$Q(\boldsymbol{\beta})$ is assumed to be twice continuously differentiable. Given any initial value of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_{(0)}$, we can perform a second-order Taylor expansion of $Q(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_{(0)}$ in order to obtain an approximation $Q^*(\boldsymbol{\beta})$ to $Q(\boldsymbol{\beta})$:

$$Q^*(\boldsymbol{\beta}) = Q(\boldsymbol{\beta}_{(0)}) + \boldsymbol{g}_{(0)}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)})^{\top}\boldsymbol{H}_{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}), \qquad (6.41)$$

where $\boldsymbol{g}(\boldsymbol{\beta})$, the **gradient** of $Q(\boldsymbol{\beta})$, is a column vector of dimension k with typical element $\partial Q(\boldsymbol{\beta})/\partial\beta_i$, and $\boldsymbol{H}(\boldsymbol{\beta})$, the **Hessian** of $Q(\boldsymbol{\beta})$, is a $k \times k$ matrix with typical element $\partial^2 Q(\boldsymbol{\beta})/\partial\beta_i\partial\beta_l$. For notational simplicity, $\boldsymbol{g}_{(0)}$ and $\boldsymbol{H}_{(0)}$ denote $\boldsymbol{g}(\boldsymbol{\beta}_{(0)})$ and $\boldsymbol{H}(\boldsymbol{\beta}_{(0)})$, respectively.

It is easy to see that the first-order conditions for a minimum of $Q^*(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{g}_{(0)} + \boldsymbol{H}_{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}) = \mathbf{0}.$$

Solving these yields a new value of $\boldsymbol{\beta}$, which we will call $\boldsymbol{\beta}_{(1)}$:

$$\boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(0)} - \boldsymbol{H}_{(0)}^{-1}\boldsymbol{g}_{(0)}. \qquad (6.42)$$

Equation (6.42) is the heart of Newton's Method. If the quadratic approximation $Q^*(\boldsymbol{\beta})$ is a strictly convex function, which it is if and only if the Hessian $\boldsymbol{H}_{(0)}$ is positive definite, then $\boldsymbol{\beta}_{(1)}$ is the global minimum of $Q^*(\boldsymbol{\beta})$. If, in addition, $Q^*(\boldsymbol{\beta})$ is a good approximation to $Q(\boldsymbol{\beta})$, $\boldsymbol{\beta}_{(1)}$ should be close to $\hat{\boldsymbol{\beta}}$, the minimum of $Q(\boldsymbol{\beta})$. Newton's Method involves using equation (6.42) repeatedly to find a succession of values $\boldsymbol{\beta}_{(1)}$, $\boldsymbol{\beta}_{(2)}, \ldots$. When the original function $Q(\boldsymbol{\beta})$ is quadratic and has a global minimum at $\hat{\boldsymbol{\beta}}$, Newton's Method evidently finds $\hat{\boldsymbol{\beta}}$ in a single step, since the quadratic approximation is then exact. When $Q(\boldsymbol{\beta})$ is approximately quadratic, as all sum-of-squares functions are when sufficiently close to their minima, Newton's Method generally converges very quickly.

Figure 6.1 illustrates how Newton's Method works. It shows the contours of the function $Q(\boldsymbol{\beta}) = \text{SSR}(\beta_1, \beta_2)$ for a regression model with two parameters. Notice that these contours are not precisely elliptical, as they would be if the function were quadratic. The algorithm starts at the point marked "0" and then jumps to the point marked "1." On the next step, it goes in almost exactly the right direction, but it goes too far, moving to "2." It then retraces its own steps to "3," which is essentially the minimum of $\text{SSR}(\beta_1, \beta_2)$. After one more step, which is too small to be shown in the figure, it has essentially converged.

Although Newton's Method works very well in this example, there are many cases in which it fails to work at all, especially if $Q(\boldsymbol{\beta})$ is not convex in the neighborhood of $\boldsymbol{\beta}_{(j)}$ for some j in the sequence. Some of the possibilities are illustrated in Figure 6.2. The one-dimensional function shown there has