CHAPTER 10 Systems of Nonlinear Equations

10.1 Introduction

A large part of the material in this book has involved the solution of systems of equations. Even so, to this point the methods have been appropriate only for systems of *linear* equations, equations in the variables $x_1, x_2, \dots, x_n$ of the form

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

for $i = 1, 2, \dots, n$. If you have wondered why we have not considered more general systems of equations, the reason is simple. It is much harder to approximate the solutions to a system of general, or *nonlinear*, equations.

Solving a system of nonlinear equations is a problem that is avoided when possible, customarily by approximating the nonlinear system by a system of linear equations. When this is unsatisfactory, the problem must be tackled directly. The most straightforward method of approach is to adapt the methods from Chapter 2 that approximate the solutions of a single nonlinear equation in one variable to apply when the single-variable problem is replaced by a vector problem that incorporates all the variables.

The principal tool in Chapter 2 was Newton's method, a technique that is generally quadratically convergent once a sufficiently accurate starting value is found. This is the first technique we modify to solve systems of nonlinear equations. Newton's method, as modified for systems of equations, is quite costly to apply, so, in Section 10.3, we describe how a modified Secant method can be used to obtain approximations more easily, although with a loss of the extremely rapid convergence that Newton's method provides.

Section 10.4 describes the method of Steepest Descent. This technique is only linearly convergent, but it does not require the accurate starting approximations needed for more rapidly-converging techniques. It is often used to find a good initial approximation for Newton's method or one of its modifications.

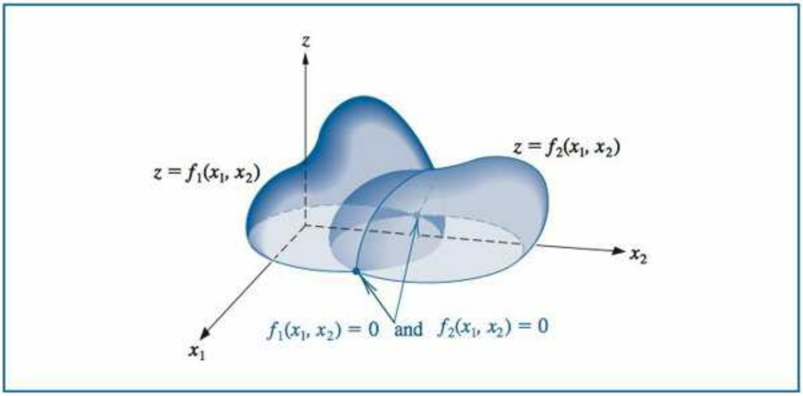
In Section 10.5, we give an introduction to continuation methods, which use a parameter to move from a problem with an easily determined solution to the solution of the original nonlinear problem.

A system of nonlinear equations has the form

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0, \end{aligned}$$

where each function f_i can be thought of as mapping a vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ of the n-dimensional space $\mathbb{R}^n$ into the real line $\mathbb{R}$. A geometric representation of a nonlinear system when $n = 2$ is given in Figure 10.1.

Figure 10.1



A general system of n nonlinear equations in n unknowns can be alternatively represented by defining a function $\mathbf{F}$, mapping $\mathbb{R}^n$ into $\mathbb{R}^n$, by

$$\mathbf{F}(x_1, x_2, \ldots, x_n) = (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_n(x_1, x_2, \ldots, x_n))^t.$$

If vector notation is used to represent the variables $x_1, x_2, \ldots, x_n$, the nonlinear system assumes the form

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}.$$

The functions $f_1, f_2, \ldots, f_n$ are the **coordinate functions** of $\mathbf{F}$.

Example 1 Place the 3×3 nonlinear system

$$\begin{aligned}
3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0,\\
x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0,\\
e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0
\end{aligned}$$

in the form $\mathbf{F}(\mathbf{x}) = \mathbf{0}$.

Solution Define the three coordinate functions f_1, f_2, and f_3 from $\mathbb{R}^3$ to $\mathbb{R}$ as

$$\begin{aligned}
f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2x_3) - \frac{1}{2},\\
f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06,\\
f_3(x_1, x_2, x_3) &= e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3}.
\end{aligned}$$

Then define $\mathbf{F}$ from $\mathbb{R}^3 \to \mathbb{R}^3$ by

$$\begin{aligned}\mathbf{F}(\mathbf{x}) &= \mathbf{F}(x_1, x_2, x_3)\\ &= (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3))^t\\ &= \left(3x_1 - \cos(x_2x_3) - \frac{1}{2}, x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06, e^{-x_1x_2} + 20x_3\right.\\ &\qquad \left. + \frac{10\pi - 3}{3}\right)^t.\end{aligned}$$

The original system is equivalent to $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. ■

Before discussing the solution of a system of nonlinear equations, we need some results concerning continuity and differentiability of functions from $\mathbb{R}^n$ into $\mathbb{R}$ and $\mathbb{R}^n$ into $\mathbb{R}^n$. These results parallel those given in Section 1.2 for a function from $\mathbb{R}$ into $\mathbb{R}$.

Let f be a function defined on a set $D \subset \mathbb{R}^n$ and mapping $\mathbb{R}^n$ into $\mathbb{R}$. The function f has the **limit** L at $\mathbf{x}_0$, written

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = L,$$

if, given any number $\varepsilon > 0$, a number $\delta > 0$ exists with the property that

$$|f(\mathbf{x}) - L| < \varepsilon \qquad \text{whenever } \mathbf{x} \in D \quad \text{and} \quad 0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta.$$

Any convenient norm can be used to satisfy the condition in this definition. The specific value of δ will depend on the norm chosen, but the existence and value of the limit L is independent of the norm.

The function f from $\mathbb{R}^n$ into $\mathbb{R}$ is **continuous** at $\mathbf{x}_0 \in D$ provided $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})$ exists and is $f(\mathbf{x}_0)$. In addition, f is **continuous on a set** D provided f is continuous at every point of D. This is expressed by writing $f \in C(D)$.

We define the limit and continuity concepts for functions from $\mathbb{R}^n$ into $\mathbb{R}^n$ by considering the coordinate functions from $\mathbb{R}^n$ into $\mathbb{R}$.

Let $\mathbf{F}$ be a function from $D \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and suppose $\mathbf{F}$ has the representation

$$\mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x}))^t,$$

where f_i is a function from $\mathbb{R}^n$ to $\mathbb{R}$ for each $i = 1, 2, \dots n$. We define the limit of $\mathbf{F}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ as

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \mathbf{F}(\mathbf{x}) = \mathbf{L} = (L_1, L_2, \dots, L_n)^t$$

if and only if $\lim_{\mathbf{x}\to\mathbf{x}_0} f_i(\mathbf{x}) = L_i$ for each $i = 1, 2, \dots, n$.

Continuity definitions for functions of n variables follow from those for a single variable by replacing, where necessary, absolute values by norms.

The function $\mathbf{F}$ is **continuous** at $\mathbf{x}_0 \in D$ provided $\lim_{\mathbf{x}\to\mathbf{x}_0} \mathbf{F}(\mathbf{x})$ exists and is $\mathbf{F}(\mathbf{x}_0)$. In addition, $\mathbf{F}$ is **continuous on the set** D if $\mathbf{F}$ is continuous at each $\mathbf{x}$ in D.

These are the basic concepts we will need for the remainder of the chapter.

10.2 Newton's Method for Systems

Newton's method for approximating the solution p to the single nonlinear equation

$$f(x) = 0$$

requires an initial approximation p_0 to p and generates a sequence defined by

$$p_k = p_{k-1} - \frac{f(p_{k-1})}{f'(p_{k-1})}, \quad \text{for } k \geq 1.$$

To modify Newton's method to find the vector solution $\mathbf{p}$ to the vector equation

$$\mathbf{F}(\mathbf{x}) = \mathbf{0},$$

we first need to determine an initial approximation vector $\mathbf{p}^{(0)}$. We must then decide how to modify the single-variable Newton's method to a vector function method that will have the same convergence properties but not require division because this operation is undefined for vectors. We also need to replace the derivative of f in the single-variable version of Newton's method with something that is appropriate for the vector function $\mathbf{F}$.

The Jacobian Matrix

The derivative $f'(x)$ of the single-variable function $f(x)$ describes how the values of the function change relative to changes in the independent variable x. The vector function $\mathbf{F}$ has n different variables, $x_1, x_2, \ldots, x_n$, and n different component functions, $f_1, f_2, \ldots, f_n$, each of which can change as any one of the variables change. The appropriate derivative modification from the single-variable Newton's method to the vector form must involve all these n^2 possible changes, and the natural way to represent n^2 items is by an $n \times n$ matrix. Each change in a component function f_i at $\mathbf{x}$ with respect to the change in the variable x_j is described by the partial derivative

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}),$$

and the $n \times n$ matrix which replaces the derivative that occurs in the single-variable case is

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

The matrix $J(\mathbf{x})$ is called the **Jacobian** matrix and has a number of applications in analysis. It might, in particular, be familiar due to its application in the multiple integration of a function of several variables over a region that requires a change of variables to be performed.

Newton's method for systems replaces division by the derivative in the single-variable case with multiplying by the inverse of the $n \times n$ Jacobian matrix in the vector situation. As a consequence, Newton's method for finding the solution $\mathbf{p}$ to the nonlinear system of equations represented by the vector equation $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ has the form

$$\mathbf{p}^{(k)} = \mathbf{p}^{(k-1)} - [J(\mathbf{p}^{(k-1)})]^{-1}\mathbf{F}(\mathbf{p}^{(k-1)}), \quad \text{for } k \geq 1,$$

given the initial approximation $\mathbf{p}^{(0)}$ to the solution $\mathbf{p}$.

A weakness in Newton's method for systems arises from the necessity of inverting the matrix $J(\mathbf{p}^{(k-1)})$ at each iteration. In practice, explicit computation of the inverse of $J(\mathbf{p}^{(k-1)})$ is avoided by performing the operation in a two-step manner. First, a vector $\mathbf{y}^{(k-1)}$ is found that will satisfy

$$J(\mathbf{p}^{(k-1)})\mathbf{y}^{(k-1)} = -\mathbf{F}(\mathbf{p}^{(k-1)}).$$

After this has been accomplished, the new approximation, $\mathbf{p}^{(k)}$, is obtained by adding $\mathbf{y}^{(k-1)}$ to $\mathbf{p}^{(k-1)}$.

The program NWTSY101 implements Newton's method for systems of nonlinear equations.

Example 1 The nonlinear system

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0, \\ x_1^2 - 81(x_2+0.1)^2 + \sin x_3 + 1.06 &= 0, \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0 \end{aligned}$$

has the approximate solution $(0.5, 0, -0.52359877)^t$. Apply Newton's method to this problem with $\mathbf{p}^{(0)} = (0.1, 0.1, -0.1)^t$.

Solution Define

$$\mathbf{F}(x_1, x_2, x_3) = (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3))^t,$$

where

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2x_3) - \frac{1}{2}, \\ f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2+0.1)^2 + \sin x_3 + 1.06, \end{aligned}$$

and

$$f_3(x_1, x_2, x_3) = e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3}.$$

The Jacobian matrix $J(\mathbf{x})$ for this system is

$$J(x_1, x_2, x_3) = \begin{bmatrix} 3 & x_3 \sin x_2x_3 & x_2 \sin x_2x_3 \\ 2x_1 & -162(x_2+0.1) & \cos x_3 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 \end{bmatrix}.$$

For $\mathbf{p}^{(0)} = (0.1, 0.1, -0.1)^t$ we have

$$\mathbf{F}(\mathbf{p}^{(0)}) = (-0.199995, -2.269833417, 8.462025346)^t$$

and

$$J(\mathbf{p}^{(0)}) = \begin{bmatrix} 3 & 9.999833334 \times 10^{-4} & 9.999833334 \times 10^{-4} \\ 0.2 & -32.4 & 0.9950041653 \\ -0.09900498337 & -0.09900498337 & 20 \end{bmatrix}.$$

Solving the linear system $J(\mathbf{p}^{(0)})\mathbf{y}^{(0)} = -\mathbf{F}(\mathbf{p}^{(0)})$ gives

$$\mathbf{y}^{(0)} = \begin{bmatrix} 0.3998696728 \\ -0.08053315147 \\ -0.4215204718 \end{bmatrix} \quad \text{and} \quad \mathbf{p}^{(1)} = \mathbf{p}^{(0)} + \mathbf{y}^{(0)} = \begin{bmatrix} 0.4998696728 \\ 0.01946684853 \\ -0.5215204718 \end{bmatrix}.$$

Continuing for $k = 2, 3, \ldots$, we have

$$\begin{bmatrix} p_1^{(k)} \\ p_2^{(k)} \\ p_3^{(k)} \end{bmatrix} = \begin{bmatrix} p_1^{(k-1)} \\ p_2^{(k-1)} \\ p_3^{(k-1)} \end{bmatrix} + \begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix},$$

where

$$\begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix} = -\left(J\left(p_1^{(k-1)}, p_2^{(k-1)}, p_3^{(k-1)}\right)\right)^{-1} \mathbf{F}\left(p_1^{(k-1)}, p_2^{(k-1)}, p_3^{(k-1)}\right).$$

At the kth step, the linear system $J(\mathbf{p}^{(k-1)})\mathbf{y}^{(k-1)} = -\mathbf{F}(\mathbf{p}^{(k-1)})$ must be solved, where

$$J(\mathbf{p}^{(k-1)}) = \begin{bmatrix} 3 & p_3^{(k-1)} \sin p_2^{(k-1)} p_3^{(k-1)} & p_2^{(k-1)} \sin p_2^{(k-1)} p_3^{(k-1)} \\ 2p_1^{(k-1)} & -162\left(p_2^{(k-1)} + 0.1\right) & \cos p_3^{(k-1)} \\ -p_2^{(k-1)} e^{-p_1^{(k-1)} p_2^{(k-1)}} & -p_1^{(k-1)} e^{-p_1^{(k-1)} p_2^{(k-1)}} & 20 \end{bmatrix},$$

$$\mathbf{y}^{(k-1)} = \begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix},$$

and

$$\mathbf{F}(\mathbf{p}^{(k-1)}) = \begin{bmatrix} 3p_1^{(k-1)} - \cos p_2^{(k-1)} p_3^{(k-1)} - \frac{1}{2} \\ \left(p_1^{(k-1)}\right)^2 - 81\left(p_2^{(k-1)} + 0.1\right)^2 + \sin p_3^{(k-1)} + 1.06 \\ e^{-p_1^{(k-1)} p_2^{(k-1)}} + 20p_3^{(k-1)} + \frac{10\pi - 3}{3} \end{bmatrix}.$$

The results using this iterative procedure are shown in Table 10.1. ▪

Table 10.1

k	$p_1^{(k)}$	$p_2^{(k)}$	$p_3^{(k)}$	$\lVert \mathbf{p}^{(k)} - \mathbf{p}^{(k-1)} \rVert_\infty$
0	0.1000000000	0.1000000000	−0.1000000000	
1	0.4998696728	0.0194668485	−0.5215204718	0.4215204718
2	0.5000142403	0.0015885914	−0.5235569638	1.788×10^{-2}
3	0.5000000113	0.0000124448	−0.5235984500	1.576×10^{-3}
4	0.5000000000	8.516×10^{-10}	−0.5235987755	1.244×10^{-5}
5	0.5000000000	-1.375×10^{-11}	−0.5235987756	8.654×10^{-10}

The previous example illustrates that Newton's method can converge very rapidly once an approximation is obtained that is near the true solution. However, it is not always easy to determine starting values that will lead to a solution, and the method is computationally expensive. In the next section, we consider a method for overcoming the latter weakness. Good starting values can usually be found by the method discussed in Section 10.4.

EXERCISE SET 10.2

1. Give an example of a function $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ that is continuous at each point of $\mathbb{R}^2$ except at $(1, 0)$.

2. Give an example of a function $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ that is continuous at each point of $\mathbb{R}^3$ except at $(1, 2, 3)$.

3. Use Newton's method with $\mathbf{p}^{(0)} = \mathbf{0}$ to compute $\mathbf{p}^{(2)}$ for each of the following nonlinear systems.

a.
$$\begin{aligned} 4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 &= 0, \\ \frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 &= 0. \end{aligned}$$

b.
$$\begin{aligned} \sin(4\pi x_1x_2) - 2x_2 - x_1 &= 0, \\ \left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 &= 0. \end{aligned}$$

c.
$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0, \\ 4x_1^2 - 625x_2^2 + 2x_2 - 1 &= 0, \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0. \end{aligned}$$

d.
$$\begin{aligned} x_1^2 + x_2 \quad\quad - 37 &= 0, \\ x_1 - x_2^2 \quad\quad - 5 &= 0, \\ x_1 + x_2 + x_3 - 3 &= 0. \end{aligned}$$

4. Use Newton's method to find a solution to the following nonlinear systems in the given domain. Iterate until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$.

a.
$$\begin{aligned} 3x_1^2 - x_2^2 &= 0, \\ 3x_1x_2^2 - x_1^3 - 1 &= 0. \end{aligned}$$
Use $\mathbf{p}^{(0)} = (1, 1)^t$.

b.
$$\begin{aligned} \ln(x_1^2 + x_2^2) - \sin(x_1x_2) &= \ln 2 + \ln \pi, \\ e^{x_1 - x_2} + \cos(x_1x_2) &= 0. \end{aligned}$$
Use $\mathbf{p}^{(0)} = (2, 2)^t$.

c.
$$\begin{aligned} x_1^3 + x_1^2x_2 - x_1x_3 + 6 &= 0, \\ e^{x_1} + e^{x_2} - x_3 &= 0, \\ x_2^2 - 2x_1x_3 &= 4. \end{aligned}$$
Use $\mathbf{p}^{(0)} = (-1, -2, 1)^t$.

d.
$$\begin{aligned} 6x_1 - 2\cos(x_2x_3) - 1 &= 0, \\ 9x_2 + \sqrt{x_1^2 + \sin x_3 + 1.06} + 0.9 &= 0, \\ 60x_3 + 3e^{-x_1x_2} + 10\pi - 3 &= 0. \end{aligned}$$
Use $\mathbf{p}^{(0)} = (0, 0, 0)^t$.

5. The nonlinear system

$$\begin{aligned} x_1^2 - x_2^2 + 2x_2 &= 0, \\ 2x_1 + x_2^2 - 6 &= 0. \end{aligned}$$

has four solutions. They are near $(-5, -4)^t$, $(2, -1)^t$, $(0.5, 2)^t$, and $(-2, 3)^t$. Use these points as initial approximations for Newton's method and iterate until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$. Do the results justify using the stated points as initial approximations?

6. The nonlinear system

$$\begin{aligned} 2x_1 - 3x_2 + x_3 - 4 &= 0, \\ 2x_1 + x_2 - x_3 + 4 &= 0, \\ x_1^2 + x_2^2 + x_3^2 - 4 &= 0. \end{aligned}$$

has a solution near $(-0.5, -1.5, 1.5)^t$.

a. Use this point as an initial approximation for Newton's method and iterate until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$.

b. Solve the first two equations for x_1 and x_3 in terms of x_2.

c. Substitute the results of (b) into the third equation to obtain a quadratic equation in x_2.

d. Solve the quadratic equation in (c) by the quadratic formula.

e. Of the solutions in (a) and (d), which is closer to the initial approximation $(-0.5, -1.5, 1.5)^t$?

7. The nonlinear system

$$3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0,$$
$$x_1^2 - 625x_2^2 - \frac{1}{4} = 0,$$
$$e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

has a singular Jacobian matrix at the solution. Apply Newton's method with $\mathbf{p}^{(0)} = (1, 1 - 1)^t$. Note that convergence may be slow or may not occur within a reasonable number of iterations.

8. The nonlinear system

$$4x_1 - x_2 + x_3 = x_1x_4,$$
$$-x_1 + 3x_2 - 2x_3 = x_2x_4,$$
$$x_1 - 2x_2 + 3x_3 = x_3x_4,$$
$$x_1^2 + x_2^2 + x_3^2 = 1$$

has six solutions.

a. Show that if $(x_1, x_2, x_3, x_4)^t$ is a solution, then $(-x_1, -x_2, -x_3, x_4)^t$ is a solution.

b. Use Newton's method three times to approximate all solutions. Iterate until $\left\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\right\|_\infty < 10^{-5}$. Use the initial vectors $(1, 1, 1, 1)^t$, $(1, 0, 0, 0)^t$, and $(1, -1, 1, -1)^t$.

9. Let A be an $n \times n$ matrix and $\mathbf{F}$ be the function from $\mathbb{R}^n$ to $\mathbb{R}^n$ defined by $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$. What is the Jacobian matrix of $\mathbf{F}$?

10. In Exercise 6 of Section 5.7, we considered the problem of predicting the population of two species that compete for the same food supply. In the problem, we made the assumption that the populations could be predicted by solving the system of equations

$$\frac{dx_1}{dt}(t) = x_1(t)\,(4 - 0.0003x_1(t) - 0.0004x_2(t))$$

and

$$\frac{dx_2}{dt}(t) = x_2(t)\,(2 - 0.0002x_1(t) - 0.0001x_2(t)).$$

In this exercise, we would like to consider the problem of determining equilibrium populations of the two species. The mathematical criteria that must be satisfied in order for the populations to be at equilibrium is that, simultaneously,

$$\frac{dx_1}{dt}(t) = 0 \quad \text{and} \quad \frac{dx_2}{dt}(t) = 0.$$

This occurs when the first species is extinct and the second species has a population of 20,000 or when the second species is extinct and the first species has a population of 13,333. Can an equilibrium occur in any other situation?

11. The amount of pressure required to sink a large, heavy object in a soft homogeneous soil that lies above a hard base soil can be predicted by the amount of pressure required to sink smaller objects in the same soil. Specifically, the amount of pressure p required to sink a circular plate of radius r a distance d in the soft soil, where the hard base soil lies a distance $D > d$ below the surface, can be approximated by an equation of the form

$$p = k_1e^{k_2r} + k_3r,$$

where k_1, k_2, and k_3 are constants, with $k_2 > 0$, depending on d and the consistency of the soil but not on the radius of the plate.

a. Find the values of k_1, k_2, and k_3 if we assume that a plate of radius 1 in. requires a pressure of 10 lb/in.^{2} to sink 1 ft in a muddy field, a plate of radius 2 in. requires a pressure of 12 lb/in.^{2} to sink 1 ft, and a plate of radius 3 in. requires a pressure of 15 lb/in.^{2} to sink this distance (assuming that the mud is more than 1 ft deep).

b. Use your calculations from (a) to predict the minimal size of circular plate that would be required to sustain a load of 500 lb on this field with sinkage of less than 1 ft.

10.3 Quasi-Newton Methods

A significant weakness of Newton's method for solving systems of nonlinear equations is the requirement that, at each iteration, a Jacobian matrix be computed and an $n \times n$ linear system solved that involves this matrix. To illustrate the magnitude of this weakness, consider the amount of computation associated with one iteration of Newton's method. The Jacobian matrix associated with a system of n nonlinear equations written in the form $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ requires that the n^2 partial derivatives of the n component functions of $\mathbf{F}$ be determined and evaluated. In most situations, the exact evaluation of the partial derivatives is inconvenient, and in many applications it is impossible. This difficulty can generally be overcome by using finite-difference approximations to the partial derivatives. For example,

$$\frac{\partial f_j}{\partial x_k}(\mathbf{x}) \approx \frac{f_j(\mathbf{x} + h\mathbf{e}_k) - f_j(\mathbf{x})}{h},$$

where h is small in absolute value and $\mathbf{e}_k$ is the vector whose only nonzero entry is a 1 in the kth coordinate.

This approximation, however, still requires that at least n^2 scalar functional evaluations be performed to approximate the Jacobian matrix and does not decrease the amount of calculation, in general $O(n^3)$, required for solving the linear system involving this approximate Jacobian. The total computational effort for just one iteration of Newton's method is, consequently, at least $n^2 + n$ scalar functional evaluations (n^2 for the evaluation of the Jacobian matrix and n for the evaluation of $\mathbf{F}$) together with $O(n^3)$ arithmetic operations to solve the linear system. This amount of computational effort can be prohibitive except for relatively small values of n and easily-evaluated scalar functions.

In this section, we consider a generalization of the Secant method to systems of nonlinear equations; in particular, a technique known as **Broyden's method** (see [Broy]). The method requires only n scalar functional evaluations per iteration and also reduces the number of arithmetic calculations to $O(n^2)$. It belongs to a class of methods known as *least-change secant updates* that produce algorithms called *quasi-Newton*. These methods replace the Jacobian matrix in Newton's method with an approximation matrix that is updated at each iteration. The disadvantage to the method is that the quadratic convergence of Newton's method is lost. It is replaced by *superlinear* convergence, which implies that

$$\lim_{i\to\infty} \frac{\|\mathbf{p}^{(i+1)} - \mathbf{p}\|}{\|\mathbf{p}^{(i)} - \mathbf{p}\|} = 0,$$

where $\mathbf{p}$ denotes the solution to $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, and $\mathbf{p}^{(i)}$ and $\mathbf{p}^{(i+1)}$ are consecutive approximations to $\mathbf{p}$. In most applications, the reduction to superlinear convergence is a more than acceptable trade-off for the decrease in the amount of computation.

An additional disadvantage of quasi-Newton methods is that, unlike Newton's method, they are not self-correcting. Newton's method, for example, will generally correct for round-off error with successive iterations, but unless special safeguards are incorporated, Broyden's method will not.

To describe Broyden's method, suppose that an initial approximation $\mathbf{p}^{(0)}$ is given to the solution $\mathbf{p}$ of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. We calculate the next approximation $\mathbf{p}^{(1)}$ in the same manner as Newton's method, or, if it is inconvenient to determine $J(\mathbf{p}^{(0)})$ exactly, we can use difference equations to approximate the partial derivatives. To compute $\mathbf{p}^{(2)}$, however, we depart from Newton's method and examine the Secant method for a single nonlinear equation. The Secant method differs from Newton's method because it uses

$$f'(p_1) \approx \frac{f(p_1) - f(p_0)}{p_1 - p_0}$$

instead of $f'(p_1)$. For nonlinear systems, $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$ is a vector, and the corresponding quotient is undefined. However, the method proceeds similarly in that we replace the matrix $J(\mathbf{p}^{(1)})$ in Newton's method by a matrix A_1 with the property that

$$A_1(\mathbf{p}^{(1)} - \mathbf{p}^{(0)}) = \mathbf{F}(\mathbf{p}^{(1)}) - \mathbf{F}(\mathbf{p}^{(0)}).$$

Any nonzero vector in $\mathbb{R}^n$ can be written as the sum of a multiple of $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$ and a multiple of a vector orthogonal to $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$. So, to uniquely define the matrix A_1, we need to specify how it acts on vectors orthogonal to $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$.

No information is available about the change in $\mathbf{F}$ in a direction orthogonal to $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$, so we simply require that no change occurs when defining A_1. That is,

$$A_1\mathbf{z} = J(\mathbf{p}^{(0)})\mathbf{z} \quad \text{whenever} \quad (\mathbf{p}^{(1)} - \mathbf{p}^{(0)})^t\mathbf{z} = 0.$$

Thus any vector orthogonal to $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$ is unaffected by the update from $J(\mathbf{p}^{(0)})$, which was used to compute $\mathbf{p}^{(1)}$, to A_1, which is used in the determination of $\mathbf{p}^{(2)}$.

These conditions uniquely define A_1 (see Exercise 8) as

$$A_1 = J(\mathbf{p}^{(0)}) + \frac{[\mathbf{F}(\mathbf{p}^{(1)}) - \mathbf{F}(\mathbf{p}^{(0)}) - J(\mathbf{p}^{(0)})(\mathbf{p}^{(1)} - \mathbf{p}^{(0)})]}{\|\mathbf{p}^{(1)} - \mathbf{p}^{(0)}\|_2^2}(\mathbf{p}^{(1)} - \mathbf{p}^{(0)})^t. \tag{10.1}$$

It is this matrix that is used in place of $J(\mathbf{p}^{(1)})$ to determine $\mathbf{p}^{(2)}$:

$$\mathbf{p}^{(2)} = \mathbf{p}^{(1)} - A_1^{-1}\mathbf{F}(\mathbf{p}^{(1)}).$$

Once $\mathbf{p}^{(2)}$ has been determined, the method can be repeated to determine $\mathbf{p}^{(3)}$, with A_1 used in place of $A_0 \equiv J(\mathbf{p}^{(0)})$ and with $\mathbf{p}^{(2)}$ and $\mathbf{p}^{(1)}$ in place of $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(0)}$, respectively. To simplify the notation we introduce the variables

$$\mathbf{s}_i = \mathbf{p}^{(i)} - \mathbf{p}^{(i-1)} \quad \text{and} \quad \mathbf{y}_i = \mathbf{F}(\mathbf{p}^{(i)}) - \mathbf{F}(\mathbf{p}^{(i-1)}).$$

Then, once $\mathbf{p}^{(i)}$ has been determined, $\mathbf{p}^{(i+1)}$ can be computed by

$$\begin{aligned} A_i &= A_{i-1} + \frac{[\mathbf{F}(\mathbf{p}^{(i)}) - \mathbf{F}(\mathbf{p}^{(i-1)})] - A_{i-1}(\mathbf{p}^{(i)} - \mathbf{p}^{(i-1)})}{\|\mathbf{p}^{(i)} - \mathbf{p}^{(i-1)}\|_2^2}(\mathbf{p}^{(i)} - \mathbf{p}^{(i-1)})^t \\ &= A_{i-1} + \frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2}\mathbf{s}_i^t. \end{aligned}$$

and

$$\mathbf{p}^{(i+1)} = \mathbf{p}^{(i)} - A_i^{-1}\mathbf{F}(\mathbf{p}^{(i)}).$$

If the method is performed as outlined, the number of scalar functional evaluations is reduced from $n^2 + n$ to n (those required for evaluating $\mathbf{F}(\mathbf{p}^{(i)})$), but the method still requires $O(n^3)$ calculations to solve the associated $n \times n$ linear system

$$A_i\mathbf{s}_{i+1} = -\mathbf{F}(\mathbf{p}^{(i)}).$$

Employing the method in this form would not be justified because of the reduction to superlinear convergence from the quadratic convergence of Newton's method. However, a significant improvement can be incorporated by employing a matrix-inversion formula.

Sherman-Morrison Formula

Sherman-Morrison Formula

If A is a nonsingular matrix and $\mathbf{x}$ and $\mathbf{y}$ are vectors with $\mathbf{y}^t A^{-1}\mathbf{x} \neq -1$, then $A + \mathbf{x}\mathbf{y}^t$ is nonsingular and

$$(A + \mathbf{x}\mathbf{y}^t)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{x}\mathbf{y}^t A^{-1}}{1 + \mathbf{y}^t A^{-1}\mathbf{x}}.$$

This formula permits A_i^{-1} to be computed directly from A_{i-1}^{-1}, eliminating the need for a matrix inversion with each iteration. This computation involves only matrix-vector multiplication at each step and therefore requires only $O(n^2)$ arithmetic calculations.

By letting $A = A_{i-1}$, $\mathbf{x} = (\mathbf{y}_i - A_{i-1}\mathbf{s}_i)/\|\mathbf{s}_i\|_2^2$, and $\mathbf{y} = \mathbf{s}_i$, the Sherman-Morrison formula implies that

$$\begin{aligned} A_i^{-1} &= \left(A_{i-1} + \frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2}\mathbf{s}_i^t\right)^{-1} \\ &= A_{i-1}^{-1} - \frac{A_{i-1}^{-1}\left(\dfrac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2}\right)\mathbf{s}_i^t A_{i-1}^{-1}}{1 + \mathbf{s}_i^t A_{i-1}^{-1}\left(\dfrac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2}\right)} \\ &= A_{i-1}^{-1} - \frac{\left(A_{i-1}^{-1}y_i - \mathbf{s}_i\right)\mathbf{s}_i^t A_{i-1}^{-1}}{\|\mathbf{s}_i\|_2^2 + \mathbf{s}_i^t A_{i-1}^{-1}\mathbf{y}_i - \|\mathbf{s}_i\|_2^2} \\ &= A_{i-1}^{-1} + \frac{\left(\mathbf{s}_i - A_{i-1}^{-1}\mathbf{y}_i\right)\mathbf{s}_i^t A_{i-1}^{-1}}{\mathbf{s}_i^t A_{i-1}^{-1}\mathbf{y}_i}. \end{aligned}$$

The program BROYM102 implements Broyden's method.

The calculation of A_i is bypassed, as is the necessity of solving the linear system.

Example 1 Use Broyden's method with $\mathbf{p}^{(0)} = (0.1, 0.1, -0.1)^t$ to approximate the solution to the nonlinear system

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0, \\ x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0, \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0. \end{aligned}$$

Solution This system was solved by Newton's method in Example 1 of Section 10.2. The Jacobian matrix for this system is

$$J(x_1, x_2, x_3) = \begin{bmatrix} 3 & x_3 \sin x_2x_3 & x_2 \sin x_2x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 \end{bmatrix}.$$

Let $\mathbf{p}^{(0)} = (0.1, 0.1, -0.1)^t$ and

$$\mathbf{F}(x_1, x_2, x_3) = (f_1(x_1, x_2, x_3),\ f_2(x_1, x_2, x_3),\ f_3(x_1, x_2, x_3))^t,$$

where

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2 x_3) - \frac{1}{2},$$
$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06,$$

and

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3}.$$

For $\mathbf{p}^{(0)} = (0.1, 0.1, -0.1)^t$ we have

$$\mathbf{F}(\mathbf{p}^{(0)}) = \begin{bmatrix} -1.199950 \\ -2.269833 \\ 8.462025 \end{bmatrix}.$$

This implies that

$$\begin{aligned} A_0 &= J(\mathbf{p}^{(0)}) \\ &= \begin{bmatrix} 3 & 9.999833 \times 10^{-4} & -9.999833 \times 10^{-4} \\ 0.2 & -32.4 & 0.9950042 \\ -9.900498 \times 10^{-2} & -9.900498 \times 10^{-2} & 20 \end{bmatrix}. \end{aligned}$$

For this first iteration, we need to find the inverse of $(J(\mathbf{p}^{(0)}))$. However, for subsequent iterations, matrix inversion is not necessary. We have

$$\begin{aligned} A_0^{-1} &= J\left(p_1^{(0)}, p_2^{(0)}, p_3^{(0)}\right)^{-1} \\ &= \begin{bmatrix} 0.3333332 & 1.023852 \times 10^{-5} & 1.615701 \times 10^{-5} \\ 2.108607 \times 10^{-3} & -3.086883 \times 10^{-2} & 1.535836 \times 10^{-3} \\ 1.660520 \times 10^{-3} & -1.527577 \times 10^{-4} & 5.000768 \times 10^{-2} \end{bmatrix}. \end{aligned}$$

So

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} - A_0^{-1}\mathbf{F}(\mathbf{p}^{(0)}) = \begin{bmatrix} 0.4998697 \\ 1.946685 \times 10^{-2} \\ -0.5215205 \end{bmatrix},$$

$$\mathbf{F}(\mathbf{p}^{(1)}) = \begin{bmatrix} -3.394465 \times 10^{-4} \\ -0.3443879 \\ 3.188238 \times 10^{-2} \end{bmatrix},$$

$$\mathbf{y}_1 = \mathbf{F}(\mathbf{p}^{(1)}) - \mathbf{F}(\mathbf{p}^{(0)}) = \begin{bmatrix} 1.199611 \\ 1.925445 \\ -8.430143 \end{bmatrix},$$

$$\mathbf{s}_1 = \begin{bmatrix} 0.3998697 \\ -8.053315 \times 10^{-2} \\ -0.4215204 \end{bmatrix},$$

$$\mathbf{s}_1^t A_0^{-1}\mathbf{y}_1 = 0.3424604,$$

$$A_1^{-1} = A_0^{-1} + \frac{1}{0.3424604}\left[(\mathbf{s}_1 - A_0^{-1}\mathbf{y}_1)\mathbf{s}_1^t A_0^{-1}\right]$$

$$= \begin{bmatrix} 0.3333781 & 1.11050\times 10^{-5} & 8.967344\times 10^{-6} \\ -2.021270\times 10^{-3} & -3.094849\times 10^{-2} & 2.196906\times 10^{-3} \\ 1.022214\times 10^{-3} & -1.650709\times 10^{-4} & 5.010986\times 10^{-2} \end{bmatrix},$$

and

$$\mathbf{p}^{(2)} = \mathbf{p}^{(1)} - A_1^{-1}\mathbf{F}(\mathbf{p}^{(1)}) = \begin{bmatrix} 0.4999863 \\ 8.737833\times 10^{-3} \\ -0.5231746 \end{bmatrix}.$$

Additional iterations are listed in Table 10.2. The 5th iteration of Broyden's method is slightly less accurate than was the 4th iteration of Newton's method given in the example at the end of the preceding section. ■

Table 10.2

k	$p_1^{(k)}$	$p_2^{(k)}$	$p_3^{(k)}$	$\Vert\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\Vert_2$
0	0.1000000	0.1000000	−0.1000000	
1	0.4998697	1.946685×10^{-2}	−0.5215205	5.93×10^{-1}
2	0.4999864	8.737839×10^{-3}	−0.5231746	1.0856×10^{-2}
3	0.5000066	8.672736×10^{-4}	−0.5235723	7.8806×10^{-3}
4	0.5000003	3.952827×10^{-5}	−0.5235977	8.2817×10^{-4}
5	0.5000000	1.934342×10^{-7}	−0.5235988	3.9351×10^{-5}

EXERCISE SET 10.3

1. Use Broyden's method with $\mathbf{p}^{(0)} = \mathbf{0}$ to compute $\mathbf{p}^{(2)}$ for each of the following nonlinear systems.

a.
$$4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0,$$
$$\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0.$$

b.
$$\sin(4\pi x_1x_2) - 2x_2 - x_1 = 0,$$
$$\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0.$$

c.
$$3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0,$$
$$4x_1^2 - 625x_2^2 + 2x_2 - 1 = 0,$$
$$e^{-x_1x_2} + 20x_3 + \frac{1}{3}(10\pi - 3) = 0.$$

d.
$$x_1^2 + x_2 - 37 = 0,$$
$$x_1 - x_2^2 - 5 = 0,$$
$$x_1 + x_2 + x_3 - 3 = 0.$$

2. Use Broyden's method to approximate solutions to the nonlinear systems in Exercise 1 using the following initial approximations $\mathbf{p}^{(0)}$ until $\Vert\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\Vert_\infty < 10^{-6}$.

a. $(0, 0)^t$ **b.** $(0, 0)^t$ **c.** $(1, 1, 1)^t$ **d.** $(2, 1, -1)^t$

3. Use Broyden's method to find a solution to the following nonlinear systems, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$.

   a. $$\begin{aligned} 3x_1^2 - x_2^2 &= 0 \\ 3x_1x_2^2 - x_1^3 - 1 &= 0 \end{aligned}$$
   Use $\mathbf{p}^{(0)} = (1, 1)^t$.

   b. $$\begin{aligned} \ln(x_1^2 + x_2^2) - \sin(x_1x_2) &= \ln 2 + \ln \pi \\ e^{x_1 - x_2} + \cos(x_1x_2) &= 0 \end{aligned}$$
   Use $\mathbf{p}^{(0)} = (2, 2)^t$.

   c. $$\begin{aligned} x_1^3 + x_1^2x_2 - x_1x_3 + 6 &= 0 \\ e^{x_1} + e^{x_2} - x_3 &= 0 \\ x_2^2 - 2x_1x_3 &= 4 \end{aligned}$$
   Use $\mathbf{p}^{(0)} = (-1, -2, 1)^t$.

   d. $$\begin{aligned} 6x_1 - 2\cos(x_2x_3) - 1 &= 0 \\ 9x_2 + \sqrt{x_1^2 + \sin x_3 + 1.06} + 0.9 &= 0 \\ 60x_3 + 3e^{-x_1x_2} + 10\pi - 3 &= 0 \end{aligned}$$
   Use $\mathbf{p}^{(0)} = (0, 0, 0)^t$.

4. The nonlinear system

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0, \\ x_1^2 - 625x_2^2 - \frac{1}{4} &= 0, \\ e^{-x_1x_2} + 20x_3 + \frac{1}{3}(10\pi - 3) &= 0 \end{aligned}$$

has a singular Jacobian matrix at the solution. Apply Broyden's method with $\mathbf{p}^{(0)} = (1, 1 - 1)^t$. Note that convergence may be slow or may not occur within a reasonable number of iterations.

5. The nonlinear system

$$\begin{aligned} 4x_1 - x_2 + x_3 &= x_1x_4, \\ -x_1 + 3x_2 - 2x_3 &= x_2x_4, \\ x_1 - 2x_2 + 3x_3 &= x_3x_4, \\ x_1^2 + x_2^2 + x_3^2 &= 1 \end{aligned}$$

has six solutions, and, as shown in Exercise 8 of Section 10.2, $(-x_1, -x_2, -x_3, x_4)$ is a solution whenever (x_1, x_2, x_3, x_4) is a solution. Use Broyden's method to approximate these solutions. Iterate until $\left\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\right\|_\infty < 10^{-5}$. Use the initial vectors $(1, 1, 1, 1)^t$, $(1, 0, 0, 0)^t$, and $(1, -1, 1, -1)^t$.

6. Show that if $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^n$, then $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$, where

$$\mathbf{z}_1 = \frac{\mathbf{y}^t\mathbf{z}}{\|\mathbf{y}\|_2^2}\mathbf{y}$$

is parallel to $\mathbf{y}$ and $\mathbf{z}_2 = \mathbf{z} - \mathbf{z}_1$ is orthogonal to $\mathbf{y}$.

7. Show that if $\mathbf{z}$ is orthogonal to $\mathbf{p}^{(1)} - \mathbf{p}^{(0)}$, then for A_1 defined in Eq. (10.1) we have $A_1 = J(\mathbf{p}^{(0)})$.

8. Let

$$A_1 = J(\mathbf{p}^{(0)}) + \frac{\left[\mathbf{F}(\mathbf{p}^{(1)}) - \mathbf{F}(\mathbf{p}^{(0)}) - J(\mathbf{p}^{(0)})(\mathbf{p}^{(1)} - \mathbf{p}^{(0)})\right](\mathbf{p}^{(1)} - \mathbf{p}^{(0)})^t}{\|\mathbf{p}^{(1)} - \mathbf{p}^{(0)}\|_2^2}.$$

   a. Show that $A_1(\mathbf{p}^{(1)} - \mathbf{p}^{(0)}) = \mathbf{F}(\mathbf{p}^{(1)}) - \mathbf{F}(\mathbf{p}^{(0)})$.

   b. Show that $A_1\mathbf{z} = J(\mathbf{p}^{(0)})\mathbf{z}$ whenever $(\mathbf{p}^{(1)} - \mathbf{p}^{(0)})^t\mathbf{z} = 0$.

9. It can be shown that if A^{-1} exists and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $(A+\mathbf{x}\mathbf{y}^t)^{-1}$ exists if and only if $\mathbf{y}^t A^{-1}\mathbf{x} \neq -1$. Use this result to verify the Sherman-Morrison formula: If A^{-1} exists and $\mathbf{y}^t A^{-1}\mathbf{x} \neq -1$, then $(A+\mathbf{x}\mathbf{y}^t)^{-1}$ exists, and

$$(A+\mathbf{x}\mathbf{y}^t)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{x}\mathbf{y}^t A^{-1}}{1+\mathbf{y}^t A^{-1}\mathbf{x}}.$$

10.4 The Steepest Descent Method

The name for the Steepest Descent method follows from the three-dimensional application of pointing in the steepest downward direction.

The advantage of the Newton and quasi-Newton methods for solving systems of nonlinear equations is their speed of convergence once a sufficiently accurate approximation is known. A weakness of these methods is that an accurate initial approximation to the solution is needed to ensure convergence. The **method of Steepest Descent** will generally converge only linearly to the solution, but it is global in nature, that is, nearly any starting value will give convergence. As a consequence, it is often used to find sufficiently accurate starting approximations for the Newton-based techniques.

The method of Steepest Descent determines a local minimum for a multivariable function of the form $g: \mathbb{R}^n \to \mathbb{R}$. The method is valuable quite apart from providing starting values for solving nonlinear systems, but we will consider only this application.

The connection between the minimization of a function from $\mathbb{R}^n$ to $\mathbb{R}$ and the solution of a system of nonlinear equations is due to the fact that a system of the form

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ f_2(x_1, x_2, \ldots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0, \end{aligned}$$

has a solution at $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ precisely when the function g from $\mathbb{R}^n$ to $\mathbb{R}$ defined by

$$g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2$$

has the minimal value zero.

The method of Steepest Descent for finding a local minimum for an arbitrary function g from $\mathbb{R}^n$ into $\mathbb{R}$ can be intuitively described as follows:

- Evaluate g at an initial approximation $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \ldots, p_n^{(0)})^t$.
- Determine a direction from $\mathbf{p}^{(0)}$ that results in a decrease in the value of g.
- Move an appropriate amount in this direction and call the new value $\mathbf{p}^{(1)}$.
- Repeat the steps with $\mathbf{p}^{(0)}$ replaced by $\mathbf{p}^{(1)}$.

The Gradient of a Function

Before describing how to choose the correct direction and the appropriate distance to move in this direction, we need to review some results from calculus. The Extreme Value Theorem implies that a differentiable single-variable function can have a relative minimum within

The root of *gradient* comes from the Latin word *gradi*, meaning "to walk". In this sense, the gradient of a surface is the rate at which it "walks uphill".

the interval only when the derivative is zero. To extend this result to multivariable functions, we need the following definition.

If $g : \mathbb{R}^n \to \mathbb{R}$, we define the **gradient** of g at $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$, $\nabla g(\mathbf{x})$, by

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}(\mathbf{x}), \frac{\partial g}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial g}{\partial x_n}(\mathbf{x}) \right)^t .$$

The gradient for a multivariable function is analogous to the derivative of a single variable function in the sense that a differentiable multivariable function can have a relative minimum at $\mathbf{x}$ only when the gradient at $\mathbf{x}$ is the zero vector.

The gradient has another important property connected with the minimization of multivariable functions. Suppose $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ is a vector in $\mathbb{R}^n$ with $\|\mathbf{v}\|_2 = 1$. The **directional derivative** of g at $\mathbf{x}$ in the direction of $\mathbf{v}$ is defined by

$$D_{\mathbf{v}} g(\mathbf{x}) = \lim_{h \to 0} \frac{1}{h} \left[g(\mathbf{x} + h\mathbf{v}) - g(\mathbf{x}) \right] = \mathbf{v}^t \cdot \nabla g(\mathbf{x}) = \sum_{i=1}^{n} v_i \frac{\partial g}{\partial x_i}(\mathbf{x}).$$

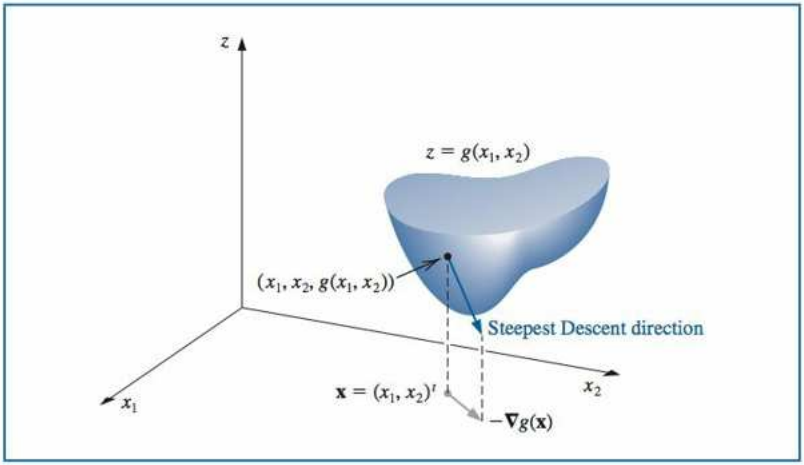
The directional derivative of g at $\mathbf{x}$ in the direction of $\mathbf{v}$ measures the change in the value of the function g relative to the change in the variable in the direction of $\mathbf{v}$.

A standard result from the calculus of multivariable functions states that the direction that produces the maximum increase for the directional derivative occurs when $\mathbf{v}$ is chosen in the direction of $\nabla g(\mathbf{x})$, provided that $\nabla g(\mathbf{x}) \neq \mathbf{0}$. So the maximum decrease will be in the direction of $-\nabla g(\mathbf{x})$.

- The direction of greatest decrease in the value of g at $\mathbf{x}$ is the direction given by $-\nabla g(\mathbf{x})$.

See Figure 10.2 for an illustration when g is a function of two variables.

Figure 10.2


z
$z = g(x_1, x_2)$
$(x_1, x_2, g(x_1, x_2))$
Steepest Descent direction
x_1
$\mathbf{x} = (x_1, x_2)^t$
x_2
$-\nabla g(\mathbf{x})$

The objective is to reduce $g(\mathbf{x})$ to its minimal value of zero, so given the initial approximation $\mathbf{p}^{(0)}$, we choose

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} - \alpha \nabla g(\mathbf{p}^{(0)}) \tag{10.2}$$

for some constant $\alpha > 0$.

The problem now reduces to choosing α so that $g(\mathbf{p}^{(1)})$ will be significantly less than $g(\mathbf{p}^{(0)})$. To determine an appropriate choice for the value α, we consider the single-variable function

$$h(\alpha) = g(\mathbf{p}^{(0)} - \alpha \nabla g(\mathbf{p}^{(0)})).$$

The value of α that minimizes h is the value needed for $\mathbf{p}^{(1)} = \mathbf{p}^{(0)} - \alpha \nabla g(\mathbf{p}^{(0)})$.

Finding a minimal value for h directly would require differentiating h and then solving a root-finding problem to determine the critical points of h. This procedure is generally too costly. Instead, we choose three numbers $\alpha_1 < \alpha_2 < \alpha_3$ that, we hope, are close to where the minimum value of $h(\alpha)$ occurs. Then we construct the quadratic polynomial $P(x)$ that interpolates h at α_1, α_2, and α_3. We define $\hat{\alpha}$ in $[\alpha_1, \alpha_3]$ so that $P(\hat{\alpha})$ is a minimum in $[\alpha_1, \alpha_3]$ and use $P(\hat{\alpha})$ to approximate the minimal value of $h(\alpha)$. Then $\hat{\alpha}$ is used to determine the new iterate for approximating the minimal value of g:

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} - \hat{\alpha} \nabla g(\mathbf{p}^{(0)}).$$

Since $g(\mathbf{p}^{(0)})$ is available, we first choose $\alpha_1 = 0$ to minimize the computation. Next a number α_3 is found with $h(\alpha_3) < h(\alpha_1)$. (Since α_1 does not minimize h, such a number α_3 does exist.) Finally, α_2 is chosen to be $\alpha_3/2$.

The minimum value $\hat{\alpha}$ of $P(x)$ on $[\alpha_1, \alpha_3]$ occurs either at the only critical point of P or at the right endpoint α_3 because, by assumption, $P(\alpha_3) = h(\alpha_3) < h(\alpha_1) = P(\alpha_1)$. The critical point is easily determined because $P(x)$ is a quadratic polynomial.

Program STPDC103 applies the method of Steepest Descent to approximate the minimal value of $g(\mathbf{x})$. To begin each iteration, the value 0 is assigned to α_1, and the value 1 is assigned to α_3. If $h(\alpha_3) \geq h(\alpha_1)$, then successive divisions of α_3 by 2 are performed and the value of α_3 is reassigned until $h(\alpha_3) < h(\alpha_1)$.

The program STPDC103 implements the Steepest Descent method.

To employ the method to approximate the solution to the system

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ f_2(x_1, x_2, \ldots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0, \end{aligned}$$

we simply replace the function g with $\sum_{i=1}^{n} f_i^2$.

Example 1 Use the Steepest Descent method with $\mathbf{p}^{(0)} = (0, 0, 0)^t$ to find a reasonable starting approximation to the solution of the nonlinear system

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2 x_3) - \frac{1}{2} = 0, \\ f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0, \\ f_3(x_1, x_2, x_3) &= e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0. \end{aligned}$$

Solution Let $g(x_1, x_2, x_3) = [f_1(x_1, x_2, x_3)]^2 + [f_2(x_1, x_2, x_3)]^2 + [f_3(x_1, x_2, x_3)]^2$. Then

$$\begin{aligned}\nabla g(x_1, x_2, x_3) \equiv \nabla g(\mathbf{x}) = \Big(&2f_1(\mathbf{x})\frac{\partial f_1}{\partial x_1}(\mathbf{x}) + 2f_2(\mathbf{x})\frac{\partial f_2}{\partial x_1}(\mathbf{x}) + 2f_3(\mathbf{x})\frac{\partial f_3}{\partial x_1}(\mathbf{x}),\\ &2f_1(\mathbf{x})\frac{\partial f_1}{\partial x_2}(\mathbf{x}) + 2f_2(\mathbf{x})\frac{\partial f_2}{\partial x_2}(\mathbf{x}) + 2f_3(\mathbf{x})\frac{\partial f_3}{\partial x_2}(\mathbf{x}),\\ &2f_1(\mathbf{x})\frac{\partial f_1}{\partial x_3}(\mathbf{x}) + 2f_2(\mathbf{x})\frac{\partial f_2}{\partial x_3}(\mathbf{x}) + 2f_3(\mathbf{x})\frac{\partial f_3}{\partial x_3}(\mathbf{x})\Big)\\ = {}& 2\mathbf{J}(\mathbf{x})^t\mathbf{F}(\mathbf{x}).\end{aligned}$$

For $\mathbf{p}^{(0)} = (0, 0, 0)^t$, we have

$$\begin{aligned}g(\mathbf{p}^{(0)}) &= f_1(0,0,0)^2 + f_2(0,0,0)^2 + f_3(0,0,0)^2\\ &= \left(-\frac{3}{2}\right)^2 + (-81(0.01) + 1.06)^2 + \left(\frac{10\pi}{3}\right)^2 = 111.975,\end{aligned}$$

and

$$z_0 = \|\nabla g(\mathbf{p}^{(0)})\|_2 = \|2\mathbf{J}(\mathbf{0})^t\mathbf{F}(\mathbf{0})\|_2 = 419.554.$$

Let

$$\mathbf{z} = \frac{1}{z_0}\nabla g(\mathbf{p}^{(0)}) = (-0.0214514, -0.0193062, 0.999583)^t.$$

With $\alpha_1 = 0$, we have $g_1 = g(\mathbf{p}^{(0)} - \alpha_1\mathbf{z}) = g(\mathbf{p}^{(0)}) = 111.975$. We arbitrarily let $\alpha_3 = 1$ so that

$$g_3 = g(\mathbf{p}^{(0)} - \alpha_3\mathbf{z}) = 93.5649.$$

Because $g_3 < g_1$, we accept α_3 and set $\alpha_2 = \alpha_3/2 = 0.5$. Evaluating g at $\mathbf{p}^{(0)} - \alpha_2\mathbf{z}$ gives

$$g_2 = g(\mathbf{p}^{(0)} - \alpha_2\mathbf{z}) = 2.53557.$$

We now find the quadratic polynomial that interpolates the data $(0, 111.975)$, $(1, 93.5649)$, and $(0.5, 2.53557)$. It is most convenient to use Newton's forward divided-difference interpolating polynomial for this purpose, which has the form

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2).$$

This interpolates

$$g(\mathbf{p}^{(0)} - \alpha\nabla g(\mathbf{p}^{(0)})) = g(\mathbf{p}^{(0)} - \alpha\mathbf{z})$$

at $\alpha_1 = 0$, $\alpha_2 = 0.5$, and $\alpha_3 = 1$ as follows:

$$\begin{aligned}&\alpha_1 = 0, && g_1 = 111.975,\\ &\alpha_2 = 0.5, && g_2 = 2.53557, && h_1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1} = -218.878,\\ &\alpha_3 = 1, && g_3 = 93.5649, && h_2 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2} = 182.059, && h_3 = \frac{h_2 - h_1}{\alpha_3 - \alpha_1} = 400.937.\end{aligned}$$

This gives

$$P(\alpha) = 111.975 - 218.878\alpha + 400.937\alpha(\alpha - 0.5) = 400.937\alpha^2 - 419.346\alpha + 111.975$$

so

$$P'(\alpha) = 801.874\alpha - 419.346$$

and $P'(\alpha) = 0$ when $\alpha = \alpha_0 = 0.522959$. Since

$$g(\mathbf{p}^{(0)} - \alpha_0\mathbf{z}) = 2.32762$$

is smaller than g_1 and g_3, we set

$$\mathbf{p}^{(1)} = \mathbf{p}^{(0)} - \alpha_0\mathbf{z} = \mathbf{p}^{(0)} - 0.522959\mathbf{z} = (0.0112182, 0.0100964, -0.522741)^t$$

and

$$g(\mathbf{p}^{(1)}) = 2.32762.$$

Table 10.3 contains the remainder of the results. A true solution is $\mathbf{p} = (0.5, 0, -0.5235988)^t$, so $\mathbf{p}^{(2)}$ would likely be adequate as an initial approximation for Newton's method or Broyden's method. One of these quicker converging techniques would be appropriate at this stage because 70 iterations of the Steepest Descent method are required to find $\|\mathbf{p}^{(k)} - \mathbf{p}\|_\infty < 0.01$. ■

Table 10.3

k	$p_1^{(k)}$	$p_2^{(k)}$	$p_3^{(k)}$	$g(p_1^{(k)}, p_2^{(k)}, p_3^{(k)})$
2	0.137860	−0.205453	−0.522059	1.27406
3	0.266959	0.00551102	−0.558494	1.06813
4	0.272734	−0.00811751	−0.522006	0.468309
5	0.308689	−0.0204026	−0.533112	0.381087
6	0.314308	−0.0147046	−0.520923	0.318837
7	0.324267	−0.00852549	−0.528431	0.287024

EXERCISE SET 10.4

1. Use the method of Steepest Descent to approximate a solution of the following nonlinear systems, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 0.05$.

   a. $4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0$
   $\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0$

   b. $3x_1^2 - x_2^2 = 0$
   $3x_1x_2^2 - x_1^3 - 1 = 0$

   c. $\ln(x_1^2 + x_2^2) - \sin(x_1x_2) = \ln 2 + \ln \pi$
   $e^{x_1 - x_2} + \cos(x_1x_2) = 0$

   d. $\sin(4\pi x_1x_2) - 2x_2 - x_1 = 0$
   $\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0$

2. Use the results in Exercise 1 and Newton's method to approximate the solutions of the nonlinear systems in Exercise 1, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$.

3. Use the method of Steepest Descent to approximate a solution of the following nonlinear systems, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 0.05$.

   a. $$\begin{aligned} 15x_1 + x_2^2 - 4x_3 &= 13 \\ x_1^2 + 10x_2 - x_3 &= 11 \\ x_2^3 - 25x_3 &= -22 \end{aligned}$$

   b. $$\begin{aligned} 10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 &= 0 \\ 8x_2^2 + 4x_3^2 - 9 &= 0 \\ 8x_2x_3 + 4 &= 0 \end{aligned}$$

   c. $$\begin{aligned} x_1^3 + x_1^2x_2 - x_1x_3 + 6 &= 0 \\ e^{x_1} + e^{x_2} - x_3 &= 0 \\ x_2^2 - 2x_1x_3 &= 4 \end{aligned}$$

   d. $$\begin{aligned} x_1 + \cos(x_1x_2x_3) - 1 &= 0 \\ (1 - x_1)^{1/4} + x_2 + 0.05x_3^2 - 0.15x_3 - 1 &= 0 \\ -x_1^2 - 0.1x_2^2 + 0.01x_2 + x_3 - 1 &= 0 \end{aligned}$$

4. Use the results of Exercise 3 and Newton's method to approximate the solutions of the nonlinear systems in Exercise 3, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 10^{-6}$.

5. Use the method of Steepest Descent to approximate minima for the following functions, iterating until $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_\infty < 0.005$.

   a. $g(x_1, x_2) = \cos(x_1 + x_2) + \sin x_1 + \cos x_2$

   b. $g(x_1, x_2) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$

   c. $g(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 + 2x_1 - 2.5x_2 - x_3 + 2$

   d. $g(x_1, x_2, x_3) = x_1^4 + 2x_2^4 + 3x_3^4 + 1.01$

6. a. Show that the quadratic polynomial that interpolates the function

$$h(\alpha) = g(\mathbf{p}^{(0)} - \alpha \nabla g(\mathbf{p}^{(0)}))$$

   at $\alpha = 0, \alpha_2$, and α_3 is

$$P(\alpha) = g(\mathbf{p}^{(0)}) + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$$

   where

$$h_1 = \frac{g(\mathbf{p}^{(0)} - \alpha_2\mathbf{z}) - g(\mathbf{p}^{(0)})}{\alpha_2},$$

$$h_2 = \frac{g(\mathbf{p}^{(0)} - \alpha_3\mathbf{z}) - g(\mathbf{p}^{(0)} - \alpha_2\mathbf{z})}{\alpha_3 - \alpha_2}, \quad \text{and} \quad h_3 = \frac{h_2 - h_1}{\alpha_3}$$

   b. Show that the only critical point of P occurs at $\alpha_0 = 0.5(\alpha_2 - h_1/h_3)$.

10.5 Homotopy and Continuation Methods

Homotopy, or *continuation*, methods for nonlinear systems embed the problem to be solved within a collection of problems. Specifically, to solve a problem of the form

$$\mathbf{F}(\mathbf{x}) = \mathbf{0},$$

which has the unknown solution $\mathbf{p}$, we consider a family of problems described using a parameter λ that assumes values in $[0, 1]$. A problem with a known solution $\mathbf{x}(0)$ corresponds to $\lambda = 0$, and the problem with the unknown solution $\mathbf{x}(1) \equiv \mathbf{p}$ corresponds to $\lambda = 1$.

Suppose $\mathbf{x}(0)$ is an initial approximation to the solution $\mathbf{p}$ of $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. Define

$$\mathbf{G} : [0, 1] \times \mathbb{R}^n \to \mathbb{R}^n$$

by

$$\mathbf{G}(\lambda, \mathbf{x}) = \lambda\mathbf{F}(\mathbf{x}) + (1 - \lambda)\,[\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}(0))] = \mathbf{F}(\mathbf{x}) + (\lambda - 1)\mathbf{F}(\mathbf{x}(0)).$$