pseudoinverse in Corollary C.

23. Show that the $n$ orthonormal eigenvectors of $A^TA$ in $U$ (in the singular-value decomposition) are a basis for the row space of the $m$-by-$n$ matrix $A$. 
Sec. 5.5 Appendix: Finding Eigenvalues and Eigenvectors

499

Hint: What is the dimension of the row space of \( A \) if its columns are linearly independent?

24. Show that the columns of \( U' \) (in the singular-value decomposition) must form a basis for the range of \( A \) (the column space of \( A \)).

Hint: Use Exercise 23, and Exercise 30 of Section 5.3.

25. Verify the singular-value decomposition (40) by showing that

\[
D_s = U'^T A U
\]

(a) First show that (40) and (*) are equivalent matrix equations.

Hint: Use the fact that (by orthonormality) \( U'^T U = U'^T U' = I \).

(b) Prove (*) by verifying the following sequence of matrix equations:

\[
U'^T A U = (A U'D_s^{-1})' = (D_s^{-1} U'^T A T)AU = D_s^{-1} U'^T (A'^T AU) = D_s^{-1} U'^T (UD_s) = D_s^{-1} D_s = D_s
\]

Appendix to Section 5.5: Finding Eigenvalues and Eigenvectors

In this appendix we present two methods for finding eigenvalues and eigenvectors. The first is a way to speed up the search for an eigenvalue \( \lambda \) and associated eigenvector \( u \) once we have a rough approximation to \( \lambda \), say, obtained by guesswork or by a few rounds of the iterative method \( A^k x \). The following basic theorems about eigenvalues are needed.

**Theorem 1**

(i) For any nonzero integer \( k \), \( A \)'s eigenvectors are eigenvectors of \( A^k \). If \( \lambda \) is an eigenvalue of \( A \), then \( \lambda^k \) is an eigenvalue of \( A^k \). In particular, \( 1/\lambda \) is an eigenvalue for \( A^{-1} \). (If \( k < 0 \), we assume that \( A^{-1} \) exists.)

(ii) For any scalar \( r \), \( A \) and \( A - rI \) have the same set of eigenvectors and \( \lambda \) is an eigenvalue for \( A \) if and only if \( \lambda - r \) is an eigenvalue of \( A - rI \).

**Proof.** We give a proof of part (i) for \( k = -1 \) [positive \( k \) and Theorem 1, part (ii) are left as exercises]. Suppose that \( u \) is an eigenvector of \( A \) with associated eigenvalue \( \lambda \). Then

\[
u = Iu = (A^{-1}A)u = A^{-1}(Au) = A^{-1}\lambda u
\]

Dividing both sides of (1) by \( \lambda \), we have
So $1/\lambda$ is an eigenvalue of $A^{-1}$ with eigenvector $u$, as claimed.

**Theorem 2.** The rate at which the iterative method $A^k x$ converges to a dominant eigenvector $\lambda_1$ of $A$ is proportional to $|\lambda_2/1/\lambda_1|$.

**Proof.** Any vector $x$ can be written as a linear combination of the eigenvectors (assuming that the $n$-by-$n$ matrix $A$ has $n$ linearly independent eigenvectors)

$$x = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

Then

$$A^k x = c_1\lambda_1^k u_1 + c_2\lambda_2^k u_2 + \cdots + c_n\lambda_n^k u_n$$

$$= \lambda_1^k \left( c_1u_1 + c_2\left( \frac{\lambda_2}{\lambda_1} \right)^k u_2 + \cdots + c_n\left( \frac{\lambda_n}{\lambda_1} \right)^k u_n \right)$$

The last line of (3) shows how the size ($\lambda_2/\lambda_1$) affects the convergence to $\lambda_1^k c_1u_1$.

Suppose that $A$ has an inverse $A^{-1}$. Since $\lambda$ is an eigenvalue of $A$ if and only if $1/\lambda$ is an eigenvalue of $A^{-1}$ and both have the same eigenvectors, we have

**Corollary.** An eigenvector $u_n$ associated with the smallest eigenvalue $\lambda_n$ of $A$ can be found (if $\lambda_n$ is unique) by applying the iterative method to find the largest eigenvalue of $A^{-1}$:

$$y^{(k)} = A^{-1}y^{(k-1)}$$

The rate of convergence will be proportional to $|\lambda_{n-1}|/|\lambda_n|$, where $\lambda_{n-1}$ is the second smallest eigenvalue of $A$.

Rather than compute the inverse of $A^{-1}$ and then use (4), we can write (4) as

$$A y^{(k)} = y^{(k-1)}$$

In (5), we are applying the regular iterative method in reverse. To find $y^{(k)}$ given $y^{(k-1)}$, we solve (5) by Gaussian elimination (saving the matrices $L$ and $U$—see Section 3.2—to use again for each successive $y^{(k)}$). We call the
method for finding the smallest eigenvalue of A and an associated eigenvector using (5) the inverse iterative method.

Surprisingly, the inverse iterative method yields a faster way to compute the dominant eigenvector than the (forward) iterative method.

Theorem 3. Shifted Inverse Iterative Method. If $\sigma$ is an approximate value for an eigenvalue of a square matrix A, an associated eigenvector $u$ can be found quickly by applying the inverse iterative method to $A - \sigma I$. The eigenvalue can then be obtained from $u$ using the Raleigh quotient. If $\lambda_p$ is the true value of the eigenvalue and $\lambda_q$ is the next-closest eigenvalue of A, the rate of convergence is $|\lambda_q - \sigma|/|\lambda_p - \sigma|$.

This method is called the shifted inverse iterative method because of the "shift" of eigenvalues caused by $-\sigma I$. Recall that from Theorem 1, part (ii), $A - \sigma I$ has the same eigenvectors as A and its eigenvalues are shifted by $\sigma$. If $\sigma$ is close to $\lambda_p$, then $\lambda_p - \sigma$ will be the smallest eigenvalue of $A - \sigma I$ by far, and the rate of convergence $|\lambda_q - \sigma|/|\lambda_p - \sigma|$ of the inverse iterative method will be very fast (i.e., two or three iterations should suffice).

We should note that computations are very unstable with $A - \sigma I$, since when $\lambda_p$ is close to $\sigma$, $1/(\lambda_p - \sigma)$ is an eigenvalue of $A^{-1}$ of immense size. This implies that $\|A^{-1}\|$, and hence the condition number of A, are very large. However, the only effect that this instability has on the computations of the inverse iterative method is a distortion of the total size of the $y^{(k)}$ but not the direction of these vectors (the total size does not concern us, since we use scaling).

Hybrid Deflation Procedure to Find All Eigenvalues and Eigenvectors of a Symmetric Matrix

Step 1. Starting with $A_1 = A$, use the (forward) iterative method a few times on $A_1$ to get an approximation $\sigma$ to the dominant eigenvalue $\lambda_i$ of $A_1$, together with an approximate eigenvector $v$.

Step 2. Starting with $v$, use the shifted inverse iterative method on $A_i - \sigma I$ to get more accurate values for the unit-length eigenvector $u_i$. Obtain $\lambda_i$ from $u_i$ by the Raleigh quotient $u_i \cdot Au_i / u_i \cdot u_i$.

Step 3. Compute $A_{i+1} = A_i - \lambda_i u_i^* u_i$, set $i = i + 1$, and if $i \leq n$, go to step 1.

We now present an almost magical procedure to find all the eigenvalues at once of a square matrix A with distinct eigenvalues.
QR Method for Finding All Eigenvalues of a Matrix with Distinct Eigenvalues

**Step 1.** Let $A_0 = A$. For successive $k$,
(a) Given $A_k$, compute the QR decomposition $A_k = Q_k R_k$ (by the Gram–Schmidt orthogonalization procedure); and then
(b) Set $A_{k+1} = R_k Q_k$ and go to step (a).
Stop when the entries below the main diagonal of $A_k$ are all almost 0 (entries above the main diagonal do not converge to 0 unless $A$ is symmetric).

**Step 2.** The entries on the main diagonal of the last $A_k$ will be approximately the eigenvalues of $A$ (in order of decreasing absolute value). Use the shifted inverse iterative method to find the eigenvector (or if the eigenvalue $\lambda$ is essentially exact, solve the homogeneous system $(A - \lambda I)u = 0$).

### Example 1. Example of QR and Shifted Inverse Iterative Method

Let us use the QR method on the matrix $L$ in the Leslie model from Example 1 of Section 4.5.

\[ x' = Lx, \quad \text{where } L = \begin{bmatrix} 0 & 4 & 1 \\ .4 & 0 & 0 \\ 0 & .6 & 0 \end{bmatrix} \]

We noted in Section 4.5 that $L^k x$ takes a long time to converge to a multiple of the dominant eigenvalue $\lambda_1$. Since convergence is slow, Theorem 2 says that the second largest eigenvalue $\lambda_2$ must be close to $\lambda_1$. The QR method is also slow for this $L$. If we run it until all below-diagonal entries are $< .001$ (then the eigenvalues are accurate to about three decimal places), it requires 60 iterations and yields

\[ A_{60} = \begin{bmatrix} 1.334 & -3.586 & -1.142 \\ .000 & -1.118 & -.394 \\ .000 & .000 & -.152 \end{bmatrix} \]

So $\lambda_1 = 1.334$, $\lambda_2 = -1.118$, $\lambda_3 = -.152$. Solving the homogeneous system $(L - \lambda_1 I) = 0$ will give a corresponding eigenvector.

Let us next try the shifted iterative method. Starting with $x = [100, 50, 30]$, we gave a table in Section 4.5 of iterates up to $L^{20}a$ (Table 4.3), at which point there was still a little cyclic behavior. Consider the ninth, tenth, and eleventh iterates:

\[ x^{(9)} = [2021, 493, 279], \quad x^{(10)} = [2250, 808, 295], \quad x^{(11)} = [3529, 900, 485] \]
When we compute the Raleigh quotients for \( k = 9 \) and \( k = 10 \), we get

\[
\begin{align*}
\frac{x^{(9)} \cdot x^{(10)}}{x^{(9)} \cdot x^{(9)}} &= \frac{5,027,899}{4,414,515} = 1.138 \\
\frac{x^{(10)} \cdot x^{(11)}}{x^{(10)} \cdot x^{(10)}} &= \frac{8,668,230}{5,802,389} = 1.493
\end{align*}
\]

The average of our two estimates is \((1.138 + 1.493)/2 = 1.32\). Presumably, we are alternating above and below the true eigenvalue, so this average value of 1.32 should be close to the true eigenvalue. Now we apply the shift step of computing \( L' = L - 1.32I \):

\[
L' = L - 1.32I = \begin{bmatrix}
0 & 4 & 1 \\
.4 & 0 & 0 \\
0 & .6 & 0
\end{bmatrix} - \begin{bmatrix}
1.32 & 0 & 0 \\
0 & 1.32 & 0 \\
0 & 0 & 1.32
\end{bmatrix} = \begin{bmatrix}
-1.32 & 4 & 1 \\
.4 & -1.32 & 0 \\
0 & .6 & -1.32
\end{bmatrix}
\]

Let us scale \( x^{(11)} \) by dividing its entries by the largest entry, 3529, to obtain \( x' = [1, .255, .137] \). We use \( x' \) as the starting vector for backward iteration with \( L' \) (in search of an eigenvector associated with the smallest eigenvalue of \( L' \)). After two rounds we have the vector (rounded to integers) \([4658, 1396, 628]\), which divided by the sum of its entries (to be like a population probability distribution) yields

\[
u_1 = [.697, .209, .094]
\]

Computing \( Lu_1 \), we obtain

\[
Lu_1 = [.930, .278, .125] = 1.334u_1
\]

so \( \lambda_1 = 1.334 \). Further, (9) confirms that \( u_1 \) is an eigenvector.

If we wanted to get all eigenvalues and associated eigenvectors for \( L \), we could use 10 or 15 rounds of the QR method to get estimates for each eigenvalue and then use the shifted inverse method to home in the associated eigenvector of each eigenvalue (any vector can be used as the starting vector for the inverse iterative method). 


The convergence is not fast, especially when two eigenvalues are close
together (in absolute value). The computation of each stage is slow (proportional to \( n^3 \)), but schemes are available to "preprocess" \( A \) so that the successive QR decompositions can be computed quickly. Other shortcuts further speed this procedure, including shifts (as in the shifted inverse iterative method). This method works in some cases where there are multiple eigenvalues (of the same absolute value).

A related method called the **LU method** reverses the matrices in the LU decomposition of a matrix the way the QR method reverses the matrices in a QR decomposition (see the books mentioned above). The LU method also converges for any matrix with distinct eigenvalues.

### Section 5.5 Appendix Exercises

#### Summary of Exercises

Exercises 1–5 relate to Theorem 1. Exercises 6–8 illustrate the hybrid deflation procedure.

1. (a) Show that for any positive integer \( k \), any eigenvector for the square matrix \( A \) is also an eigenvector for \( A^k \).
   
   (b) Also show that if \( A^{-1} \) exists, part (a) is true for negative integers.

2. (a) Show that if \( A^{-1} \) exists, any eigenvector for \( A^k \) is also an eigenvector for \( A \).
   
   (b) The existence of \( A^{-1} \) is essential for the result in part (a). Verify
   this by showing that for \( A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \), \( 1 \) is an eigenvector for \( A^2 \) but not for \( A \).

3. (a) Show that for any positive integer, if \( \lambda \) is an eigenvalue of the square matrix \( A \), then \( \lambda^k \) is an eigenvalue of \( A^k \).
   
   (b) Also show that if \( A^{-1} \) exists, part (a) is true for negative integers.

4. Show that for any scalar \( r \), \( A \) and \( A - rI \) have the same set of eigenvectors.

5. Show that for any scalar \( r \), \( \lambda \) is an eigenvalue for \( A \) if and only if \( \lambda - r \) is an eigenvalue for \( A - rI \).

6. Use the inverse power method to find the dominant eigenvalue and
   eigenvector for the following matrices.
   
   \[
   \begin{align*}
   (a) \quad & \begin{bmatrix} 4 & 0 \\ 2 & 2 \end{bmatrix} \\
   (b) \quad & \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
   (c) \quad & \begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix} \\
   (d) \quad & \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}
   \end{align*}
   \]

7. Use the hybrid deflation procedure to find all eigenvalues and associated
   eigenvectors for the following symmetric matrices.
   
   \[
   \begin{align*}
   (a) \quad & \begin{bmatrix} 1 & 6 \\ 6 & 1 \end{bmatrix} \\
   (b) \quad & \begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix} \\
   (c) \quad & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
   \end{align*}
   \]
8. Apply the QR method (using the program in Exercise 10 or a software package) to find the eigenvalues and associated eigenvectors for the matrices in Exercise 7.

Programming Projects
9. Write a program to implement the hybrid deflation procedure.

10. Write a program to implement the QR method.