APPLICATIONS OF A HYBRID DERIVATIVE-FREE ALGORITHM FOR LOCATING EXTREMA

Dudley J. Benton
Tennessee Valley Authority
Engineering Laboratory
Norris, Tennessee

presented at the
SIAM-SEAS Annual Meeting
April 12-13, 1991
Cullowhee, North Carolina

ABSTRACT

Applications of a hybrid derivative-free algorithm for locating extrema of nonlinear functions of several variables based on Broyden's method is presented in which the problems of starting values and extraneous entrapment are addressed. The principal intended application of the algorithm is to find solutions to simultaneous nonlinear equations. The main objective of the algorithm is to minimize the number of function evaluations for problems where the equations are computationally intensive or partial derivatives cannot be determined analytically. Four examples drawn from diverse fields are given for illustration. Comparisons are made to the Newton-Raphson, conjugate-gradient, and steepest-descent methods.

NOMENCLATURE

A=rectangular matrix having ***M*** columns and ***N*** rows
B=column matrix having ***M*** elements
F=a function of several variables
M=the number of residuals ($M \geq N$)
N=the number of unknowns
R=residual column matrix having ***M*** elements
X=unknown column matrix having ***N*** elements

superscript
T=matrix transpose

subscripts
N=new or current value
O=old or previous value

INTRODUCTION

Many practical problems can be cast into the form of a search for extrema of a function of several variables. A common function is the sum of squared residuals, in which case the extrema of interest are the roots of simultaneous equations. Methods abound which require knowledge of the partial derivatives. Many of these derivative-based methods can be adapted by using finite differences to solve problems where the partial derivatives cannot be analytically determined. Such implementations are impractical when the function is computationally intensive.

Derivative-based methods such as the Newton-Raphson discard at each step all information previously learned about the behavior of the function except the current location. Even the Conjugate-Gradient method when applied to nonlinear problems may only preserve one previous direction of search. When the function evaluation is computationally intensive it is essential that as much information as possible about the behavior of the function learned from previous evaluations be preserved and utilized.

Broyden's method is very attractive when considered from this perspective. It does not require knowledge of the partial derivatives, nor does it attempt to compute them directly. Furthermore, Broyden's method preserves all of the information learned about the behavior of the function for the last ***N+1*** steps where ***N*** is the number of unknowns.

Four enhancements to Broyden's method were made to arrive at the present algorithm: a method for selecting starting values, step length control, hybrid search algorithm, and a method for escaping from extraneous entrapment.

THE BASIC METHOD

Given a set of ***N*** unknowns represented by the column matrix ***X*** and a corresponding set of ***M*** residuals represented by the column matrix ***R***, the least-squares function would be $F=R^TR$. The extrema of ***F*** occur at the locations where $\partial F/\partial X=0$. If the residuals, ***R***, were linear functions of the unknowns, ***X***, then the function, ***F***, would be quadratic and its contours would plot as ellipsoids. This linear case could be described by Equations 1 and 2

$$R = A^T X + B \qquad \textbf{(1)}$$

$$F = R^T R = X^T A A^T X + 2 B^T A^T X + B^T B \qquad \textbf{(2)}$$

where ***A*** is a rectangular matrix having ***M*** columns and ***N*** rows and ***B*** is a column matrix having ***M*** elements.

Broyden (1969) reasoned that ***A*** and ***B*** should be selected such that exact agreement would be preserved for the previous ***N+1*** steps. Assuming that no two of the previous ***N+1 Xs*** are the same, there should be a unique solution to the resulting ***M(N+1)*** equations for the elements of ***A*** and ***B***. Ignoring for the moment how this sequence of ***Xs*** might be obtained, the matrices ***A*** and ***B*** can be sequentially updated using the following algorithm

$$A_N = \frac{A_O + \left[(R_N - R_O) - A_O (X_N - X_O) \right] (X_N - X_O)^T}{(X_N - X_O)^T (X_N - X_O)}$$

(3)

$$B_N = R_N - A_N^T X_N \quad \textbf{(4)}$$

where the subscripts ***N*** and ***O*** refer to *new* and *old* respectively--or the current step and the previous one. Equations 3 and 4 can be verified by substitution into Equation 1 with the *new* and *old* subscripts added. If ***A*** and ***B*** are initialized to zero and ***N+1*** unique starting values of ***X*** are selected, then after ***N+1*** function evaluations and updates, matrices ***A*** and ***B*** will be uniquely defined and the search for a solution could proceed.

The interesting property of Equation 3 which led Broyden to this selection is that the change in ***A*** is only in the direction of the last step in ***X***. That is, the only information about the behavior of the function which is added to ***A*** at each step is its variation along the current search direction. All of the information about the function in the ***N-1*** directions orthogonal to the current search direction remains intact; thus, it is a *rank-one* update method.

Broyden used this algorithm for obtaining and updating matrices *A* and **B**, along with Newton's method to search for the extrema. Thus in its original form, Broyden's is a quasi-Newton method (Morè and Sorensen (1984) discuss Newton and quasi-Newton methods in some detail.). The following calculus can be applied to the matrices in order to illustrate this procedure.

$$\frac{\partial R}{\partial X} = A^T \quad \textbf{(5)}$$

$$R_O = A_O^T (X_O - X_N) \quad \textbf{(6)}$$

Matrix ***A*** contains the partial derivatives of the residuals, ***R***, with respect to the unknowns, ***X***. Thus matrix ***A*** is the *Jacobian* of ***R*** with respect to ***X***.

As indicated by Equation 6, the gradient of the function lies along the direction ***AR***; therefore, most any gradient search method could be implemented and updated using Broyden's method for determining the Jacobian. Ortega and Rheinboldt (1970) discuss on an analytical level a number of methods which could be applied at this point. Actual selection of a practical method which will produce satisfactory stable results for a wide range of problems is quite another matter.

THE MODIFIED METHOD

Nonlinear simultaneous equations may have no solution, one solution, or many solutions. The most helpful physical analogy is that of a relief map of the Earth's surface where the unknowns are latitude and longitude and the function is the elevation with respect to mean sea level. No conceivable practical method could hope to locate Mt. Everest or the Marianas Trench regardless of the starting values. While it is reasonable to search for local extrema, it is fortuitous to locate the global extremum--assuming one does exist. Given this analogy it is understandable that no practical algorithm can be expected to locate even a local extremum in every case. Fletcher (1987) discusses these and other problems related to locating extrema in more detail.

Selection of Starting Values

This geographical analogy illustrates the necessity of limiting the region to be searched for extrema. In the present algorithm, a minimum and maximum value for each element in ***X*** must be supplied. This not only provides an extent to the range of ***X***, but it also serves as an indication of the scale. Any change in ***X*** which is on the order of the machine precision when compared to the range of ***X*** is considered negligible. One logical choice for the ***N+1*** starting values of ***X*** would be the center plus the ***N*** evenly distributed surrounding values inside the hypercube defined by the specified range of ***X***.

If the function at the central point is greater than at the surrounding points, then the first iteration would direct the search outside of the range of ***X***. If this occurs the range is bisected such that the new center point is mid way between the previous center and the surrounding point corresponding to the least value of the function. If this bisection is unsuccessful after sufficient attempts so as to diminish the subrange of any element of ***X*** to the previously determined negligible level, the search is abandoned.

Steplength Control

The unmodified method often results in unstable iterations. Not only is it necessary to confine ***X*** to the specified range, it is also necessary to damp the iteration or, as in this case, apply a steplength control algorithm. Ortega and Rheinboldt discuss several steplength algorithms. The parabola method defined by the current location, one *close* point, and the next step prescribed by the unmodified method has proven to be as successful as any tested. Using the unmodified point as an outer limit on the steplength arises from the observation that the unmodified method has a strong tendency to overshoot.

Hybrid Search Algorithm

Because Broyden's is basically a quasi-Newton method, the search proceeds in much the same direction as with *Newton-Raphson* (*NR*). In cases where the *NR* method would fail to locate an extremum, most likely Broyden would also. Broyden's method can also be viewed as a means by which to obtain the Jacobian (matrix A). If the Newton iteration is not successful, the Jacobian can be used to implement other methods. The method of *steepest descent* (*SD*) is more robust, but converges less rapidly than *NR*. When the *NR* iteration fails to result in a reduction of the residual, the direction of steepest descent is searched.

In the present algorithm, the *Conjugate-Gradient* (*CG*) method with the restart procedure recommended by Powell (1977) is also used to supplement the *NR/SD* iteration. The only information added to the Jacobian by Broyden's update is along the search direction. Information about the character of nonlinear functions in directions orthogonal to the search direction can be essential to locating extrema. The *CG* method provides a systematic procedure for searching other directions. In the present algorithm, the *NR*, *SD*, and *CG* methods are used alternately as each ceases to provide continual reduction of the residual.

Escape from Extraneous Entrapment

If $\boldsymbol{N}$ directions have been searched without further improvement, then either a local extremum has been found or extraneous entrapment has occurred. Whether the current location is a local extremum or a nuisance of finite-precision arithmetic can be partially determined by examining the history of matrix $\boldsymbol{A}$. For nonlinear problems the character of $\boldsymbol{A}$ can change substantially as the search proceeds.

The unmodified Broyden update to $\boldsymbol{A}$ replaces the information along the direction of the current step--thus discarding the previous information along this direction. If an *old* copy of $\boldsymbol{A}$ is retained along with the *new* copy and the search direction indicated by the old $\boldsymbol{A}$ is away from that indicated by the new $\boldsymbol{A}$ (viz. the dot product of the column matrices is less than or equal to zero), then the iteration may have skipped over an inflection point. In this case a search is conducted along the direction connecting these two provisional new values of $\boldsymbol{X}$.

Extraneous entrapment can sometimes be corrected by arbitrarily perturbing the solution away from the current location to see if it will return to the same point. After this perturbation has been attempted without success in $\boldsymbol{N}$ directions the procedure is abandoned.

Extension to Least-Squares

In the case where $\boldsymbol{M}>\boldsymbol{N}$, Equation 7 must be pre-multiplied by $\boldsymbol{A}$. The simultaneous nonlinear equations are then solved in the least-squares sense. For most problems the stability of the method also improves when this multiplication is performed even in the case of $\boldsymbol{M}=\boldsymbol{N}$. Therefore, in the present algorithm it is done regardless of the values of $\boldsymbol{N}$ and $\boldsymbol{M}$.

COMPARISON TO OTHER METHODS

The present derivative-free *enhanced* Broyden (*EB*) method was compared to the Newton-Raphson (*NR*) and Conjugate-Gradient (*CG*) methods. The results are listed in Table 1. All three methods have step-size control and for the test cases were required to obtain essentially the same solution. All three methods were given the same starting values (initial guess) so that there was no advantage of one over the other in these respects.

Table 1 lists the number of variables (independent unknowns and dependent residuals), the number of function evaluations, and relative performance. The relative performance is the number of CPU-seconds required for the *NR* divided by the number required for the particular method (thus, *NR* will always have a relative performance of 1.0).

Test Case 1

The first test case is a nonlinear constrained curve fitting problem. The best fitting single branch of a hyperbola was sought which would not only agree with the data (in this case experimental film boiling droplet area as a function of time), but would also have asymptotic characteristics conforming to the observed phenomena. The resulting curve fit must have one and only one root. The root must lie outside the range of the data and the derivative must be infinite at that point. The problem is nonlinear because of the constraints and the form (a rational polynomial). The partial derivatives of the residual cannot be determined analytically as these result in yet another set of simultaneous nonlinear equations. This test case was selected as being typical from among a set of 125.

Test Case 2

The second test case is similar to a nonlinear unconstrained curve fitting problem. The values of hydraulic conductivity and storativity (groundwater analogs of electrical conductance and capacitance) were sought which would best characterize a measured field response. A field test was conducted by pumping water from a well and measuring the change in the water table in a nearby well. An analytical expression for the ideal response of an aquifer contains these two unknown parameters which must be selected so as to best agree with the measured response. This problem is nonlinear, however the partial derivatives of the residual can be computed analytically (see note * in the table). This test case was selected as being typical from among a set of 33.

Test Case 3

The third test case is the determination of four *calibration factors* (mass transfer and pressure drop coefficients characterizing a particular type of plastic media) which are needed to run a large finite-integral code (numerical model of a cooling tower). Forty-nine sets of field data were collected for this plastic media. What was sought are the calibration factors which when input to the model will best reproduce the measured results. The finite-integral code itself was repeatedly run to provide the residuals. Needless to say, this was a very computationally intensive process--one in which minimizing the number of function evaluations was crucial. This test case, which is actually a type of inverse mass transfer problem, was selected as being typical from among a set of 6.

Test Case 4

The fourth test case is the determination of 4 phase lags and 4 corresponding weights which would best characterize the transient response of a dammed reservoir. A linear model was sought for the cross-sectionally averaged transient flow at a specific location (adjacent to a large power plant) within a reservoir bounded by two dams which are used for peaking (i.e., they discharge water only during times of peak electrical demand). This linear model was to become part of a larger linear systems optimization code used for long-range planning and resource management. An existing dynamic fluid flow model was used along with historical dam operations to produce a target data set. This test case was selected as being typical from among a set of 4.

DISCUSSION

The focus of these four test cases is not on many variables, but on non-analytically differentiable residuals and lengthy function evaluations. In each case there is some physical phenomenon which provides the basis for the residuals. Because these test cases are based on physical phenomena, the approximate bounds on the solution are also known. In each case lengthy graphical or cumbersome numerical techniques exist for finding extrema. The advantages to using the present algorithm in these cases are convenience and speed.

In the first test case (fitting a hyperbola with constraints and later taking its derivative) has been done for years using hand-drawn curves and a drafting protractor. The second test case (determining hydraulic conductance and capacitance) has also been done for years by graphical means and more recently by asymptotic extension to separate the coupled influence of the unknowns. The third test case (determining calibration factors for mass transfer and pressure drop) has typically been done by assuming half of the unknowns to be the same as a similar media and computing the others by *trial-and-error*.

For these test cases the average performance of the *EB* method is about 4 times the *NR* and *CG* methods. As mentioned previously, these are not isolated examples, but *working problems* from a variety of fields which were the impetus for developing the method. The *EB* method utilizes the best features of the *NR*, *CG*, and *SD* methods along with avoiding direct calculation of the Jacobian. The relative advantage of the *EB* method was most dramatic for Test Case 3 where the difference in runtime was a matter of days (on a 33MHz-80386/7 machine).

A two-variable problem is best suited to illustrate the searching procedure graphically. Figures 1 through 3 show the contours of the function in Test Case 2 and the first few steps in the search path for the *NR*, *CG*, and *EB* methods respectively. The Z-axis or the contours is percent total residual (in 20% intervals). The dark (dense dot) region is close to the extremum and the light (sparse dot) region is far from the extremum. This graphical format was selected in order to give a *bulls-eye* appearance.

In this case the *CG* essentially follows the gradient inward to the center of the *bulls-eye* (see the dark path line in Figure 2). The *CG* path is almost perpendicular to the contours as it crosses each one. The *SD* path if it were shown would differ little from the *CG*. The *NR* and *EB* paths differ markedly from the *CG* (compare the dark path lines in Figures 1 and 3 to Figure 2). The *NR* and *EB* methods reach the vicinity of the extremum (i.e., penetrate the darkest inner contour) in significantly fewer steps than does the *CG* method.

The hybrid implementation of the present method can be seen by comparing the second step in the *NR* and *EB* paths. The line connecting the second and third points on the *NR* line (Figure 1) is almost parallel to the contour next to it (i.e., this step is almost perpendicular to the gradient). The line connecting the second and third points on the *EB* path (Figure 3) is almost perpendicular to the contour (i.e., almost in line with the gradient at the point where it crosses the inner contour). This illustrates how the *EB* method checks the search direction corresponding to all three methods (*NR*, *CG*, and *SD*) to see which is more advantageous at a particular location.

Table 1. Comparison of Methods for Locating Extrema

test	variables		function evaluations			performance index		
case	N	M	N-R	C-G	E-B	N-R	C-G	E-B
1	4	50	417	462	76	1.0	0.8	1.9
2	2	47	35†	51†	61	1.0	0.8	3.1
3	4	98	137	298	14	1.0	0.5	10.4
4	8	8760	103	344	18	1.0	0.4	2.4

†Note: function calls reporting analytical partial derivatives are more time consuming than those which do not.

CONCLUSIONS

Broyden's derivative-free method for solving nonlinear simultaneous equations has been presented along with four enhancements. These enhancements include: a method for selecting starting values, step length control, hybrid search algorithm, and a method for escaping from extraneous entrapment. A significant performance improvement over the Newton-Raphson and Conjugate-Gradient methods is shown for four test cases taken from varied fields. Part of this performance improvement is a consequence of the derivative-free method. The hybrid search algorithm used in this enhanced Broyden method further improves the performance by utilizing the strengths of three other methods (the Newton-Raphson, Conjugate-Gradient, and Steepest-Descent).

REFERENCES

Broyden, C., 1969, "A New Method of Solving Nonlinear Simultaneous Equations," *Computational Journal*, Vol. 12, pp. 94-99.

Fletcher, R., 1987, *Practical Methods of Optimization*, John Wiley and Sons, New York, NY.

Morè, J. J. and D. C. Sorensen, 1984, "Newton's Method," *Studies in Numerical Analysis*, G. H. Golub, ed., The Mathematical Association of America, pp. 29-82.

Ortega, J. M. and W. C. Rheinboldt, 1970, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

Powell, M. J. D., 1977, "Restart Procedures for the Conjugate Gradient Method," *Mathematical Programming*, Vol. 12, pp. 241-254.