IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 16, Issue 3 Ser. III (May – June 2020), PP 15-25
www.iosrjournals.org

Various Solution Techniques for Unconstrained Non-Linear Optimization Problems – A Survey

Ubon Akpan Abasiekwere[*1], Imoh Udo Moffat[2]
[1]*Department of Mathematics, University of Uyo, Uyo, Akwa Ibom State, Nigeria*
[2]*Department of Statistics, University of Uyo, Uyo, Akwa Ibom State, Nigeria*

Abstract: *The performances of five traditional methods for unconstrained nonlinear optimization problems are evaluated using a test problem. Efficiency index is based on convergence rate, application to a wider class of functions and ease of manual application. It is seen that optimization techniques that make use of the gradient vector are better-off than those that do not involve it in their operations. The former is observed to converge quadratically.*

Keywords: *Optimization, nonlinear, unconstrained, minimize, objective function*

Date of Submission: 14-05-2020 Date of Acceptance: 29-05-2020

I. Introduction

Generally speaking, operation research is an approach to the analysis of the operation that to a greater of lesser extent adopts scientific methods (Observation, hypothesis, deduction and experimentation as far as possible) and an explicit formulation of complex relationships. Operation research can be said to be the application of the methods of science to complex problems arising in the direction of management of large system of men, machines and materials.

An optimization problem in operation research is that which seek to minimize or maximize a specific quantity called the objective function which depends on a finite number of input variables.

We see optimization techniques discussed in artificial neural network, clustering and classifications, constraint-handling, queueing theory, support vector machine and multi-objective optimization, evolutionary computation, nature-inspired algorithms and many other topics. Even in the area of statistics, the theory of stochastic optimization has been growing rapidly in popularity over the last few decades, with a number of methods now becoming "industry standard" approaches for solving challenging optimization problems ([8], [9] and the references therein).

This research work on comparison of various methods of solving unconstrained non-linear problems has been thought of as being of immense importance following the fact that, in many situations, assumption of linearity as applied to a real world process might be questionable. It considers other various methods of solving unconstrained minimization problems, applies each method manually or by means of a computer program written in BASIC programming language to the sample problem:

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_1^2$.

This is done with the aim to recommend a method of solution which is more efficient and yet easier to go about. It suggests possible modifications for easier convergence of functions.

An unconstrained minimization problem is one where a value of the design vector

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

is sought that minimizes the function $f(X)$. This problem can be considered as a particular case of the general (constrained) non-linear programming problem. The special characteristic of this problem is that the solution vector *X* need not satisfy any constraint. It is true that rarely a particular design problem would be constrained.

Several methods are available for solving an unconstrained minimization problem. These methods can be classified into two broad categories as direct search methods and descent methods as shown below:

Direct search methods	Descent methods
1. Random search method	1. Steepest descent method
2. Univariate method	2. Conjugate gradient method (Fletcher-Reeves)
2. Pattern search methods a. Powell's method b. Hooke and Jeeves method	3. Newton's method
4. Rosenbrock's method of rotating coordinate	4. Variable metric method (Davidon-Fletcher-Powell)
5. Simplex method	

The direct search methods require only objective function evaluations and do not involve the partial derivatives of the function in finding the required minimum and hence are often called the non-gradient methods [1]. The descent techniques require, in addition to objective function evaluations, the evaluation of first and possibly higher order derivatives of the objective function

This work selectively investigates five methods of solving unconstrained non-linear optimization problems namely, the Univariate method, Powell's method, Hooke and Jeeves method, Conjugate gradient method (Fletcher-Reeves) and the Variable metric method (Davidon-Fletcher-Powell), carries out an analysis of each of them with a view to observing their rate of convergence, application to a wider class of functions and ease of manual application.All the unconstrained minimization methods are interactive in nature and hence start from an initial trial solution and proceed towards the minimum point in a sequential manner ([1], [2], [3], [4], [7]).

II. Univariate Method

The univariate method as a direct search method involves the changing of only one variable at a time and trying to produce a sequence of improved approximations to the minimum point being sought. By starting at a base point $X_i$in the *i*th iteration, we fix values of $n-1$ variables and vary the remaining variable. Since only one variable is changed, the problem therefore becomes a one – dimensional minimization problem and the method of producing a new base point X_{i+1} is illustrated as in the sample below [10]. This search is now continued in a new direction. The choice of the direction and the step length in the univariate method for an n – dimensional problem can be summarized as follows.

i. *Choose the starting point X_i and set i=1*

ii. *Find the search direction S_i as*

$$S_i^T = \begin{cases} (1,0,0,\ldots,0) & for \quad i = 1, n+1, 2n+1, \\ (0,1,0,\ldots,0) & for \quad i = 2, n+2, 2n+2 \\ (0,0,1,\ldots,0) & for \quad i = 3, n+3, 2n+3 \\ (0,0,0,\ldots,1) & for \quad i = 4, n+4, 2n+4 \end{cases}.$$

iii. *Determine whether X_i should be positive or negative.This means that for the current direction S_i, we find whether the function value decreases in the positive or negative direction. For this, we take a small probe length e and evaluate $f = f(X_i)$, $f^+ = f(X_i + eS_i)$ and $f^- = f(X_i - eS_i)$. If $f^+ < f_1$, S_i will be the correct direction for decreasing the value of f and if $f^- < f_i$, $-S_i$ will be the correct one. If both f^+ and f^- are greater than f_i, we take X_i as the minimum along the direction S_i.*

iv. *Find the optimal step length λ_i^* such that $f(X_i \pm \lambda_i^* S_i) = \min_{\lambda_i}(X_i \pm \lambda_i^* S_i)$, where + or – has to be used depending upon whether S_i or $-S_i$ is the direction for decreasing the function value.*

v. *Set $X_{i+1} = X_i \pm \lambda_i^* S_i$ depending on the direction of decreasing the function value, and $f_{i+1} = f(X_{i+1})$.*

vi. *Set the new value of i = i +1, and go to step (ii). Continue this procedure until no significant change is achieved in the value of the objective function.*

Application: We minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ with the starting point as $X_1 = (0,0)$. We will take the probe length (e) as 0.01. it is important to note here that the method of differential calculus will be used to find the optimum step length λ_i^*.

Iteration*i =1:*

The search direction is chosen as $S_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

To find whether the value of *f*decreases along S_1 or $-S_1$, we use the probe length *e*.

Since $f_1 = f(X_1) = f(0,0) = 0$,

$$f^+ = f(X_1 + eS_1) = f(e,0) = 0.01 - 0 + 2(0.0001) + 0 + 0 = 0.0102 > f_1,$$

$$f^- = f(X_1 + eS_1) = f(-e,0) = -0.01 - 0 + 2(0.0001) + 0 + 0 = -0.9998 < f_1.$$

Therefore, $-S_1$ is the correct direction for minimizing *f* from X_1. To find the optimum step λ_i^*, we minimize

$$f_1(X_1, -\lambda S_1) = f(-\lambda_1) = -0 + 2(-\lambda_1)^2 + 0 + 0 = 2 - \lambda_1^{\,2} - \lambda_1$$

We now set $\frac{\partial f}{\partial \lambda_1} = 0$ i.e, $(4\lambda_1 - 1 = 0)$

$\lambda_1^* = ¼$

Set $X_2 = X_1 - \lambda_1^* S_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{4} \\ 0 \end{pmatrix}$

$f_2 = f(-¼, 0) = -\frac{1}{8}$

Iteration $i = 2$:

Choose the direction $S_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Since $f_2 = (X_2) = -0.125$,

$$f^+ = f(X_2 + eS_2) = f(-0.25, 0.01) = -0.1399 < f_2,$$

$$f^- = f(X_2 - eS_2) = f(-0.25, 0.01) = -0.1099 > f_2.$$

Therefore, S_2 is the correct direction minimize *f* from X_2.

We shall now minimize $f(X_2, -\lambda_2 S_2)$ to find λ_2^*.

Here,

$$f(X_2, -\lambda_2 S_2) = f(-0.25, \lambda_2) = -0.25 - \lambda_2 + 2(0.25)^2 - 2(0.25)\lambda_2 + \lambda_2^{\,2} = \lambda_2^{\,2} - 1.5\lambda_2 - 0.125,$$

$\frac{\partial f}{\partial \lambda_2} = 2\lambda_2 - 1.5 = 0$ at $\lambda_2^* = 0.75$.

Set $X_3 = X^2 + \lambda_2^* S_2 = \begin{pmatrix} -0.25 \\ 0 \end{pmatrix} + 0.75\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.25 \\ 0.75 \end{pmatrix}$,

$f_3 = f(X_3) = -0.6875$.

As a result of the clumsy nature of the manual application of the univariate method, a computer program in BASIC codes wasimplemented. The convergence of $f(X^*)$ is found to be -1.25.

III. Hooke and Jeeves Method

The pattern search method of Hooke and Jeeves is a sequential technique, each step of which consists of two kinds of moves, one called the exploratory move and the other called the pattern move. The general procedure can be described by the following steps [11]:

i. *Start with an arbitrary chosen point* $X_1 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ *called the starting base point, and prescribed step length* ΔX_i *in each of the coordinate directions* U_i, $i = 1, 2, \ldots, n$

ii. *Set k = 1 and compute $f_k = f(X_k)$. Set* $i = 1, Y_{k_0} = x_k$ *and start the explanatory more as stated in the next step.*

iii. *The variable X_i is perturbed about the current base point* $Y_{k,i-1}$ *to obtain the new temporary base point as*

$$Y_{k,i} = \begin{cases} Y_{k,i-1} + \Delta X_i U_i\,, & if \quad f^+ = \left(Y_{k,i-1} + \Delta X_i U_i\right) < f = f(Y_{k,i-1}) \\ Y_{k,i} - \Delta X_i U_i\,, & if \quad f^- = (Y_{k,i-1} - \Delta X_i U_i) < f = f(Y_{k,i-1}) < f^+ = \left(Y_{k,i-1} + \Delta X_i U_i\right) \\ Y_{k,i}\,, & if \quad f == (Y_{k,i-1}) < min(f^+, f^-) \end{cases}$$

The process of finding the new temporary base point is continued for i = 1, 2,… until $X_n$is perturbed to find $Y_{k,n}$.

iv. *If the point* $Y_{k,n}$ *remains the same as X_k, reduce the step length* ΔX_i *(say by a factor of 2), set i = 1 and go to step 3.*

If $Y_{k,n}$ *is different from X_k, obtain the new base point* $X_{k+1} = Y_{k,n}$ *and go to step 5.*

v. *With the help of the base point X_k and X_{k+1}establish a pattern direction as* $S = X_{k+1} - X_k$ *and find a point $Y_{k+1,0}$ as* $Y_{k+1,0} = X_{k+1} + \lambda S$*, where λ is the step length which can be taken as 1 for* simplicity.

vi. *Set k = k + 1, $f_k = f(Y_{k,0})$, i=1 and repeat step 3. If at the end of step 3,* $f\left(Y_{k,n}\right) < f\left(X_k\right)$*, the new base point is taken as X_{k+1} and* $Y_{k,n}$ *and control is transferred to step 5. On the other hand, if* $f\left(Y_{k,n}\right) \geq f\left(X_k\right)$*, set $X_{k+1} = X_k$, reduce the step length ΔX_i, set k = k + 1 and go to step 2.*

vii. *The process is assumed to have converged whenever the step lengths fall below a small certain number e. Thus, the process is terminated if* $\max_i(\Delta X_i) < e$.

Application: Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from the point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Take $\Delta X_1 = \Delta X_2 = 0.8$ and $e = 0.01$.

Step 1:We take starting point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and step length as $\Delta X_1 = \Delta X_2 = 0.8$, along the coordinate directions U_1 and U_2 respectively. Set $k = 1$.

Step 2:$f_1 = f(X_1) = 0$, $i = 1$ and $Y_{10} = X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Step 3:To find new temporary base point, we set i=1 and evaluate $f = f(Y_{10}) = 0.0$.

$$f^+ = f(Y_{10} + \Delta X_i U_i) = f(0.8, 0.0) = 2.08,$$

$$f^- = f(Y_{10} - \Delta X_i U_i) = f(-0.8, 0.0) = -0.408.$$

Since $f < min(f^+, f)$ we take $Y_{11} = X_1$. Next, we set $i = 2$ and evaluate $f = f(Y_{11}) = 0.0$.

$$f^+ = f(Y_{11} + \Delta X_2 U_2) = f(0.0, 0.8) = -0.16\,.$$

Since $f^+ < f$, we set $Y_{12} = \begin{pmatrix} 0.0 \\ 0.8 \end{pmatrix}$.

Step 4:As Y_{12} is different from X_1, the new base point is taken as $X_2 = \begin{pmatrix} 0.0 \\ 0.8 \end{pmatrix}$

Step 5: A pattern direction is established as $S = X_2 - X_1 = \begin{pmatrix} 0.0 \\ 0.8 \end{pmatrix} - \begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix} = \begin{pmatrix} 0.0 \\ 0.8 \end{pmatrix}$.

The optimal step length λ^* is found by minimizing $f(X_2 - \lambda S) = f(0.0, 0.8 + 0.8\lambda) = 0.6\lambda^2 + 0.48\lambda - 0.16\,.$

$$\frac{\delta f}{\delta \lambda}=0 \Rightarrow 1.28\lambda+0.48=0 \Rightarrow \lambda^*=-0.375.$$

We obtain the point Y_{20} as

$$Y_{20}=X_2+\lambda^* S=\begin{Bmatrix}0.0\\0.8\end{Bmatrix}-0.375\begin{Bmatrix}0.0\\0.8\end{Bmatrix}=\begin{Bmatrix}0.0\\0.5\end{Bmatrix}.$$

Step 6: Set $k = 2, f = f_2 = f(Y_{20}) = -0.025$ and repeat step 3. Thus with $i = 1$, we evaluate

$$f^+=f(Y_{20}+\Delta X_1 U_1)=f(0.8,0.5)=2.63,$$
$$f^-=f(Y_{20}-\Delta X_1 U_1)=f(-0.8,0.5)=-0.57.$$

Since $f^- < f < f^+$, we take $Y_{21}=\begin{pmatrix}-0.8\\0.5\end{pmatrix}$.

Next we set $i = 2$ and evaluate $f = f(Y_{21}) = -0.57$ and $f^+ < f\left(Y_{21}-\Delta X_2 U_2\right)=f\left(-0.8,1.3\right)=-1.21$.As $f^+ < f$, we take $Y_{22}=\begin{pmatrix}-0.8\\1.3\end{pmatrix}$.

Since $f(Y_{22}) = 1.21 < f(X_2) = -0.25$, we take the new base point as $X_3=Y_{22}=\begin{pmatrix}-0.8\\1.3\end{pmatrix}$ and go to step 5

Step 5:A pattern direction S is established as

$$S=X_3-X_2=\begin{pmatrix}-0.8\\1.3\end{pmatrix}-\begin{pmatrix}0.0\\0.8\end{pmatrix}=\begin{pmatrix}-0.8\\0.5\end{pmatrix}.$$

$$f(X_3-\lambda S)=f(-0.8-0.8\lambda,1.3+0.5\lambda)=0.73\lambda^2-0.32\lambda-1.21.$$

$$\frac{\partial f}{\partial \lambda}=0 \Rightarrow 1.46\lambda-0.32=0 \Rightarrow \lambda^*=-0.219.$$

We obtain the point Y_{30} as $Y_{30}=X_3+\lambda^* S=\begin{pmatrix}-0.8\\1.3\end{pmatrix}-0.219\begin{pmatrix}-0.8\\0.5\end{pmatrix}=\begin{pmatrix}-0.975\\1.410\end{pmatrix}.$

Step 6: Set $k = 3, f = f_2 = f(Y_{30}) = -1.235$ and repeat step 3. By setting $i = 1$, we evaluate

$$f^+=f(Y_{30}+\Delta X_1 U_1)=f(-0.175,1.410)=-0.018$$
$$f^-=f(Y_{30}-\Delta X_1 U_1)=f(-1.775,1.410)=0.105.$$

Since $f < min(f^+, f)$, we set $Y_{31}=Y_{30}=\begin{pmatrix}-0.975\\1.410\end{pmatrix}$.

Next we set $i = 2$to obtain

$$f=f(Y_{31})=-1.235$$
$$f^+=f(Y_{31}+\Delta X_2 U_2)=f(-0.975,2.210)=-0.695$$
$$f^-=f(Y_{31}-\Delta X_2 U_2)=f(-0.975,0.610)=0.501.$$

Since $f < min(f^+, f)$, we set $Y_{32}=Y_{31}=\begin{pmatrix}-0.975\\1.410\end{pmatrix}$, and as $Y_{32}$is different from X_3, we take the new base point as $X_4=Y_{32}=\begin{pmatrix}-0.975\\1.410\end{pmatrix}$ and go to step 5.

Step 5: A pattern direction S is established as

$$S=X_4-X_3=\begin{pmatrix}-0.975\\1.410\end{pmatrix}-\begin{pmatrix}-0.8\\1.3\end{pmatrix}=\begin{pmatrix}-0.175\\0.11\end{pmatrix},$$

$$f(X_4+\lambda S)=f(-0.975-0.175\lambda,1.41+0.11\lambda)=-1.237-0.22\lambda+0.0341\lambda_2,$$

$$\frac{\partial f}{\partial \lambda} = 0.6\lambda + 0.22 = 0 \Rightarrow \lambda^* = -0.22/0.66 = -3.67,$$

$$Y_{40} = x_4 + \lambda^* S = \begin{pmatrix} 0.64 \\ 0.40 \end{pmatrix}.$$

Step 6:$f = f_4 = f(Y_{40}) = 1.7312$. Set $k = 4$ and repeat step 3. By setting $k = 4$, we evaluate

$$f^+ = f(Y_{40} + \Delta X_1 U_1) = f(1.44, 0.4) = 6.5$$

$$f^- = f(Y_{40} - \Delta X_1 U_1) = f(-0.16, 0.4) = 0.4768.$$

Since $f^+ < f < f$, we take $Y_{41} = \begin{pmatrix} -0.16 \\ 0.4 \end{pmatrix}$. We then set $i = 2$ and evaluate $f = f(Y_{41}) = -0.4768$,

$f^+ = f(Y_{41} + \Delta X_2 U_2) = f(-0.98, 1.48) = -1.25$.

As $f^+ < f$, we take $Y_{42} = \begin{pmatrix} -0.98 \\ 1.48 \end{pmatrix}$, $f(Y_{42}) = -1.25$ and $f(X_4) = -1.245$.

Since $f(Y_{42}) < f(X_4)$, we take the new base point as $X_5 = Y_{42} = \begin{pmatrix} -1.0 \\ 1.5 \end{pmatrix}$ which can be seen as the optimum point.

IV. Powell's method

Theorem 4.1. If a quadratic function $Q(X) = \frac{1}{2} X^T A X + B^T X + C$ is minimized sequentially, once along each direction of a set of n – linearly independent, A –conjugate directions, the global minimum of Qwill be located at or before the n^{th} step regardless of the starting point.

This method is an extension of the basic pattern search method and is known to be a method that makes use of conjugate directions.According to Theorem4.1, it is known that a conjugate direction method minimizes a quadratic function in a finite number of steps.

Definition4.1 Let A be an $n \times n$ symmetric matrix. A set of n vectors (or directions) {S} is said to be conjugate (more accurately A – conjugate) if $S_i^T A S_j = 0$ for all $i \neq j$, $i = 1,2,...,n$, $j = 1,2,...n$.

Definition 4.2 If a minimization always locates the minimum of the general quadratic function in no more than a pre-determined number of operations and if the limiting number of operations is directly related to the number of variables n, then the method is said to be quadratically convergent.

The quadratic convergent property of Powell's method has ranked it as the most efficient direct search method. This is not because we often get quadratic functions for minimization, but because of the fact that most of the function can be approximated very closely by a quadratic function near their minima [5].

Application: Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from the point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

f is minimized along $S_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ from X_1.

$$f_1 = f(X_1) = 0.0$$

$$f^+ = f(X_1 + eS_1) = f(0.0, 0.01) = -0.0099 < f_1.$$

f decreases along the direction S_2 and so $f(X_1 + \lambda S_2) = f(0.0, \lambda) = \lambda^2 - \lambda$.

$$\frac{\partial f}{\partial \lambda} = 2\lambda - 1 = 0 \Rightarrow \lambda^* = 1/2.$$

$X_2 = X_1 + \lambda^* S_2 = \begin{pmatrix} 0.0 \\ 0.5 \end{pmatrix}$ minimize *f* along $S_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ from X_2.

$$f_2 = f(X_2) = f(0.0, 0.5) = -0.25$$

$$f^+ = f(X_2 + eS_1) = f(0.01, 0.50) = -0.2298 > f_2$$

$$f^- = f(X_2 + eS_1) = f(-0.01, 0.50) = -2.689 < f_2$$

$$f(X_2 + \lambda S_1) = f(-\lambda, 0.50) = 2\lambda^2 - 2\lambda - 0.25$$

$$\frac{\partial f}{\partial \lambda} = 4\lambda - 2 = 0 \Rightarrow \lambda^* = \frac{1}{2}.$$

$X_3 = X_2 + \lambda^* S_1 = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$ minimizes*f* along $S_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ from X_3.

$$f_3 = f(X_3) = -0.75$$

$$f^+ = f(X_3 + eS_2) = f(-0.5, 0.51) = -0.7599 < f_3$$

$$f(X_3 + \lambda S_2) = f(-0.5, 0.5 + \lambda) = \lambda^2 - \lambda - 0.75$$

$$\frac{\partial f}{\partial \lambda} = 2\lambda - 1 = 0 \Rightarrow \lambda^* = \frac{1}{2}.$$

$$X_4 = X_3 + \lambda^* S_2 = \begin{pmatrix} -0.5 \\ 1.5 \end{pmatrix}$$

The first pattern direction is found to be $S_p^{(1)} = x_4 - x_2 = \begin{pmatrix} -\frac{1}{2} \\ 1.5 \end{pmatrix} - \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$.

We now minimize*f* along $S_P^{(1)}$ from X_4

$$f_4 = f(X_4) = -1.0$$

$$f^+ = f(X_4 + eS_p^{(1)}) = f(-0.5 - 0.005, 1 + 0.005) = f(-0.505, 1.005) = -1.004875 < f_4$$

$$f(X_4 + \lambda S_p^{(1)}) = f(-0.5 - 0.05, 1 + 0.5\lambda) = 0.25\lambda^2 - 0.05\lambda - 1$$

$$\frac{\partial f}{\partial \lambda} = 0.5\lambda - 0.5 = 0 \Rightarrow \lambda^* = 1$$

$X_5 = X_4 + \lambda^* S_P^{(1)} = \begin{pmatrix} -1 \\ \frac{1}{2} \end{pmatrix}$, so we minimize *f* along S_2 from X_5

$$f_5 = f(X_5) = -1.25$$

$$f^+ = f(X_5 + eS_2)$$

$$f^- = f(X_5 + eS_2)$$

This implies that *f* cannot be minimized along S_2 and therefore $X_5 = \begin{pmatrix} -1 \\ \frac{1}{2} \end{pmatrix}$ is the optimum point.

V. Conjugate Gradient Method

The conjugate gradient method as a method makes use of conjugate direction which leads to fast convergence. Any minimization method that makes use of conjugate direction is quadratically convergence which is a very useful property because it ensures that the method will minimize a quadratic function in two steps or less. This also applies to non-quadratic functions because they can be approximated by a quadratic near the optimum point and any quadratically convergent function is supposed to converge to an optimum within a finite number of iterations [12].The algorithm is as follows:

Begin with an initial point X_1
Let $S_1 = -\Delta f(X_1) = -\Delta f_1$
$X_2 = X_3 + \lambda_1 {}^* S_1$, set i =2 and proceed
Compute $\Delta f_i = \Delta f X x_i$) and set $S_i = -\Delta f_i + \left\{ |\Delta f_i|^2 \big/ |\Delta f_{i-1}|^2 \right\} S_{i-1}$
The optimum length λ_i^* *is computed in the S_i direction and the new point is given as* $X_{i+1} = X_i + \lambda_i^* S_i$
Test for optimality of the point x_{i+1} (x_{1+1} is optimum if there is no search direction to reduce f further). If optimality is not satisfied, set I = I + 1 and repeat steps (iv), (v) and (vi) until optimality condition is met.

Application: We want to minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from the point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\Delta f = (1 + 4x_1 + 2x_2, -1 + 2x_1 + 2x_2)$ and $S_1 = \Delta f(X_1) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

We now minimize $f(X_1 + \lambda_1 S_1)$ with respect to Δ_1.

$$f(X_1 + \lambda_1 S_1) = f(\Delta_1, \lambda_1) = \Delta_1^2 - 2\lambda_1$$

$$\frac{\partial f}{\partial \lambda} = 2\lambda_1 - 2 = 0 \Rightarrow \lambda_1^* = 1$$

$$X_2 = X_1 + \lambda_1^* S_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Iteration *i = 2*

$$\Delta f_2 = \Delta f(X_2) = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

$S_2 = \Delta f_2 + \frac{|\Delta f_2|^2}{|\Delta f_1|^2} S_1$, where $|\Delta f_2|^2 = \sqrt{\left(\frac{\partial f_2}{\partial x_1}\right)^2 + \left(\frac{\partial f_2}{\partial x_1}\right)^2}$ *and* $|\Delta f_1|^2 = \sqrt{\left(\frac{\partial f_1}{\partial x_1}\right)^2 + \left(\frac{\partial f_1}{\partial x_1}\right)^2}$.

$|\Delta f_2|^2$ and $|\Delta f_1|^2 = 2$.

$$S_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

$$f(X_2 + \lambda_2 S_2) = f(-1, 1 + 2\lambda_2) = 4\lambda_2^2 - 2\lambda_2 - 1$$

$$\frac{\partial f}{\partial \lambda} = 8\lambda_2 - 2 = 0 \Rightarrow \lambda_1^* = \frac{1}{4}$$

$$X_3 = X_2 + \lambda_2^* S_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}.$$

Iteration *i = 3*

$\Delta f_3 = \Delta f(X_3) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, |\Delta f_2|^2 = 2$ and $|\Delta f_3|^2 = 0$.

$$S_3 = -\Delta f_3 + \left\{ |\Delta f_3|^2 \big/ |\Delta f_2|^2 \right\} S_2 = -\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This shows that there is no search direction to reduce *f* further and hence $X_3 = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$ is the optimum point.

VI. Variable Metric Method

This is a method for determining numerically local minima of differentiable functions of several variables. In the process of locating each minimum, a matrix which characterizes the behavior of the function about the minimum is determined. For a region in which the function depends quadratically on the variables, no more than *N* iterations are required, where *N* is the number of variables [6].

This method makes use of the derivatives that are currently available and has the following algorithm.

i. *Start with an initial point X_1 and an n x n positive semi-definite matrix H_1. For simplicity, H_1 can be taken as the identity matrix I. set I = 1*

ii. $S_i = -H_i \Delta f_i(x_i)$

iii. $x_{i+1} = x_i + \lambda_i^* S_i$

iv. *Test x_{i+1} for optimality (x_{i+1} is optimal if $\Delta f(x_{i+1}) = 0$) if not optimal proceed.*

v. *Update matrix H as follows:*

$$Q_i = \Delta f(x_{1+1}) - \Delta f(x_i) = \Delta f_{i+1} - \Delta f_i$$

$$N_i = \frac{-(H_i Q_i)(H_i Q_i)^T}{Q_i^T H_i Q_i}$$

$$M_i = \lambda_i^* \frac{S_i S_i^T}{S_i Q_i}$$

$$H_{i+!} = H_i + M_i + N_i$$

vi. *Set I = I + 1 and go to step 2*

Application:We minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from the point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Let $H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

$$\Delta f = \Delta f(X_1) = (1 + 4x_1 + 2x_2, -1 + 2x_1 + 2x_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$S_1 = -H_1 \Delta f_1(X_1) = -\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \times \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$f(X_1 + \lambda_1 S_1) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$$

$$\partial f \big/ \partial \lambda_1 = 8\lambda_2 - 2 = 0 \Rightarrow \lambda_1^* = 1$$

$$X_2 = X_1 + \lambda_1^* S_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\Delta f(X_2) = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

For iteration *i = 2*

$$Q = \Delta f(X_2) - \Delta f(X_1) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$M_1 = \frac{\lambda_1^* S_1 S_1^T}{S_1^T Q_1} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$Q_1^T H_1 Q_1 = (-2 \quad 0)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = 4$$

$$H_1 Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$N_i = \frac{-(H_i Q_i)(H_i Q_i)^T}{Q_i^T H_i Q_i} = \frac{-\begin{pmatrix} -2 \\ 0 \end{pmatrix}(-2 \quad 0)}{4} = -\frac{1}{4}\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$$

$$H_2 = H_1 + M_1 + N_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

$$S_2 = -H_2 \Delta f_2 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$f(X_2 + \lambda_2 S_2) = f(-1, 1 + \lambda_2) = \lambda_2^2 - \lambda_2 - 1$$

$$\frac{\partial f}{\partial \lambda_1} = 2\lambda_2 - 1 = 0 \Rightarrow \lambda_0^* = \frac{1}{2}$$

$$X_3 = X_2 + \lambda_2^* S_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1.0 \\ 1.5 \end{bmatrix}$$

$$\Delta f(X_3) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Therefore $X_3 = \begin{bmatrix} -1 \\ 1.5 \end{bmatrix}$ is the optimum point.

VII. Conclusion

There is at present no universal optimization routine which will solve any given problem more efficiently than any other method and it should always be borne in mind that there are some problems for which those algorithms usually considered to be inefficient may prove very useful. The initial choice lies between direct search and gradient methods, and gradient methods should only be used if the derivatives of the function can be obtained analytically.It is obvious from the investigation carried out so far that methods which make use of the gradient function are much more effective than those void of the gradient function, which simply knocks off the direct search methods. The variable metric method (descent method) is observed to be the most efficient of all the methods of solving unconstrained non-linear optimization problems considered in this work since it converges to the optimum in a considerably smaller number of iterations and it is also very stable and cuts across a much wider class of functions. If analytic derivatives don't exist, then a direct search method should be used and of these Powell's algorithm is probably the most effective,

Reference

[1]. Kreyzig, E. *Advanced Engineering Mathematics*, John Wiley & Sons, 2010.
[2]. Rao, S. S. *Optimization Theory and Applications*, Wiley Eastern Ltd, 1979.
[3]. Hillier, F. S. and Lieberman, G. J. *Introduction to Operations Research*, McGraw-Hill, New York, 2009.
[4]. Taha, H. A., *Operation Research: An Introduction*, Pearson Prentice Hall Inc, New Jersey, U.S.A . 2017.
[5]. Powell, M. J. D. An Efficient Method of Finding the Minimum of a Function of Several Variables Without Calculating Derivatives, *The Computer Journal*, Vol 7 (2), 1964, pp. 155–162.

[6]. Davidon, W. C. Variable Metric Method for Minimization, *Siam Journal of Optimization*, Vol. 1 (1), 1991, pp. 1-17.

[7]. Box, M. J., Davies, D. and Swann, W. H. Non-linear Optimization Techniques, *I.C.I. Monograph No. 5* (Oliver and Boyd, Edinburgh, 1969).

[8]. Alrefaei, M. H. and Andradottir, S. (2001), A Modification of the Stochastic Ruler Method forDiscrete Stochastic Optimization, *European Journal of Operational Research*, 133(1), 160–182.

[9]. Ankenman, B., Nelson, B. L. and Staum, J. (2010), Stochastic Kriging for Simulation Meta-Modeling, *Operations Research*, 58(2), 371–382.

[10]. Wilson, P. and Mantooth, H. A. (2013), *Model-Based Engineering for Complex Electronic Systems*, Newnes, United States.

[11]. Gottfried, B. S. and Weisman, J. *Introduction to Optimization Theory*, Prentice Hall, Englewood Cliffs, NJ, 1973.

[12]. Djordjevic, S. S.*Unconstrained Optimization Methods: Conjugate Gradient Methods and Trust-Region Methods,* IntechOpen, 2019. doi: 10.5772/intechopen.84374.

Ubon Akpan Abasiekwere, et. al. "Various Solution Techniques for Unconstrained Non-Linear Optimization Problems – A Survey." *IOSR Journal of Mathematics (IOSR-JM),* 16(3), (2020): pp. 15-25.