CHAPTER

4

Least Squares

The global positioning system (GPS) is a satellite-based location technology that provides accurate positioning at any time, from any point on earth. In just a few years, GPS has gone from a special-purpose navigation technology used by pilots, ship captains, and hikers to everyday use in automobiles, cellphones, and PDAs.

The system consists of 24 satellites following precisely regulated orbits, emitting synchronized signals. An earth-based receiver picks up the satellite signals, finds its distance from all visible satellites, and uses the data to triangulate its position.

Reality ✓ Check Reality Check 4 on page 238 shows the use of equation solvers and least squares calculations to do the location estimation.

The concept of least squares dates from the pioneering work of Gauss and Legendre in the early 19th century. Its use permeates modern statistics and mathematical modeling. The key techniques of regression and parameter estimation have become fundamental tools in the sciences and engineering.

In this chapter, the normal equations are introduced and applied to a variety of data-fitting problems. Later, a more sophisticated approach, using the QR factorization, is explored, followed by a discussion of nonlinear least squares problems.

4.1 LEAST SQUARES AND THE NORMAL EQUATIONS

The need for least squares methods comes from two different directions, one each from our studies of Chapters 2 and 3. In Chapter 2, we learned how to find the solution of $Ax = b$ when a solution exists. In this chapter, we find out what to do when there is no solution. When the equations are inconsistent, which is likely if the number of equations exceeds the number of unknowns, the answer is to find the next best thing: the least squares approximation.

Chapter 3 addressed finding polynomials that exactly fit data points. However, if the data points are numerous, or the data points are collected only within some margin of error, fitting a high-degree polynomial exactly is rarely the best approach. In such cases, it is more reasonable to fit a simpler model that may only approximate the data points. Both problems, solving inconsistent systems of equations and fitting data approximately, are driving forces behind least squares.

4.1.1 Inconsistent systems of equations

It is not hard to write down a system of equations that has no solutions. Consider the following three equations in two unknowns:

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3. \end{aligned} \tag{4.1}$$

Any solution must satisfy the first and third equations, which cannot both be true. A system of equations with no solution is called **inconsistent**.

What is the meaning of a system with no solutions? Perhaps the coefficients are slightly inaccurate. In many cases, the number of equations is greater than the number of unknown variables, making it unlikely that a solution can satisfy all the equations. In fact, m equations in n unknowns typically have no solution when $m > n$. Even though Gaussian elimination will not give us a solution to an inconsistent system $Ax = b$, we should not completely give up. An alternative in this situation is to find a vector x that comes the closest to being a solution.

If we choose this "closeness" to mean close in Euclidean distance, there is a straightforward algorithm for finding the closest x. This special x will be called the **least squares solution**.

We can get a better picture of the failure of system (4.1) to have a solution by writing it in a different way. The matrix form of the system is $Ax = b$, or

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}. \tag{4.2}$$

The alternative view of matrix/vector multiplication is to write the equivalent equation

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}. \tag{4.3}$$

In fact, any $m \times n$ system $Ax = b$ can be viewed as a vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b, \tag{4.4}$$

which expresses b as a linear combination of the columns v_i of A, with coefficients $x_1, \ldots, x_n$. In our case, we are trying to hit the target vector b as a linear combination of two other three-dimensional vectors. Since the combinations of two three-dimensional vectors form a plane inside R^3, equation (4.3) has a solution only if the vector b lies in that plane. This will always be the situation when we are trying to solve m equations in n unknowns, with $m > n$. Too many equations make the problem overspecified and the equations inconsistent.

Figure 4.1(b) shows a direction for us to go when a solution does not exist. There is no pair x_1, x_2 that solves (4.1), but there is a point in the plane Ax of all possible candidates that

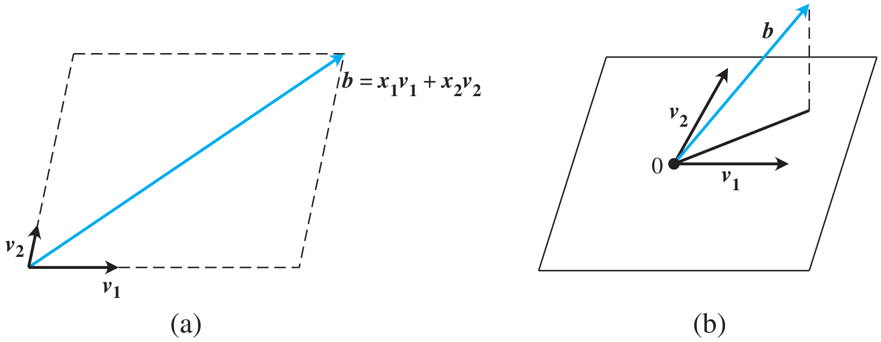


Figure 4.1 Geometric solution of a system of three equations in two unknowns. (a) Equation (4.3) requires that the vector b, the right-hand side of the equation, is a linear combination of the columns vectors v_1 and v_2. (b) If b lies outside of the plane defined by v_1 and v_2, there will be no solution. The least squares solution $\overline{x}$ makes the combination vector $A\overline{x}$ the one in the plane Ax that is nearest to b in the sense of Euclidean distance.

is closest to b. This special vector $A\overline{x}$ is distinguished by the following fact: The residual vector $b - A\overline{x}$ is perpendicular to the plane $\{Ax | x \in R^n\}$. We will exploit this fact to find a formula for $\overline{x}$, the least squares "solution."

First we establish some notation. Recall the concept of the **transpose** A^T of the $m \times n$ matrix A, which is the $n \times m$ matrix whose rows are the columns of A and whose columns are the rows of A, in the same order. The transpose of the sum of two matrices is the sum of the transposes, $(A + B)^T = A^T + B^T$. The transpose of a product of two matrices is the product of the transposes in the reverse order—that is, $(AB)^T = B^T A^T$.

To work with perpendicularity, recall that two vectors are at right angles to one another if their dot product is zero. For two m-dimensional column vectors u and v, we can write the dot product solely in terms of matrix multiplication by

$$u^T v = [u_1, \ldots, u_m] \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}. \tag{4.5}$$

The vectors u and v are perpendicular, or **orthogonal**, if $u^T \cdot v = 0$, using ordinary matrix multiplication.

Now we return to our search for a formula for $\overline{x}$. We have established that

$$(b - A\overline{x}) \perp \{Ax | x \in R^n\}.$$

Expressing the perpendicularity in terms of matrix multiplication, we find that

$$(Ax)^T (b - A\overline{x}) = 0 \text{ for all } x \text{ in } R^n.$$

Using the preceding fact about transposes, we can rewrite this expression as

$$x^T A^T (b - A\overline{x}) = 0 \text{ for all } x \text{ in } R^n,$$

SPOTLIGHT ON

Orthogonality Least squares is based on orthogonality. The shortest distance from a point to a plane is carried by a line segment orthogonal to the plane. The normal equations are a computational way to locate the line segment, which represents the least squares error.

meaning that the n-dimensional vector $A^T(b - A\overline{x})$ is perpendicular to every vector x in R^n, including itself. There is only one way for that to happen:

$$A^T(b - A\overline{x}) = 0.$$

This gives a system of equations that defines the least squares solution,

$$A^T A\overline{x} = A^T b. \tag{4.6}$$

The system of equations (4.6) is known as the **normal equations**. Its solution $\overline{x}$ is the so-called least squares solution of the system $Ax = b$.

Normal equations for least squares

Given the inconsistent system

$$Ax = b,$$

solve

$$A^T A\overline{x} = A^T b$$

for the least squares solution $\overline{x}$ that minimizes the Euclidean length of the residual $r = b - Ax$.

▶ **EXAMPLE 4.1** Use the normal equations to find the least squares solution of the inconsistent system (4.1).

The problem in matrix form $Ax = b$ has

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

The components of the normal equations are

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

and

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

The normal equations

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

can now be solved by Gaussian elimination. The tableau form is

$$\left[\begin{array}{cc|c} 3 & 1 & 6 \\ 1 & 3 & 4 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 3 & 1 & 6 \\ 0 & 8/3 & 2 \end{array}\right],$$

which can be solved to get $\overline{x} = (\overline{x}_1, \overline{x}_2) = (7/4, 3/4)$. ◀

Substituting the least squares solution into the original problem yields

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

To measure our success at fitting the data, we calculate the residual of the least squares solution $\overline{x}$ as

$$r = b - A\overline{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.0 \\ 0.5 \end{bmatrix}.$$

If the residual is the zero vector, then we have solved the original system $Ax = b$ exactly. If not, the Euclidean length of the residual vector is a backward error measure of how far $\overline{x}$ is from being a solution.

There are at least three ways to express the size of the residual. The Euclidean length of a vector,

$$||r||_2 = \sqrt{r_1^2 + \cdots + r_m^2}, \tag{4.7}$$

is a norm in the sense of Chapter 2, called the **2-norm**. The **squared error**

$$\text{SE} = r_1^2 + \cdots + r_m^2,$$

and the **root mean squared error** (the root of the mean of the squared error)

$$\text{RMSE} = \sqrt{\text{SE}/m} = \sqrt{(r_1^2 + \cdots + r_m^2)/m}, \tag{4.8}$$

are also used to measure the error of the least squares solution. The three expressions are closely related; namely

$$\text{RMSE} = \frac{\sqrt{\text{SE}}}{\sqrt{m}} = \frac{||r||_2}{\sqrt{m}},$$

so finding the $\overline{x}$ that minimizes one, minimizes all. For Example 4.1, the SE $= (.5)^2 + 0^2 + (-.5)^2 = 0.5$, the 2-norm of the error is $||r||_2 = \sqrt{0.5} \approx 0.707$, and the RMSE $= \sqrt{0.5/3} = 1/\sqrt{6} \approx 0.408$.

▶ **EXAMPLE 4.2** Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$.

The normal equations $A^T Ax = A^T b$ are

$$\begin{bmatrix} 9 & 6 \\ 6 & 29 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 45 \\ 75 \end{bmatrix}.$$

The solution of the normal equations are $\overline{x}_1 = 3.8$ and $\overline{x}_2 = 1.8$. The residual vector is

$$r = b - A\overline{x} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 3.8 \\ 1.8 \end{bmatrix}$$

$$= \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} - \begin{bmatrix} -3.4 \\ 13 \\ 11.2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 2 \\ -2.2 \end{bmatrix},$$

which has Euclidean norm $||e||_2 = \sqrt{(0.4)^2 + 2^2 + (-2.2)^2} = 3$. This problem is solved in an alternative way in Example 4.14. ◀

4.1.2 Fitting models to data

Let $(t_1, y_1), \ldots, (t_m, y_m)$ be a set of points in the plane, which we will often refer to as the "data points." Given a fixed class of models, such as all lines $y = c_1 + c_2 t$, we can seek to locate the specific instance of the model that best fits the data points in the 2-norm. The core of the least squares idea consists of measuring the residual of the fit by the squared errors of the model at the data points and finding the model parameters that minimize this quantity. This criterion is displayed in Figure 4.2.

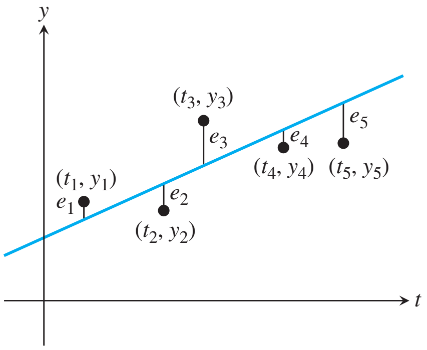


Figure 4.2 Least squares fitting of a line to data. The best line is the one for which the squared error $e_1^2 + e_2^2 + \cdots + e_5^2$ is as small as possible among all lines $y = c_1 + c_2 t$.

▶ EXAMPLE 4.3 Find the line that best fits the three data points $(t, y) = (1, 2), (-1, 1)$, and $(1, 3)$ in Figure 4.3.

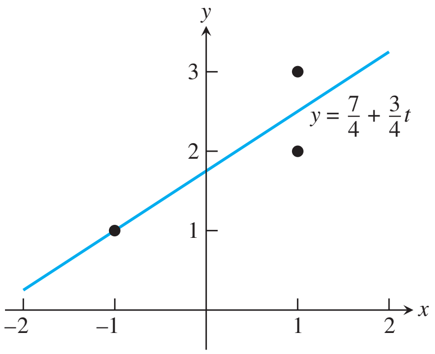


Figure 4.3 Best line in Example 4.3. One each of the data points lies above, on, and below the best line.

The model is $y = c_1 + c_2 t$, and the goal is to find the best c_1 and c_2. Substitution of the data points into the model yields

$$\begin{aligned} c_1 + c_2(1) &= 2 \\ c_1 + c_2(-1) &= 1 \\ c_1 + c_2(1) &= 3, \end{aligned}$$

or, in matrix form,

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

We know this system has no solution (c_1, c_2) for two separate reasons. First, if there is a solution, then the $y = c_1 + c_2 t$ would be a line containing the three data points. However, it is easily seen that the points are not collinear. Second, this is the system of equation (4.2) that we discussed at the beginning of this chapter. We noticed then that the first and third equations are inconsistent, and we found that the best solution in terms of least squares is $(c_1, c_2) = (7/4, 3/4)$. Therefore, the best line is $y = 7/4 + 3/4t$. ◀

We can evaluate the fit by using the statistics defined earlier. The residuals at the data points are

t	y	line	error
1	2	2.5	−0.5
−1	1	1.0	0.0
1	3	2.5	0.5

and the RMSE is $1/\sqrt{6}$, as seen earlier.

The previous example suggests a three-step program for solving least squares data-fitting problems.

Fitting data by least squares

Given a set of m data points $(t_1, y_1), \ldots, (t_m, y_m)$.

STEP 1. Choose a model. Identify a parameterized model, such as $y = c_1 + c_2 t$, which will be used to fit the data.

STEP 2. Force the model to fit the data. Substitute the data points into the model. Each data point creates an equation whose unknowns are the parameters, such as c_1 and c_2 in the line model. This results in a system $Ax = b$, where the unknown x represents the unknown parameters.

STEP 3. Solve the normal equations. The least squares solution for the parameters will be found as the solution to the system of normal equations $A^T Ax = A^T b$.

These steps are demonstrated in the following example:

▶ **EXAMPLE 4.4** Find the best line and best parabola for the four data points $(-1, 1), (0, 0), (1, 0), (2, -2)$ in Figure 4.4.

In accordance with the preceding program, we will follow three steps: (1) Choose the model $y = c_1 + c_2 t$ as before. (2) Forcing the model to fit the data yields

SPOTLIGHT ON

Compression Least squares is a classic example of data compression. The input consists of a set of data points, and the output is a model that, with a relatively few parameters, fits the data as well as possible. Usually, the reason for using least squares is to replace noisy data with a plausible underlying model. The model is then often used for signal prediction or classification purposes.

In Section 4.2, various models are used to fit data, including polynomials, exponentials, and trigonometric functions. The trigonometric approach will be pursued further in Chapters 10 and 11, where elementary Fourier analysis is discussed as an introduction to signal processing.

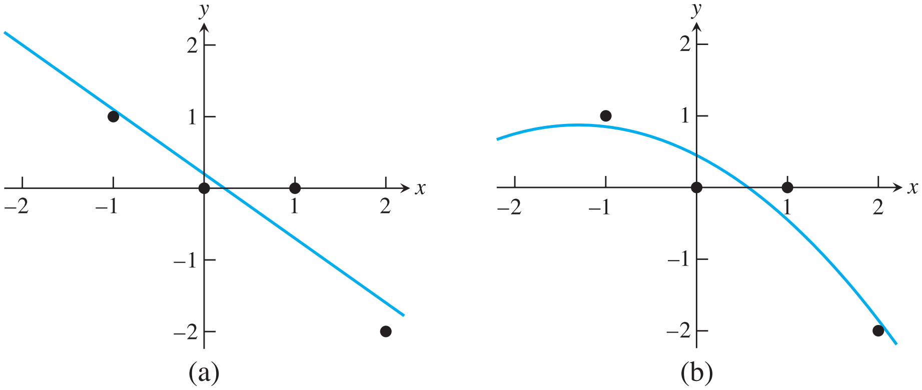


Figure 4.4 Least Squares Fits to Data Points in Example 4.4. (a) Best line $y = 0.2 - 0.9t$. RMSE is 0.418. (b) Best parabola $y = 0.45 - 0.65t - 0.25t^2$. RMSE is 0.335.

$$
\begin{aligned}
c_1 + c_2(-1) &= 1 \\
c_1 + c_2(0) &= 0 \\
c_1 + c_2(1) &= 0 \\
c_1 + c_2(2) &= -2,
\end{aligned}
$$

or, in matrix form,

$$
\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}
\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}
=
\begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix}.
$$

(3) The normal equations are

$$
\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}
\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}
=
\begin{bmatrix} -1 \\ -5 \end{bmatrix}.
$$

Solving for the coefficients c_1 and c_2 results in the best line $y = c_1 + c_2 t = 0.2 - 0.9t$.

The residuals are

t	y	line	error
−1	1	1.1	−0.1
0	0	0.2	−0.2
1	0	−0.7	0.7
2	−2	−1.6	−0.4

The error statistics are squared error SE $= (-.1)^2 + (-.2)^2 + (.7)^2 + (-.4)^2 = 0.7$ and RMSE $= \sqrt{.7}/\sqrt{4} = 0.418$.

Next, we extend this example by keeping the same four data points, but changing the model. Set $y = c_1 + c_2 t + c_3 t^2$ and substitute the data points to yield

$$
\begin{aligned}
c_1 + c_2(-1) + c_3(-1)^2 &= 1 \\
c_1 + c_2(0) + c_3(0)^2 &= 0 \\
c_1 + c_2(1) + c_3(1)^2 &= 0 \\
c_1 + c_2(2) + c_3(2)^2 &= -2,
\end{aligned}
$$

SPOTLIGHT ON

Conditioning Since input data is assumed to be subject to errors in least squares problems, it is especially important to reduce error magnification. We have presented the normal equations as the most straightforward approach to solving the least squares problem, and it is fine for small problems. However, the condition number cond($A^T A$) is approximately the square of the original cond(A), which will greatly increase the possibility that the problem is ill-conditioned. More sophisticated methods allow computing the least squares solution directly from A without forming $A^T A$. These methods are based on the QR-factorization, introduced in Section 4.3, and the singular value decomposition of Chapter 12.

or, in matrix form,

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix}.$$

This time, the normal equations are three equations in three unknowns:

$$\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ -7 \end{bmatrix}.$$

Solving for the coefficients results in the best parabola $y = c_1 + c_2 t + c_3 t^2 = 0.45 - 0.65t - 0.25t^2$. The residual errors are given in the following table:

t	y	parabola	error
-1	1	0.85	0.15
0	0	0.45	-0.45
1	0	-0.45	0.45
2	-2	-1.85	-0.15

The error statistics are squared error SE $= (.15)^2 + (-.45)^2 + (.45)^2 + (-.15)^2 = 0.45$ and RMSE $= \sqrt{.45}/\sqrt{4} \approx 0.335$. ◀

The MATLAB commands `polyfit` and `polyval` are designed not only to interpolate data, but also to fit data with polynomial models. For n input data points, `polyfit` used with input degree $n - 1$ returns the coefficients of the interpolating polynomial of degree $n - 1$. If the input degree is less than $n - 1$, `polyfit` will instead find the best least squares polynomial of that degree. For example, the commands

```
>> x0=[-1 0 1 2];
>> y0=[1 0 0 -2];
>> c=polyfit(x0,y0,2);
>> x=-1:.01:2;
>> y=polyval(c,x);
>> plot(x0,y0,’o’,x,y)
```

find the coefficients of the least squares degree-two polynomial and plot it along with the given data from Example 4.4.

Example 4.4 shows that least squares modeling need not be restricted to finding best lines. By expanding the definition of the model, we can fit coefficients for any model as long as the coefficients enter the model in a linear way.

4.1.3 Conditioning of least squares

We have seen that the least squares problem reduces to solving the normal equations $A^T A\overline{x} = A^T b$. How accurately can the least squares solution $\overline{x}$ be determined? This is a question about the forward error of the normal equations. We carry out a double precision numerical experiment to test this question, by solving the normal equations in a case where the correct answer is known.

▶ **EXAMPLE 4.5** Let $x_1 = 2.0, x_2 = 2.2, x_3 = 2.4, \ldots, x_{11} = 4.0$ be equally spaced points in $[2,4]$, and set $y_i = 1 + x_i + x_i^2 + x_i^3 + x_i^4 + x_i^5 + x_i^6 + x_i^7$ for $1 \le i \le 11$. Use the normal equations to find the least squares polynomial $P(x) = c_1 + c_2 x + \cdots + c_8 x^7$ fitting the (x_i, y_i).

A degree 7 polynomial is being fit to 11 data points lying on the degree 7 polynomial $P(x) = 1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7$. Obviously, the correct least squares solution is $c_1 = c_2 = \cdots = c_8 = 1$. Substituting the data points into the model $P(x)$ yields the system $Ac = b$:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^7 \\ 1 & x_2 & x_2^2 & \cdots & x_2^7 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^7 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{11} \end{bmatrix}.$$

The coefficient matrix A is a **Van der Monde matrix**, a matrix whose jth column consists of the elements of the second column raised to the $(j-1)$st power. We use MATLAB to solve the normal equations:

```
>> x = (2+(0:10)/5)';
>> y = 1+x+x.^2+x.^3+x.^4+x.^5+x.^6+x.^7;
>> A = [x.^0 x x.^2 x.^3 x.^4 x.^5 x.^6 x.^7];
>> c = (A'*A)\(A'*y)

c=
    1.5134
   -0.2644
    2.3211
    0.2408
    1.2592
    0.9474
    1.0059
    0.9997

>> cond(A'*A)

ans=
  1.4359e+019
```

Solving the normal equations in double precision cannot deliver an accurate value for the least squares solution. The condition number of $A^T A$ is too large to deal with in double precision arithmetic, and the normal equations are ill-conditioned, even though the original least squares problem is moderately conditioned. There is clearly room for improvement in the normal equations approach to least squares. In Example 4.15, we revisit this problem after developing an alternative that avoids forming $A^T A$. ◀

4.1 Exercises

1. Solve the normal equations to find the least squares solution and 2-norm error for the following inconsistent systems:

$$\text{(a)} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 2 \end{bmatrix}$$

2. Find the least squares solutions and RMSE of the following systems:

$$\text{(a)} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 2 \end{bmatrix}$$

3. Find the least squares solution of the inconsistent system

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}.$$

4. Let $m \geq n$, let A be the $m \times n$ identity matrix (the principal submatrix of the $m \times m$ identity matrix), and let $b = [b_1, \ldots, b_m]$ be a vector. Find the least squares solution of $Ax = b$ and the 2-norm error.

5. Prove that the 2-norm is a vector norm. You will need to use the Cauchy–Schwarz inequality $|u \cdot v| \leq ||u||_2||v||_2$.

6. Let A be an $n \times n$ nonsingular matrix. (a) Prove that $(A^T)^{-1} = (A^{-1})^T$. (b) Let b be an n-vector; then $Ax = b$ has exactly one solution. Prove that this solution satisfies the normal equations.

7. Find the best line through the set of data points, and find the RMSE:
(a) $(-3, 3), (-1, 2), (0, 1), (1, -1), (3, -4)$ (b) $(1, 1), (1, 2), (2, 2), (2, 3), (4, 3)$.

8. Find the best line through each set of data points, and find the RMSE:
(a) $(0, 0), (1, 3), (2, 3), (5, 6)$ (b) $(1, 2), (3, 2), (4, 1), (6, 3)$ (c) $(0, 5), (1, 3), (2, 3), (3, 1)$.

9. Find the best parabola through each data point set in Exercise 8, and compare the RMSE with the best-line fit.

10. Find the best degree 3 polynomial through each set in Exercise 8. Also, find the degree 3 interpolating polynomial, and compare.

11. Assume that the height of a model rocket is measured at four times, and the measured times and heights are $(t, h) = (1, 135), (2, 265), (3, 385), (4, 485)$, in seconds and meters. Fit the model $h = a + bt - 4.905t^2$ to estimate the eventual maximum height of the object and when it will return to earth.

12. Given data points $(x, y, z) = (0, 0, 3), (0, 1, 2), (1, 0, 3), (1, 1, 5), (1, 2, 6)$, find the plane in three dimensions (model $z = c_0 + c_1x + c_2y$) that best fits the data.

4.1 Computer Problems

1. Form the normal equations, and compute the least squares solution and 2-norm error for the following inconsistent systems:

$$\text{(a)}\quad \begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix} \qquad \text{(b)}\quad \begin{bmatrix} 4 & 2 & 3 & 0 \\ -2 & 3 & -1 & 1 \\ 1 & 3 & -4 & 2 \\ 1 & 0 & 1 & -1 \\ 3 & 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 2 \\ 0 \\ 5 \end{bmatrix}$$

2. Consider the world oil production data of Computer Problem 3.2.3. Find the best least squares (a) line, (b) parabola, and (c) cubic curve through the 10 data points and the RMSE of the fits. Use each to estimate the 2010 production level. Which fit best represents the data in terms of RMSE?

3. Consider the world population data of Computer Problem 3.1.1. Find the best least squares (a) line, (b) parabola through the data points, and the RMSE of the fit. In each case, estimate the 1980 population. Which fit gives the best estimate?

4. Consider the carbon dioxide concentration data of Exercise 3.1.13. Find the best least squares (a) line, (b) parabola, and (c) cubic curve through the data points and the RMSE of the fit. In each case, estimate the 1950 CO_2 concentration.

5. A company test-markets a new soft drink in 22 cities of approximately equal size. The selling price (in dollars) and the number sold per week in the cities are listed as follows:

city	price	sales/week
1	0.59	3980
2	0.80	2200
3	0.95	1850
4	0.45	6100
5	0.79	2100
6	0.99	1700
7	0.90	2000
8	0.65	4200
9	0.79	2440
10	0.69	3300
11	0.79	2300

city	price	sales/week
12	0.49	6000
13	1.09	1190
14	0.95	1960
15	0.79	2760
16	0.65	4330
17	0.45	6960
18	0.60	4160
19	0.89	1990
20	0.79	2860
21	0.99	1920
22	0.85	2160

(a) First, the company wants to find the "demand curve": how many it will sell at each potential price. Let P denote price and S denote sales per week. Find the line $S = c_1 + c_2 P$ that best fits the data from the table in the sense of least squares. Find the normal equations and the coefficients c_1 and c_2 of the least squares line. Plot the least squares line along with the data, and calculate the root mean square error.

(b) After studying the results of the test marketing, the company will set a single selling price P throughout the country. Given a manufacturing cost of \$0.23 per unit, the total profit (per city, per week) is $S(P - 0.23)$ dollars. Use the results of the preceding least squares approximation to find the selling price for which the company's profit will be maximized.

6. What is the "slope" of the parabola $y = x^2$ on $[0, 1]$? Find the best least squares line that fits the parabola at n evenly spaced points in the interval for (a) $n = 10$ and (b) $n = 20$. Plot the

parabola and the lines. What do you expect the result to be as $n \to \infty$? (c) Find the minimum of the function $F(c_1, c_2) = \int_0^1 (x^2 - c_1 - c_2 x)^2 \, dx$, and explain its relation to the problem.

7. Find the least squares (a) line (b) parabola through the 13 data points of Figure 3.5 and the RMSE of each fit.

8. Let A be the $10 \times n$ matrix formed by the first n columns of the 10×10 Hilbert matrix. Let c be the n-vector $[1, \ldots, 1]$, and set $b = Ac$. Use the normal equations to solve the least squares problem $Ax = b$ for (a) $n = 6$ (b) $n = 8$, and compare with the correct least squares solution $\overline{x} = c$. How many correct decimal places can be computed? Use condition number to explain the results. (This least squares problem is revisited in Computer Problem 4.3.7.)

9. Let $x_1, \ldots, x_{11}$ be 11 evenly spaced points in $[2, 4]$ and $y_i = 1 + x_i + x_i^2 + \cdots + x_i^d$. Use the normal equations to compute the best degree d polynomial, where (a) $d = 5$ (b) $d = 6$ (c) $d = 8$. Compare with Example 4.5. How many correct decimal places of the coefficients can be computed? Use condition number to explain the results. (This least squares problem is revisited in Computer Problem 4.3.8.)

10. The following data, collected by US Bureau of Economic Analysis, lists the year-over-year percent change in mean disposable personal income in the United States during 15 election years. Also, the proportion of the U.S. electorate that voted for the incumbent party's presidential candidate is listed. The first line of the table says that income increased by 1.49% from 1951 to 1952, and that 44.6% of the electorate voted for Adlai Stevenson, the incumbent Democratic party's candidate for president. Find the best least squares linear model for incumbent party vote as a function of income change. Plot this line along with the 15 data points. How many percentage points of vote can the incumbent party expect for each additional percent of change in personal income?

year	% income change	% incumbent vote
1952	1.49	44.6
1956	3.03	57.8
1960	0.57	49.9
1964	5.74	61.3
1968	3.51	49.6
1972	3.73	61.8
1976	2.98	49.0
1980	−0.18	44.7
1984	6.23	59.2
1988	3.38	53.9
1992	2.15	46.5
1996	2.10	54.7
2000	3.93	50.3
2004	2.47	51.2
2008	−0.41	45.7

4.2 A SURVEY OF MODELS

The previous linear and polynomial models illustrate the use of least squares to fit data. The art of data modeling includes a wide variety of models, some derived from physical principles underlying the source of the data and others based on empirical factors.

4.2.1 Periodic data

Periodic data calls for periodic models. Outside air temperatures, for example, obey cycles on numerous timescales, including daily and yearly cycles governed by the rotation of the earth and the revolution of the earth around the sun. As a first example, hourly temperature data are fit to sines and cosines.

▶ EXAMPLE 4.6 Fit the recorded temperatures in Washington, D.C., on January 1, 2001, as listed in the following table, to a periodic model:

time of day	t	temp (C)
12 mid.	0	-2.2
3 am	$\frac{1}{8}$	-2.8
6 am	$\frac{1}{4}$	-6.1
9 am	$\frac{3}{8}$	-3.9
12 noon	$\frac{1}{2}$	0.0
3 pm	$\frac{5}{8}$	1.1
6 pm	$\frac{3}{4}$	-0.6
9 pm	$\frac{7}{8}$	-1.1

We choose the model $y = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t$ to match the fact that temperature is roughly periodic with a period of 24 hours, at least in the absence of longer-term temperature movements. The model uses this information by fixing the period to be exactly one day, where we are using days for the t units. The variable t is listed in these units in the table.

Substituting the data into the model results in the following overdetermined system of linear equations:

$$\begin{aligned}
c_1 + c_2 \cos 2\pi(0) + c_3 \sin 2\pi(0) &= -2.2 \\
c_1 + c_2 \cos 2\pi\left(\frac{1}{8}\right) + c_3 \sin 2\pi\left(\frac{1}{8}\right) &= -2.8 \\
c_1 + c_2 \cos 2\pi\left(\frac{1}{4}\right) + c_3 \sin 2\pi\left(\frac{1}{4}\right) &= -6.1 \\
c_1 + c_2 \cos 2\pi\left(\frac{3}{8}\right) + c_3 \sin 2\pi\left(\frac{3}{8}\right) &= -3.9 \\
c_1 + c_2 \cos 2\pi\left(\frac{1}{2}\right) + c_3 \sin 2\pi\left(\frac{1}{2}\right) &= 0.0 \\
c_1 + c_2 \cos 2\pi\left(\frac{5}{8}\right) + c_3 \sin 2\pi\left(\frac{5}{8}\right) &= 1.1 \\
c_1 + c_2 \cos 2\pi\left(\frac{3}{4}\right) + c_3 \sin 2\pi\left(\frac{3}{4}\right) &= -0.6 \\
c_1 + c_2 \cos 2\pi\left(\frac{7}{8}\right) + c_3 \sin 2\pi\left(\frac{7}{8}\right) &= -1.1
\end{aligned}$$

SPOTLIGHT ON

Orthogonality The least squares problem can be simplified considerably by special choices of basis functions. The choices in Examples 4.6 and 4.7, for instance, yield normal equations already in diagonal form. This property of orthogonal basis functions is explored in detail in Chapter 10. Model (4.9) is a Fourier expansion.

The corresponding inconsistent matrix equation is $Ax = b$, where

$$A = \begin{bmatrix} 1 & \cos 0 & \sin 0 \\ 1 & \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ 1 & \cos\frac{\pi}{2} & \sin\frac{\pi}{2} \\ 1 & \cos\frac{3\pi}{4} & \sin\frac{3\pi}{4} \\ 1 & \cos\pi & \sin\pi \\ 1 & \cos\frac{5\pi}{4} & \sin\frac{5\pi}{4} \\ 1 & \cos\frac{3\pi}{2} & \sin\frac{3\pi}{2} \\ 1 & \cos\frac{7\pi}{4} & \sin\frac{7\pi}{4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \sqrt{2}/2 & \sqrt{2}/2 \\ 1 & 0 & 1 \\ 1 & -\sqrt{2}/2 & \sqrt{2}/2 \\ 1 & -1 & 0 \\ 1 & -\sqrt{2}/2 & -\sqrt{2}/2 \\ 1 & 0 & -1 \\ 1 & \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} -2.2 \\ -2.8 \\ -6.1 \\ -3.9 \\ 0.0 \\ 1.1 \\ -0.6 \\ -1.1 \end{bmatrix}.$$

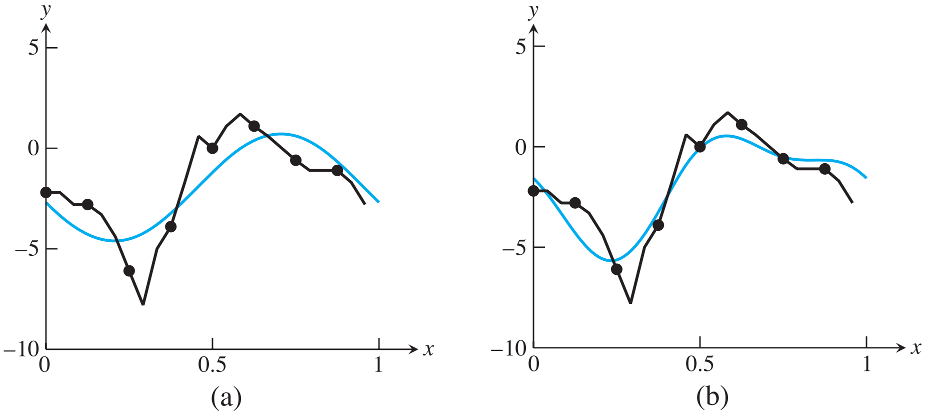
The normal equations $A^T Ac = A^T b$ are

$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \end{bmatrix},$$

which are easily solved as $c_1 = -1.95$, $c_2 = -0.7445$, and $c_3 = -2.5594$. The best version of the model, in the sense of least squares, is $y = -1.9500 - 0.7445\cos 2\pi t - 2.5594\sin 2\pi t$, with RMSE ≈ 1.063. Figure 4.5(a) compares the least squares fit model with the actual hourly recorded temperatures. ◀

▶ **EXAMPLE 4.7** Fit the temperature data to the improved model

$$y = c_1 + c_2\cos 2\pi t + c_3 \sin 2\pi t + c_4 \cos 4\pi t. \tag{4.9}$$



(a) (b)

Figure 4.5 Least Squares Fits to Periodic Data in Example 4.6. (a) Sinusoid model $y = -1.95 - 0.7445\cos 2\pi t - 2.5594\sin 2\pi t$ shown in bold, along with recorded temperature trace on Jan 1, 2001. (b) Improved sinusoid $y = -1.95 - 0.7445\cos 2\pi t - 2.5594\sin 2\pi t + 1.125\cos 4\pi t$ fits the data more closely.

The system of equations is now

$$\begin{aligned}
c_1 + c_2\cos 2\pi(0) + c_3\sin 2\pi(0) + c_4\cos 4\pi(0) &= -2.2\\
c_1 + c_2\cos 2\pi\left(\frac{1}{8}\right) + c_3\sin 2\pi\left(\frac{1}{8}\right) + c_4\cos 4\pi\left(\frac{1}{8}\right) &= -2.8\\
c_1 + c_2\cos 2\pi\left(\frac{1}{4}\right) + c_3\sin 2\pi\left(\frac{1}{4}\right) + c_4\cos 4\pi\left(\frac{1}{4}\right) &= -6.1\\
c_1 + c_2\cos 2\pi\left(\frac{3}{8}\right) + c_3\sin 2\pi\left(\frac{3}{8}\right) + c_4\cos 4\pi\left(\frac{3}{8}\right) &= -3.9\\
c_1 + c_2\cos 2\pi\left(\frac{1}{2}\right) + c_3\sin 2\pi\left(\frac{1}{2}\right) + c_4\cos 4\pi\left(\frac{1}{2}\right) &= 0.0\\
c_1 + c_2\cos 2\pi\left(\frac{5}{8}\right) + c_3\sin 2\pi\left(\frac{5}{8}\right) + c_4\cos 4\pi\left(\frac{5}{8}\right) &= 1.1\\
c_1 + c_2\cos 2\pi\left(\frac{3}{4}\right) + c_3\sin 2\pi\left(\frac{3}{4}\right) + c_4\cos 4\pi\left(\frac{3}{4}\right) &= -0.6\\
c_1 + c_2\cos 2\pi\left(\frac{7}{8}\right) + c_3\sin 2\pi\left(\frac{7}{8}\right) + c_4\cos 4\pi\left(\frac{7}{8}\right) &= -1.1,
\end{aligned}$$

leading to the following normal equations:

$$\begin{bmatrix} 8 & 0 & 0 & 0\\ 0 & 4 & 0 & 0\\ 0 & 0 & 4 & 0\\ 0 & 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} c_1\\ c_2\\ c_3\\ c_4 \end{bmatrix} = \begin{bmatrix} -15.6\\ -2.9778\\ -10.2376\\ 4.5 \end{bmatrix}.$$

The solutions are $c_1 = -1.95$, $c_2 = -0.7445, c_3 = -2.5594$, and $c_4 = 1.125$, with RMSE ≈ 0.705. Figure 4.5(b) shows that the extended model $y = -1.95 - 0.7445\cos 2\pi t - 2.5594\sin 2\pi t + 1.125\cos 4\pi t$ substantially improves the fit. ◀

4.2.2 Data linearization

Exponential growth of a population is implied when its rate of change is proportional to its size. Under perfect conditions, when the growth environment is unchanging and when the population is well below the carrying capacity of the environment, the model is a good representation.

The **exponential model**

$$y = c_1 e^{c_2 t} \tag{4.10}$$

cannot be directly fit by least squares because c_2 does not appear linearly in the model equation. Once the data points are substituted into the model, the difficulty is clear: The set of equations to solve for the coefficients are nonlinear and cannot be expressed as a linear system $Ax = b$. Therefore, our derivation of the normal equations is irrelevant.

There are two ways to deal with the problem of nonlinear coefficients. The more difficult way is to directly minimize the least square error, that is, solve the nonlinear least squares problem. We return to this problem in Section 4.5. The simpler way is to change the problem. Instead of solving the original least squares problem, we can solve a different problem, which is related to the original, by "linearizing" the model.

In the case of the exponential model (4.10), the model is linearized by applying the natural logarithm:

$$\ln y = \ln(c_1 e^{c_2 t}) = \ln c_1 + c_2 t. \tag{4.11}$$

Note that for an exponential model, the graph of $\ln y$ is a linear plot in t. At first glance, it appears that we have only traded one problem for another. The c_2 coefficient is now linear in the model, but c_1 no longer is. However, by renaming $k = \ln c_1$, we can write

$$\ln y = k + c_2 t. \tag{4.12}$$

Now both coefficients k and c_2 are linear in the model. After solving the normal equations for the best k and c_2, we can find the corresponding $c_1 = e^k$ if we wish.

It should be noted that our way out of the difficulty of nonlinear coefficients was to change the problem. The original least squares problem we posed was to fit the data to (4.10)—that is, to find c_1, c_2 that minimize

$$(c_1 e^{c_2 t_1} - y_1)^2 + \cdots + (c_1 e^{c_2 t_m} - y_m)^2, \tag{4.13}$$

the sum of squares of the residuals of the equations $c_1 e^{c_2 t_i} = y_i$ for $i = 1, \ldots, m$. For now, we solve the revised problem minimizing least squares error in "log space"—that is, by finding c_1, c_2 that minimizes

$$(\ln c_1 + c_2 t_1 - \ln y_1)^2 + \cdots + (\ln c_1 + c_2 t_m - \ln y_m)^2, \tag{4.14}$$

the sum of squares of the residuals of the equations $\ln c_1 + c_2 t_i = \ln y_i$ for $i = 1, \ldots, m$. These are two different minimizations and have different solutions, meaning that they generally result in different values of the coefficients c_1, c_2.

Which method is correct for this problem, the nonlinear least squares of (4.13) or the model-linearized version (4.14)? The former is least squares, as we have defined it. The latter is not. However, depending on the context of the data, either may be the more natural choice. To answer the question, the user needs to decide which errors are most important to minimize, the errors in the original sense or the errors in "log space." In fact, the log model is linear, and it may be argued that only after log-transforming the data to a linear relation is it natural to evaluate the fitness of the model.

▶ EXAMPLE 4.8

Use model linearization to find the best least squares exponential fit $y = c_1 e^{c_2 t}$ to the following world automobile supply data:

year	cars ($\times 10^6$)
1950	53.05
1955	73.04
1960	98.31
1965	139.78
1970	193.48
1975	260.20
1980	320.39

The data describe the number of automobiles operating throughout the world in the given year. Define the time variable t in terms of years since 1950. Solving the linear least squares problem yields $k_1 \approx 3.9896$, $c_2 \approx 0.06152$. Since $c_1 \approx e^{3.9896} \approx 54.03$, the model

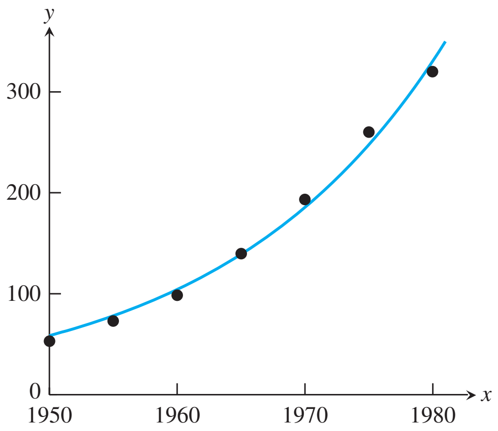


Figure 4.6 Exponential fit of world automobile supply data, using linearization. The best least squares fit is $y = 54.03e^{0.06152t}$. Compare with Figure 4.14.

is $y = 54.03e^{0.06152t}$. The RMSE of the log-linearized model in log space is ≈ 0.0357, while RMSE of the original exponential model is ≈ 9.56. The best model and data are plotted in Figure 4.6. ◀

▶ **EXAMPLE 4.9** The number of transistors on Intel central processing units since the early 1970s is given in the table that follows. Fit the model $y = c_1 e^{c_2 t}$ to the data.

CPU	year	transistors
4004	1971	2,250
8008	1972	2,500
8080	1974	5,000
8086	1978	29,000
286	1982	120,000
386	1985	275,000
486	1989	1,180,000
Pentium	1993	3,100,000
Pentium II	1997	7,500,000
Pentium III	1999	24,000,000
Pentium 4	2000	42,000,000
Itanium	2002	220,000,000
Itanium 2	2003	410,000,000

Parameters will be fit by using model linearization (4.11). Linearizing the model gives

$$\ln y = k + c_2 t.$$

We will let $t = 0$ correspond to the year 1970. Substituting the data into the linearized model yields

$$\begin{aligned} k + c_2(1) &= \ln 2250 \\ k + c_2(2) &= \ln 2500 \\ k + c_2(4) &= \ln 5000 \\ k + c_2(8) &= \ln 29000, \end{aligned} \qquad (4.15)$$

and so forth. The matrix equation is $Ax = b$, where $x = (k, c_2)$,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ \vdots & \vdots \\ 1 & 33 \end{bmatrix}, \text{ and } \quad b = \begin{bmatrix} \ln 2250 \\ \ln 2500 \\ \ln 5000 \\ \ln 29000 \\ \vdots \\ \ln 410000000 \end{bmatrix}. \tag{4.16}$$

The normal equations $A^T Ax = A^T b$ are

$$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} \begin{bmatrix} k \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix},$$

which has solution $k \approx 7.197$ and $c_2 \approx 0.3546$, leading to $c_1 = e^k \approx 1335.3$. The exponential curve $y = 1335.3e^{0.3546t}$ is shown in Figure 4.7 along with the data. The doubling time for the law is $\ln 2/c_2 \approx 1.95$ years. Gordon C. Moore, cofounder of Intel, predicted in 1965 that over the ensuing decade, computing power would double every 2 years. Astoundingly, that exponential rate has continued for 40 years. There is some evidence in Figure 4.7 that this rate has accelerated since 2000.

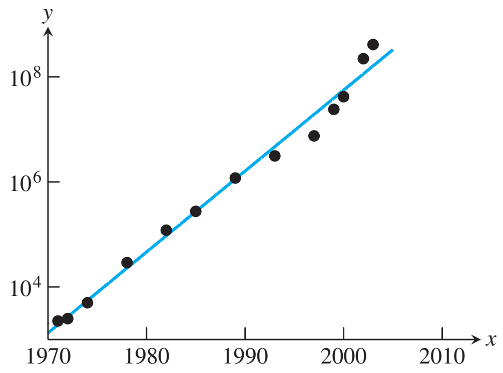


Figure 4.7 Semilog Plot of Moore's Law. Number of transistors on CPU chip versus year.

◀

Another important example with nonlinear coefficients is the **power law** model $y = c_1 t^{c_2}$. This model also can be simplified with linearization by taking logs of both sides:

$$\begin{aligned} \ln y &= \ln c_1 + c_2 \ln t \\ &= k + c_2 \ln t. \end{aligned} \tag{4.17}$$

Substitution of data into the model will give

$$k + c_2 \ln t_1 = \ln y_1 \tag{4.18}$$

$$\vdots$$

$$k + c_2 \ln t_n = \ln y_n, \tag{4.19}$$

resulting in the matrix form

$$A = \begin{bmatrix} 1 & \ln t_1 \\ \vdots & \vdots \\ 1 & \ln t_n \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} \ln y_1 \\ \vdots \\ \ln y_n \end{bmatrix}. \tag{4.20}$$

The normal equations allow determination of k and c_2, and $c_1 = e^k$.

▸ **EXAMPLE 4.10** Use linearization to fit the given height–weight data with a power law model.

The mean height and weight of boys ages 2–11 were collected in the U.S. National Health and Nutrition Examination Survey by the Centers for Disease Control (CDC) in 2002, resulting in the following table:

age (yrs.)	height (m)	weight (kg)
2	0.9120	13.7
3	0.9860	15.9
4	1.0600	18.5
5	1.1300	21.3
6	1.1900	23.5
7	1.2600	27.2
8	1.3200	32.7
9	1.3800	36.0
10	1.4100	38.6
11	1.4900	43.7

Following the preceding strategy, the resulting power law for weight versus height is $W = 16.3H^{2.42}$. The relationship is graphed in Figure 4.8. Since weight is a proxy for volume, the coefficient $c_2 \approx 2.42$ can be viewed as the "effective dimension" of the human body.

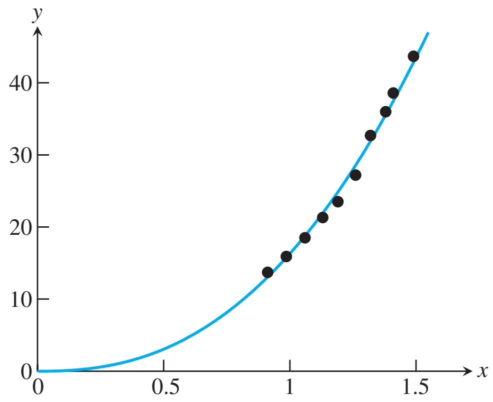


Figure 4.8 Power law of weight versus height for 2–11-year-olds. The best fit formula is $W = 16.3H^{2.42}$.

◂

The time course of drug concentration y in the bloodstream is well described by

$$y = c_1 t e^{c_2 t}, \tag{4.21}$$

where t denotes time after the drug was administered. The characteristics of the model are a quick rise as the drug enters the bloodstream, followed by slow exponential decay. The **half-life** of the drug is the time from the peak concentration to the time it drops to half that level. The model can be linearized by applying the natural logarithm to both sides, producing

$$\begin{aligned} \ln y &= \ln c_1 + \ln t + c_2 t \\ k + c_2 t &= \ln y - \ln t, \end{aligned}$$

where we have set $k = \ln c_1$. This leads to the matrix equation $Ax = b$, where

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} \ln y_1 - \ln t_1 \\ \vdots \\ \ln y_m - \ln t_m \end{bmatrix}. \tag{4.22}$$

The normal equations are solved for k and c_2, and $c_1 = e^k$.

▶ EXAMPLE 4.11 Fit the model (4.21) with the measured level of the drug norfluoxetine in a patient's bloodstream, given in the following table:

hour	concentration (ng/ml)
1	8.0
2	12.3
3	15.5
4	16.8
5	17.1
6	15.8
7	15.2
8	14.0

Solving the normal equations yields $k \approx 2.28$ and $c_2 \approx -0.215$, and $c_1 \approx e^{2.28} \approx 9.77$. The best version of the model is $y = 9.77te^{-0.215t}$, plotted in Figure 4.9. From the model, the timing of the peak concentration and the half-life can be estimated. (See Computer Problem 5.)

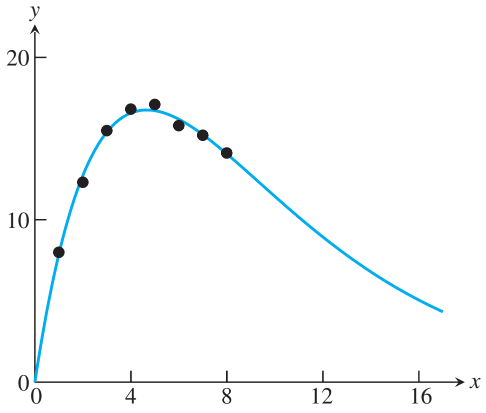


Figure 4.9 Plot of drug concentration in blood. Model (4.21) shows exponential decay after initial peak.

◀

It is important to realize that model linearization changes the least squares problem. The solution obtained will minimize the RMSE with respect to the linearized problem, not necessarily the original problem, which in general will have a different set of optimal parameters. If they enter the model nonlinearly, they cannot be computed from the normal equations, and we need nonlinear techniques to solve the original least squares problem. This is done in the Gauss–Newton Method in Section 4.5, where we revisit the automobile supply data and compare fitting the exponential model in linearized and nonlinearized forms.

4.2 Exercises

1. Fit data to the periodic model $y = F_3(t) = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t$. Find the 2-norm error and the RMSE.

(a)

t	y
0	1
1/4	3
1/2	2
3/4	0

(b)

t	y
0	1
1/4	3
1/2	2
3/4	1

(c)

t	y
0	3
1/2	1
1	3
3/2	2

2. Fit the data to the periodic models $F_3(t) = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t$ and $F_4(t) = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t + c_4 \cos 4\pi t$. Find the 2-norm errors $||e||_2$ and compare the fits of F_3 and F_4.

(a)

t	y
0	0
1/6	2
1/3	0
1/2	−1
2/3	1
5/6	1

(b)

t	y
0	4
1/6	2
1/3	0
1/2	−5
2/3	−1
5/6	3

3. Fit data to the exponential model by using linearization. Find the 2-norm of the difference between the data points y_i and the best model $c_1 e^{c_2 t_i}$.

(a)

t	y
−2	1
0	2
1	2
2	5

(b)

t	y
0	1
1	1
1	2
2	4

4. Fit data to the exponential model by using linearization. Find the 2-norm of the difference between the data points y_i and the best model $c_1 e^{c_2 t_i}$.

(a)

t	y
−2	4
−1	2
1	1
2	1/2

(b)

t	y
0	10
1	5
2	2
3	1

5. Fit data to the power law model by using linearization. Find the RMSE of the fit.

(a)

t	y
1	6
2	2
3	1
4	1

(b)

t	y
1	2
1	4
2	5
3	6
5	10

6. Fit data to the drug concentration model (4.21). Find the RMSE of the fit.

(a)

t	y
1	3
2	4
3	5
4	5

(b)

t	y
1	2
2	4
3	3
4	2

4.2 Computer Problems

1. Fit the monthly data for Japan 2003 oil consumption, shown in the following table, with the periodic model (4.9), and calculate the RMSE:

month	oil use (10^6 bbl/day)
Jan	6.224
Feb	6.665
Mar	6.241
Apr	5.302
May	5.073
Jun	5.127
Jul	4.994
Aug	5.012
Sep	5.108
Oct	5.377
Nov	5.510
Dec	6.372

2. The temperature data in Example 4.6 was taken from the Weather Underground website `www.wunderground.com`. Find a similar selection of hourly temperature data from a location and date of your choice, and fit it with the two sinusoidal models of the example.

3. Consider the world population data of Computer Problem 3.1.1. Find the best exponential fit of the data points by using linearization. Estimate the 1980 population, and find the estimation error.

4. Consider the carbon dioxide concentration data of Exercise 3.1.17. Find the best exponential fit of the difference between the CO_2 level and the background (279 ppm) by using linearization. Estimate the 1950 CO_2 concentration, and find the estimation error.

5. (a) Find the time at which the maximum concentration is reached in model (4.21). (b) Use an equation solver to estimate the half-life from the model in Example 4.11.

6. The bloodstream concentration of a drug, measured hourly after administration, is given in the accompanying table. Fit the model (4.21). Find the estimated maximum and the half-life. Suppose that the therapeutic range for the drug is 4–15 ng/ml. Use the equation solver of your choice to estimate the time the drug concentration stays within therapeutic levels.

hour	concentration (ng/ml)
1	6.2
2	9.5
3	12.3
4	13.9
5	14.6
6	13.5
7	13.3
8	12.7
9	12.4
10	11.9

7. The file `windmill.txt`, available from the textbook website, is a list of 60 numbers which represent the monthly megawatt-hours generated from Jan. 2005 to Dec. 2009 by a wind turbine owned by the Minnkota Power Cooperative near Valley City, ND. The data is currently available at http://www.minnkota.com. For reference, a typical home uses around 1 MWh per month.

   (a) Find a rough model of power output as a yearly periodic function. Fit the data to equation (4.9),

   $$f(t) = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t + c_4 \cos 4\pi t$$

   where the units of t are years, that is $0 \le t \le 5$, and write down the resulting function.

   (b) Plot the data and the model function for years $0 \le t \le 5$. What features of the data are captured by the model?

8. The file `scrippsy.txt`, available from the textbook website, is a list of 50 numbers which represent the concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded at Mauna Loa, Hawaii, each May 15 of the years 1961 to 2010. The data is part of a data collection effort initiated by Charles Keeling of the Scripps Oceanographic Institute (Keeling et al. [2001]). Subtract the background level 279 ppm as in Computer Problem 4, and fit the data to an exponential model. Plot the data along with the best fit exponential function, and report the RMSE.

9. The file `scrippsm.txt`, available from the textbook website, is a list of 180 numbers which represent the concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded monthly at Mauna Loa from Jan. 1996 to Dec. 2010, taken from the same Scripps study as Computer Problem 8.

   (a) Carry out a least squares fit of the CO_2 data using the model

   $$f(t) = c_1 + c_2 t + c_3 \cos 2\pi t + c_4 \sin 2\pi t$$

   where t is measured in months. Report the best fit coefficients c_i and the RMSE of the fit. Plot the continuous curve from Jan. 1989 to the end of this year, including the 180 data points in the plot.

   (b) Use your model to predict the CO_2 concentration in May 2004, Sept. 2004, May 2005, and Sept. 2005. These months tend to contain the yearly maxima and minima of the CO_2 cycle. The actual recorded values are 380.63, 374.06, 382.45, and 376.73 ppv, respectively. Report the model error at these four points.

   (c) Add the extra term $c_5 \cos 4\pi t$ and redo parts (a) and (b). Compare the new RMSE and four model errors.

(d) Repeat part (c) using the extra term $c_5 t^2$. Which term leads to more improvement in the model, part (c) or (d)?

(e) Add both terms from (c) and (d) and redo parts (a) and (b). Prepare a table summarizing your results from all parts of the problem, and try to provide an explanation for the results.

See the website `http://scrippsco2.ucsd.edu` for much more data and analysis of the Scripps carbon dioxide study.

4.3 QR FACTORIZATION

In Chapter 2, the LU factorization was used to solve matrix equations. The factorization is useful because it encodes the steps of Gaussian elimination. In this section, we develop the QR factorization as a way to solve least squares calculations that is superior to the normal equations.

After introducing the factorization by way of Gram–Schmidt orthogonalization, we return to Example 4.5, for which the normal equations turned out to be inadequate. Later in this section, Householder reflections are introduced as a more efficient method of computing Q and R.

4.3.1 Gram–Schmidt orthogonalization and least squares

The Gram–Schmidt method orthogonalizes a set of vectors. Given an input set of m-dimensional vectors, the goal is to find an orthogonal coordinate system for the subspace spanned by the set. More precisely, given n linearly independent input vectors, it computes n mutually perpendicular unit vectors spanning the same subspace as the input vectors. The unit length is with respect to the Euclidean or 2-norm (4.7), which is used throughout Chapter 4.

Let $A_1, \ldots, A_n$ be linearly independent vectors from R^m. Thus $n \le m$. The Gram–Schmidt method begins by dividing A_1 by its length to make it a unit vector. Define

$$y_1 = A_1 \quad \text{and} \quad q_1 = \frac{y_1}{||y_1||_2}. \tag{4.23}$$

To find the second unit vector, subtract away the projection of A_2 in the direction of q_1, and normalize the result:

$$y_2 = A_2 - q_1(q_1^T A_2), \quad \text{and} \quad q_2 = \frac{y_2}{||y_2||_2}. \tag{4.24}$$

Then $q_1^T y_2 = q_1^T(A_2 - q_1(q_1^T A_2)) = q_1^T A_2 - q_1^T A_2 = 0$, so q_1 and q_2 are pairwise orthogonal, as shown in Figure 4.10.

At the jth step, define

$$y_j = A_j - q_1(q_1^T A_j) - q_2(q_2^T A_j) - \ldots - q_{j-1}(q_{j-1}^T A_j) \quad \text{and} \quad q_j = \frac{y_j}{||y_j||_2}. \tag{4.25}$$

It is clear that q_j is orthogonal to each of the previously produced q_i for $i = 1, \ldots, j-1$, since (4.25) implies

$$\begin{aligned} q_i^T y_j &= q_i^T A_j - q_i^T q_1 q_1^T A_j - \ldots - q_i^T q_{j-1} q_{j-1}^T A_j \\ &= q_i^T A_j - q_i^T q_i q_i^T A_j = 0, \end{aligned}$$

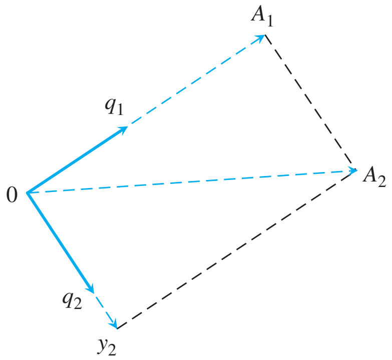


Figure 4.10 Gram–Schmidt orthogonalization. The input vectors are A_1 and A_2, and the output is the orthonormal set consisting of q_1 and q_2. The second orthogonal vector q_2 is formed by subtracting the projection of A_2 in the direction of q_1 from A_2, followed by normalizing.

where by induction hypothesis, the q_i are pairwise orthogonal for $i < j$. Geometrically, (4.25) corresponds to subtracting from A_j the projections of A_j onto the previously determined orthogonal vectors $q_i, i = 1, \ldots, j-1$. What remains is orthogonal to the q_i and, after dividing by its length to become a unit vector, is used as q_j. Therefore, the set $\{q_1, \ldots, q_n\}$ consists of mutually orthogonal vectors spanning the same subspace of R^m as $\{A_1, \ldots, A_n\}$.

The result of Gram–Schmidt orthogonalization can be put into matrix form by introducing new notation for the dot products in the above calculation. Define $r_{jj} = ||y_j||_2$ and $r_{ij} = q_i^T A_j$. Then (4.23) and (4.24) can be written

$$\begin{aligned} A_1 &= r_{11}q_1 \\ A_2 &= r_{12}q_1 + r_{22}q_2, \end{aligned}$$

and the general case (4.25) translates to

$$A_j = r_{1j}q_1 + \cdots + r_{j-1,j}q_{j-1} + r_{jj}q_j.$$

Therefore, the result of Gram–Schmidt orthogonalization can be written in matrix form as

$$(A_1|\cdots|A_n) = (q_1|\cdots|q_n)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}, \tag{4.26}$$

or $A = QR$, where we consider A to be the matrix consisting of the columns A_j. We call this the **reduced QR factorization**; the full version is just ahead. The assumption that the vectors A_j are linearly independent guarantees that the main diagonal coefficients r_{jj} are nonzero. Conversely, if A_j lies in the span of $A_1, \ldots, A_{j-1}$, then the projections onto the latter vectors make up the entire vector, and $r_{jj} = ||y_j||_2 = 0$.

▶ **EXAMPLE 4.12** Find the reduced QR factorization by applying Gram–Schmidt orthogonalization to the columns of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

Set $y_1 = A_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Then $r_{11} = ||y_1||_2 = \sqrt{1^2 + 2^2 + 2^2} = 3$, and the first unit vector is

$$q_1 = \frac{y_1}{||y_1||_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

To find the second unit vector, set

$$y_2 = A_2 - q_1 q_1^T A_2 = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} 2 = \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$$

and

$$q_2 = \frac{y_2}{||y_2||_2} = \frac{1}{5} \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}.$$

Since $r_{12} = q_1^T A_2 = 2$ and $r_{22} = ||y_2||_2 = 5$, the result written in matrix form (4.26) is

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 \\ 2/3 & 1/3 \\ 2/3 & 2/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = QR.$$

◀

We use the term "classical" for this version of Gram–Schmidt, since we will provide an upgraded, or "modified," version at the end of this section.

Classical Gram–Schmidt orthogonalization

```
      Let $A_j, j = 1, \ldots, n$ be linearly independent vectors.
for $j = 1, 2, \ldots, n$
    $y = A_j$
    for $i = 1, 2, \ldots, j - 1$
        $r_{ij} = q_i^T A_j$
        $y = y - r_{ij} q_i$
    end
    $r_{jj} = ||y||_2$
    $q_j = y / r_{jj}$
end
```

When the method is successful, it is customary to fill out the matrix of orthogonal unit vectors to a complete basis of R^m, to achieve the "full" QR factorization. This can be done, for example, by adding $m - n$ extra vectors to the A_j, so that the m vectors span R^m, and carrying out the Gram–Schmidt method. In terms of the basis of R^m formed by $q_1, \ldots, q_m$, the original vectors can be expressed as

$$(A_1|\cdots|A_n) = (q_1|\cdots|q_m)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{4.27}$$

This matrix equation is the **full QR factorization** of the matrix $A = (A_1|\cdots|A_n)$, formed by the original input vectors. Note the matrix sizes in the full QR factorization: A is $m \times n$, Q is a square $m \times m$ matrix, and the upper triangular matrix R is $m \times n$, the same size as A. The matrix Q in the full QR factorization has a special place in numerical analysis and is given a special definition.

DEFINITION 4.1 A square matrix Q is **orthogonal** if $Q^T = Q^{-1}$. ❐

Note that a square matrix is orthogonal if and only if its columns are pairwise orthogonal unit vectors (Exercise 9). Therefore, a full QR factorization is the equation $A = QR$, where Q is an orthogonal square matrix and R is an upper triangular matrix the same size as A.

The key property of an orthogonal matrix is that it preserves the Euclidean norm of a vector.

LEMMA 4.2 If Q is an orthogonal $m \times m$ matrix and x is an m-dimensional vector, then $||Qx||_2 = ||x||_2$. ■

Proof. $||Qx||_2^2 = (Qx)^T Qx = x^T Q^T Qx = x^T x = ||x||_2^2$. ❐

The product of two orthogonal $m \times m$ matrices is again orthogonal (Exercise 10). The QR factorization of an $m \times m$ matrix by the Gram–Schmidt method requires approximately m^3 multiplication/divisions, three times more than the LU factorization, plus about the same number of additions (Exercise 11).

▶ EXAMPLE 4.13 Find the full QR factorization of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

SPOTLIGHT ON

Orthogonality In Chapter 2, we found that the LU factorization is an efficient means of encoding the information of Gaussian elimination. In the same way, the QR factorization records the orthogonalization of a matrix, namely, the construction of an orthogonal set that spans the space of column vectors of A. Doing calculations with orthogonal matrices is preferable because (1) they are easy to invert by definition, and (2) by Lemma 4.2, they do not magnify errors.

In Example 4.12, we found the orthogonal unit vectors $q_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$ and $q_2 = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}$. Adding a third vector $A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ leads to

$$\begin{aligned} y_3 &= A_3 - q_1 q_1^T A_3 - q_2 q_2^T A_3 \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \frac{1}{3} - \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ -\frac{2}{15} \end{bmatrix} \left(-\frac{14}{15} \right) = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix} \end{aligned}$$

and $q_3 = y_3/||y_3|| = \begin{bmatrix} \frac{2}{15} \\ \frac{10}{15} \\ -\frac{11}{15} \end{bmatrix}$. Putting the parts together, we obtain the full QR factorization

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 & 2/15 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & 2/15 & -11/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = QR.$$

Note that the choice of A_3 was arbitrary. Any third column vector linearly independent of the first two columns could be used. Compare this result with the reduced QR factorization in Example 4.12. ◀

The MATLAB command `qr` carries out the QR factorization on an $m \times n$ matrix. It does not use Gram–Schmidt orthogonalization, but uses more efficient and stable methods that will be introduced in a later subsection. The command

```
>> [Q,R]=qr(A,0)
```

returns the reduced QR factorization, and

```
>> [Q,R]=qr(A)
```

returns the full QR factorization.

There are three major applications of the QR factorization. We will describe two of them here; the third is the QR algorithm for eigenvalue calculations, introduced in Chapter 12.

First, the QR factorization can be used to solve a system of n equations in n unknowns $Ax = b$. Just factor $A = QR$, and the equation $Ax = b$ becomes $QRx = b$ and $Rx = Q^T b$. Assuming that A is nonsingular, the diagonal entries of the upper triangular matrix R are nonzero, so that R is nonsingular. A triangular back substitution yields the solution x. As mentioned before, this approach is about three times more expensive in terms of complexity when compared with the LU approach.

The second application is to least squares. Let A be an $m \times n$ matrix with $m \geq n$. To minimize $||Ax - b||_2$, rewrite as $||QRx - b||_2 = ||Rx - Q^T b||_2$ by Lemma 4.2.

The vector inside the Euclidean norm is

$$\begin{bmatrix} e_1 \\ \vdots \\ e_n \\ \hdashline e_{n+1} \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ \hdashline 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} d_1 \\ \vdots \\ d_n \\ \hdashline d_{n+1} \\ \vdots \\ d_m \end{bmatrix} \tag{4.28}$$

where $d = Q^T b$. Assume that $r_{ii} \neq 0$. Then the upper part $(e_1, \ldots, e_n)$ of the error vector e can be made zero by back substitution. The choice of the x_i makes no difference for the lower part of the error vector; clearly, $(e_{n+1}, \ldots, e_m) = (-d_{n+1}, \ldots, -d_m)$. Therefore, the least squares solution is minimized by using the x from back-solving the upper part, and the least squares error is $||e||_2^2 = d_{n+1}^2 + \cdots + d_m^2$.

Least squares by QR factorization

Given the $m \times n$ inconsistent system

$$Ax = b,$$

find the full QR factorization $A = QR$ and set

$$\hat{R} = \text{upper } n \times n \text{ submatrix of } R$$
$$\hat{d} = \text{upper } n \text{ entries of } d = Q^T b$$

Solve $\hat{R}\overline{x} = \hat{d}$ for least squares solution $\overline{x}$.

▶ EXAMPLE 4.14 Use the full QR factorization to solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$.

We need to solve $Rx = Q^T b$, or

$$\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ \hdashline 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 5 & 10 & 10 \\ -14 & 5 & 2 \\ 2 & 10 & -11 \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ \hdashline 3 \end{bmatrix}.$$

The least squares error will be $||e||_2 = ||(0, 0, 3)||_2 = 3$. Equating the upper parts yields

$$\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \end{bmatrix},$$

whose solution is $\overline{x}_1 = 3.8, \overline{x}_2 = 1.8$. This least squares problem was solved by the normal equations in Example 4.2. ◀

Finally, we return to the problem in Example 4.5 that led to an ill-conditioned system of normal equations.

SPOTLIGHT ON

Conditioning In Chapter 2, we found that the best way to handle ill-conditioned problems is to avoid them. Example 4.15 is a classic case of that advice. While the normal equations of Example 4.5 are ill-conditioned, the QR approach solves least squares without constructing $A^T A$.

▶ EXAMPLE 4.15 Use the full QR factorization to solve the least squares problem of Example 4.5.

The normal equations were notably unsuccessful in solving this least squares problem of 11 equations in 8 variables. We use the MATLAB `qr` command to carry out an alternative approach:

```
>> x=(2+(0:10)/5)';
>> y=1+x+x.^2+x.^3+x.^4+x.^5+x.^6+x.^7;
>> A=[x.^0 x x.^2 x.^3 x.^4 x.^5 x.^6 x.^7];
>>  [Q,R]=qr(A);
>> b=Q'*y;
>> c=R(1:8,1:8)\b(1:8)

c=
    0.99999991014308
    1.00000021004107
    0.99999979186557
    1.00000011342980
    0.99999996325039
    1.00000000708455
    0.99999999924685
    1.00000000003409
```

Six decimal places of the correct solution $c = [1, \ldots, 1]$ are found by using QR factorization. This approach finds the least squares solution without forming the normal equations, which have a condition number of about 10^{19}. ◀

4.3.2 Modified Gram–Schmidt orthogonalization

A slight modification to Gram–Schmidt turns out to enhance its accuracy in machine calculations. The new algorithm called modified Gram–Schmidt is mathematically equivalent to the original, or "classical" Gram–Schmidt algorithm.

Modified Gram–Schmidt orthogonalization

Let A_j, $j = 1, \ldots, n$ be linearly independent vectors.

```
for $j = 1, 2, \ldots, n$
    $y = A_j$
    for $i = 1, 2, \ldots, j-1$
        $r_{ij} = q_i^T y$
        $y = y - r_{ij} q_i$
    end
    $r_{jj} = ||y||_2$
    $q_j = y / r_{jj}$
end
```

The only difference from classical Gram–Schmidt is that A_j is replaced by y in the innermost loop. Geometrically speaking, when projecting away the part of vector A_j in the direction of q_2, for example, one should subtract away the projection of the remainder y of A_j with the q_1 part already removed, instead of the projection of A_j itself on q_2. Modified Gram–Schmidt is the version that will be used in the GMRES algorithm in Section 4.4.

► **EXAMPLE 4.16** Compare the results of classical Gram–Schmidt and modified Gram–Schmidt, computed in double precision, on the matrix of almost-parallel vectors

$$\begin{bmatrix} 1 & 1 & 1 \\ \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta \end{bmatrix}$$

where $\delta = 10^{-10}$.

First, we apply classical Gram–Schmidt.

$$y_1 = A_1 = \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} \qquad \text{and} \qquad q_1 = \frac{1}{\sqrt{1+\delta^2}} \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix}.$$

Note that $\delta^2 = 10^{-20}$ is a perfectly acceptable double precision number, but $1 + \delta^2 = 1$ after rounding. Then

$$y_2 = \begin{bmatrix} 1 \\ 0 \\ \delta \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} q_1^T A_2 = \begin{bmatrix} 1 \\ 0 \\ \delta \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\delta \\ \delta \\ 0 \end{bmatrix} \quad \text{and} \quad q_2 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$$

after dividing by $||y_2||_2 = \sqrt{\delta^2 + \delta^2} = \sqrt{2}\delta$. Completing classical Gram–Schmidt,

$$y_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \delta \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} q_1^T A_3 - \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} q_2^T A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \delta \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\delta \\ 0 \\ \delta \end{bmatrix} \quad \text{and} \quad q_3 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Unfortunately, due to the double precision rounding done in the first step, q_2 and q_3 turn out to be not orthogonal:

$$q_2^T q_3 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{2}.$$

On the other hand, modified Gram–Schmidt does much better. While q_1 and q_2 are calculated the same way, q_3 is found as

$$y_3^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \delta \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} q_1^T A_3 = \begin{bmatrix} 0 \\ -\delta \\ 0 \\ \delta \end{bmatrix},$$

$$y_3 = y_3^1 - \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} q_2^T y_3^1 = \begin{bmatrix} 0 \\ -\delta \\ 0 \\ \delta \end{bmatrix} - \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \frac{\delta}{\sqrt{2}}$$

$$= \begin{bmatrix} 0 \\ -\frac{\delta}{2} \\ -\frac{\delta}{2} \\ \delta \end{bmatrix} \quad \text{and} \quad q_3 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}.$$

Now $q_2^T q_3 = 0$ as desired. Note that for both classical and modified Gram–Schmidt, $q_1^T q_2$ is on the order of δ, so even modified Gram–Schmidt leaves room for improvement. Orthogonalization by Householder reflectors, described in the next section, is widely considered to be more computationally stable. ◀

4.3.3 Householder reflectors

Although the modified Gram–Schmidt orthogonalization method is an improved way to calculate the QR factorization of a matrix, it is not the best way. An alternative method using Householder reflectors requires fewer operations and is more stable, in the sense of amplification of rounding errors. In this section, we will define the reflectors and show how they are used to factorize a matrix.

A Householder reflector is an orthogonal matrix that reflects all m-vectors through an $m-1$ dimensional plane. This means that the length of each vector is unchanged when multiplied by the matrix, making Householder reflectors ideal for moving vectors. Given a vector x that we would like to relocate to a vector w of equal length, the recipe for Householder reflectors gives a matrix H such that $Hx = w$.

The origin of the recipe is clear in Figure 4.11. Draw the $m-1$ dimensional plane bisecting x and w, and perpendicular to the vector connecting them. Then reflect all vectors through the plane.

LEMMA 4.3 Assume that x and w are vectors of the same Euclidean length, $||x||_2 = ||w||_2$. Then $w - x$ and $w + x$ are perpendicular. ■

Proof. $(w-x)^T(w+x) = w^Tw - x^Tw + w^Tx - x^Tx = ||w||^2 - ||x||^2 = 0.$ ❐

Define the vector $v = w - x$, and consider the projection matrix

$$P = \frac{vv^T}{v^Tv}. \tag{4.29}$$

A **projection matrix** is a matrix that satisfies $P^2 = P$. Exercise 13 asks the reader to verify that P in (4.29) is a symmetric projection matrix and that $Pv = v$. Geometrically, for any vector u, Pu is the projection of u onto v. Figure 4.11 hints that if we subtract twice the projection Px from x, we should get w. To verify this, set $H = I - 2P$. Then

$$\begin{aligned} Hx &= x - 2Px \\ &= w - v - \frac{2vv^Tx}{v^Tv} \\ &= w - v - \frac{vv^Tx}{v^Tv} - \frac{vv^T(w-v)}{v^Tv} \\ &= w - \frac{vv^T(w+x)}{v^Tv} \\ &= w, \end{aligned} \tag{4.30}$$

the latter equality following from Lemma 4.3, since $w + x$ is orthogonal to $v = w - x$.

The matrix H is called a **Householder reflector**. Note that H is a symmetric (Exercise 14) and orthogonal matrix, since

$$\begin{aligned} H^TH = HH &= (I-2P)(I-2P) \\ &= I - 4P + 4P^2 \\ &= I. \end{aligned}$$

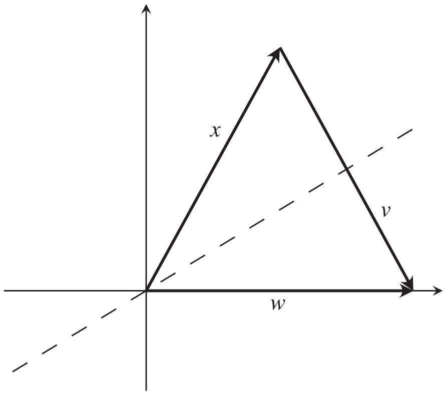


Figure 4.11 Householder reflector. Given equal length vectors x and w, reflection through the bisector of the angle between them (dotted line) exchanges them.

These facts are summarized in the following theorem:

THEOREM 4.4 **Householder reflectors.** Let x and w be vectors with $||x||_2 = ||w||_2$ and define $v = w - x$. Then $H = I - 2vv^T/v^Tv$ is a symmetric orthogonal matrix and $Hx = w$. ■

► **EXAMPLE 4.17** Let $x = [3, 4]$ and $w = [5, 0]$. Find a Householder reflector H that satisfies $Hx = w$.

Set

$$v = w - x = \begin{bmatrix} 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix},$$

and define the projection matrix

$$P = \frac{vv^T}{v^Tv} = \frac{1}{20}\begin{bmatrix} 4 & -8 \\ -8 & 16 \end{bmatrix} = \begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix}.$$

Then

$$H = I - 2P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.4 & -0.8 \\ -0.8 & 1.6 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}.$$

Check that H moves x to w and vice versa:

$$Hx = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} = w$$

and

$$Hw = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}\begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = x.$$ ◄

As a first application of Householder reflectors, we will develop a new way to do the QR factorization. In Chapter 12, we apply Householder to the eigenvalue problem, to put matrices into upper Hessenberg form. In both applications, we will use reflectors for a single purpose: to move a column vector x to a coordinate axis as a way of putting zeros into a matrix.

We start with a matrix A that we want to write in the form $A = QR$. Let x_1 be the first column of A. Let $w = \pm(||x_1||_2, 0, \dots, 0)$ be a vector along the first coordinate axis of identical Euclidean length. (Either sign works in theory. For numerical stability, the sign is often chosen to be the opposite of the sign of the first component of x to avoid the possibility of subtracting nearly equal numbers when forming v.) Create the Householder reflector H_1 such that $H_1 x = w$. In the 4 × 3 case, multiplying H_1 by A results in

$$H_1 A = H_1 \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}.$$

We have introduced some zeros into A. We want to continue in this way until A becomes upper triangular; then we will have R of the QR factorization. Find the Householder reflector $\hat{H}_2$ that moves the $(m-1)$-vector x_2 consisting of the lower $m-1$ entries in column 2 of $H_1 A$ to $\pm(||x_2||_2, 0, \dots, 0)$. Since $\hat{H}_2$ is an $(m-1) \times (m-1)$-matrix, define H_2 to be the $m \times m$ matrix formed by putting $\hat{H}_2$ into the lower part of the identity matrix. Then

$$\left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & \hat{H}_2 & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|cc} \times & \times & \times \\ \hline 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{array}\right) = \left(\begin{array}{c|cc} \times & \times & \times \\ \hline 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{array}\right)$$

The result $H_2 H_1 A$ is one step from upper triangularity. One more step gives

$$\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & & \\ 0 & 0 & & \hat{H}_3 \end{array}\right) \left(\begin{array}{cc|c} \times & \times & \times \\ 0 & \times & \times \\ \hline 0 & 0 & \times \\ 0 & 0 & \times \end{array}\right) = \left(\begin{array}{cc|c} \times & \times & \times \\ 0 & \times & \times \\ \hline 0 & 0 & \times \\ 0 & 0 & 0 \end{array}\right)$$

and the result

$$H_3 H_2 H_1 A = R,$$

an upper triangular matrix. Multiplying on the left by the inverses of the Householder reflectors allows us to rewrite the result as

$$A = H_1 H_2 H_3 R = QR,$$

where $Q = H_1 H_2 H_3$. Note that $H_i^{-1} = H_i$ since H_i is symmetric orthogonal. Computer Problem 3 asks the reader to write code for the factorization via Householder reflectors.

▸ EXAMPLE 4.18 Use Householder reflectors to find the QR factorization of

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 3 \end{bmatrix}.$$

We need to find a Householder reflector that moves the first column [3, 4] onto the x-axis. We found such a reflector H_1 in Example 4.17, and

$$H_1 A = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 0 & -1 \end{bmatrix}.$$

Multiplying both sides on the left by $H_1^{-1} = H_1$ yields

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 0 & -1 \end{bmatrix} = QR,$$

where $Q = H_1^T = H_1$. ◂

▶ EXAMPLE 4.19 Use Householder reflectors to find the QR factorization of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

We need to find a Householder reflector that moves the first column $x = [1, 2, 2]$ to the vector $w = [||x||_2, 0, 0]$. Set $v = w - x = [3, 0, 0] - [1, 2, 2] = [2, -2, -2]$. Referring to Theorem 4.4, we have

$$H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{2}{12}\begin{bmatrix} 4 & -4 & -4 \\ -4 & 4 & 4 \\ -4 & 4 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

and

$$H_1 A = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & -3 \\ 0 & -4 \end{bmatrix}.$$

The remaining step is to move the vector $\hat{x} = [-3, -4]$ to $\hat{w} = [5, 0]$. Calculating $\hat{H}_2$ from Theorem 4.4 yields

$$\begin{bmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{bmatrix}\begin{bmatrix} -3 \\ -4 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix},$$

leading to

$$H_2 H_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.6 & -0.8 \\ 0 & -0.8 & 0.6 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = R.$$

Multiplying both sides on the left by $H_1^{-1} H_2^{-1} = H_1 H_2$ yields the QR factorization:

$$\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = H_1 H_2 R = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.6 & -0.8 \\ 0 & -0.8 & 0.6 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 & -14/15 & -2/15 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & 2/15 & 11/15 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = QR.$$

Compare this result with the factorization from Gram–Schmidt orthogonalization in Example 4.13. ◀

The QR factorization is not unique for a given $m \times n$ matrix A. For example, define $D = \mathrm{diag}(d_1, \ldots, d_m)$, where each d_i is either $+1$ or -1. Then $A = QR = QDDR$, and we check that QD is orthogonal and DR is upper triangular.

Exercise 12 asks for an operation count of QR factorization by Householder reflections, which comes out to $(2/3)m^3$ multiplications and the same number of additions—lower complexity than Gram–Schmidt orthogonalization. Moreover, the Householder method is known to deliver better orthogonality in the unit vectors and has lower memory requirements. For these reasons, it is the method of choice for factoring typical matrices into QR.

4.3 Exercises

1. Apply classical Gram–Schmidt orthogonalization to find the full QR factorization of the following matrices:

$$\text{(a)} \begin{bmatrix} 4 & 0 \\ 3 & 1 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 2 & 1 \end{bmatrix} \quad \text{(d)} \begin{bmatrix} 4 & 8 & 1 \\ 0 & 2 & -2 \\ 3 & 6 & 7 \end{bmatrix}$$

2. Apply classical Gram–Schmidt orthogonalization to find the full QR factorization of the following matrices:

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

3. Apply modified Gram–Schmidt orthogonalization to find the full QR factorization of the matrices in Exercise 1.

4. Apply modified Gram–Schmidt orthogonalization to find the full QR factorization of the matrices in Exercise 2.

5. Apply Householder reflectors to find the full QR factorization of the matrices in Exercise 1.

6. Apply Householder reflectors to find the full QR factorization of the matrices in Exercise 2.

7. Use the QR factorization from Exercise 2, 4, or 6 to solve the least squares problem.

$$\text{(a)} \begin{bmatrix} 2 & 3 \\ -2 & -6 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 6 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 0 \end{bmatrix}$$

8. Find the QR factorization and use it to solve the least squares problem.

$$\text{(a)} \begin{bmatrix} 1 & 4 \\ -1 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ -3 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 2 & 4 \\ 0 & -1 \\ 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

9. Prove that a square matrix is orthogonal if and only if its columns are pairwise orthogonal unit vectors.

10. Prove that the product of two orthogonal $m \times m$ matrices is again orthogonal.

11. Show that the Gram–Schmidt orthogonalization of an $m \times m$ matrix requires approximately m^3 multiplications and m^3 additions.

12. Show that the Householder reflector method for the QR factorization requires approximately $(2/3)m^3$ multiplications and $(2/3)m^3$ additions.

13. Let P be the matrix defined in (4.29). Show (a) $P^2 = P$ (b) P is symmetric (c) $Pv = v$.

14. Prove that Householder reflectors are symmetric matrices.

15. Verify that classical and modified Gram–Schmidt are mathematically identical (in exact arithmetic).

4.3 Computer Problems

1. Write a MATLAB program that implements classical Gram–Schmidt to find the reduced QR factorization. Check your work by comparing factorizations of the matrices in Exercise 1 with the MATLAB `qr(A,0)` command or equivalent. The factorization is unique up to signs of the entries of Q and R.

2. Repeat Computer Problem 1, but implement modified Gram–Schmidt.

3. Repeat Computer Problem 1, but implement Householder reflections.

4. Write a MATLAB program that implements (a) classical and (b) modified Gram–Schmidt to find the full QR factorization. Check your work by comparing factorizations of the matrices in Exercise 1 with the MATLAB `qr(A)` command or equivalent.

5. Use the MATLAB QR factorization to find the least squares solutions and 2-norm error of the following inconsistent systems:

$$\text{(a)}\quad \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 5 \end{bmatrix} \qquad \text{(b)}\quad \begin{bmatrix} 1 & 2 & 2 \\ 2 & -1 & 2 \\ 3 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 10 \\ 3 \end{bmatrix}$$

6. Use the MATLAB QR factorization to find the least squares solutions and 2-norm error of the following inconsistent systems:

$$\text{(a)}\quad \begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix} \qquad \text{(b)}\quad \begin{bmatrix} 4 & 2 & 3 & 0 \\ -2 & 3 & -1 & 1 \\ 1 & 3 & -4 & 2 \\ 1 & 0 & 1 & -1 \\ 3 & 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 2 \\ 0 \\ 5 \end{bmatrix}$$

7. Let A be the $10 \times n$ matrix formed by the first n columns of the 10×10 Hilbert matrix. Let c be the n-vector $[1, \dots, 1]$, and set $b = Ac$. Use the QR factorization to solve the least squares problem $Ax = b$ for (a) $n = 6$ (b) $n = 8$, and compare with the correct least squares solution $\overline{x} = c$. How many correct decimal places can be computed? See Computer Problem 4.1.8, where the normal equations are used.

8. Let $x_1, \dots, x_{11}$ be 11 evenly spaced points in $[2, 4]$ and $y_i = 1 + x_i + x_i^2 + \cdots + x_i^d$. Use the QR factorization to compute the best degree d polynomial, where (a) $d = 5$ (b) $d = 6$ (c) $d = 8$. Compare with Example 4.5 and Computer Problem 4.1.9. How many correct decimal places of the coefficients can be computed?

4.4 Generalized Minimum Residual (GMRES) Method

In Chapter 2, we saw that the Conjugate Gradient Method can be viewed as an iterative method specially designed to solve the matrix system $Ax = b$ for a symmetric square matrix A. If A is not symmetric, the conjugate gradient theory fails. However, there are several alternatives that work for the nonsymmetric problem. One of the most popular is the Generalized Minimum Residual Method, or GMRES for short. This method is a good choice for the solution of large, sparse, nonsymmetric linear systems $Ax = b$.

At first sight, it might seem strange to be discussing a method for solving linear systems in the chapter on least squares. Why should orthogonality matter to a problem that has