Therefore,

$$\mathbf{X}_4 = \mathbf{X}_3 + \lambda_3^* \mathbf{S}_3 = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix} + 1.0 \begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix} = \begin{Bmatrix} -1.0 \\ 1.4 \end{Bmatrix}$$

The gradient at $\mathbf{X}_4$ is given by

$$\nabla f_4 = \begin{Bmatrix} -0.20 \\ -0.20 \end{Bmatrix}$$

Since $\nabla f_4 \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$, $\mathbf{X}_4$ is not optimum and hence we have to proceed to the next iteration. This process has to be continued until the optimum point, $\mathbf{X}^* = \begin{Bmatrix} -1.0 \\ 1.5 \end{Bmatrix}$, is found.

Convergence Criteria. The following criteria can be used to terminate the iterative process:

1. When the change in function value in two consecutive iterations is small:

$$\left| \frac{f(\mathbf{X}_{i+1}) - f(\mathbf{X}_i)}{f(\mathbf{X}_i)} \right| \leq \varepsilon_1 \tag{6.71}$$

2. When the partial derivatives (components of the gradient) of f are small:

$$\left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon_2, \qquad i = 1,2,\ldots,n \tag{6.72}$$

3. When the change in the design vector in two consecutive iterations is small:

$$|\mathbf{X}_{i+1} - \mathbf{X}_i| \leq \varepsilon_3 \tag{6.73}$$

6.12 CONJUGATE GRADIENT (FLETCHER-REEVES) METHOD

The convergence characteristics of the steepest descent method can be improved greatly by modifying it into a conjugate gradient method (which can be considered as a conjugate directions method involving the use of the gradient of the function). We saw (in Section 6.7) that any minimization method that makes use of the conjugate directions is quadratically convergent. This property of quadratic convergence is very useful because it ensures that the method will minimize a quadratic function in n steps or less. Since any general function can be approximated reasonably well by a quadratic near the optimum point, any quadratically convergent method is expected to find the optimum point in a finite number of iterations.

We have seen that Powell's conjugate direction method requires n single-variable minimizations per iteration and sets up a new conjugate direction at the end of each iteration. Thus it requires, in general, n^2 single-variable minimizations to find the minimum of a quadratic function. On the other hand, if we can evaluate the gradients of the objective function, we can set up a new conjugate direction after every one-dimensional minimization, and hence we can achieve faster convergence. The construction of conjugate directions and development of the Fletcher–Reeves method are discussed in this section.

6.12.1 Development of the Fletcher–Reeves Method

Consider the development of an algorithm by modifying the steepest descent method applied to a quadratic function $f(\mathbf{X}) = \frac{1}{2}\mathbf{X}^T\mathbf{A}\mathbf{X} + \mathbf{B}^T\mathbf{X} + C$ by imposing the condition that the successive directions be mutually conjugate. Let $\mathbf{X}_1$ be the starting point for the minimization and let the first search direction be the steepest descent direction:

$$\mathbf{S}_1 = -\nabla f_1 = -\mathbf{A}\mathbf{X}_1 - \mathbf{B} \tag{6.74}$$

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^*\mathbf{S}_1 \tag{6.75}$$

or

$$\mathbf{S}_1 = \frac{\mathbf{X}_2 - \mathbf{X}_1}{\lambda_1^*} \tag{6.76}$$

where λ_1^* is the minimizing step length in the direction $\mathbf{S}_1$, so that

$$\mathbf{S}_1^T \nabla f|_{\mathbf{X}_2} = 0 \tag{6.77}$$

Equation (6.77) can be expanded as

$$\mathbf{S}_1^T[\mathbf{A}(\mathbf{X}_1 + \lambda_1^*\mathbf{S}_1) + \mathbf{B}] = 0 \tag{6.78}$$

from which the value of λ_1^* can be found as

$$\lambda_1^* = \frac{-\mathbf{S}_1^T(\mathbf{A}\mathbf{X}_1 + \mathbf{B})}{\mathbf{S}_1^T\mathbf{A}\mathbf{S}_1} = -\frac{\mathbf{S}_1^T}{\mathbf{S}_1^T}\frac{\nabla f_1}{\mathbf{A}\mathbf{S}_1} \tag{6.79}$$

Now express the second search direction as a linear combination of $\mathbf{S}_1$ and $-\nabla f_2$:

$$\mathbf{S}_2 = -\nabla f_2 + \beta_2\mathbf{S}_1 \tag{6.80}$$

where β_2 is to be chosen so as to make $\mathbf{S}_1$ and $\mathbf{S}_2$ conjugate. This requires that

$$\mathbf{S}_1^T\mathbf{A}\mathbf{S}_2 = 0 \tag{6.81}$$

Substituting Eq. (6.80) into Eq. (6.81) leads to

$$\mathbf{S}_1^T\mathbf{A}\,(-\nabla f_2 + \beta_2\mathbf{S}_1) = 0 \tag{6.82}$$

Equations (6.76) and (6.82) yield

$$-\frac{(\mathbf{X}_2 - \mathbf{X}_1)^T}{\lambda_1^*}\,\mathbf{A}\,(\nabla f_2 - \beta_2\mathbf{S}_1) = 0 \tag{6.83}$$

The difference of the gradients $(\nabla f_2 - \nabla f_1)$ can be expressed as

$$(\nabla f_2 - \nabla f_1) = (\mathbf{A}\mathbf{X}_2 + \mathbf{B}) - (\mathbf{A}\mathbf{X}_1 + \mathbf{B}) = \mathbf{A}(\mathbf{X}_2 - \mathbf{X}_1) \tag{6.84}$$

With the help of Eq. (6.84), Eq. (6.83) can be written as

$$(\nabla f_2 - \nabla f_1)^T(\nabla f_2 - \beta_2\mathbf{S}_1) = 0 \tag{6.85}$$

where the symmetricity of the matrix $\mathbf{A}$ has been used. Equation (6.85) can be expanded as

$$\nabla f_2^T\,\nabla f_2 - \nabla f_1^T\,\nabla f_2 - \beta_2\,\nabla f_2^T\,\mathbf{S}_1 + \beta_2\,\nabla f_1^T\,\mathbf{S}_1 = 0 \tag{6.86}$$

Since $\nabla f_1^T\,\nabla f_2 = -\mathbf{S}_1^T\,\nabla f_2 = 0$ from Eq. (6.77), Eq. (6.86) gives

$$\beta_2 = -\frac{\nabla f_2^T\,\nabla f_2}{\nabla f_1^T\,\mathbf{S}_1} = \frac{\nabla f_2^T\,\nabla f_2}{\nabla f_1^T\,\nabla f_1} \tag{6.87}$$

Next we consider the third search direction as a linear combination of $\mathbf{S}_1$, $\mathbf{S}_2$, and $-\nabla f_3$ as

$$\mathbf{S}_3 = -\nabla f_3 + \beta_3\mathbf{S}_2 + \delta_3\mathbf{S}_1 \tag{6.88}$$

where the values of β_3 and δ_3 can be found by making $\mathbf{S}_3$ conjugate to $\mathbf{S}_1$ and $\mathbf{S}_2$. By using the condition $\mathbf{S}_1^T\mathbf{A}\mathbf{S}_3 = 0$, the value of δ_3 can be found to be zero (see Problem 6.40). When the condition $\mathbf{S}_2^T\mathbf{A}\mathbf{S}_3 = 0$ is used, the value of β_3 can be obtained as (see Problem 6.41)

$$\beta_3 = \frac{\nabla f_3^T\,\nabla f_3}{\nabla f_2^T\,\nabla f_2} \tag{6.89}$$

so that Eq. (6.88) becomes

$$\mathbf{S}_3 = -\nabla f_3 + \beta_3\mathbf{S}_2 \tag{6.90}$$

where β_3 is given by Eq. (6.89). In fact, Eq. (6.90) can be generalized as

$$\mathbf{S}_i = -\nabla f_i + \beta_i \mathbf{S}_{i-1} \tag{6.91}$$

where

$$\beta_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_{i-1}^T \nabla f_{i-1}} \tag{6.92}$$

Equations (6.91) and (6.92) define the search directions used in the Fletcher–Reeves method [6.13].

6.12.2 Fletcher–Reeves Method

The iterative procedure of Fletcher–Reeves method can be stated as follows:

1. Start with an arbitrary initial point $\mathbf{X}_1$.
2. Set the first search direction $\mathbf{S}_1 = -\nabla f(\mathbf{X}_1) = -\nabla f_1$.
3. Find the point $\mathbf{X}_2$ according to the relation

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^* \mathbf{S}_1$$

   where λ_1^* is the optimal step length in the direction $\mathbf{S}_1$. Set $i = 2$ and go to the next step.
4. Find $\nabla f_i = \nabla f(\mathbf{X}_i)$, and set

$$\mathbf{S}_i = -\nabla f_i + \frac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2} \mathbf{S}_{i-1} \tag{6.93}$$

5. Compute the optimum step length λ_i^* in the direction $\mathbf{S}_i$, and find the new point

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda_i^* \mathbf{S}_i \tag{6.94}$$

6. Test for the optimality of the point $\mathbf{X}_{i+1}$. If $\mathbf{X}_{i+1}$ is optimum, stop the process. Otherwise, set the value of $i = i + 1$ and go to step 4.

Remarks:

1. The Fletcher–Reeves method was originally proposed by Hestenes and Stiefel [6.14] as a method for solving systems of linear equations derived from the stationary conditions of a quadratic. Since the directions $\mathbf{S}_i$ used in this method are **A**-conjugate, the process should converge in n cycles or less for a quadratic function. However, for ill-conditioned quadratics (whose contours are highly eccentric and distorted), the method may

require much more than n cycles for convergence. The reason for this has been found to be the cumulative effect of rounding errors. Since $\mathbf{S}_i$ is given by Eq. (6.93), any error resulting from the inaccuracies involved in the determination of λ_i^*, and from the round-off error involved in accumulating the successive $|\nabla f_i|^2 \mathbf{S}_{i-1}/|\nabla f_{i-1}|^2$ terms, is carried forward through the vector $\mathbf{S}_i$. Thus the search directions $\mathbf{S}_i$ will be progressively contaminated by these errors. Hence it is necessary, in practice, to restart the method periodically after every, say, m steps by taking the new search direction as the steepest descent direction. That is, after every m steps, $\mathbf{S}_{m+1}$ is set equal to $-\nabla f_{m+1}$ instead of the usual form. Fletcher and Reeves have recommended a value of $m = n + 1$, where n is the number of design variables.

2. Despite the limitations indicated above, the Fletcher–Reeves method is vastly superior to the steepest descent method and the pattern search methods, but it turns out to be rather less efficient than the Newton and the quasi-Newton (variable metric) methods discussed in the latter sections.

Example 6.10 Minimize $f(x_1,x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from the point $\mathbf{X}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$.

SOLUTION

Iteration 1

$$\nabla f = \begin{Bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{Bmatrix} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}$$

$$\nabla f_1 = \nabla f(\mathbf{X}_1) = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

The search direction is taken as $\mathbf{S}_1 = -\nabla f_1 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$. To find the optimal step length λ_1^* along $\mathbf{S}_1$, we minimize $f(\mathbf{X}_1 + \lambda_1 \mathbf{S}_1)$ with respect to λ_1. Here

$$f(\mathbf{X}_1 + \lambda_1 \mathbf{S}_1) = f(-\lambda_1, +\lambda_1) = \lambda_1^2 - 2\lambda_1$$

$$\frac{df}{d\lambda_1} = 0 \quad \text{at} \quad \lambda_1^* = 1$$

Therefore,

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^* \mathbf{S}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + 1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

Iteration 2

Since $\nabla f_2 = \nabla f(\mathbf{X}_2) = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix}$, Eq. (6.93) gives the next search direction as

$$\mathbf{S}_2 = -\nabla f_2 + \frac{|\nabla f_2|^2}{|\nabla f_1|^2}\mathbf{S}_1$$

where

$$|\nabla f_1|^2 = 2 \quad \text{and} \quad |\nabla f_2|^2 = 2$$

Therefore,

$$\mathbf{S}_2 = -\begin{Bmatrix} -1 \\ -1 \end{Bmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ +2 \end{Bmatrix}$$

To find λ_2^*, we minimize

$$\begin{aligned} f(\mathbf{X}_2 + \lambda_2 \mathbf{S}_2) &= f(-1, 1 + 2\lambda_2) \\ &= -1 - (1 + 2\lambda_2) + 2 - 2(1 + 2\lambda_2) + (1 + 2\lambda_2)^2 \\ &= 4\lambda_2^2 - 2\lambda_2 - 1 \end{aligned}$$

with respect to λ_2. As $df/d\lambda_2 = 8\lambda_2 - 2 = 0$ at $\lambda_2^* = \frac{1}{4}$, we obtain

$$\mathbf{X}_3 = \mathbf{X}_2 + \lambda_2^* \mathbf{S}_2 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + \frac{1}{4}\begin{Bmatrix} 0 \\ 2 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1.5 \end{Bmatrix}$$

Thus the optimum point is reached in two iterations. Even if we do not know this point to be optimum, we will not be able to move from this point in the next iteration. This can be verified as follows.

Iteration 3

Now

$$\nabla f_3 = \nabla f(\mathbf{X}_3) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \quad |\nabla f_2|^2 = 2, \quad \text{and} \quad |\nabla f_3|^2 = 0.$$

Thus

$$\mathbf{S}_3 = -\nabla f_3 + (|\nabla f_3|^2/|\nabla f_2|^2)\mathbf{S}_2 = -\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

This shows that there is no search direction to reduce f further, and hence $\mathbf{X}_3$ is optimum.

6.13 NEWTON'S METHOD

Newton's method presented in Section 5.12.1 can be extended for the minimization of multivariable functions. For this, consider the quadratic approximation of the function $f(\mathbf{X})$ at $\mathbf{X} = \mathbf{X}_i$ using the Taylor's series expansion

$$f(\mathbf{X}) = f(\mathbf{X}_i) + \nabla f_i^T(\mathbf{X} - \mathbf{X}_i) + \tfrac{1}{2}(\mathbf{X} - \mathbf{X}_i)^T[J_i](\mathbf{X} - \mathbf{X}_i) \quad (6.95)$$

where $[J_i] = [J]|_{\mathbf{X}_i}$ is the matrix of second partial derivatives (Hessian matrix) of f evaluated at the point $\mathbf{X}_i$. By setting the partial derivatives of Eq. (6.95) equal to zero for the minimum of $f(\mathbf{X})$, we obtain

$$\frac{\partial f(\mathbf{X})}{\partial x_j} = 0, \qquad j = 1,2,\ldots,n \quad (6.96)$$

Equations (6.96) and (6.95) give

$$\nabla f = \nabla f_i + [J_i](\mathbf{X} - \mathbf{X}_i) = \mathbf{0} \quad (6.97)$$

If $[J_i]$ is nonsingular, Eqs. (6.97) can be solved to obtain an improved approximation ($\mathbf{X} = \mathbf{X}_{i+1}$) as

$$\mathbf{X}_{i+1} = \mathbf{X}_i - [J_i]^{-1}\nabla f_i \quad (6.98)$$

Since higher-order terms have been neglected in Eq. (6.95), Eq. (6.98) is to be used iteratively to find the optimum solution $\mathbf{X}^*$.

The sequence of points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i+1}$ can be shown to converge to the actual solution $\mathbf{X}^*$ from any initial point $\mathbf{X}_1$ sufficiently close to the solution $\mathbf{X}^*$, provided that $[J_1]$ is nonsingular. It can be seen that Newton's method uses the second partial derivatives of the objective function (in the form of the matrix $[J_i]$) and hence is a second-order method.

Example 6.11 Show that the Newton's method finds the minimum of a quadratic function in one iteration.

SOLUTION Let the quadratic function be given by

$$f(\mathbf{X}) = \tfrac{1}{2}\mathbf{X}^T[A]\mathbf{X} + \mathbf{B}^T\mathbf{X} + C$$

The minimum of $f(\mathbf{X})$ is given by

$$\nabla f = [A]\mathbf{X} + \mathbf{B} = \mathbf{0}$$

or

$$\mathbf{X}^* = -[A]^{-1}\mathbf{B}$$

The iterative step of Eq. (6.98) gives

$$\mathbf{X}_{i+1} = \mathbf{X}_i - [A]^{-1}([A]\mathbf{X}_i + \mathbf{B}) \qquad (E_1)$$

where $\mathbf{X}_i$ is the starting point for the ith iteration. Thus Eq. (E_1) gives the exact solution

$$\mathbf{X}_{i+1} = \mathbf{X}^* = -[A]^{-1}\mathbf{B}$$

Figure 6.17 illustrates this process.

Example 6.12 Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ by taking the starting point as $\mathbf{X}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$.

SOLUTION To find $\mathbf{X}_2$ according to Eq. (6.98), we require $[J_1]^{-1}$, where

$$[J_1] = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\, \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2\, \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{\mathbf{X}_1} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

Therefore,

$$[J_1]^{-1} = \frac{1}{4}\begin{bmatrix} +2 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

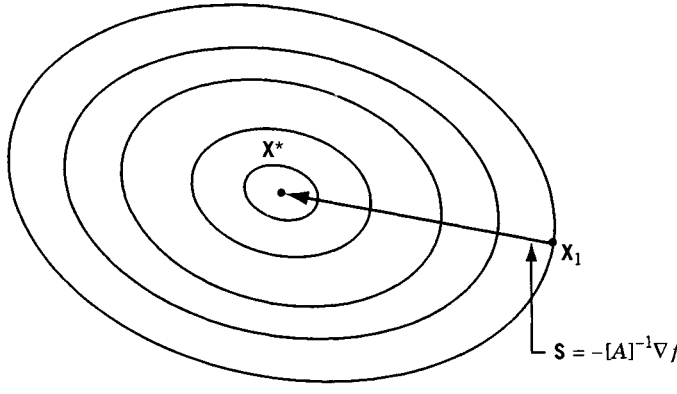

Figure 6.17 Minimization of a quadratic function in one step.

As

$$\mathbf{g}_1 = \begin{Bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{Bmatrix}_{\mathbf{X}_1} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 3x_2 \end{Bmatrix}_{(0,0)} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

Equation (6.98) gives

$$\mathbf{X}_2 = \mathbf{X}_1 - [J_1]^{-1}\mathbf{g}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} - \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ \frac{3}{2} \end{Bmatrix}$$

To see whether or not $\mathbf{X}_2$ is the optimum point, we evaluate

$$\mathbf{g}_2 = \begin{Bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{Bmatrix}_{\mathbf{X}_2} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}_{(-1,3/2)} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

As $\mathbf{g}_2 = \mathbf{0}$, $\mathbf{X}_2$ is the optimum point. Thus the method has converged in one iteration for this quadratic function.

If $f(\mathbf{X})$ is a nonquadratic function, Newton's method may sometimes diverge, and it may converge to saddle points and relative maxima. This problem can be avoided by modifying Eq. (6.98) as

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda_i^* \mathbf{S}_i = \mathbf{X}_i - \lambda_i^* [J_i]^{-1} \nabla f_i \tag{6.99}$$

where λ_i^* is the minimizing step length in the direction $\mathbf{S}_i = -[J_i]^{-1} \nabla f_i$. The modification indicated by Eq. (6.99) has a number of advantages. First, it will find the minimum in lesser number of steps compared to the original method. Second, it finds the minimum point in all cases, whereas the original method may not converge in some cases. Third, it usually avoids convergence to a saddle point or a maximum. With all these advantages, this method appears to be the most powerful minimization method. Despite these advantages, the method is not very useful in practice, due to the following features of the method:

1. It requires the storing of the $n \times n$ matrix $[J_i]$.
2. It becomes very difficult and sometimes, impossible to compute the elements of the matrix $[J_i]$.
3. It requires the inversion of the matrix $[J_i]$ at each step.
4. It requires the evaluation of the quantity $[J_i]^{-1} \nabla f_i$ at each step.

These features make the method impractical for problems involving a complicated objective function with a large number of variables.

6.14 MARQUARDT METHOD

The steepest descent method reduces the function value when the design vector $\mathbf{X}_i$ is away from the optimum point $\mathbf{X}^*$. The Newton method, on the other hand, converges fast when the design vector $\mathbf{X}_i$ is close to the optimum point $\mathbf{X}^*$. The Marquardt method [6.15] attempts to take advantage of both the steepest descent and Newton methods. This method modifies the diagonal elements of the Hessian matrix, $[J_i]$, as

$$[\tilde{J}_i] = [J_i] + \alpha_i[I] \tag{6.100}$$

where $[I]$ is an identity matrix and α_i is a positive constant that ensures the positive definiteness of $[\tilde{J}_i]$ when $[J_i]$ is not positive definite. It can be noted that when α_i is sufficiently large (on the order of 10^4), the term $\alpha_i[I]$ dominates $[J_i]$ and the inverse of the matrix $[J_i]$ becomes

$$[\tilde{J}_i]^{-1} = [[J_i] + \alpha_i[I]]^{-1} \approx [\alpha_i[I]]^{-1} = \frac{1}{\alpha_i}[I] \tag{6.101}$$

Thus if the search direction $\mathbf{S}_i$ is computed as

$$\mathbf{S}_i = -[\tilde{J}_i]^{-1} \nabla f_i \tag{6.102}$$

$\mathbf{S}_i$ becomes a steepest descent direction for large values of α_i. In the Marquardt method, the value of α_i is taken to be large at the beginning and then reduced to zero gradually as the iterative process progresses. Thus, as the value of α_i decreases from a large value to zero, the characteristics of the search method change from those of a steepest descent method to those of the Newton method. The iterative process of a modified version of Marquardt method can be described as follows.

1. Start with an arbitrary initial point $\mathbf{X}_1$ and constants α_1 (on the order of 10^4), c_1 ($0 < c_1 < 1$), c_2 ($c_2 > 1$), and ε (on the order of 10^{-2}). Set the iteration number as $i = 1$.
2. Compute the gradient of the function, $\nabla f_i = \nabla f(\mathbf{X}_i)$.
3. Test for optimality of the point $\mathbf{X}_i$. If $\|\nabla f_i\| = \|\nabla f(\mathbf{X}_i)\| \leq \varepsilon$, $\mathbf{X}_i$ is optimum and hence stop the process. Otherwise, go to step 4.
4. Find the new vector $\mathbf{X}_{i+1}$ as

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \mathbf{S}_i = \mathbf{X}_i - [[J_i] + \alpha_i[I]]^{-1} \nabla f_i \tag{6.103}$$

5. Compare the values of f_{i+1} and f_i. If $f_{i+1} < f_i$, go to step 6. If $f_{i+1} \geq f_i$, go to step 7.