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

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

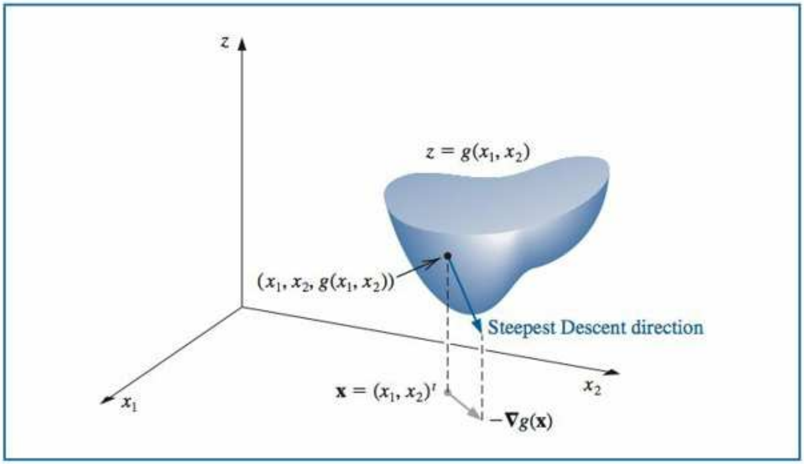
The directional derivative of g at $\mathbf{x}$ in the direction of $\mathbf{v}$ measures the change in the value of the function g relative to the change in the variable in the direction of $\mathbf{v}$.

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



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.

The program STPDC103 implements the Steepest Descent method.

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 $h(\alpha_3) < h(\alpha_1)$.

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

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

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

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