AN OVERVIEW ON CLASSIFICATION AND METHODS USED IN UNCONSTRAINED OPTIMIZATION

An Overview on Classification and Methods Used in Unconstrained Optimization

Shashi Sharma*

Mathematics Department, D. A. V. College, Muzaffarnagar-251001

Abstract – The Numerical Optimization of general nonlinear multivariable target functions requires productive and robust techniques. Effectiveness is significant on the grounds that these issues require an iterative arrangement technique, and experimentation ends up unrealistic for more than three or four variables. Robustness (the capacity to accomplish an answer) is alluring in light of the fact that a general nonlinear function is capricious in its conduct; there might be relative maxima or minima, saddle focuses, locales of convexity, concavity, etc. In certain areas the optimization algorithm may advance all around gradually toward the optimum, requiring over the top PC time. Luckily, we can draw on broad involvement in testing nonlinear programming algorithms for unconstrained functions to assess different methodologies proposed for the optimization of such functions. In this Article, we studied about the Unconstrained Optimization, its classification and Methods used in it.

-------------------------◆----------------------------

1.   INTRODUCTION

The analytics of varieties is worried about the assurance of extreme (maxima and minima) or stationary estimations of functional. A functional can be characterized as a function of a few different functions. The analytics of varieties is an intense technique for the arrangement of problems in optimal monetary activity of intensity frameworks. In this segment we present the subject of variational math through a derivation of the Euler equations and related transversality conditions.

In the unconstrained optimization problem, we have to discover the estimation of the vector $X = [x_1,....,x_n]T$ that limits the function:

$$f(x_1, ...x_n)$$

given that the function f is continuous and has a first-arrange derivative.

To get the minimum as well as maximum of the function f we set its first derivative as for x to zero

$$\frac{\partial f(x_1, ...x_n)}{\partial x_1} = 0$$
$$\frac{\partial f(x_1, ...x_n)}{\partial x_2} = 0$$
$$\cdot$$
$$\frac{\partial f(x_1, ...x_n)}{\partial x_1} = 0$$

Above equations speak to n equations in n questions. The arrangement of these equations produces competitor arrangement focuses. In the event that the function f has second halfway derivatives, at that point we ascertain the Hessian matrix,

$$H = \frac{\partial^2 f(x_1, ...x_n)}{\partial x_i^2}$$

On the off chance that the matrix H is sure unequivocal, at that point the function f is a minimum at the applicant points, yet in the event that the matrix H is negative unmistakable then f is a maximum at the candidate points.

2. CLASSIFICATIONS OF UNCONSTRAINED OPTIMIZATION METHODS

A few methods are accessible for taking care of an unconstrained minimization problem. These methods can be grouped into two general classifications as direct search methods and descent methods. The direct search methods require just the target function values yet not the partial derivatives of the function in finding the minimum and consequently are regularly called the non-gradient methods. The direct search methods are otherwise called zeroth-order methods since they utilize zeroth-order derivatives of the function. These methods are most reasonable for basic problems including a generally modest number of variables. These methods are, by and large, less effective than the descent methods. The descent methods require, notwithstanding the function values, the first and sometimes the second derivatives of the

goal function. Since more data about the function being limited is utilized (using derivatives), descent methods are for the most part more effective than direct search procedures. The descent methods are known as gradient methods. Among the gradient methods, those requiring just first derivatives of the function are called first-order methods; those requiring both first and second derivatives of the function are named second-order methods.

3. UNCONSTRAINED OPTIMALITY CONDITIONS

All the unconstrained minimization methods are iterative in nature and thus they begin from an underlying preliminary arrangement and continue toward the minimum point in a consecutive way. The iterative procedure is given by

$$X_{i+1} = X_i + \lambda_i^* S_i$$

Where X_i is the starting point, Si is the search direction, λ_i^* is the optimal step length, and X_{i+1} is the last point in emphasis I. Note that all the unconstrained minimization methods (1) require an underlying point X1 to begin the iterative methodology, and (2) vary from each other just in the method of creating the new point X_{i+1} (from X_i) and in testing the point X_{i+1} for optimality.

3.1 Rate of Convergence

Distinctive iterative optimization methods have diverse rates of union. When all is said in done, and optimization method is said to have joining of order p if

$$\frac{\| X_{i+1} - X^* \|}{\| X_i - X^* \|^p} \leq k, \quad k \geq 0, \quad p \geq 1$$

where Xi and Xi+1 denote the points obtained at the end of iterations i and i + 1, respectively, X^* speaks to the ideal point, and $\|X\|$ indicates the length or standard of the vector X:

$$\| X \| = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$$

If $p = 1$ and $0 \leq k \leq 1$, the method is said to be straightly united (relates to moderate joining). In the event that p = 2, the method is said to be quadratically merged (relates to quick joining). An optimization method is said to have super linear meeting (compares to quick joining) if

$$\lim_{i \to \infty} \frac{\| X_{i+1} - X^* \|}{\| X_i - X^* \|} \to 0$$

The meanings of rates of convergence given in equations are pertinent to single-variable and in addition multivariable optimization problems. On account of single-variable problems, the vector, X_i, for example, degenerates to a scalar, x_i.

3.2 Scaling of Design Variables

The rate of convergence of most unconstrained minimization methods can be enhanced by scaling the design variables. For a quadratic target function, the scaling of the design variables changes the condition number† of the Hessian matrix. At the point when the condition number of the Hessian matrix is 1, the steepest descent method, for instance, finds the minimum of a quadratic target function in a cycle:

If $f = \frac{1}{2} X^T [A] X$ denotes a quadratic term, a transformation of the form

$$X = [R] Y \quad or \quad \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix}$$

Can be used to obtain a new quadratic term as

$$Y = [S] Z \quad or \quad \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = \begin{bmatrix} s_{11} & 0 \\ 0 & s_{22} \end{bmatrix} \begin{Bmatrix} z_1 \\ z_2 \end{Bmatrix}$$

Where the matrix [S] is given by

$$[S] = \begin{bmatrix} s_{11} = \dfrac{1}{\sqrt{\tilde{a}_{11}}} & 0 \\ 0 & s_{22} = \dfrac{1}{\sqrt{\tilde{a}_{22}}} \end{bmatrix}$$

4. Methods Used in Unconstrained Optimization

4.1 Grid Search Method

This method includes setting up a reasonable network in the design space, assessing the target function at all the brace points, and finding the framework direct comparing toward the most minimal function value. For instance, if the lower and upper limits on the ith design variable are known to be l_i and u_i, respectively, we can divide the range (l_i, u_i) into pi − 1 equal parts so that x (1) i , x (2) i , . . . , x (pi) i denote the grid points along the xi axis (i = 1, 2, . . . , n). This leads to a total of $p_1 p_2 \cdots$ pn grid points in the design space. A lattice with pi = 4 is appeared in a two-dimensional design space. The lattice points can likewise be picked in view of methods of exploratory design. It can be seen that the framework method requires restrictively vast number of function assessments in most pragmatic problems. For instance, for a problem with 10 design variables (n = 10), the quantity of matrix points will be 3^{10} = 59,049 with pi = 3 and 4^{10} = 1,048,576 with pi = 4. Notwithstanding, for problems with few design variables, the grid method can be utilized helpfully to locate an inexact minimum. Additionally, the grid method can be utilized to locate

Shashi Sharma*

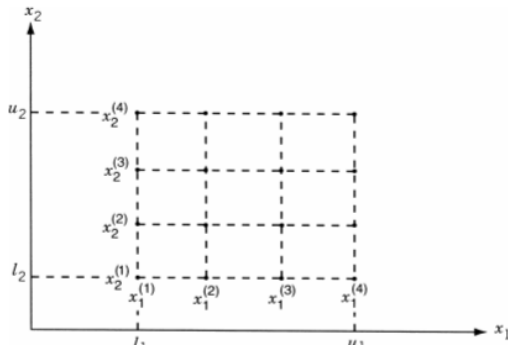a decent beginning stage for one of the more effective methods.

Figure 1: Grid with p_i = 4.

4.2 Newton's Method

Newton's method, with every one of its varieties, is the most critical method in unconstrained optimization. Let f: $IR^n \to IR$ be a given function and assume that $\nabla^2 f$ is continuous. Newton's method for the minimization of f can be inferred expecting that, given x_k, the point xk+1 are acquired limiting a quadratic guess of f. As f is two times differentiable, it is conceivable to compose:

$$f(x_k + s) = f(x_k) + \nabla f(x_k)`s + \frac{1}{2} s`\nabla^2 f(x_k)s + \beta(x_k, s),$$

In which

$$\lim_{\|s\| \to 0} \frac{\beta(x_k, s)}{\| s \|^2} = 0$$

For ||s|| adequately little, it is conceivable to rough $f(xk + s)$ with its quadratic approximation

$$q(s) = f(x_k) + \nabla f(x_k)s`\nabla^2 f(x_k)s$$

If $\nabla 2 f(xk) > 0$, the value of s minimizing q(s) can be gotten setting to zero the gradient of q(s), i.e

$$\nabla q(s) = \nabla f(x_k) + \nabla^2 f(x_k)s = 0$$

Yielding

$$s = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

The point xk+1 is thus given by

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Finally, Newton's method can be described by the simple scheme.

Step 0: Given $x_0 \in IR^n$.

Step 1: Set k = 0

Step 2: Compute

$$s = -[\nabla^2 f(x_k)]^2 \nabla f(x_k)$$

Step 3:

Set $xk + 1 = xk + s, k = k + 1$. Go to Step 2.

4.3 Quasi-Newton Method

By and large, methods that use gradient data look to locate a stationary purpose of f by finding a zero of the gradient ∇f. A general class of methods, semi Newton methods, tries to do this by utilizing Newton's method to discover a base of ∇f. The hidden supposition in these methods is that the function f can locally be approximated by a quadratic.

Standard Newton's method refreshes candidate arrangements at every iteration via

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Where $\nabla 2 f(x)$ indicates the Hessian, or the second derivative of f . Updates can be extremely costly since we should locate the reverse of a n x n matrix at each iteration. To ease computational cost, approximations to the Hessian and its backwards are utilized. There are various ways the Hessian can be approximated, one method that is widely utilized is from the Broyden family which utilizes a raised mix of Daviodon– Fletcher– Powell and BFGS refreshes.

4.4 Univariate Method

In this method we change just a single variable at any given moment and try to deliver an arrangement of enhanced approximations to the minimum point. By beginning at a base point Xi in the ith iteration, we settle the values of n − 1 variables and fluctuate the staying variable. Since just a single variable is changed, the problem turns into a one-dimensional minimization problem can be utilized to create another base point X_{i+1}. The search is presently preceded in another direction.

This new direction is acquired by changing any of the n − 1 variables that were settled in the past iteration. Truth be told, the search method is proceeded by taking each organize direction thusly. After all the n

Shashi Sharma*

directions are searched successively, the principal cycle is finished and henceforth we rehash the whole procedure of consecutive minimization. The technique is preceded until the point when no further change is conceivable in the target function in any of the n directions of a cycle.

The univariate method is exceptionally basic and can be executed effectively. Be that as it may, it won't merge quickly to the ideal arrangement, as it tends to waver with consistently diminishing advancement toward the ideal. Henceforth it will be smarter to stop the calculations sooner or later close to the ideal point instead of endeavoring to locate the exact ideal point. In theory, the univariate method can be connected to locate the minimum of any function that has continuous derivatives. In any case, if the function has a precarious valley, the method may not by any means focalize. For instance, think about the forms of a function of two variables with a valley. In the event that the univariate search begins at point P, the function value can't be diminished either in the direction $\pm S_1$ or in the direction $\pm S_2$. In this manner the search stops and one might be deluded to take the point P, which is surely not the ideal point, as the ideal point. This circumstance emerges at whatever point the value of the test length ε required for distinguishing the best possible direction $(\pm S_1 \text{ or } \pm S_2)$ happens to be not as much as the quantity of huge figures utilized as a part of the calculations.

4.5 Steepest Descent Method

The utilization of the negative of the gradient vector as a direction for minimization was first made by Cauchy in 1847. In this method we begin from an underlying preliminary point X1 and iteratively move along the steepest descent directions until the point that the ideal point is found. The steepest descent method can be condensed by the accompanying advances:

1. Start with an arbitrary initial point X_1. Set the iteration number as i = 1.

2. Find the search direction S_i as

$$S_i = -\nabla f_i = -\nabla f(X_i)$$

3. Determine the optimal step length λ_i^* in the direction Si and set

$$X_{i+1} = X_i + \lambda_i^* S_i = X_i - \lambda_i^* \nabla f_i$$

4. Test the new point, X_{i+1} for optimality. If X_{i+1} are optimum, stop the process. Otherwise, go to step 5.

5. Set the new iteration number i = i + 1 and go to step 2.

The method of steepest descent may have all the earmarks of being the best unconstrained minimization system since every one-dimensional search begins in the "best" direction. Nonetheless, inferable from the way that the steepest descent direction is a neighborhood property, the method isn't generally viable in many problems.

4.6 Random Search Methods

Random search methods depend on the utilization of random numbers in finding the minimum point. Since a large portion of the PC libraries have random number generators, these methods can be utilized helpfully. A portion of the best known random search methods are introduced in this segment.

5. CONCLUSION

In most structural design problems the objective is to limit a function with many design variables, yet the investigation of minimization of functions of a solitary design variable is significant for a few reasons. Substituting projection methods have been broadly used to locate the nearest point, to a given point, in the convergence of a few given sets that have a place with a Hilbert space. One of the attributes of these plans is the moderate union that can be seen in practical applications. Albeit most structural optimization problems include limitations that bound the design space, investigation of the methods of unconstrained optimization is significant for a few reasons. Above all else, on the off chance that the design is at a phase where no requirements are dynamic, at that point the way toward deciding a look bearing and travel remove for limiting the objective function includes an unconstrained function minimization algorithm. Obviously in such a case one has always to look for requirement violations amid the move in design space.

6. REFERENCES

1. Kamat, M.P. and Hayduk, R.J., (Hayduk. 2011) "Recent Developments in Quasi–Newton Methods for Structural Analysis and Synthesis," AIAA J., 20 (5), pp. 672–679,

2. Avriel, M., Nonlinear (Nonlinear, 2006) Programming: Analysis and Methods. Prentice–Hall, Inc.,

3. Powell, M.J.D., (M.J.D.2014) "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," Computer J., 7, pp. 155–162, Kiefer, J., "Sequential Minmax Search for a Maximum," Proceedings of the American Mathematical Society, 4, pp. 502–506, 2013.

Shashi Sharma*

4.   Nelder, J. A. and Mead, R., (Nelder & Mead. 2015) "A Simplex Method for Function Minimization," Computer J., 7, pp. 308–313,

5.   Chen, D. H., Saleem, Z., and Grace, D. W.,(Saleem, 2016) "A New Simplex Procedure for Funct   ion Minimization," Int. J. of Modelling & Simulation, 6, 3, pp. 81–85,

7.   Cauchy, A., (Cauchy. 2007.) "Methode Generale pour la Resolution des Systemes D'equations Simultanees," Comp. Rend. l'Academie des Sciences Paris, 5, pp 536–538,

8.   Hestenes, M.R. and Stiefel, E.,(Hestenes & Stiefel 2012.) "Methods of Conjugate Gradients for Solving Linear Systems," J. Res. Nat. Bureau Stand., 49, pp.409–436,

9.   Fletcher, R. and Reeves, C.M.,(Fletcher & Reeves 2014.)"Function Minimization by Conjugate Gradients," Computer J., 7, pp.149–154,

10. Gill, P.E. and Murray, (Murray. 2009.)W.,"Conjugate-Gradient Methods for Large Scale Nonlinear Optimization," Technical Report 79-15; Systems Optimization Lab., Dept. of Operations Res., Stanford Univ., pp. 10–12,

Corresponding Author

Shashi Sharma*

Mathematics Department, D.   A. V. College, Muzaffarnagar-251001