within the matrix. (a) Print the nonzero structure `spy(A)`. (b) Let x_e be the vector of n ones. Set $b = Ax_e$, and apply the Conjugate Gradient Method, without preconditioner, with the Jacobi preconditioner, and with the Gauss–Seidel preconditioner. Compare errors of the three runs in a plot versus step number.

2. Let $n = 1000$. Start with the $n \times n$ matrix A from Computer Problem 1, and add the nonzero entries $A(i, 2i) = A(2i, i) = 1/2$ for $1 \le i \le n/2$. Carry out steps (a) and (b) as in that problem.

3. Let $n = 500$, and let A be the $n \times n$ matrix with entries $A(i, i) = 2$, $A(i, i + 2) = A(i + 2, i) = 1/2$, $A(i, i + 4) = A(i + 4, i) = 1/2$ for all i, and $A(500, i) = A(i, 500) = -0.1$ for $1 \le i \le 495$. Carry out steps (a) and (b) as in Computer Problem 1.

4. Let A be the matrix from Computer Problem 3, but with the diagonal elements replaced by $A(i, i) = \sqrt[3]{i}$. Carry out parts (a) and (b) as in that problem.

5. Let C be the 195×195 matrix block with $C(i, i) = 2$, $C(i, i + 3) = C(i + 3, i) = 0.1$, $C(i, i + 39) = C(i + 39, i) = 1/2$, $C(i, i + 42) = C(i + 42, i) = 1/2$ for all i. Define A to be the $n \times n$ matrix with $n = 780$ formed by four diagonally arranged blocks C, and with blocks $\frac{1}{2}C$ on the super- and subdiagonal. Carry out steps (a) and (b) as in Computer Problem 1 to solve $Ax = b$.

4.5 NONLINEAR LEAST SQUARES

The least squares solution of a linear system of equations $Ax = b$ minimizes the Euclidean norm of the residual $||Ax - b||_2$. We have learned two methods to find the solution $\overline{x}$, one based on the normal equations and another on the QR factorization.

Neither method can be applied if the equations are nonlinear. In this section, we develop the Gauss–Newton Method for solving nonlinear least squares problems. In addition to illustrating the use of the method to solve circle intersection problems, we apply Gauss–Newton to fitting models with nonlinear coefficients to data.

4.5.1 Gauss–Newton Method

Consider the system of m equations in n unknowns

$$\begin{aligned} r_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ r_m(x_1, \ldots, x_n) &= 0. \end{aligned} \tag{4.31}$$

The sum of the squares of the errors is represented by the function

$$E(x_1, \ldots, x_n) = \frac{1}{2}(r_1^2 + \cdots + r_m^2) = \frac{1}{2} r^T r,$$

where $r = [r_1, \ldots, r_m]^T$. The constant $1/2$ has been included in the definition to simplify later formulas. To minimize E, we set the gradient $F(x) = \nabla E(x)$ to zero:

$$0 = F(x) = \nabla E(x) = \nabla \left(\frac{1}{2} r(x)^T r(x) \right) = r(x)^T Dr(x). \tag{4.32}$$

Observe that we have used the dot product rule for the gradient (see Appendix A).

We begin by recalling Multivariate Newton's Method, and apply it to the function viewed as a column vector $F(x)^T = (r^T Dr)^T = (Dr)^T r$. The matrix/vector product rule (see Appendix A) can be applied to yield

$$DF(x)^T = D((Dr)^T r) = (Dr)^T \cdot Dr + \sum_{i=1}^{m} r_i \, Dc_i,$$

where c_i is the ith column of Dr. Note that $Dc_i = H_{r_i}$, the matrix of second partial derivatives, or **Hessian**, of r_i:

$$H_{r_i} = \begin{bmatrix} \frac{\partial^2 r_i}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 r_i}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_n \partial x_n} \end{bmatrix}.$$

The application of Newton's Method can be simplified by dropping some of the terms. Without the above m-term summation, we have the following.

Gauss–Newton Method

To minimize

$$r_1(x)^2 + \cdots + r_m(x)^2.$$

Set $x^0 =$ initial vector,
for $k = 0, 1, 2, \ldots$

$$A = Dr(x^k) \tag{4.33}$$
$$A^T A v^k = -A^T r(x^k)$$
$$x^{k+1} = x^k + v^k \tag{4.34}$$

end

Notice that each step of the Gauss–Newton Method is reminiscent of the normal equations, where the coefficient matrix has been replaced by Dr. The Gauss–Newton Method solves for a root of the gradient of the squared error. Although the gradient must be zero at the minimum, the converse is not true, so it is possible for the method to converge to a maximum or a neutral point. Caution must be used in interpreting the algorithm's result.

The following three examples illustrate use of the Gauss-Newton Method, as well as Multivariate Newton's Method of Chapter 2. Two intersecting circles intersect in one or two points, unless the circles coincide. Three circles in the plane, however, typically have no points of common intersection. In such a case, we can ask for the point in the plane that comes closest to being an intersection point in the sense of least squares. For three circles, this is a question of three nonlinear equations in the two unknowns x, y.

Example 4.21 shows how the Gauss-Newton Method solves this nonlinear least squares problem. Example 4.22 defines the best point in a different way: Find the unique point of intersection of the 3 circles, allowing their radii to be changed by a common amount K. This is a question of three equations in three unknowns x, y, K, not a least squares problem, and is solved using Multivariate Newton's Method.

Finally, Example 4.23 adds a fourth circle. The solution of four equations in the three unknowns x, y, K is again a least squares problem that requires Gauss-Newton. This last formulation is relevant to calculations in GPS, as shown in Reality Check 4.

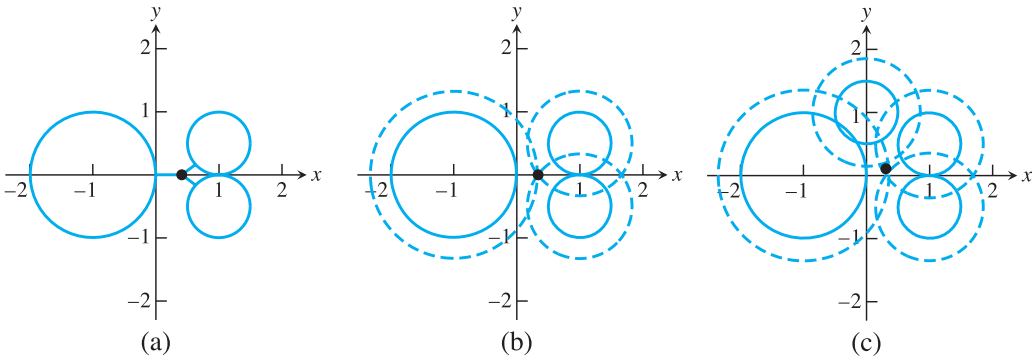

Figure 4.13 Near-intersection points of three circles. (a) The least squares near-intersection point, found by the Gauss–Newton Method. (b) Expanding the radii by a common amount gives a different type of near-intersection point by Multivariate Newton's Method. (c) The four circles of Example 4.23 with least squares solution point found by the Gauss-Newton Method.

► EXAMPLE 4.21 Consider the three circles in the plane with centers $(x_1, y_1) = (-1, 0), (x_2, y_2) = (1, 1/2)$, $(x_3, y_3) = (1, -1/2)$ and radii $R_1 = 1, R_2 = 1/2, R_3 = 1/2$, respectively. Use the Gauss–Newton Method to find the point for which the sum of the squared distances to the three circles is minimized.

The circles are shown in Figure 4.13(a). The point (x, y) in question minimizes the sum of the squares of the residual errors:

$$r_1(x, y) = \sqrt{(x - x_1)^2 + (y - y_1)^2} - R_1$$

$$r_2(x, y) = \sqrt{(x - x_2)^2 + (y - y_2)^2} - R_2$$

$$r_3(x, y) = \sqrt{(x - x_3)^2 + (y - y_3)^2} - R_3.$$

This follows from the fact that the distance from a point (x, y) to a circle with center (x_1, y_1) and radius R_1 is $|\sqrt{(x - x_1)^2 + (y - y_1)^2} - R_1|$ (see Exercise 3). The Jacobian of $r(x, y)$ is

$$Dr(x, y) = \begin{bmatrix} \frac{x-x_1}{S_1} & \frac{y-y_1}{S_1} \\ \frac{x-x_2}{S_2} & \frac{y-y_2}{S_2} \\ \frac{x-x_3}{S_3} & \frac{y-y_3}{S_3} \end{bmatrix},$$

where $S_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ for $i = 1, 2, 3$. The Gauss–Newton iteration with initial vector $(x^0, y^0) = (0, 0)$ converges to $(\overline{x}, \overline{y}) = (0.412891, 0)$ within six correct decimal places after seven steps. ◄

A related problem for three circles gives a different type of answer. Instead of looking for points that most resemble intersection points, we can expand (or contract) the circles' radii by a common amount until they have a common intersection. This is equivalent to solving the system

$$r_1(x, y, K) = \sqrt{(x - x_1)^2 + (y - y_1)^2} - (R_1 + K) = 0$$
$$r_2(x, y, K) = \sqrt{(x - x_2)^2 + (y - y_2)^2} - (R_2 + K) = 0$$
$$r_3(x, y, K) = \sqrt{(x - x_3)^2 + (y - y_3)^2} - (R_3 + K) = 0. \qquad (4.35)$$

The point (x, y) identified in this way is in general different from the least squares solution of Example 4.21.

► EXAMPLE 4.22 Solve the system (4.35) for (x, y, K), using the circles from Example 4.21.

The system consists of three nonlinear equations in three unknowns, calling for Multivariate Newton's Method. The Jacobian is

$$Dr(x, y, K) = \begin{bmatrix} \frac{x-x_1}{S_1} & \frac{y-y_1}{S_1} & -1 \\ \frac{x-x_2}{S_2} & \frac{y-y_2}{S_2} & -1 \\ \frac{x-x_3}{S_3} & \frac{y-y_3}{S_3} & -1 \end{bmatrix}.$$

Newton's Method yields the solution $(x, y, K) = (1/3, 0, 1/3)$ in three steps. The intersection point $(1/3, 0)$ and the three circles with radii expanded by $K = 1/3$ appear in Figure 4.13(b). ◄

Examples 4.21 and 4.22 show two different viewpoints on the meaning of the "near-intersection point" of a group of circles. Example 4.23 combines the two different approaches.

► EXAMPLE 4.23 Consider the four circles with centers $(-1, 0), (1, 1/2), (1, -1/2), (0, 1)$ and radii $1, 1/2, 1/2, 1/2$, respectively. Find the point (x, y) and constant K for which the sum of the squared distances from the point to the four circles with radii increased by K (thus $1 + K, 1/2 + K, 1/2 + K, 1/2 + K$, respectively) is minimized.

This is a straightforward combination of the previous two examples. There are four equations in the three unknowns x, y, K. The least squares residual is similar to (4.35), but with four terms, and the Jacobian is

$$Dr(x, y, K) = \begin{bmatrix} \frac{x-x_1}{S_1} & \frac{y-y_1}{S_1} & -1 \\ \frac{x-x_2}{S_2} & \frac{y-y_2}{S_2} & -1 \\ \frac{x-x_3}{S_3} & \frac{y-y_3}{S_3} & -1 \\ \frac{x-x_4}{S_4} & \frac{y-y_4}{S_4} & -1 \end{bmatrix}.$$

The Gauss–Newton Method provides the solution $(\overline{x}, \overline{y}) = (0.311385, 0.112268)$ with $\overline{K} = 0.367164$, pictured in Figure 4.13(c). ◄

The analogue of Example 4.23 for spheres in three dimensions forms the mathematical foundation of the Global Positioning System (GPS). See Reality Check 4.

4.5.2 Models with nonlinear parameters

An important application of the Gauss–Newton Method is to fit models that are nonlinear in the coefficients. Let $(t_1, y_1), \ldots, (t_m, y_m)$ be data points and $y = f_c(x)$ the function to be

fit, where $c = [c_1, \ldots, c_p]$ is a set of parameters to be chosen to minimize the sum of the squares of the residuals

$$\begin{aligned} r_1(c) &= f_c(t_1) - y_1 \\ &\vdots \\ r_m(c) &= f_c(t_m) - y_m. \end{aligned}$$

This particular case of (4.31) is seen commonly enough to warrant special treatment here.

If the parameters $c_1, \ldots, c_p$ enter the model in a linear way, then this is a set of linear equations in the c_i, and the normal equations, or QR-factorization solution, gives the optimal choice of parameters c. If the parameters c_i are nonlinear in the model, the same treatment results in a system of equations that is nonlinear in the c_i. For example, fitting the model $y = c_1 t^{c_2}$ to the data points (t_i, y_i) yields the nonlinear equations

$$\begin{aligned} y_1 &= c_1 t_1^{c_2} \\ y_2 &= c_1 t_2^{c_2} \\ &\vdots \\ y_m &= c_1 t_m^{c_2}. \end{aligned}$$

Because c_2 enters the model nonlinearly, the system of equations cannot be put in matrix form.

In Section 4.2, we handled this difficulty by changing the problem: We "linearized the model" by taking log of both sides of the model and minimized the error in these log-transformed coordinates by least squares. In cases where the log-transformed coordinates are really the proper coordinates in which to be minimizing error, this is appropriate.

To solve the original least squares problem, however, we turn to the Gauss–Newton Method. It is used to minimize the error function E as a function of the vector of parameters c. The matrix Dr is the matrix of partial derivatives of the errors r_i with respect to the parameters c_j, which are

$$(Dr)_{ij} = \frac{\partial r_i}{\partial c_j} = f_{c_j}(t_i).$$

With this information, the Gauss–Newton Method (4.33) can be implemented.

▶ EXAMPLE 4.24

Use the Gauss–Newton Method to fit the world automobile supply data of Example 4.8 with a (nonlinearized) exponential model.

Finding the best least squares fit of the data to an exponential model means finding c_1, c_2 that minimize the RMSE for errors $r_i = c_1 e^{c_2 t_i} - y_i$, $i = 1, \ldots, m$. Using model linearization in the previous section, we minimized the RMSE for the errors of the log model $\ln y_i - (\ln c_1 + c_2 t_i)$. The values of c_i that minimize the RMSE in the two different senses are different in general.

To compute the best least squares fit by the Gauss–Newton Method, define

$$r = \begin{bmatrix} c_1 e^{c_2 t_1} - y_1 \\ \vdots \\ c_1 e^{c_2 t_m} - y_m \end{bmatrix},$$

and take derivatives with respect to the parameters c_1 and c_2 to get

$$Dr = -\begin{bmatrix} e^{c_2 t_1} & c_1 t_1 e^{c_2 t_1} \\ \vdots & \vdots \\ e^{c_2 t_m} & c_1 t_m e^{c_2 t_m} \end{bmatrix}.$$

SPOTLIGHT ON

Convergence Nonlinearity in least squares problems causes extra challenges. The normal equations and QR approach find the single solution as long as the coefficient matrix A has full rank. On the other hand, Gauss–Newton iteration applied to a nonlinear problem may converge to one of several different relative minima of the least squares error. Using a reasonable approximation for the initial vector, if available, aids convergence to the absolute minimum.

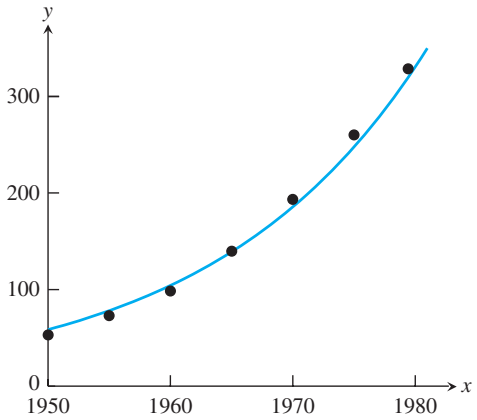

Figure 4.14 Exponential fit of world automobile supply data, without using linearization. The best least squares fit is $y = 58.51e^{0.05772t}$.

This model is fit with the world automobile supply data, where t is measured in years since 1970, and cars in millions. Five steps of the Gauss–Newton Method (4.33) from initial guess $(c_1, c_2) = (50, 0.1)$ yields $(c_1, c_2) \approx (58.51, 0.05772)$ with four digits of precision. The best least squares exponential model for the data is

$$y = 58.51e^{0.05772t}. \tag{4.36}$$

The RMSE is 7.68, meaning an average modeling error, in the least squares sense, of 7.68 million cars (see Figure 4.14).

The best model (4.36) can be compared with the best linearized exponential model

$$y = 54.03e^{0.06152t}$$

calculated in Example 4.8. This was obtained from the normal equations applied to the linearized model $\ln y = \ln c_1 + c_2 t$. The RMSE of the errors r_i of the linearized model is 9.56, greater than the RMSE of (4.36), as necessary. However, the linearized model minimizes the RMSE of the errors $\ln y_i - (\ln c_1 + c_2 t_i)$, giving a value of 0.0357, lower than the corresponding value 0.0568 for model (4.36), also as required. Each of the models is the optimal fit in its data space.

The moral is that there are computational algorithms for solving either problem. Minimizing the r_i is the standard least squares problem, but the user must decide on the basis of the data context whether it is more appropriate to minimize errors or log errors. ◀

4.5.3 The Levenberg–Marquardt Method.

Least squares minimization is especially challenging when the coefficient matrix turns out to be ill-conditioned. In Example 4.5, large errors were encountered in the least squares solution of $Ax = b$ when using the normal equations, since $A^T A$ had large condition number.

The problem is often worse for nonlinear least squares minimization. Many plausible model definitions yield poorly conditioned Dr matrices. The Levenberg–Marquardt Method uses a "regularization term" to partially remedy the conditioning problem. It can be thought of as a mixture of Gauss–Newton and the steepest descent method, which will be introduced for general optimization problems in Chapter 13.

The algorithm is a simple modification of the Gauss–Newton Method.

Levenberg–Marquardt Method

To minimize

$$r_1(x)^2 + \cdots + r_m(x)^2.$$

Set $x^0 =$ initial vector, $\lambda =$ constant
for $k = 0, 1, 2, \ldots$

$$\begin{aligned} A &= Dr(x^k) \\ (A^T A + \lambda \operatorname{diag}(A^T A))v^k &= -A^T r(x^k) \\ x^{k+1} &= x^k + v^k \end{aligned}$$

end

The $\lambda = 0$ case is identical to Gauss–Newton. Increasing the regularization parameter λ accentuates the effect of the diagonal of the matrix $A^T A$, which improves the condition number and generally allows the method to converge from a broader set of initial guesses x_0 than Gauss–Newton.

► EXAMPLE 4.25 Use Levenberg–Marquardt to fit the model $y = c_1 e^{-c_2(t-c_3)^2}$ to the data points $(t_i, y_i) = \{(1, 3), (2, 5), (2, 7), (3, 5), (4, 1)\}$.

We must find the c_1, c_2, c_3 that minimize the RMSE for error vector

$$r = \begin{bmatrix} c_1 e^{-c_2(t_1-c_3)^2} - y_1 \\ \vdots \\ c_1 e^{-c_2(t_5-c_3)^2} - y_5 \end{bmatrix}.$$

The derivative of r evaluated at the five data points is the 5×3 matrix

$$Dr = \begin{bmatrix} e^{-c_2(t_1-c_3)^2} & -c_1(t_1 - c_3)^2 e^{-c_2(t_1-c_3)^2} & 2c_1c_2(t_1 - c_3)e^{-c_2(t_1-c_3)^2} \\ \vdots & \vdots & \vdots \\ e^{-c_2(t_5-c_3)^2} & -c_1(t_5 - c_3)^2 e^{-c_2(t_5-c_3)^2} & 2c_1c_2(t_5 - c_3)e^{-c_2(t_5-c_3)^2} \end{bmatrix}.$$

Levenberg–Marquardt with initial guess $(c_1, c_2, c_3) = (1, 1, 1)$ and λ fixed at 50 converges to the best least squares model

$$y = 6.301e^{-0.5088(t-2.249)^2}.$$

The best model is plotted along with the data points in Figure 4.15. The corresponding Gauss–Newton Method diverges to infinity from this initial guess. ◄

The method originated by a suggestion in Levenberg [1944] to add λI to $A^T A$ in Gauss–Newton to improve its conditioning. Several years later, D. Marquardt, a statistician at DuPont, improved on Levenberg's suggestion by replacing the identity matrix with the diagonal of $A^T A$ (Marquardt [1963]).

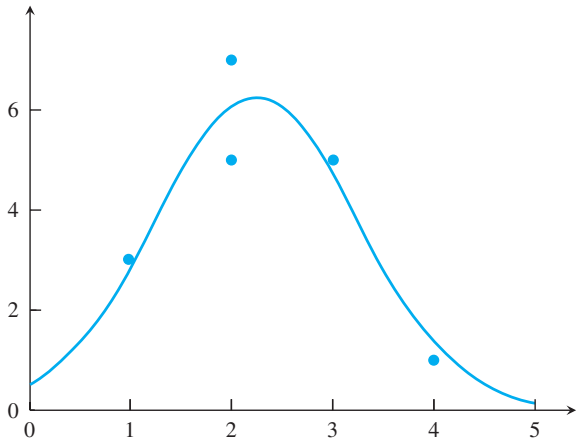

Figure 4.15 Model Fit of Example 4.25. The Levenberg–Marquardt Method is used to find the best least squares model $y = 6.301e^{-0.5088(t-2.249)^2}$, plotted as the solid curve.

Although we have treated λ as a constant for simplicity, the method is often applied adaptively with a varying λ. A common strategy is to continue to decrease λ by a factor of 10 on each iteration step as long as the residual sum of squared errors is decreased by the step, and if the sum increases, to reject the step and increase λ by a factor of 10.

4.5 Exercises

1. The Gauss–Newton Method can be applied to find the point $\overline{x}, \overline{y}$ for which the sum of the squared distances to the three circles is minimized. Using initial vector $(x_0, y_0) = (0, 0)$, carry out the first step to find (x_1, y_1) (a) centers $(0, 1), (1, 1), (0, -1)$ and all radii 1 (b) centers $(-1, 0), (1, 1), (1, -1)$ and all radii 1. (Computer Problem 1 asks for $(\overline{x}, \overline{y})$.)

2. Carry out the first step of Multivariate Newton's Method applied to the system (4.35) for the three circles in Exercise 1. Use $(x_0, y_0, K_0) = (0, 0, 0)$. (Computer Problem 2 asks for the solution (x, y, K).)

3. Prove that the distance from a point (x, y) to a circle $(x - x_1)^2 + (y - y_1)^2 = R_1^2$ is $|\sqrt{(x - x_1)^2 + (y - y_1)^2} - R_1|$.

4. Prove that the Gauss–Newton Method applied to the linear system $Ax = b$ converges in one step to the solution of the normal equations.

5. Find the matrix Dr needed for the application of Gauss–Newton iteration to the model-fitting problem with three data points $(t_1, y_1), (t_2, y_2), (t_3, y_3)$, (a) power law $y = c_1 t^{c_2}$ (b) $y = c_1 t e^{c_2 t}$.

6. Find the matrix Dr needed for the application of Gauss–Newton iteration to the model-fitting problem with three data points $(t_1, y_1), (t_2, y_2), (t_3, y_3)$ (a) translated exponential $y = c_3 + c_1 e^{c_2 t}$ (b) translated power law $y = c_3 + c_1 t^{c_2}$

7. Prove that the number of real solutions (x, y, K) of (4.35) is either infinity or at most two.

4.5 Computer Problems

1. Apply the Gauss–Newton Method to find the point $(\overline{x}, \overline{y})$ for which the sum of the squared distances to the three circles is minimized. Use initial vector $(x_0, y_0) = (0, 0)$. (a) Centers $(0, 1), (1, 1), (0, -1)$ and all radii 1. (b) Centers $(-1, 0), (1, 1), (1, -1)$ and all radii 1.