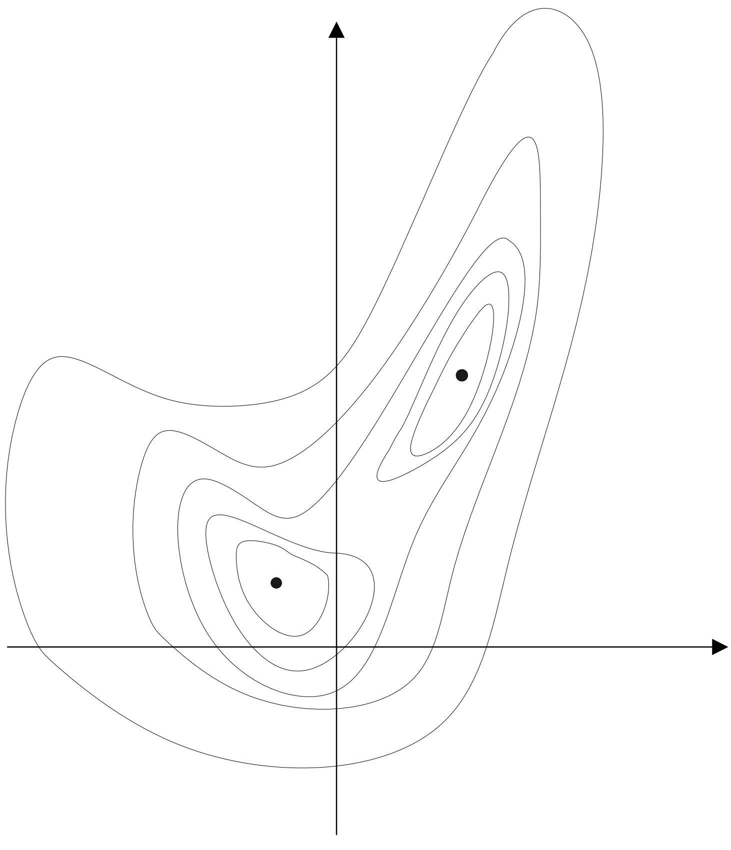
CHAPTER 8. MULTIDIMENSIONAL UNCONSTRAINED PROBLEMS

8.1 INTRODUCTION

Multidimensional unconstrained problems consist in the optimization of an objective function $f(x)$ of several variables $x = [x_1, x_2, \ldots, x_n]^T$ in the absence of constraints on x, i.e. each x_i is defined on the real axis, $[-\infty, \infty]$. Functions of two variables $f(x_1, x_2)$ can be conveniently visualized using contours plots. A contour plot is found by plotting x_2 versus x_1 for constant values of f i.e. $f=c$. Figure 8-1 shows an example of contour plots. It can be seen that the functions has two optimum points: A and B.



x_2

Figure 8-1: Contour plot of a function of two variables with two optimum points

Similarly to the single variable case we can establish the conditions for the existence of a local optimum. But before that, let first define two quantities that extend the notion of first and second derivatives to a function of several variables.

- The first derivative $\frac{df}{dx}$ is a vector defined as

$$\frac{df}{dx} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right]^T \tag{8.1}$$

It is called the gradient and is denoted by $\nabla f(x)$. The gradient vector $\nabla f(x)$ has the important characteristic that it is always orthogonal to the contour of f at point x, as shown in Figure 8-2.

- The second derivative $\frac{d^2 f}{dx^2}$ is a matrix denoted $\nabla^2 f$ and called the *Hessian* $H(x)$. It is defined as

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{8.2}$$

Note that the Hessian is a symmetric matrix.

8.2 UNIMODAL FUNCTIONS

Similarly to the single variable case, a multivariable function is said to be unimodal if it has one optimum point. It is multimodal if more than one optimum exist. The property of unimodality can be checked for a function of two variables $f(x_1,x_2)$ by plotting the contours as was shown in Figure 8-1. This property is however, difficult to check for more than two variables.

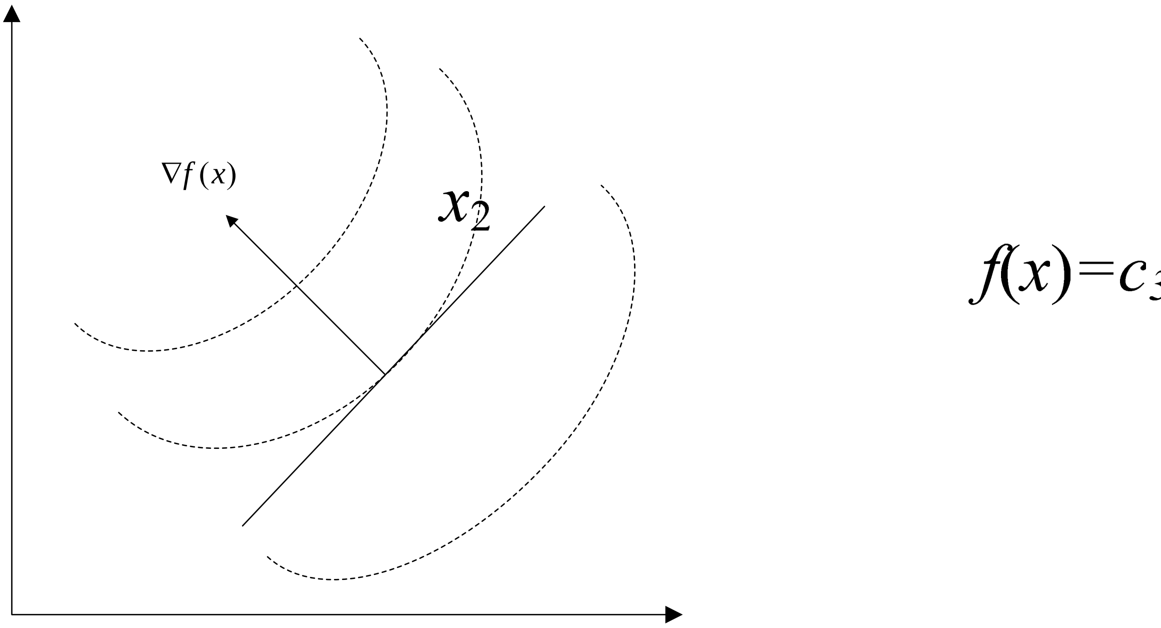


Figure 8-2: Location of the gradient

8.3 CONVEX FUNCTIONS

Similarly to the single variable case, the properties of the second order derivative (the Hessian) $H(x)$ are closely related to the convexity of the function. It can be shown that

$$f(x) \text{ is convex if } H(x) \text{ is positive semi - definite}$$
$$f(x) \text{ is concave if } H(x) \text{ is negative semi - definite}$$

The symmetric matrix H is said to be positive semi-definite (resp. negative semi-definite) if any of the following equivalent conditions hold:

1. $x^T Hx \geq 0 \quad (\text{resp.} \leq 0) \forall x \neq 0$
2. All the diagonal elements are positive (resp. $\leq$ 0) as well as the determinant of all leading principal minors (including the matrix H). Recall that a leading principal minor of order k of matrix is a submatrix found by deleting the last n-k columns and rows.
3. All the eigenvalues of $H(x)$ are positive (≥ 0), (resp. ≤ 0)

If none of these properties is satisfied then the matrix is said to be indefinite. Moreover, if the conditions above are strict (strictly positive), the matrix is said to be positive definite (resp. negative definite). In this case the function f is strictly convex (resp. strictly concave).

Example 8.1: Optimums of a function of two variables

Consider the following function

$$f(x_1,x_2) = x_1^2 + 4x_2^2 - 3x_1 x_2 + x_2 + 2 \tag{8.3}$$

To form the Hessian the second order derivatives are determined as follows:

$$\frac{\partial f}{\partial x_1} = 2x_1 - 3x_2 \tag{8.4}$$

$$\frac{\partial f}{\partial x_2} = 8x_2 - 3x_1 + 1 \tag{8.5}$$

$$\frac{\partial f^2}{\partial x_1^2} = 2 \tag{8.6}$$

$$\frac{\partial f^2}{\partial x_2^2} = 8 \tag{8.7}$$

$$\frac{\partial f^2}{\partial x_1 \partial x_2} = -3 \tag{8.8}$$

The Hessian $H(x)$ is, therefore

$$H(x) = \begin{bmatrix} 2 & -3 \\ -3 & 8 \end{bmatrix} \tag{8.9}$$

Applying the second test, we have that all the diagonal elements are positive and that the leading principal minors are:

$$M_1\ (order\ 1) = 2, \text{ and } det(M_1) = 2 > 0 \tag{8.10}$$

$$M_2\ (order\ 2) = H, \text{ and } det(M_2) = 7 > 0 \tag{8.11}$$

Thus, it can be concluded that $f(x)$ is convex. In fact it is strictly convex.

8.4 OPTIMALITY CRITERIA

Similarly to the single variable case, in order to derive the necessary conditions for an optimum to exist we start with a Taylor expansion around a potential optimum x^*. We also assume that f is twice differentiable at x^*. We have then

$$f(x) = f(x^*) + \frac{d^T f(x^*)}{dx}(x - x^*) + \frac{1}{2!}(x - x^*)^T \frac{d^2 f(x^*)}{d^2 x}(x - x^*) + \cdots \tag{8.12}$$

Using the notation adopted for the gradients and Hessian results in:

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2!}(x - x^*)^T H(x^*)(x - x^*) + \cdots \tag{8.13}$$

The analysis carried out in sec.7.2.3 for the single variable case can be readily applied to Eq.(8.13). A necessary condition for the occurrence of an optimum is that:

$$\nabla f(x^*) = 0 \tag{8.14}$$

That is the gradient $\nabla f(x^*)$ should vanish. This point is also called a stationary point. The sign of the second term:

$$\frac{1}{2!}(x - x^*)^T H(x^*)(x - x^*) \tag{8.15}$$

determines the character of the stationary point minimum, maximum or saddle point. A sufficient condition for the stationary point to be a minimum (resp. maximum) is that the $(x-x^*)^T H(x^*)(x-x^*)$ is positive (resp. negative). This is equivalent to the Hessian, *H*(*x*), being positive-semi definite (resp. negative-semi definite). The results of the optimality criteria are summarized in Table 8-1.

Table 8-1: Optimality Criteria

	Minimum	Maximum
Necessary	$\nabla f(x^*)=0$	$\nabla f(x^*)=0$
Sufficient	*H*(*x*) positive semi-definite	*H*(*x*) negative semidefinite

A stationary point for which *H* is indefinite is a saddle point.

Example 8.2: Optimization of compressor work

Consider the problem of finding the minimum work for a reversible adiabatic compression of ideal gas in three compressors with inter cooling, as shown in Figure 8-3. The work required for compression from pressure P_1 to P_4 can be shown to be

$$W=\frac{kRT}{k-1}\left[\left(\frac{P_2}{P_1}\right)^{\frac{k-1}{k}}+\left(\frac{P_3}{P_2}\right)^{\frac{k-1}{k}}+\left(\frac{P_4}{P_3}\right)^{\frac{k-1}{k}}-3\right] \tag{8.16}$$

where *k* is the isentropic coefficient, *T* the process temperature (assumed constant) and *R* the ideal gas constant. The inlet and outlet pressures P_1 and P_4 are known while P_2 and P_3 are the unknown variables. Since *k*, *R* and *T* are constants, the objective function to minimize can be put in the form:

$$f(P_2,P_3)=\frac{W(k-1)}{kRT}=\left[\left(\frac{P_2}{P_1}\right)^{\frac{k-1}{k}}+\left(\frac{P_3}{P_2}\right)^{\frac{k-1}{k}}+\left(\frac{P_4}{P_3}\right)^{\frac{k-1}{k}}-3\right] \tag{8.17}$$

To determine the necessary conditions for the existence of a minimum the gradient vector is computed:

$$\nabla f = \left[\frac{\partial f}{\partial P_2}, \ \frac{\partial f}{\partial P_3} \right]^T \tag{8.18}$$

We have then:

$$\frac{\partial f}{\partial P_2} = \frac{\zeta}{P_2} \left[\left(\frac{P_2}{P_1} \right)^{\zeta} - \left(\frac{P_3}{P_2} \right)^{\zeta} \right] \tag{8.19}$$

$$\frac{\partial f}{\partial P_3} = \frac{\zeta}{P_3} \left[\left(\frac{P_3}{P_2} \right)^{\zeta} - \left(\frac{P_4}{P_3} \right)^{\zeta} \right] \tag{8.20}$$

with ζ = (k-1)/k. Solving simultaneously these equations yield the optimum values of P_2 and P_3

$$P_2^{\ 2} = P_1 \, P_3 \,, \qquad P_3^{\ 2} = P_4 \, P_2 \tag{8.21}$$

Solving for P_2 and P_3 yields

$$P_2 = (P_1^2 P_4)^{1/3}, \qquad P_3 = (P_4^2 P_1)^{1/3} \tag{8.22}$$

To characterize the nature of this stationary point, the Hessian needs to be computed. The second derivatives are given by

$$\frac{\partial^2 f}{\partial P_2^2} = \frac{\zeta(\zeta - 1)}{P_2^2} \left(\frac{P_2}{P_1} \right)^{\zeta} + \frac{\zeta(\zeta + 1)}{P_2^2} \left(\frac{P_3}{P_2} \right)^{\zeta} \tag{8.23}$$

$$\frac{\partial^2 f}{\partial P_3^2} = \frac{\zeta(\zeta - 1)}{P_3^2} \left(\frac{P_3}{P_2} \right)^{\zeta} + \frac{\zeta(\zeta + 1)}{P_3^2} \left(\frac{P_4}{P_3} \right)^{\zeta} \tag{8.24}$$

$$\frac{\partial^2 f}{\partial P_2 \partial P_3} = \frac{\partial^2 f}{\partial P_3 \partial P_2} = -\zeta^2 \left(\frac{P_3}{P_2} \right)^{\zeta} \frac{1}{P_2 P_3} \tag{8.25}$$

Substituting for the optimum values of P_2 and P_3 (Eq.8.21) into Eqs.8.23-8.25 yields

$$\frac{\partial^2 f}{\partial P_2^2} = 2\zeta^2 P_1^{\frac{-4-\zeta}{3}} P_4^{\frac{\zeta-2}{3}} \tag{8.26}$$

$$\frac{\partial^2 f}{\partial P_3^2} = 2\zeta^2 P_1^{\frac{-2-\zeta}{3}} P_4^{\frac{\zeta-4}{3}} \tag{8.27}$$

$$\frac{\partial^2 f}{\partial P_2\, \partial P_3} = -\zeta^2 P_1^{\frac{-3-\zeta}{3}} P_4^{\frac{\zeta-3}{3}} \tag{8.28}$$

Since $\zeta > 0$ we see that that the diagonal elements $\frac{\partial^2 f}{\partial P_2^2}$ and $\frac{\partial^2 f}{\partial P_3^2}$ are always positive. The determinant $\Delta = \frac{\partial^2 f}{\partial P_2^2} \frac{\partial^2 f}{\partial P_3^2} - [\frac{\partial^2 f}{\partial P_2\, \partial P_3}]^2$ of the Hessian is equal, after some manipulations, to

$$\Delta = 3\zeta^4 P_1^{\frac{-2\zeta-6}{3}} P_4^{\frac{2\zeta-6}{3}} \qquad \text{98.29)}$$

which is also positive. The stationary point defined by Eqs.8.22 is therefore a minimum. But generally it is difficult to check analytically whether the Hessian is positive or negative definite. This example will be revisited in a later section and will be solved numerically.

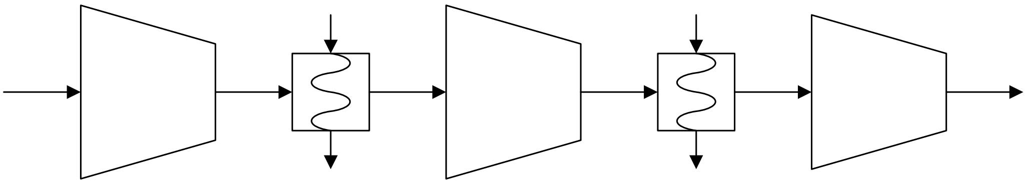


Figure 8-3: Ideal compression for adiabatic flow

8.5 NUMERICAL TECHNIQUES

Most numerical techniques for the solution of the multivariable optimization problems are iterative in nature . At each iteration k starting with the vector $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})^T$, the general procedure consists in

- Choosing a search direction $s^{(k)}$. A search direction is a vector whose coordinates are known.
- The new solution vector is

$$x^{(k+1)} = x^{(k)} + \lambda\, s^{(k)} \tag{8.30}$$

where λ is a scalar that determines the step length in the direction $s^{(k)}$

Starting with an initial guess $x^{(0)} = (x_1^{(0)}, x_1^{(0)}, \ldots, x_n^{(0)})^T$, these iterations are repeated until convergence or a conclusion can be reached. The various optimization methods differ in the selection of the search direction and also whether they use the derivatives in determining the search direction. We can distinguish between three classes of techniques.

1. Direct search methods that use only the value of the function $f(x)$.
2. Indirect methods of first order that require the use of first derivative $f'(x)$.
3. Indirect methods of second order that require the use of all the above in addition to the second derivative $f''(x)$.

Each of the three classes contains various methods and each of these classes have advantages and limitations. Before we present some common methods for each class, we first discuss the selection of the step length λ. There are a number of techniques used for the selection of λ . As it will be seen in coming sections, some of the numerical methods fix the value of λ while for other methods it is necessary to compute the value of λ .

Once the search direction $s^{(k)}$ is selected (or calculated), the goal in a minimization problem is to calculate λ that reduces the value of the function f to a certain degree. This is equivalent to requiring that

$$f(x^{(k)} + \lambda s^{(k)}) < f(x^{(k)}) \tag{8.31}$$

One simple idea consists in taking a constant value λ for all the iterations. The drawback is that generally there is no idea on the appropriate value to choose. Choosing the wrong value may increase the iterations for the optimization problem. A better idea is to select the value of λ that optimizes the function in that direction i.e. find λ such that $f(x^{(k)} + \lambda s^{(k)})$ is optimum. Since at iteration (k) both $x^{(k)}$ and $s^{(k)}$ are known, this type of search is a one dimensional optimization problem (also called a line search) over λ. It can carried out by any of the one dimensional search techniques discussed in Chapter 7. An other idea is to approximate locally the objective function by a quadratic approximation. The approximation comes from applying second order Taylor expansion around $x^{(k)}$

Let $x = x^{(k)} + \Delta x$ where $\Delta x = \lambda s^{(k)}$. The second order Taylor expansion is

$$f(x) = f(x^{(k)} + \Delta x) = f(x^{(k)}) + \nabla^T f(x^{(k)})\lambda s^{(k)} + \frac{1}{2}(\lambda s^{(k)})^T H(x^{(k)})\lambda s^{(k)} \tag{8.32}$$

The optimum of $f(x)$ over λ is obtained by setting the derivative $df/d\lambda$ to zero. Taking the derivative yields

$$\frac{df}{d\lambda} = \nabla^T f(x^{(k)})s^{(k)} + \lambda(s^{(k)})^T H(x^{(k)})s^{(k)} = 0 \tag{8.33}$$

Therefore the optimum value of λ is given by

$$\lambda = -\frac{\nabla^T f(x^{(k)}) s^{(k)}}{(s^{(k)})^T H(x^{(k)}) s^{(k)}} \qquad (8.34)$$

Note that this method requires the evaluation of both the gradients and the Hessian. Now we go back to presenting some of the common numerical methods of optimization of unconstrained problems.

8.5.1 Powell's Conjugate Direction Method

Powell's method belongs to the class of direct methods that do not require the evaluation of any derivatives of the objective function *f*. Starting from an initial guess this method finds the minimum of $f(x_1,x_2,\ldots,x_n)$ through uni-dimensional search along a set of directions called conjugates. A set of *n* linearly independent search directions $s^0,s^1,\ldots,s^{n-1}$ are said to be conjugate with respect to a positive definite matrix *Q* if

$$(s^i)^T Q(s^j) = 0, \; 0 \le i, \; j \le n-1 \text{ for } i \neq j \qquad (8.35)$$

It can be seen that when *Q* is the identity matrix (Q=I) then we have

$$(s^i)^T (s^j) = 0, \quad 0 \le i, \; j \le n-1 \text{ for } i \neq j \qquad (8.36)$$

and we recover the traditional concept of orthogonality. What is special about the conjugate search directions is that experience has shown that when the matrix *Q* is taken to be the Hessian *H*, the search for optimum along the conjugate vectors is quite efficient. Moreover there is a simple way to construct search directions that are conjugate with respect to $H(x)$. Let s^0 be a given search direction, and we would like to find a search direction conjugate to s^0. Starting with an initial point x^0, we first find the point *y* that is the minimum of $f(x)$ in the direction of s^0. Starting from an other initial point x^1 we find the point *z* that is also the minimum of $f(x)$ in the direction of s^0. It can be shown that the vectors $(e:=z-y)$ and s^0 are conjugates with respect to $H(x)$.

Organigram 8.1: Powell Method

The algorithm for the Powell's method is summarized in the following steps.

Step 1: Define a starting vector $x^{(0)}$ and define also a set of n linearly independent directions $s^{(i)}$. Usually these are the independent coordinates, i.e.

$$s^{(1)} = [1,0,\ldots,0,0,\ldots]^T,\ s^{(2)} = [0,1,\ldots,0,0,\ldots]^T,\ \ldots \tag{8.37}$$

Step 2: Perform a minimization along (n+1) directions in this order:

- Minimize along the n^{th} direction $s^{(n)}$, i.e. Find λ such that $f\,(x^{(0)} + \lambda s^{(n)})$ is minimum. Set

$$x^{(1)} = x^{(0)} + \lambda s^{(n)} \tag{8.38}$$

- Using $x^{(1)}$, minimize along the first direction $s^{(1)}$ i.e., Find λ such that $f\,(x^{(1)} + \lambda s^{(1)})$ is minimum. Set

$$x^{(2)} = x^{(1)} + \lambda s^{(1)} \tag{8.39}$$

- These steps are repeated and minimization is done form 2 to $n-1$ to obtain $x^{(3)}$, $x^{(4)}$, $x^{(n)}$
- In the last step, minimize again in the n^{th} direction $s^{(n)}$, i.e. Find λ such that $f\,(x^{(n)} + \lambda s^{(n)})$ is minimum. Set

$$x^{(n+1)} = x^{(n)} + \lambda s^{(n)} \tag{8.40}$$

Step 3: Using the values $x^{(1)}$, $x^{(2)}$,…, $x^{(n)}$, $x^{(n+1)}$ compute the new conjugate direction $s^{(n+1)} = x^{(n+1)} - x^{(1)}$. The new search direction is normalized, i.e.

$$s^{(n+1)} = \frac{s^{(n+1)}}{\left\| s^{(n+1)} \right\|_2} \tag{8.41}$$

Step 4: Replace the direction $s^{(1)}$ with $s^{(2)}$, $s^{(2)}$ with $s^{(3)}$ and so on. Replace the direction $s^{(n)}$ with the new conjugate direction $s^{(n+1)}$

Repeat steps 2 to 4 until the procedure converges or a conclusion is reached.

The algorithm is illustrated in Figure 8-4. The algorithm will terminate if, at the end of each stage, the change in each independent variable is less than required accuracy, e.g.:

$$\frac{\left\| x^{(k)} - x^{(k-1)} \right\|}{\left\| x^{(k-1)} \right\|} \le \varepsilon \tag{8.42}$$

Or if the iteration exceeds a predefined maximum value.

It should be noted that if the function of n variables to minimize is quadratic then Powell method will converge in n steps if the Hessian is a constant matrix. If the function is not quadratic some steps will lead to directions vector that are not linear. Additional tests are required to ensure that the direction sets are independent.

Example 8.3: Powell method applied to example 8.2

Consider the previous problem (Eq. 8.16) of finding the pressures P_2 and P_3 that minimize the compression work. To get numerical results, the following values are taken: P_1 = 0.200 kbar P_4 = 1.3 kbar and k = 1.4. Figure [8.5] shows the contours of the objective function and the location of the minimum. The iterations steps are summarized below. Starting with $x_0 = [0.3,0.4]^T$ with $f(x_0) = 0.60889$, the steps indicated by the algorithm is performed as follows:

Step 1: Let $s^{(1)} = [1,0]^T$ and $s^{(2)} = [0,1]^T$

Step2 :

- Minimize along $s^{(2)}$. Find λ such that $f\,(x^{(0)} + \lambda s^{(2)})$ is minimum. The result is λ = 0.12282. Therefore $x^{(1)} = [0.3,0.4]^T + 0.12882[0,1]^T = [0.3,0.52282]^T$ with $f(x^{(1)})$ = 0.59207.
- Minimize along $s^{(1)}$. Find λ such that $f\,(x^{(1)} + \lambda s^{(1)})$ is minimum. The result is λ = 0.02092 Therefore, $x^{(2)} = [0.3,0.52282]^T + 0.02092[1,0]^T = [0.32092,0.52282]^T$ with $f(x^{(2)}) = 0.59154$

- Minimize again along $s^{(2)}$. Find λ such that $f(x^{(2)} + \lambda s^{(2)})$ is minimum. The result is $\lambda = 0.09115$. Therefore, $x^{(3)} = [0.32092, 0.61397]^T$ with $f(x^{(3)}) = 0.58731$.

Step 3: Let $s^{(3)} = x^{(3)} - x^{(1)} = [0.02093, 0.09115]^T$. The norm of $s^{(3)}$ is $\|s^{(3)}\|_2 = \sqrt{0.02093^2 + 0.09115^2} = 0.09352$ Normalize $s^{(3)}$ by the norm

$$s^{(3)} = \frac{s^{(3)}}{\|s^{(3)}\|_2} \tag{8.43}$$

lead to

$$s^{(3)} = [0.22380, 0.97464]^T \tag{8.44}$$

Step 4: Let $s^{(1)} = s^{(2)}$; $s^{(2)} = s^{(3)}$. Repeat step 2, i.e., find λ such that $f(x^{(3)} + \lambda s^{(2)})$ is minimum. The result is λ =0.06751. That is: $x^{(4)} = [0.33603, 0.67977]^T$ $f(x^{(4)}) = 0.58631$.

Some iterations are shown in Table 8-2. The optimum solution to a 10^{-4} precision is $x_1 = 0.37325$, $x_2 = 0.69658$ with $f(x_1, x_2) = 0.58543$.

Table 8-2: Powell's method iteration for the minimization of compression work

k	$s^{(k)}$	λ	$x^{(k)}$	$f(x^{(k)})$
1	[0.00000E+00 0.10000E+01]	0.122818	[0.30000E+00 0.52282E+00]	0.592069
2	[0.10000E+01 0.00000E+00]	0.020925	[0.32093E+00 0.52282E+00]	0.591548
3	[0.00000E+00 0.10000E+01]	0.091156	[0.32093E+00 0.61397E+00]	0.587347
4	[0.22380E+00 0.97464E+00]	0.067509	[0.33603E+00 0.67977E+00]	0.586314
5	[0.00000E+00 0.10000E+01]	-0.019646	[0.33603E+00 0.66012E+00]	0.586236
6	[0.22380E+00 0.97464E+00]	0.031620	[0.34311E+00 0.69094E+00]	0.586063
7	[0.53536E+00 0.84463E+00]	0.038741	[0.36385E+00 0.72366E+00]	0.585734
8	[0.22380E+00 0.97464E+00]	-0.032165	[0.35665E+00 0.69232E+00]	0.585611
9	[0.53536E+00 0.84463E+00]	0.021966	[0.36841E+00 0.71087E+00]	0.585515
10	[0.33582E+00-0.94192E+00]	0.015017	[0.37346E+00 0.69672E+00]	0.585433
11	[0.53536E+00 0.84463E+00]	-0.000310	[0.37329E+00 0.69646E+00]	0.585433
12	[0.33582E+0 -0.94192E+00]	-0.000122	[0.37325E+00 0.69658E+00]	0.585433
13	[-0.83201E+00 -0.55475E+00]	-0.000003	[0.37325E+00 0.69658E+00]	0.585433

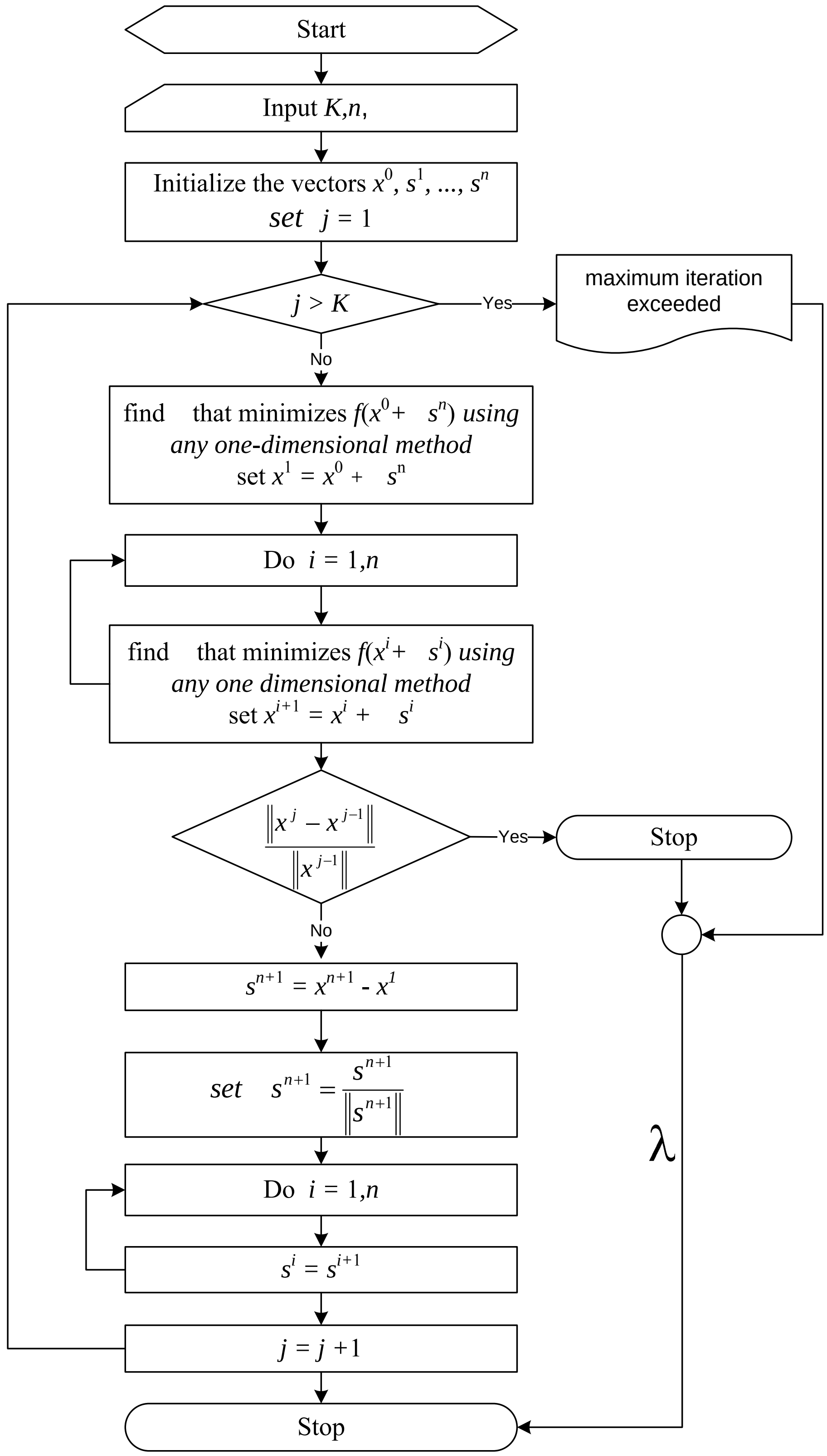


Figure 8-4: Powell optimization method

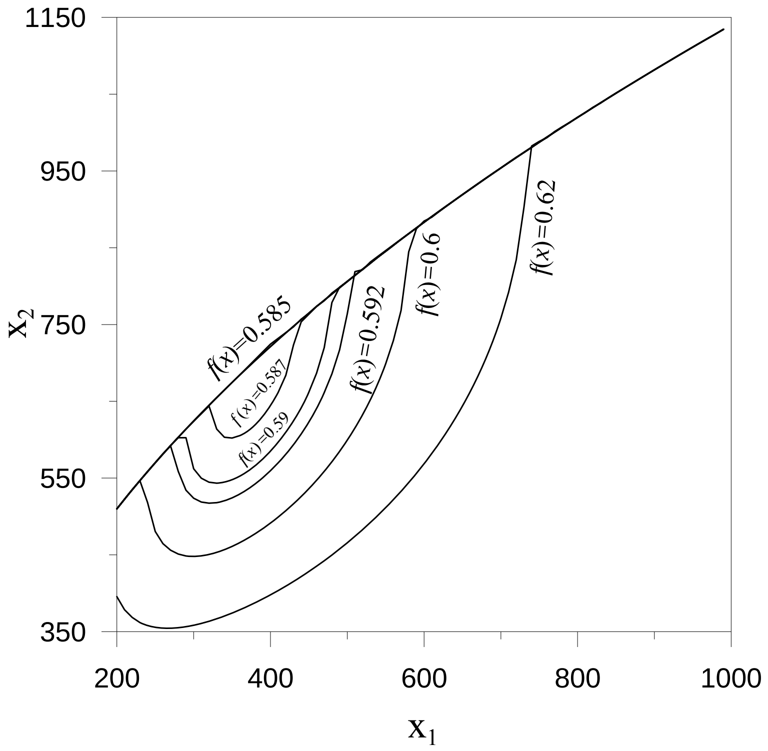


Figure 8-5: Contours plot for the compression work objective function

8.5.2 Gradient Method

The gradient method belongs to the class of indirect methods that require the evaluation of the first derivative. Starting from an initial guess $x^{(0)}$, in order to find minimum (resp. maximum) the gradient method chooses the search direction $s=x^{(1)}-x^{(0)}$ that produces the most local descent (resp. acsent). To see that consider a first order Taylor expansion around $x^{(0)}$

$$f(x^{(1)}) = f(x^{(0)}) + \nabla^T f(x^{(0)})(x^{(1)} - x^{(0)}) \tag{8.43}$$

That is

$$f(x^{(1)}) = f(x^{(0)}) + \nabla^T f(x^{(0)})s \tag{8.44}$$

To make $f(x^{(1)}) < f(x^{(0)})$ (for a minimization problem) , the search direction (s) should be selected such that it makes the term $\nabla f(x^{(0)})s$ negative. Such a direction is called a descent (resp. ascent for a maximization problem). The maximum descent is achieved if the direction s is choosed to be the negative of the gradient i.e. $s= -\nabla f(x^{(0)})$. Recall that the gradient $\nabla f(x^{(0)})$ is a vector orthogonal to the contour at the point $x^{(0)}$.

∇

In a practical implementation of the gradient method, the amount of change from $x^{(k-1)}$ to $x^{(k)}$ in the direction $s^{(k)} = -\nabla f(x^{(k-1)})$ is given by $x^{(k)} = x^{(k)} + \lambda s$. where λ can be constant or computed through a line search such that $f(x^{(k-1)} + \lambda s)$ is minimum.

Organigram 8.2: The Gradient Method

The gradient method algorithm for a minimization problem is explained by the following steps:

Step 1: At iteration k, determine the gradient $\nabla f(x^{(k)})$ and the search direction $s^{(k)} = -\nabla f(x^{(k)})$

Step 2: Set $\lambda^{(k)}$ constant or find $\lambda^{(k)}$ such that $f(x^{(k)} + \lambda^{(k)} s^{(k)})$ is a minimum

Step 3: Set $x^{(k+1)} = x^{(k)} + \lambda^{(k)} s^{(k)}$

Step 4: If the stopping criteria are satisfied then stop, otherwise repeat step 1 to 3.

The stopping criteria can be one of the following

- The norm $\left\|\nabla f(x^{(k)})\right\|$ is less than specified tolerance
- The size $\left|\lambda^{(k)}\right|$ is less than a specified tolerance
- The maximum number of iterations is exceeded.

The gradient based method can terminate on a saddle point .Recall that a saddle point also admits zero gradient. The Hessian must be then calculated to see if it is positive semi definite (for a minimum) or indefinite (for a saddle) . The algorithm for the gradient method is illustrated in Figure 8-6.

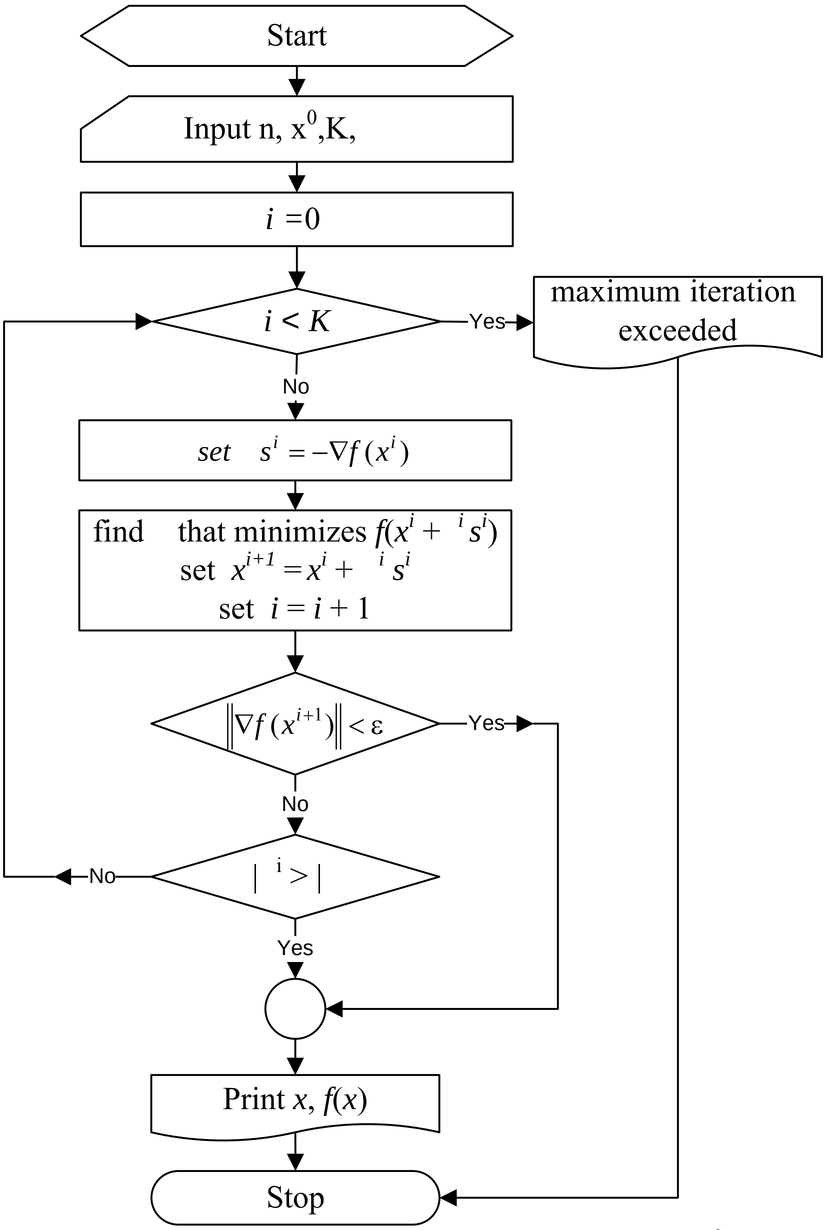


Figure 8-6: Gradient Method

Example 8.4: Gradient method applied to example 8.2

Consider again the previous example of minimizing compression work. Starting from the same initial condition $x^{(0)} = [0.3,0.4]^T$ with $f(x^{(0)}) = 0.60889$, the following steps of the algorithm are performed using a constant step length λ=1.

- Step 1: The initial search direction is $s^{(0)} = -\nabla f(x^{(0)}) = [-0.15644, -0.28600]^T$. Thus

$$x^{(1)} = |0.3, 0.4|^T + 1[-0.1564 \qquad 0.28600]^T = [\, 0.24356 \quad 0.68600] \ ^T$$

The rest of iterations are shown in Table 8-3 .

Table 8-3: Gradient method iteration for the minimization of compression work

k	$s^{(k)}$	$x^{(k)}$	$f(x^{(k)})$
1	[-0.35388E-01 0.22481E+00]	[0.26461 0.62481]	0.594364
2	[0.21051E+00 -0.20750E-01]	[0.47512 0.60406]	0.596276
3	[-0.12595E+00 0.82208E-01]	[0.34917 0.68627]	0.585790
4	[0.33029E-01 -0.52892E-02]	[0.38220 0.68098]	0.585590
5	[-0.17823E-01 0.98530E-02]	[0.36438 0.69083]	0.585476
10	[0.10403E-02 0.29384E-03]	[0.37305 0.69499]	0.585433
15	[0.26115E-05 0.12484E-03]	[0.37314 0.69629]	0.585433
20	[0.88798E-05 0.21855E-04]	[0.37323 0.69652]	0.585433
25	[0.14170E-05 0.44342E-05]	[0.37325 0.69657]	0.585433
30	[0.29387E-06 0.87240E-06]	[0.37325 0.69658]	0.585433
35	[0.57573E-07 0.17273E-06]	[0.37325 0.69658]	0.585433
40	[0.11409E-07 0.34155E-07]	[0.37325 0.69658]	0.585433

8.5.3 Newton's Method

Newton's method belongs to the class of indirect methods that require the evaluation of the function, its first order and its second order derivatives. Starting with $x^{(k)}$ let consider the second order Taylor expansion around $x^{(k)}$

$$f(x) = f(x^k) + \nabla f^T(x^k)(x - x^k) + \frac{1}{2!}(x - x^k)^T H(x^k)(x - x^k) + \cdots \tag{8.45}$$

Let $\tilde{f}(x)$ be the approximation to $f(x)$ after dropping the terms of order 3 and higher

$$\tilde{f}(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2!}(x - x^k)^T H(x^k)(x - x^k) \tag{8.46}$$

$\tilde{f}(x)$ is, hence a quadratic approximation of $f(x)$ around $x^{(k)}$. The next best point $x^{(k+1)}$ that reduces the value of $f(x)$ is found by forcing the gradient of the approximation $\tilde{f}(x)$ to vanish at $x^{(k+1)}$. Taking the gradient of the approximation (Eq. 8.46) yields:

$$\nabla\tilde{f}(x) = \nabla f(x^{(k)}) + H(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0 \tag{8.47}$$

Solving for $x^{(k+1)}$, the following Newton's iteration is obtained:

$$x^{(k+1)} = x^{(k)} + s^{(k)} \qquad with \qquad \left(H(x^{(k)})\right) s^{(k)} = -\nabla f(x^{(k)}) \tag{8.48}$$

Figure 8-7 shows the algorithm for Newton method . The stopping criteria can be one of the following:

- The norm $\left\|\nabla f(x^{(k)})\right\|$ is less than specified tolerance
- The change in each independent variable is less than required accuracy, e.g.:

$$\frac{\left\|x^{(k)} - x^{(k-1)}\right\|}{\left\|x^{(k-1)}\right\|} \le \varepsilon \tag{8.49}$$

- The maximum number of iterations is exceeded.

Note that for Newton method the step length λ is fixed and is equal to 1. Since Newton's method is based on a quadratic approximation of the function around each point $x^{(k)}$, the method is of quadratic convergence. In fact if the function is quadratic then only one iteration is needed to reach the optimum. The advantage of Newton method is that it performs better than other methods near the optimum. There are on the other hand some disadvantages to the method. The method requires the evaluation of second order derivatives as well as the solution of linear system (Eq.8.48). Both these problems can be attenuated through the use of finite difference approximation for the derivatives and the selection of appropriate codes for linear systems solutions. However the major problem of the method is that there is no guarantee that, as the iterations proceed, the Hessian remains positive definite. In the next section we present a method that precisely overcomes this weakness.

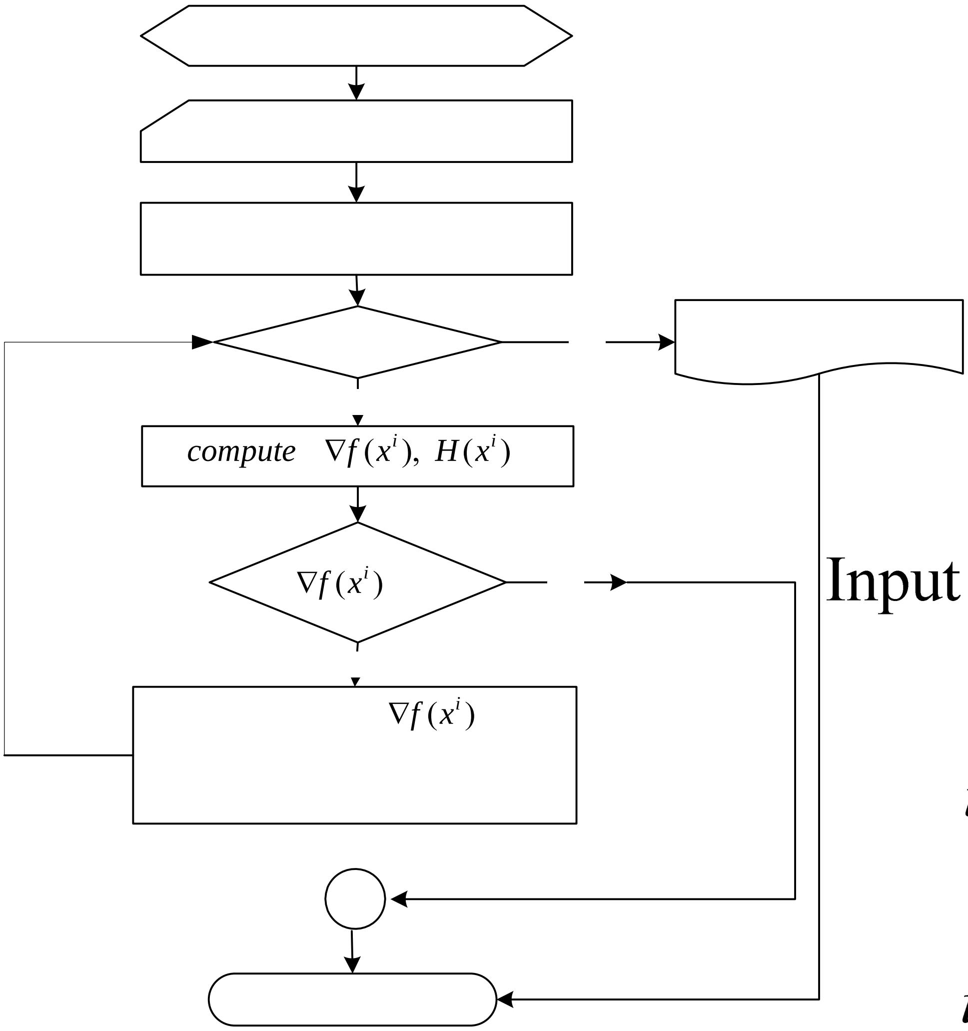


Figure 8-7: Newton Method

Example 8.5: Newton method applied to example 8.2

The previous example is reexamined with the same initial conditions $x^{(0)}$ = $[0.3,0.4]^T$. The elements of the Hessian were already evaluated in Eqs.(8.26-8.28). Table 8-4 shows the iterations for this example.

8.5.4 Marquart's Method

This procedure uses the Newton's iteration (Eq. 8.48), and at each iteration it adds the matrix μI to the Hessian $H(x)$, where I is the identity matrix. The new Hessian is:

$$\widetilde{H}(x) = H(x) + \mu I \tag{8.50}$$

μ is selected large enough to guarantee the Hessian is positive definite, i.e, all its eignevalues are positive.The search direction is, therefore

$$s^{(k)} = -\left(H^{(k)} + \mu^{(k)} I\right)^{-1} \nabla f(x^{(k)}) \tag{8.51}$$

Table 8-4: Newton method iterations for the minimization of compression work

K	$s^{(k)}$	$x^{(k)}$	$f(x^{(k)})$
1	[0.34853E-01 0.15487E+00]	[0.33485E+00 0.55487E+00]	0.589263
2	[0.28932E-01 0.10420E+00]	[0.36379E+00 0.65907E+00]	0.585658
3	[0.87861E-02 0.36044E-01]	[0.37257E+00 0.69512E+00]	0.585433
4	[0.67736E-03 0.74734E-03]	[0.37325E+00 0.69587E+00]	0.585433
5	[0.15991E-05 0.11867E-02]	[0.37325E+00 0.69705E+00]	0.585433
10	[0.63914E-08 -0.15538E-03]	[0.37325E+00 0.69652E+00]	0.585433
15	[0.11042E-09 0.20466E-04]	[0.37325E+00 0.69659E+00]	0.585433
20	[0.19158E-11 -0.26951E-05]	[0.37325E+00 0.69658E+00]	0.585433
25	[0.33243E-13 0.35491E-06]	[0.37325E+00 0.69658E+00]	0.585433
30	[0.53702E-15 -0.46737E-07]	[0.37325E+00 0.69658E+00]	0.585433

and the new search is :

$$x^{(k+1)} = x^{(k)} + s^{(k)} \tag{8.52}$$

The proposed procedure is due to Maquardt. It should be noted that if μ is chosen very large then:

$$\left(H^{(k)} + \mu^{(k)} I\right)^{-1} \approx \left[\mu^{(k)} I\right]^{-1} = \frac{1}{\mu^{(k)}} I \tag{8.53}$$

Therefore, the search direction becomes:

$$s^{(k)} = -\frac{1}{\mu^{(k)}} \nabla f(x^{(k)}) \tag{8.54}$$

which is the gradient method. As μ decreases, approaching closely to the solution, the method goes from gradient to Newton's. The Marquardt method has, therefore, the advantage that it combines the advantages of both the gradient method and the Newton's method . The gradient provides the direction of the most local ascent or descent and therefore provides a good reduction in the objective when the initial guess $x^{(0)}$ is far from the optimum x^* The Newton's method generates good search directions near the solution. Furthermore the procedure does not require a line search .

The selection of μ can be done in the following way. We can start with a large value of μ such as $\mu^0 = 10^3$ and test if $f(x^{(1)}) < f(x^{(0)})$, then decrease μ, (take for example $\mu^1 = \mu^0/2$). Otherwise set μ to a higher value (take for example $\mu^1 = 2\mu^0$) and repeat the step. The algorithm is shown in the following section.

Organigram 8.3: The Marquardt Method

The Marquardt procedure for a minimization is shown in Figure 8-8. The following stopping criteria can be used:

- The norm of the gradient of f is small
- The maximum number of iterations is exceeded.

Example 8.6: Marquardt method applied to example 8.2

Reconsider the previous problem of minimizing the compression work. Let us start with $x^{(0)} = [0.3,0.4]^T$ and $\mu^{(0)} = 10^3$, the following steps are performed:

- Step 1 : The gradient was calculated in the previous section and it is

$$\nabla f(x^{(0)}) = [0.03538, -0.22481]^T$$

- Step 2 : Compute the Hessian and then solve the linear problem $[H^{(0)} + \mu^{(0)} I] s^{(0)} = -\nabla f(x^{(0)})$ to find the search direction $s^{(0)} =$ [-3.51555E-05, 2.24373E-04]^T thus,

$x^{(1)} = x^{(0)} + s^{(0)} = [0.29996 \quad 0.40022\ ]^T$ and $f(x^{(1)}) = 0.60884$

- Step 3: Since $f(x^1) < f(x^0)$, set $\mu^1 = (1/2)\,\mu$. Go to step 2

The rest of iterations are shown in Table 8-5.

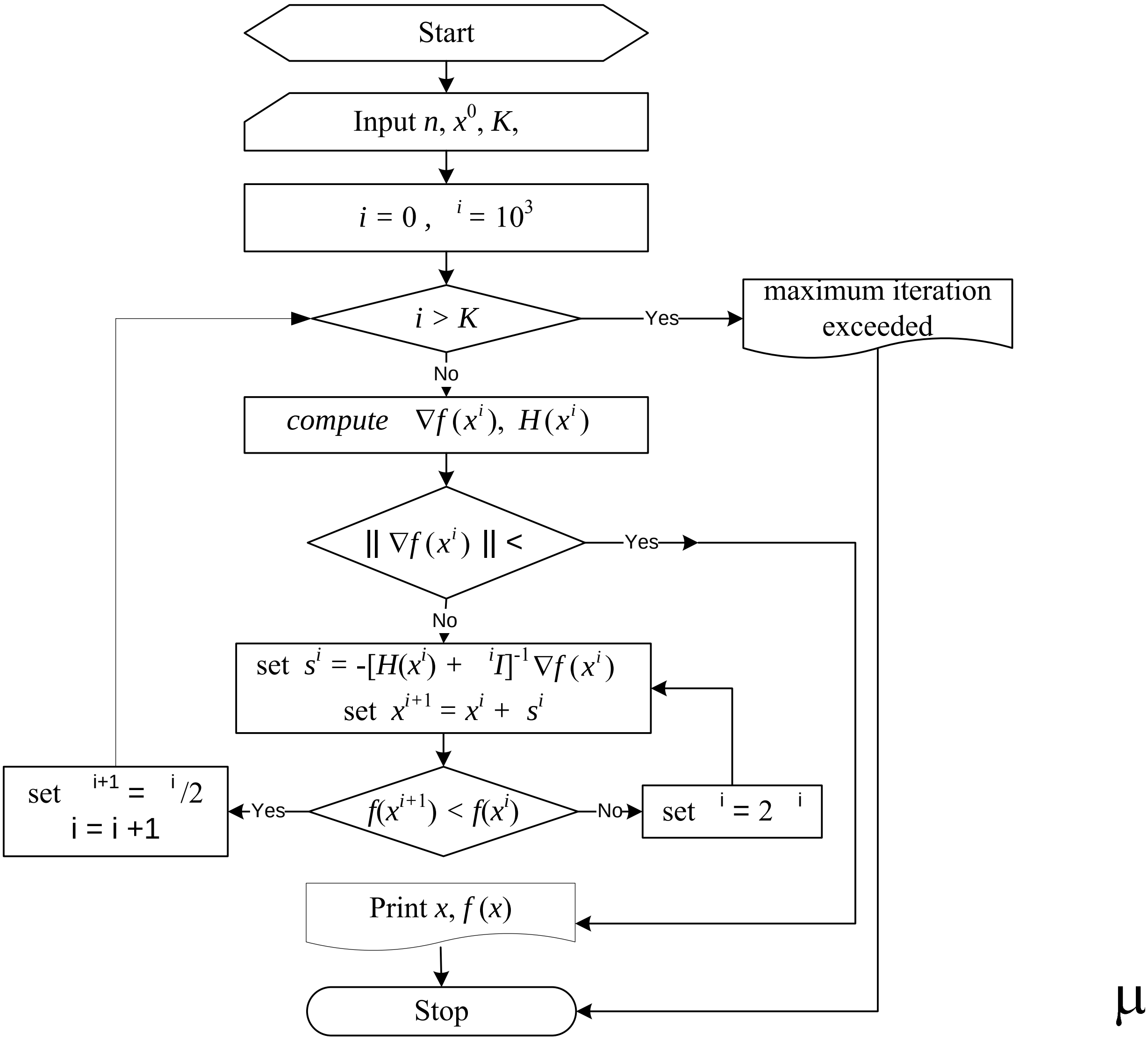


Figure 8-8: Marquardt Method

Table 8-5: Marquardt method iterations for the minimization of compression work

K	$s^{(k)}$	$\mu^{(k)}$	$x^{(k)}$	$f(x^{(k)})$
1	[-0.35156E-040.22437E-03]	0.10000E+04	[0.29996E+00 0.40022E+00]	0.608841
2	[-0.69389E-04 0.44701E-03]	0.50000E+03	[0.29990E+00 0.40067E+00]	0.608738
3	[-0.13514E-03 0.88716E-03]	0.25000E+03	[0.29976E+00 0.40156E+00]	0.608536
4	[-0.25609E-03 0.17475E-02]	0.12500E+03	[0.29950E+00 0.40331E+00]	0.608143
5	[-0.45826E-03 0.33929E-02]	0.62500E+02	[0.29905E+00 0.40670E+00]	0.607399
10	[0.97755E-02 0.40243E-01]	0.19531E+01	[0.31097E+00 0.51379E+00]	0.592399
15	[0.35213E-02 0.10757E-01]	0.61035E-01	[0.37240E+00 0.69396E+00]	0.585434
20	[0.17267E-08 0.51721E-08]	0.19073E-02	[0.37325E+00 0.69658E+00]	0.585433
25	[-0.88882E-17 0.75188E-17]	0.59605E-04	[0.37325E+00 0.69658E+00]	0.585433
30	[-0.88883E-17 0.75198E-17]	0.18626E-05	[0.37325E+00 0.69658E+00]	0.585433

8.6 OTHER SOLUTION TECHNIQUES

This chapter has covered the basic numerical techniques. Other solutions methods are very much beyond the scope of this book.

8.7 IMSL ROUTINES

Some IMSL routine to solve multivariable unconstrained optimization problems are as follows:

Routine	Features
UMINF	Find the minimum of a multivariable function a quasi-Newton method and a finite-difference gradient. Only function evaluations are needed.
UMING	Same as UMINF but the user provides an analytical Jacobean
UMIDH	Used a modified Newton method and a finite-difference Hessian.
UMIAH	Same as UMIDH but the user provides an analytical Jacobean and Hessian
UMCGF	Uses conjugate gradient algorithm and a finite-difference gradient.
UMCGG	Uses conjugate gradient algorithm and a user-supplied gradient.