ANALYSIS OF SOURCE LOCATION ALGORITHMS
Part II: Iterative methods

MAOCHEN GE
Pennsylvania State University, University Park PA 16802

Abstract

Iterative algorithms are of particular importance in source location as they provide a much more flexible means to solve nonlinear equations, which is essential in order to deal with a wide range of practical problems. The most important iterative algorithms are Geiger's method and the Simplex method. This article provides an overview of iterative algorithms as well as an in-depth analysis of several major methods.

1. Introduction

In Part I, we discussed the non-iterative location methods. A restriction that severely limits the application of these methods is the assumption of a single velocity for all channels. This assumption is not arbitrary or just for convenience; it is necessary in order to keep the source location equations in the simplest form so that they can be solved non-iteratively. In other words, the assumption is a reflection of an inherent difficulty associated with non-iterative location methods. If a single velocity model is not suitable, as in most cases, we have to turn to an iterative method.

While the associated searching strategies may vary significantly, iterative methods in general refer to those algorithms that start from an initial solution, commonly called a guess or trial solution. This trial solution is tested by the given conditions, and then updated by the predefined schemes, which subsequently forms a new trial solution. Therefore, iterative methods, in essence, are a process of testing and updating trial solutions.

An iterative method is distinguished by its updating scheme, which determines its efficiency as well as its other major characters. While there are many different schemes, the ones that are truly significant both theoretically and practically are few for the purpose of source location. The object of this article is to provide an in-depth analysis of these methods. But first we will give a brief review on the basic searching approaches.

2. An Overview on Basic Searching Approaches

The iterative methods used for source location fall into several basic categories, which are derivative, sequential, genetic and Simplex. A brief discussion of these approaches is given as follow.

Derivative approach

Derivative approach refers to those methods that use the derivative information to update trial solutions. Derivative approach is a classical means in mathematics for solving nonlinear problems and is probably one of the most widely used numerical approaches for this purpose. "*Root finding*" by derivatives, an elementary topic in calculus, is such an example.

Derivative methods update their trial solutions based on the nonlinear behavior information at the trial solution as given by derivatives. This searching mechanism makes derivative methods far more efficient than any other iterative methods.

The derivative algorithms used for source location include Geiger's method (Geiger, 1910, 1912) and Thurber's method (Thurber, 1985). The difference between these two methods is that the Geiger's method uses the first order derivatives while the Thurber's method uses both the first and second order derivatives. Mathematically, Geiger's method is an application of Gauss-Newton's method (Lee and Stewart, 1981) and Thurber's method an application of Newton's method (Thurber, 1985).

Geiger's method is probably the most important source location method, and has been used almost exclusively for local earthquake locations. Understanding this method is important for both theoretical and practical reasons and we will give a detailed discussion on this method. We will also discuss Thurber's method. In addition to the fact that Thurber's method constitutes a significant derivative approach, the discussion of this method will help us understand the derivative approach as a whole.

Sequential searching algorithms

Sequential searching algorithms here refer to those methods that partition the monitoring volume into smaller blocks and study these blocks sequentially. With these methods, each block, which is represented by either the center of the block or other feature locations, is considered as a potential AE source. The block may be further refined in the searching process. While the approach is extremely simple, easy to use, and easy to modify, the main problem is inefficiency, which essentially blocks its application potential for problems requiring the good location accuracy. For instance, if we would like to achieve 1 mm location accuracy on a 1 cubic meter block, the points that to be searched are in the order of one billion. In contrast, it may take only a few iterations for Geiger's method to achieve this accuracy.

Simplex algorithm

A notable problem with high efficiency derivative algorithms is divergence, which could become very severe if the associated system is not stable. Sequential searching algorithms, on the other hand, exhibit very stable solution process, although they are just too slow. A method that comes in between is the Simplex algorithm, which is quite efficient while showing very stable characteristics.

The Simplex algorithm is a robust curve fitting technique developed by Nelder and Mead (1965). It was introduced for the source location purpose in late 1980s by Prugger and Gendzwill (Prugger and Gendzwill, 1988; Gendzwill and Prugger, 1989). The mathematical procedures and related concepts in error estimation for this method were further discussed by Ge (1995). Because of the rare combination of efficiency and stability, the Simplex algorithm is suited for a wide range of problems and has rapidly become a primary source location method. We will give a detailed discussion on this method.

Genetic Algorithm

The genetic algorithm was developed by Holland (1975). It is an optimization technique that simulates natural selection in that only the "fittest" solutions survive so that they can create even better answers in the process of reproduction. The algorithm was applied by a number of researchers for earthquake locations (Kennett and Sambridge, 1992; Sambridge and Gallagher,

1993; Billings et al., 1994, Xie et al., 1996). While the algorithm seems very flexible for incorporating various source location conditions, it is less conclusive on its efficiency and accuracy. The viability of the algorithm for source location will largely depend on how these questions are answered.

3 Geiger's Method

Geiger's method, developed at the beginning of the last century (Geiger, 1910, 1912), is the classical source location method by all accounts. In addition to its long history, the method is the best known and most widely used source location method. In seismology, it is used almost universally for local earthquake locations.

3.1 Algorithm

Geiger's method (Geiger, 1910, 1912) is an example of the Gauss-Newton's method (Lee and Stewart, 1981), a classical algorithm for solving nonlinear problems. The method is discussed here in terms of the first-degree Taylor polynomials and the least-squares solution to an inconsistent linear system.

Let $f_i(\boldsymbol{x})$ represent the arrival time function associated with the ith sensor, where $\boldsymbol{x}$ denotes the hypocenter parameters:

$$\boldsymbol{x} = (x, y, z, t)^T . \tag{1}$$

The unknowns, x, y and z, are the coordinates of an event and t is the origin time of this event.

Expand $f_i(\boldsymbol{x})$ at a nearby location, $\boldsymbol{x_0}$, and express the expansion by the first-degree Taylor polynomial:

$$f_i(\boldsymbol{x}) = f_i(\boldsymbol{x_0} + \delta\boldsymbol{x}) = f_i(\boldsymbol{x_0}) + \frac{\partial f_i}{\partial x}\delta x + \frac{\partial f_i}{\partial y}\delta y + \frac{\partial f_i}{\partial z}\delta z + \frac{\partial f_i}{\partial t}\delta t \tag{2}$$

where

$$\boldsymbol{x} = \boldsymbol{x_0} + \boldsymbol{\delta x},$$

$$\boldsymbol{x_0} = (x_0, y_0, z_0, t_0)^T ,$$

and

$$\delta\boldsymbol{x} = (\delta x, \delta y, \delta z, \delta t)^T . \tag{3}$$

Eq. (2) may also be expressed in vector notation:

$$f_i(\boldsymbol{x}) = f_i(\boldsymbol{x_0} + \delta\boldsymbol{x}) = f_i(\boldsymbol{x_0}) + g_i^T \delta\,\boldsymbol{x} \tag{4}$$

where g_i^T is the transpose of the gradient vector g_i and is defined by

$$g_i^T = \nabla f_i(\boldsymbol{x}) = \left(\frac{\partial f_i}{\partial x}, \frac{\partial f_i}{\partial y}, \frac{\partial f_i}{\partial z}, \frac{\partial f_i}{\partial t}\right) \tag{5}$$

In source location, the nearby location, $\boldsymbol{x_0}$, is conventionally called *guess* or *trial* solution. Since the trial solution is either assigned by users or generated from the previous iteration, it is always known at the beginning of each iteration. As such, $f_i(\boldsymbol{x_0})$ is also a known quantity and is called *calculated arrival time*. The term *calculated arrival time* reflects the fact that this quantity is obtained by calculation, assuming the trial solution, $\boldsymbol{x_0}$, as the hypocenter.

The term on the left hand side of Eq. (2), $f_i(\boldsymbol{x_0} + \delta\boldsymbol{x})$, represents the arrival time recorded at the *i*th sensor, which is conventionally termed *observed arrival time*. As such, the physical meaning of Eq. (2) is that an observed arrival time is expressed by the arrival time calculated from a nearby location, and by

$$\frac{\partial f_i}{\partial x}\delta x + \frac{\partial f_i}{\partial y}\delta y + \frac{\partial f_i}{\partial z}\delta z + \frac{\partial f_i}{\partial t}\delta t ,$$

a correction factor, which is a function of the partial derivatives of the hypocenter parameters. All the partial derivatives of the arrival time function are known quantities here as they can be numerically evaluated based on the trial solution.

In solving a system of the equations defined by Eq. (2), our goal is to find an $\boldsymbol{x_0}$, such that the calculated arrival times will best match the observed arrival times so that $\boldsymbol{x_0}$ can be considered as the hypocenter of the event. This is done in a self correction process: the trial solution is updated at the beginning of each iteration by adding $\delta\boldsymbol{x}$, known as the correction vector, obtained from the previous iteration.

For the convenience of the solution for $\delta\boldsymbol{x}$, we rearrange Eq. (2) into the form:

$$\frac{\partial f_i}{\partial x}\delta x + \frac{\partial f_i}{\partial y}\delta y + \frac{\partial f_i}{\partial z}\delta z + \frac{\partial f_i}{\partial t}\delta t = \gamma_i \tag{6}$$

where

$$\gamma_i = t_{oi} - t_{ci} , \tag{7}$$

$$t_{oi} = f_i(\boldsymbol{x}) ,$$

and $t_{ci} = f_i(\boldsymbol{x_0})$.

Here, γ_i is known as *channel residual*.

In matrix notation, a system defined by Eq. (6) can be written:

$$\boldsymbol{A}\delta\boldsymbol{x} = \boldsymbol{\gamma} \tag{8}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial t} \\ \vdots & & & \vdots \\ \frac{\partial f_m}{\partial x} & \frac{\partial f_m}{\partial y} & \frac{\partial f_m}{\partial z} & \frac{\partial f_m}{\partial t} \end{bmatrix} \qquad \delta\boldsymbol{x} = \begin{bmatrix} \delta x \\ \delta y \\ \delta z \\ \delta t \end{bmatrix} \qquad \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{bmatrix}$$

The least squares solution to the system defined by Eq. (8) satisfies (Strang, 1980)

$$\boldsymbol{A^T A}\delta\boldsymbol{x} = \boldsymbol{A^T\gamma} \tag{9}$$

or

$$\delta\boldsymbol{x} = (\boldsymbol{A^T A})^{-1}\boldsymbol{A^T\gamma}$$

The total effect of the mismatch between the observed and calculated arrival times is called the event residual or simply residual. The event residual that is defined by the least-squares solution is (Ge, 1985):

$$Res = \sqrt{\frac{\boldsymbol{\gamma}^T \boldsymbol{\gamma}}{m-q}} \tag{10}$$

where m is the number of equations and q is the degree of freedom. For the number of hypocenter parameters defined by Eq. (1), the degree of freedom is 4.

Now the correction vector, $\boldsymbol{\delta x}$, has been found and it can be added to the previous trial solution to form a new trial solution. This process is repeated until the given error criterion is fulfilled, and the final trial solution is then regarded as the true source.

3.2 Implementation

The algorithm of Geiger's method discussed in the previous section was developed for general arrival time functions, that is, we can use this algorithm for any arrival time functions as far as the functions and their partial derivatives can be evaluated. To further enhance our understanding of this algorithm, we now discuss the implementation of Geiger's method through examples.

The implementation of Geiger's method is a three-step process:

- establishing arrival time functions,
- data preparation, and
- solving a system of simultaneous equations.

Establishing arrival time functions

The first task in implementing the Geiger's method is establishing arrival time functions. Arrival times are affected by many factors. Categorically, there are three major ones: structure and composition of media where stress waves propagate, source mechanism and relative orientation of the source and sensors, and the shape and geometry of the structure under study. While real travel time models are complicated in nature, the arrival time functions that are used to describe a model have to be simplified for either theoretical and/or practical reasons. As an example, the following is an arrival time function for a homogeneous velocity model:

$$f_i(\boldsymbol{x}) = f_i(x, y, z, t) = t + \frac{1}{v_i}\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \tag{11}$$

where the unknowns, x, y and z, are the coordinates of an AE event; t is the origin time of this event; x_i, y_i and z_i, are the coordinates of the ith sensor, and v_i is the velocity of the stress wave.

We note the difference of the velocity model used here and the velocity model assumed for those non-iterative methods discussed in the preceding paper. For those non-iterative methods, we have to assume a single velocity for all arrival time data. With the homogeneous velocity model used in this example, each equation can have its own velocity. This allows us to assign the velocity based on the arrival type, which is critical for accurate source location.

The arrival time functions used for Geiger's method can be much more complicated than the one used in this example. In fact, Geiger's method posts no restrictions on the velocity model to be utilized as far as arrival time functions can be established and their first-order derivatives can be evaluated.

Data preparation

Once arrival time functions are established, the next step is preparing data. It is know from Eq. (2) that there are four types of data that we have to prepare, which are: trial solution, observed arrival time, calculated arrival time, and partial derivatives.

Trial solution At the beginning of the iteration process, a trial solution has to be assigned manually by users or generated automatically by the location code. After this, it is updated automatically by adding the new correction vector.

A question that is frequently asked is: is it necessary for the trial solution to be very close to the true event location? While it would never hurt to have a close guess, it may not be achievable in many cases, especially in the situation where source location is carried out automatically. Fortunately, the answer to this question is "no". In general, it will be good enough if a trial solution is within the monitoring area. A practice that is frequently adopted by the author is to use the location of the first triggered senor as the trial solution if we do not have any prior knowledge on event locations.

There is a perception, however, that the choice of the trial solution is important. While it is possible that one has to "play" the initial trial solutions in order to get the right answer, it usually indicates that the associated system is unstable, a far more serious problem than the choice of the trial solution. When this is the case, the confidence that one can put on the final solution is significantly diminished.

Observed arrival time The observed arrival times are the data provided externally. Since the physics of source location is to find a location that its associated arrival times best match the observed arrival times, the accuracy of the observed arrival times has to be compatible with the required accuracy. For instance, if the required location accuracy is 1 mm and the travel velocity of the stress wave under study is 1 km/sec, then the allowed timing error is 1/1000000 = 1 μs.

Calculated arrival time The calculated arrival times are obtained by substituting the trial solution into the arrival time functions, such as Eq. (11).

Partial derivatives The partial derivatives defined by $\boldsymbol{A}$ in Eq. (8) have to be fulfilled. This is a two-step procedure: deriving the general expressions of the partial derivatives and numerical evaluation of these partial derivatives in terms of the trial solution. As a demonstration, the following are the general expressions of the partial derivatives of the arrival time function given by Eq. (11):

$$\frac{\partial f_i}{\partial x} = -\frac{x_i - x_0}{v_i R}$$

$$\frac{\partial f_i}{\partial y} = -\frac{y_i - y_0}{v_i R}$$

$$\frac{\partial f_i}{\partial z} = -\frac{z_i - z_0}{v_i R}$$

$$\frac{\partial f_i}{\partial t} = 1$$

$$R = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}$$

Solving a system of simultaneous equations

The least squares solution to an inconsistent system is given by Eq. (9). Usually, the size of correction vector, $\boldsymbol{\delta x}$, will decrease rapidly and reach a prescribed accuracy within a few iterations. However, it is possible that $\boldsymbol{\delta x}$ will not converge: it may oscillate or even increase beyond control. The problem of the divergence is a sign of the instability of the associated mathematical system, which is usually the result of poor array geometry.

3.3 Mechanics of iteration by first order derivatives

Although Geiger's method is relatively straight forward from a computational point of view, conceptually, the method is still quite confusing despite its enormous popularity and long history. For instance, it is a generally accepted perception in seismology that Geiger's method is a linear approximation of nonlinear source location problems (Thurber, 1985). The implication of this perception is that the method is unable to take into account of the nonlinear behavior of arrival time functions. This is a serous mistake. It affects not only our theoretical understanding of the method, but also its applications.

While the confusions that surround Geiger's method may be attributed to many causes, fundamentally, it is the lack of the correct understanding of the mechanics of derivative methods. In order to solve this problem, there are two issues we have to discuss further: Taylor's theorem and the function of first-order derivatives

Taylor's theorem and formulation principle

The key element in developing Geiger's method is the expansion of arrival time functions into the first-degree Taylor polynomials, and we begin our discussion with Taylor's theorem. The focus of this discussion is whether the expansion of a function by the Taylor polynomials is an approximation of that function.

The Taylor's theorem states that a function at a point may be evaluated by the Taylor polynomial of the function at its neighboring point and the error for this approximation can be evaluated by the associated error function. The key here is that, when the Taylor polynomials are used for the function evaluation at the location by its neighbors, the accuracy of this approximation is the function of the size of this neighborhood and, unless demonstrated otherwise, it has to be assumed very small. Therefore, the Taylor polynomial in general is a highly localized function in that it changes with each neighboring point that is being chosen and there is no unique Taylor polynomial that can represent a function for its entire domain.

Furthermore, the Taylor polynomials used for the purpose of numerical computations are mostly associated with a very low degree, typically first or second. Under this condition, it is virtually impossible to approximate a function by polynomials. If we consider the fact that Geiger's method only uses the first order derivatives, it is impossible to represent arrival time functions by these polynomials.

When an arrival time function is expanded in the form of the Taylor polynomial, the resulting equation, such as Eq. (2), no longer represents the original arrival time function. The original arrival time function is the function of hypocenter parameters and the new equation is the function of a correction vector. With this new equation, the observed arrival time is represented by the arrival time calculated for a nearby point and a correction factor. As it has been discussed earlier, the calculated arrival time is the result of the evaluation of the original arrival time function in terms of the trial solution, which eventually represents the hypocenter. The correction factor

determines how the trial solution is to be changed in the next iteration. As such, none of these terms can be regarded as the linear portion of the original arrival time function.

It is understood from the above analysis that the first-degree Taylor polynomials used in Geiger's method are not the approximation of original arrival time functions and, therefore, this expansion process cannot be characterized by linearization. The analysis of the components of the expanded function also shows that there is no physical evidence to characterize this process by linearization.

So far we have demonstrated that the Taylor polynomials used in Geiger's method are not the approximations of arrival time functions, and, therefore, linearization is not an appropriate term to characterize the formulation of Geiger's method. We now discuss the question whether the searching process can be termed as linear because Geiger's method uses only first order derivatives.

The answer to this question is actually quite simple: any derivative method is a nonlinear searching method. This is because derivatives, regardless of their orders, are used to catch up with the nonlinear behavior of functions at trial locations and correction vectors are determined by this information. Therefore, the question with a derivative method is not whether it is a nonlinear searching method; the question is the type of the nonlinear behavior that is utilized. We now demonstrate the geometric meaning of the searching process associated with Geiger's method.

Consider the general form of a nonlinear system:

$$\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0},$$

where

$$\boldsymbol{F}(\boldsymbol{x}) = (\, f_1(\boldsymbol{x}),\, f_2(\boldsymbol{x}), \ldots,\, f_m(\boldsymbol{x})),$$

$$f_1(\boldsymbol{x}) = 0\,,\; f_2(\boldsymbol{x}) = 0\,,\; \ldots\ldots,\; f_m(\boldsymbol{x}) = 0\,,$$

and $\quad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T\,.$

The solution of this system is to find the common intersection of the functions of $\boldsymbol{F}(\mathbf{x})$ at $\boldsymbol{F}(\mathbf{x}) = 0$. Because of the inherent difficulty to solve a nonlinear system analytically, it is usually done numerically, and the corresponding process is commonly known as root finding. One of the best known methods for this purpose is the Gauss-Newton's method, and Geiger's method is an application of this method (Lee and Stewart, 1981). The Gauss-Newton's method uses the first-order derivative information to determine the correction vector.

Gradient vectors, such as the one given by Eq. (5), represent the directions of the steepest slopes. The key to understand the Gauss-Newton's method, and therefore Geiger's method, is how these first order derivatives are used to determine correction vectors. Although it is impossible to demonstrate the geometrical meaning of gradient vectors for problems with more than two independent variables, it is fortunate that the mechanics remains the same for all dimensions. As such, we will use the Newton-Raphson method, the Gauss-Newton's method in one variable, to illustrate how first order derivatives are used to determine correction vectors.

Newton-Raphson method

The Newton-Raphson method is one of the most powerful numerical methods for finding a root of $f(x) = 0$, and, yet, both the concept and the procedure are extremely simple. The method begins with the first-degree Taylor polynomial for $f(x)$, expanded about x_0,

$$f(x) = f(x_0) + (x - x_0) f'(x_0)$$

Since we are finding the root of f, so that $f(x) = 0$ and the above equation becomes:

$$0 = f(x_0) + (x - x_0) f'(x_0)$$

Solving for x in this equation gives:

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

where x should be a better approximation of the root. This sets the stage for the Newton-Raphson method, which involves generating the sequence $\{x_n\}$ in an iteration process:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

The geometry of the Newton-Raphson method is shown in Fig. 1. At each iteration stage, we determine the first-order derivative of the function at the trial solution. Geometrically, it represents the tangent line of the function at this location. The intersection of this line with the x-axis defines the new trial solution, x_{n+1}. The corresponding correction vector is $\delta x = x_n - x_{n+1}$. From the figure, it is easy to verify the relation:

$$f'(x_n) = \frac{f(x_n)}{x_n - x_{n+1}}.$$

The above equation is another form of the preceding equation, but with much clearer geometrical meaning: the slope of the tangent line represented by $f'(x_n)$ on the left-hand side of the equation is the ratio of the function value at the trial solution and the correction vector.

In summary, the Newton-Raphson method can be viewed as a procedure that we use the intersection of the tangent line at the x-axis to approximate the root of the function and the direction of the tangent line is defined by the first order derivative of the function at the trial solution.

3.4 Stopping criteria

When an iterative method is used, the calculation has to be stopped at a certain point. There are three commonly used stopping criteria for derivative methods, and their applicability for the problem of source location is discussed next.

Residual as stopping criterion

When residual is utilized as the stopping criterion, we stop the calculation soon after the residual is smaller than a pre-defined tolerance, such that:

$$\text{Res} < \varepsilon, \qquad \text{for } \varepsilon > 0,$$

where *Res* is the event residual and ε is the tolerance. The event residual is a measurement of the location error in terms of the total effect of the mismatch between the observed and calculated arrival times. Mathematically, it is defined by the regression method being used. For the least-squares method, the event residual is (Ge, 1996) given by Eq. (10):

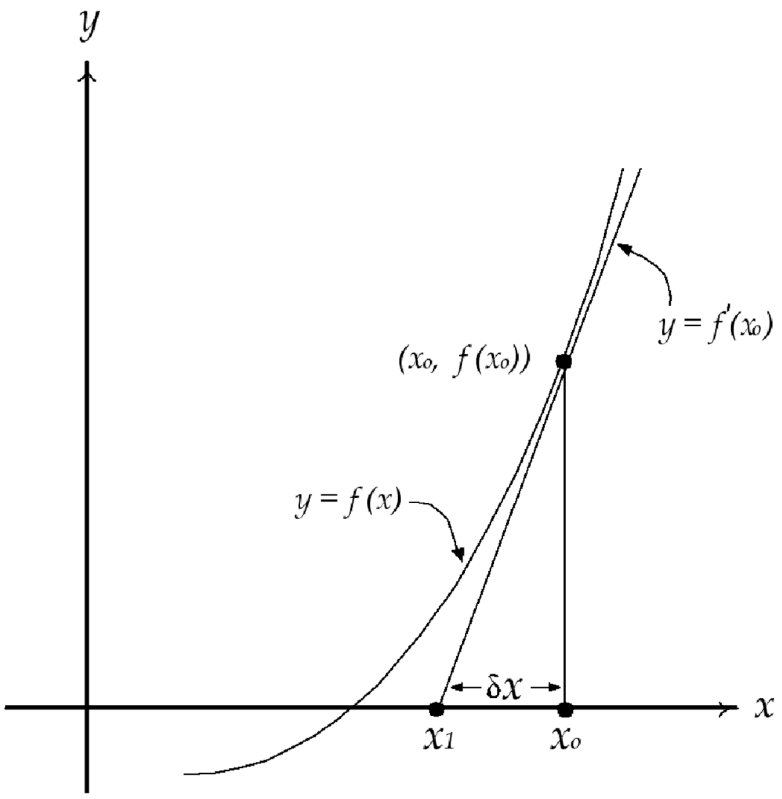

Fig. 1 Geometry of Newton-Raphson method

$$Res = \sqrt{\frac{\boldsymbol{\gamma}^T \boldsymbol{\gamma}}{m-q}} \tag{10}$$

as defined previously.

Although the approach appears quite natural for many applications as the residual is a mathematical measurement of the computational error, the problem in the case of source location is that it may vary in a very wide range for events under similar conditions, say, covered by the same array. If the tolerance is set too low, many solutions may never be able to reach that level. If it is set too high, we lose the accuracy. Therefore, it is difficult to implement this approach for source location problems.

Size of correction vector as stopping criterion

The second criterion is the size of correction vectors, $\boldsymbol{\delta x}$. In the case of source location, correction vectors include both spatial and time components. This creates a problem for calculating the size of $\boldsymbol{\delta x}$ because of the difference in units. A solution to this problem is to represent the size of correction vectors by $\boldsymbol{\delta d}$, the size of the spatial components only, such as

$$\delta \boldsymbol{d} = \sqrt{\delta x^2 + \delta y^2 + \delta z^2}$$

and the corresponding stopping criterion is:

$$\boldsymbol{\delta d} < \varepsilon$$

When the size of correction vectors is used as a stopping criterion, the underlying assumption is that it is a sign of the convergency when $\boldsymbol{\delta d} \rightarrow 0$ as the number of iterations increases to infinity.

Relative size of correction vector as stopping criterion

The third criterion is the relative size of correction vectors. Again, if we only consider the spatial components, this third criterion can be expressed:

$$\frac{\delta \boldsymbol{d}}{\sqrt{x^2 + y^2 + z^2}} < \varepsilon$$

where x, y and z are the coordinates of the event location determined at the current iteration. The relative error is a better measurement for many applications, although it has two serious drawbacks in the case of source location. First, the relative error would be the function of the coordinate system. For instance, the relative error for an array described by three digits would be drastically larger than the same array that is described by five digits. Secondly, the approach would make the error as the function of event locations. If we consider the fact that the size of an array is normally much larger than the location error, typically of the order of 2 or higher, the event location will significantly affect the outcome of this approach. For these reasons, relative size is not a suitable stopping criterion.

In comparison of these three criteria, the size of correction vectors appears most suitable for the purpose of source location. In any case another stopping criterion is always necessary: the number of iterations. This is especially true for derivative methods, with which divergence may occur.

3.5 Problem of divergence

The major problem associated with the Geiger's method is divergence. Divergence is a common problem associated with many iterative algorithms. The cause of divergence is complex and a detailed discussion of the problem is beyond the scope of this article. However, there are two points we wish to make. First, although divergence can be caused by many different technical reasons, it is fundamentally governed by the stability of the associated mathematical system. In the case of source location, this stability is determined by the sensor array geometry relative to the actual event location. For instance, because of the inherent difficulty to spread the seismographs in the vertical direction, the location accuracy of local earthquakes in the depth direction is in general poorer than in the other directions. Therefore, the best means to alleviate the divergence problem is to optimize the sensor array geometry, if it is possible. The second point is that divergence is a phenomenon closely related to the convergence rate. For those methods with a very high convergence rate, such as the Geiger's method, the chances to develop the divergence problem is also considerably higher.

Many research studies were carried out in seismology on the improvement of the convergence character of the source location algorithm, and some remedial measures were proposed (Smith, 1976; Buland, 1976; Lee and Stewart, 1981; Anderson, 1982; Lienert and Frazer, 1983; Thurber, 1985). In general, the efficiency of these remedial measures is very limited as none of them could fundamentally address the two inherent difficulties discussed earlier.

4. Thurber's Method

Geiger's method uses the first order derivatives to catch up the nonlinear behavior of arrival time functions at the trial solutions. We now discuss another significant derivative approach, Thurber's method (Thurber, 1985), which uses the information of both the first and second derivative to determine correction vectors.

4.1 Algorithm

Similar to our discussion on Geiger's method, let $f_i(\boldsymbol{x})$ be the general form of the arrival time function associated with the ith sensor, where $\boldsymbol{x}$ denotes the hypocenter parameters defined by Eq. (1). Expand $f_i(\boldsymbol{x})$ at a nearby location, $\boldsymbol{x_0}$, and express the expansion by the second-degree Taylor polynomial:

$$f_i(\boldsymbol{x}) = f_i(\boldsymbol{x_0} + \delta\boldsymbol{x}) = f_i(\boldsymbol{x_0}) + g_i^T \delta\boldsymbol{x} + \frac{1}{2}\delta\boldsymbol{x}^T H_i \delta\boldsymbol{x} \tag{12}$$

where g_i^T is the transpose of the gradient vector g_i defined by Eq. (5) and H_i is the Hessian matrix:

$$H_i = \begin{bmatrix} \frac{\partial^2 f_i}{\partial x^2} & \frac{\partial^2 f_i}{\partial x \partial y} & \frac{\partial^2 f_i}{\partial x \partial z} & \frac{\partial^2 f_i}{\partial x \partial t} \\ \frac{\partial^2 f_i}{\partial y \partial x} & \frac{\partial^2 f_i}{\partial y^2} & \frac{\partial^2 f_i}{\partial y \partial z} & \frac{\partial^2 f_i}{\partial y \partial t} \\ \frac{\partial^2 f_1}{\partial z \partial x} & \frac{\partial^2 f_i}{\partial z \partial y} & \frac{\partial^2 f_i}{\partial z^2} & \frac{\partial^2 f_i}{\partial z \partial t} \\ \frac{\partial^2 f_i}{\partial t \partial x} & \frac{\partial^2 f_i}{\partial t \partial y} & \frac{\partial^2 f_i}{\partial t \partial z} & \frac{\partial^2 f_i}{\partial t^2} \end{bmatrix} \tag{13}$$

The physical meanings of $f_i(\boldsymbol{x_0} + \delta\boldsymbol{x})$ and $f_i(\boldsymbol{x_0})$ in Eq. (12) remains the same as in Eq. (2), which are the observed and calculated arrival times, respectively.

Eq. (12) is a quadratic function of the correction vector. With the partial differentiation of the equation and setting the resulting equation to zero, we have

$$g_i + H_i \delta\boldsymbol{x} = 0 \ . \tag{14}$$

For convenience, the hypocenter notation given by Eq. (1) may also be expressed as

$$\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T \ . \tag{15}$$

With the hypocenter notation given by Eq. (15), an alternative form of Eq. (14) is

$$\frac{\partial f_i(x)}{\partial x_k} + \sum_{j=1}^{4} \delta x_j \frac{\partial^2 f_i(x)}{\partial x_j \partial x_k} = 0 \,, \qquad k = 1, 2, 3, 4, \tag{16}$$

which gives us a detailed view of the content of Eq. (14).

The least squares solution for the system defined by Eq. (14) is (Thurber, 1985):

$$\Delta\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A} - (\nabla\boldsymbol{A}^T)\boldsymbol{r})^{-1}\boldsymbol{A}^T\boldsymbol{r} \tag{17}$$

where both $\boldsymbol{A}$ and $\boldsymbol{r}$ are defined in Eq. (8). If we use $\boldsymbol{N}$ to represent $(\nabla\boldsymbol{A}^T)\boldsymbol{r}$ in Eq. (17), such that

$$\boldsymbol{N} = (\nabla\boldsymbol{A}^T)\boldsymbol{r} \,, \tag{18}$$

and the entry $\boldsymbol{N}_{ij}$ of this 4 x 4 matrix is given by

$$N_{ij} = \sum_{k=1}^{m} \frac{\partial^2 f_k}{\partial x_i \partial x_j} \gamma_k \qquad i, j = 1, 2, 3 \tag{19}$$

$$N_{ij} = 0 \qquad i, j = 4$$

The step that holds the key to understand Newton's method, and therefore Thurber's method, is the transition from Eq. (12) to Eq. (14). Eq. (12) is a quadratic function. By the partial differentiation of this equation and setting the resulting equation to zero, Eq. (14) defines the extreme of this quadratic function. This extreme in Newton's method is regarded as the optimum correction vector for the trial solution, and therefore, the solution for Eq. (12).

4.2 Mechanics of iteration by first and second order derivatives

Following the approach that the Newton-Raphson's method was used to illustrate the geometric meaning of the first order derivatives, we now use Newton's method in one variable to demonstrate geometrically how the first and second order derivatives are used to determine the correction vector.

Finding the root with first and second derivatives

Consider the second-degree Taylor polynomial for $f(x)$, expanded about x_0,

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0) \tag{20}$$

Our goal is to determine x so that $f(x) = 0$. Note that $f(x)$ in this case is expressed by a quadratic function of x and the best approximation of $f(x) = 0$ is to find the minimum of the function and this can be done by taking the first derivative of the function with respect to x,

$$\frac{df(x)}{dx} = \frac{df(x_0)}{dx} + \frac{d((x - x_0) f'(x_0))}{dx} + \frac{1}{2}\frac{d((x - x_0)^2 f''(x_0))}{dx}$$

$$= f'(x_0) + (x - x_0) f''(x_0)$$

and let the resulting equation equal to zero,

$$f'(x_0) + (x - x_0) f''(x_0) = 0$$

Solving the equation for x,

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

the final solution can be approached iteratively by the sequence $\{x_i\}$ defined by

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)} \qquad i \geq 0 \tag{21}$$

The geometry of this solution process is illustrated in Fig. 2. The second-degree Taylor polynomial, denoted by $p(x_0)$ in the figure, is used to represent the function at the neighborhood of x_0. The extreme of this polynomial, x_1, is considered by Newton's method as the optimum solution which becomes the new trial solution. Eq. (21) is a mathematical definition of this extreme.

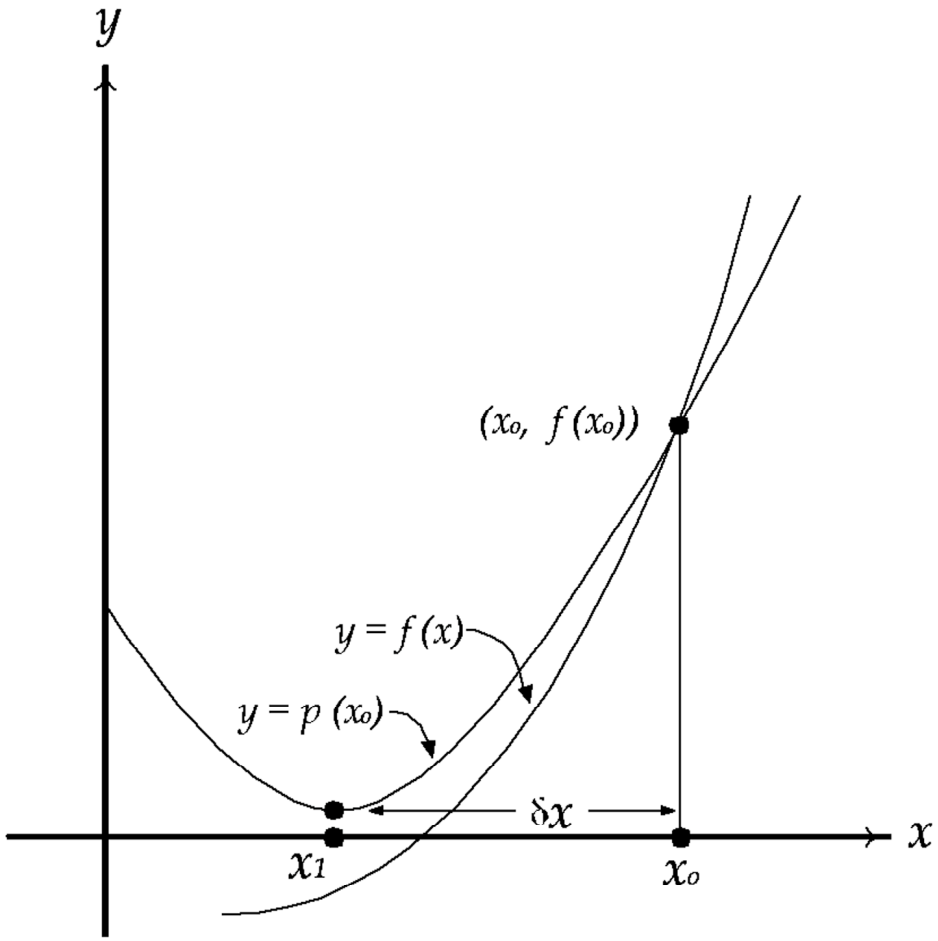

Fig. 2 Geometry of Newton's method in one variable.

A very interesting point shown by Fig. 2 is the difference between $f(x)$, the original function, and $p(x_0)$, the second-degree Taylor polynomial, which clearly demonstrates that the Taylor polynomial is not the approximation of that function.

4.3 Discussion

The essence of derivative methods is using nonlinear characters of functions at the trial solution, described by derivatives, to determine correction vectors. The difference among derivative methods is hence the type of nonlinear characters that are used. The simplest derivative method is Geiger's method. The nonlinear character used by this method is gradient, or the steepest slope at the trial solution, which are represented by first order derivatives. Thurber's method uses both the first and second order derivatives and its solution is geometrically represented by the extreme of a quadratic function.

Thurber's method is a more sophisticated derivative method than Geiger's method in that it utilizes the quadratic model for optimization. Although the method was demonstrated with the improved stability in several cases (Thurber, 1985), it is not clear whether it can be regarded as a general solution to the problem. A major uncertainty associated with the quadratic model based methods is whether the model is positive definite. While it is not a requirement, an underlying assumption for these methods is that the model is positive definite. If the assumption is not fulfilled, the performance of these methods is much more difficult to predict.

5. Simplex Method

The Simplex method is a relatively new method developed by Nelder and Mead (1965). It searches the minimums of mathematical functions through function comparison. The method was introduced for the source location purpose in late 1980s by Prugger and Gendzwill (Prugger and Gendzwill, 1988; Gendzwill and Prugger, 1989). The mathematical procedures and related concepts in error estimation for this method were further discussed by Ge (1995). Readers should be aware that the Simplex method discussed here is different from the one used in linear programming where it refers an algorithm for a special type of linear problems.

5.1 Method concept

The Simplex method starts from the idea that, for a given set of arrival times associated with a set of sensors, an error can be calculated for any point by comparing the observed and the calculated arrival times. An error space is thus the one in which every point is defined by an associated error, and the point with the minimal error is the event location.

The process of searching for the minimal point with the Simplex method is unique. The solution is said to be found when a Simplex figure falls into the depression of the error space. The Simplex is a geometric figure which contains one more vertex than the dimension of the space in which it is used. A simplex for a two dimensional space is a triangle, a simplex for a three dimensional space is a tetrahedron, etc.

The search for the final solution with the Simplex source location method involves four general steps:

i) setting an initial Simplex figure;
ii) calculating errors for vertices;
iii) moving Simplex figures; and
iv) examining the status of convergency.

At the beginning of search, an initial Simplex figure has to be set, which is then rolling through the error space, expanding, contracting, shrinking and turning towards the minimal error point of the space. The movement of the Simplex figure is governed by the error distribution at its vertex, which is calculated each time when the Simplex figure is reshaped.

5.2 Simplex figure and its initial setting

It is understood from the earlier discussion that the Simplex is a geometric figure that contains one more vertex than the dimension of the space in which it is used. Because a source location problem is spanned by not only its geometric dimensions, but also time dimension, the Simplex figure will be tetrahedron and five-vertices for two and three geometrical problems, respectively, if we apply the Simplex method directly to our source location problems.

A more convenient approach is to consider spatial coordinates and origin time "separately" in that error spaces are already optimized in terms of origin time. With this treatment, the dimension of the Simplex figures is solely determined by the geometric dimension of the source location problems: a triangle for two-dimensional problems and a tetrahedron for three-dimensional problems. We will discuss the mathematical procedure of this approach in the next section.

There is no rigid guideline regarding how to set the initial Simplex figure. For the purpose of efficiency, one may want to set it near the potential location with a reasonable size. Care must

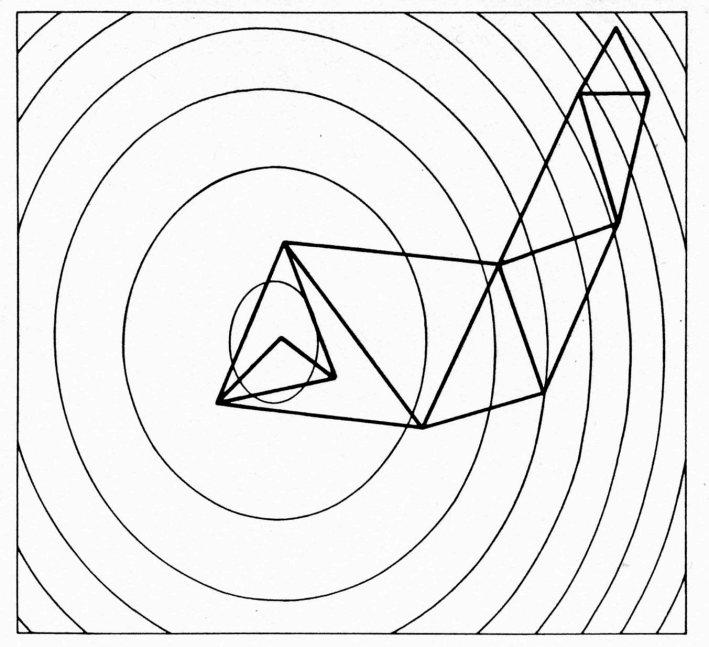

Fig. 3 An example of the Simplex moving on the error space contour plot (after Caceci and Cacheris, 1984).

be taken that the figure is not dimensionally reduced. A practice by the author is to set the initial Simplex figure near the first triggered sensor with the size about one third to one half of the monitoring array. Because of the very robust nature of the Simplex method, the size, shape and location of the initial figure can be quite arbitrary.

5.3 Error calculation and origin time

The location errors are defined by the associated optimization method. Two most popularly employed such methods are the least squares method (L2 norm) and the absolute value method (L1 norm) method. The implementation of these methods for the Simplex source location method is discussed as follows.

Error estimation by the Least squares method

Assume that the observed arrival time at the ith station is t_{oi} and the calculated arrival time at this station is t_{ci}, the station residual as given by Eq. (7) is

$$\gamma_i = t_{oi} - t_{ci}, \tag{7}$$

where γ_i is station residual. A calculated arrival time consists of two parts: origin time, t, and travel time from the vertex under concern to the ith station, tt_i, which are related by the following equation:

$$t_{ci} = tt_i + t. \tag{22}$$

It is noted that origin time is a source parameter, and hence an unknown here. Substituting Eq. (22) into Eq. (7), we express the station residual as a function of origin time:

$$\gamma_i = t_{oi} - (tt_i + t), \tag{23}$$

Now the goal is to find the best estimate of the origin time such that the total error will be minimized. With the least squares method, the total error is defined by

$$\sum \gamma_i^2$$

and it is minimized when the origin time is determined by the equation

$$\frac{d(\sum \gamma_i^2)}{dt} = 0$$

Solving the above equation, the best estimate of the origin time is

$$t = \frac{\sum t_i}{n} - \frac{\sum tt_i}{n} \tag{24}$$

By substituting Eq. (24) into Eq. (23), we express the station residual in terms of observed arrival time and calculated travel time:

$$\gamma_i = (t_i - \frac{\sum t_i}{n}) - (tt_i - \frac{\sum tt_i}{n}) \tag{25}$$

The event residual is

$$Res = \sqrt{\frac{\sum \gamma_i^2}{m-q}} \quad . \tag{26}$$

Eq. (26) is the equation that is used for the error calculation for each vertex if the least squares method is used. Note that the origin time with this approach is given by Eq. (24).

Error estimation by the absolute value method

For the absolute value method, the total error is defined by

$$\sum |\gamma_i| \, .$$

The best estimate of the origin time has to satisfy the objective function

$$\text{Minimize } (\sum |\gamma_i|).$$

Substituting Eq. (23) into the above expression, we have

$$\text{Minimize } (\sum |t_i - (tt_i + t)|).$$

For a set of linear equations with the form of

$$x = b_i,$$

where x is the variable to be estimated and b_i is a constant, the analytical solution of the best fit for x defined by the L1 misfit norm is the median of b_is (Chavatal, 1983). Accordingly, the best fit of the origin time is the median of all $(t_i - tt_i)$. Let us denote it as t_m. The station residual defined by the absolute value method is therefore

$$\gamma_i = t_i - tt_i - t_m \tag{27}$$

and the total error is

$$\sum |\gamma_i| = \sum |t_i - tt_i - t_m| \, . \tag{28}$$

Eq. (28) is the equation that is used for the error calculation for each vertex if the absolute value method is used. Note that the origin time with this approach is the median of all $(t_i - tt_i)$.

5.3 Rules for moving Simplex figures

The movement of Simplex figures is realized by deformation. There are four deformation mechanisms: reflection, expansion, contraction, and shrinkage, which are explained in Fig. 4.

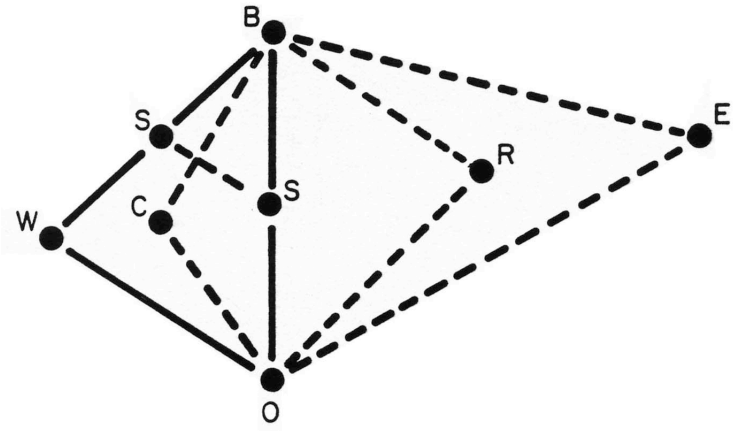

Fig. 4 An illustration of four deformation mechanisms: reflection, expansion, contraction and shrinkage. BWO = the Simplex figure prior to deformation, B = best vertex, W = worst vertex, E = expanded vertex, C = contracted vertex, and S = shrunken vertexes (after Caceci and Cacheris, 1984).

Triangle BWO in the figure represents a two-dimensional Simplex figure. Assuming that W and B represent the worst and best vertex, respectively. R is then called reflected vertex, a mirror image of W about the midpoint of all the other vertexes. E is expanded vertex, which doubles the image distance. C is contracted vertex, located a halfway from W to the midpoint. S represents shrunken vertexes, the middle locations between the best vertex and the others.

The logic to apply these mechanisms is illustrated in the flow diagram in Fig. 5. The following is a summary of this logic:

i) calculating the location error for each vertex,
ii) determining the vertices with the maximum and minimum errors,
iii) locating the reflected vertex and calculating its location error,
iv) if this error is less than the old maximum but more than the minimum, a new Simplex figure is formed by replacing the old maximum by the reflected vertex, and go back to i) to restart the process.
v) if the error at the reflected vertex is less than the minimum, locating the expanded vertex and calculating its location error. If the error is less than the minimum, a new Simplex figure is formed by replacing the old maximum by the expanded vertex, and go back to i) to restart the process. If the error is larger than the minimum, a new Simplex figure is formed by replacing the old maximum by the reflected vertex, and go back to i) to restart the process,
vi) if the error at the reflected vertex is larger than the old maximum, locating the contracted vertex and calculating its location error. If this error is less than the old maximum, replacing the old maximum by the contracted vertex. Otherwise, shrinking the Simplex figure by moving all vertices other than the old minimum to the shrunken vertices. Go back to i) to restart the process.

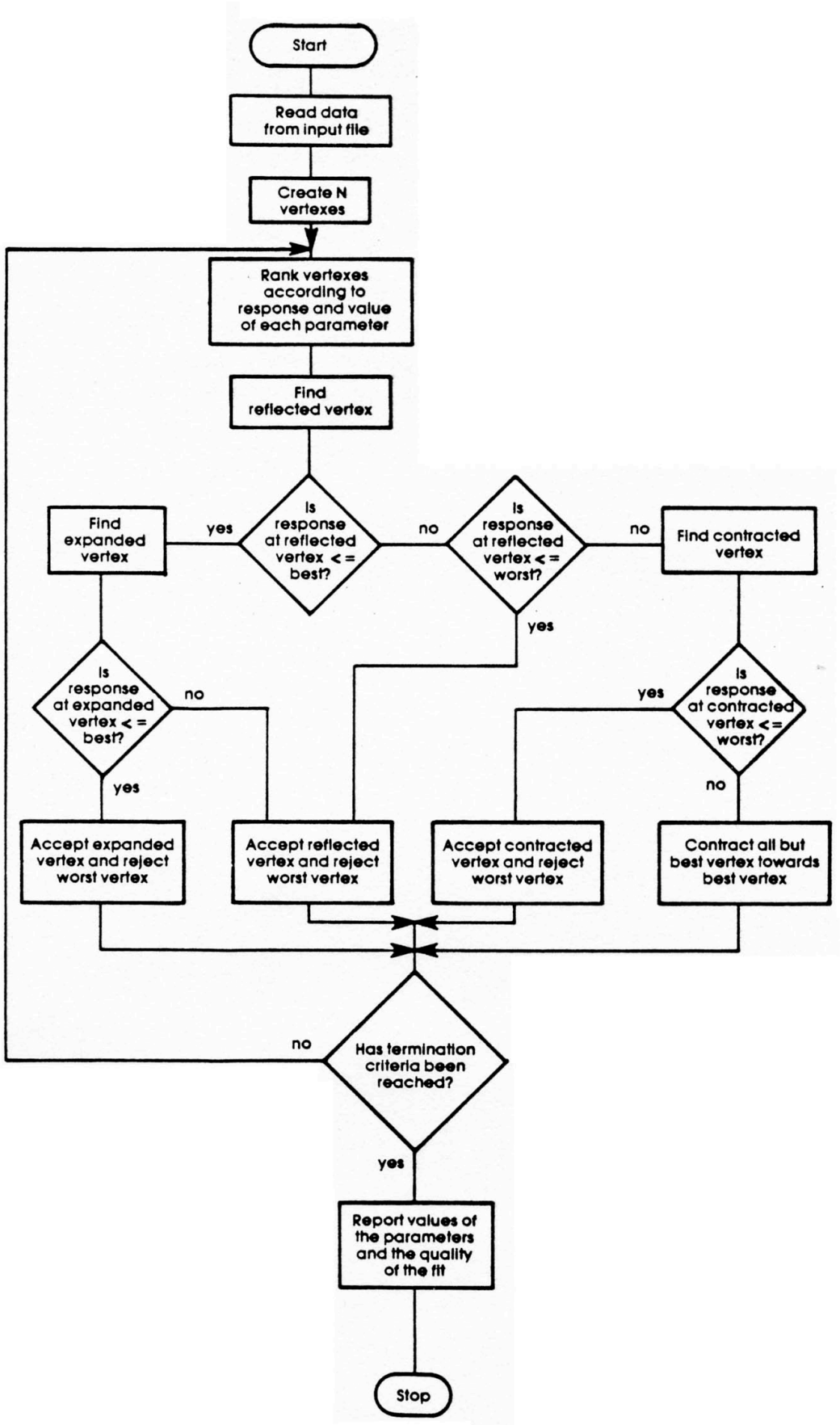

Fig. 5. Flowchart of the Simplex algorithm (after Caceci and Cacheris, 1984).

5.5 Examining the status of convergency

In section 3.4 we discussed the stopping criterion for derivative methods and determined that the size of correction vectors would be most appropriate for the purpose of source location. It is for the same reason that we use the size of Simplex figures as the measurement of the convergency status of Simplex solutions. We accept the solution if the size of the Simplex figure is less than the tolerance. The size may be defined differently. For instance, the average distance from a vertex to others would be a convenient and representative measurement of the size.

5.6 Discussion

The Simplex method offers several distinctive advantages over derivative methods. The most important one is that divergence is impossible. The author manually examined several thousands of location data by the Simplex method and did not observe a single divergence case. In fact, the same Simplex code developed by the author has been installed at a numerous mine sites for the continuous monitoring for many years and there has been no divergence problems ever reported. The reason for this robust performance is that the Simplex figure will not leave the lowest error point which has been found unless a better one is located. Therefore, the Simplex method will always keep the best solution that has been found, whereas for others it may be lost in iteration processes. This character is especially important for the automated monitoring.

The other important advantage of the method is its flexibility. Unlike derivative methods, with which arrival time functions have to be established prior to the analysis, arrival time functions used in the Simplex method can be established during the data processing, which is a necessary condition in order to handle many sophisticated problems.

A frequently mentioned advantage of the Simplex method is that we avoid many computational problems associated with partial derivatives and matrix inversions. It is, however, important to understand that this is not equivalent to say that the underlying problems are also gone. For instance, an ill-conditioned matrix in source location is a reflection of the instability of the associated mathematical system. It would be a serious mistake to expect the disappearance of the problem by using the Simplex method. The truth is that the problem is physical; its existence is independent of the solution methods.

6. Conclusions

Iterative methods are of primary importance in source location methods. This is because of their flexibility in dealing with arrival time functions, which is essential for realistically approaching a great majority of source location problems. Non-derivative methods have to assume a single velocity for all channels, which severely limits their application.

The best known group of iterative methods is derivative algorithms. With derivative methods, we approach the final solution through a continuous updating process of trial solution, and this is done by adding the correction vector found from the previous iteration at each stage. The correction vector is determined by the assessment of nonlinear behavior of arrival time functions at the trial solution. The difference among derivative methods is therefore the type of the nonlinear behavior that is utilized.

The nonlinear behavior used by Geiger's method is the gradient of arrival time functions, or the steepest direction of these functions, represented by the first-order derivatives. The one util

ized by Thurber's method is the extreme of the quadratic functions, which is the second-degree of Taylor polynomials, including both the first- and second-order derivatives..

Because the nonlinear behaviors are used decisively for the calculation of correction vectors, derivative methods, such as Geiger's method and Thurber's method, are inherently efficient, and offer superior performance to other iterative methods. The final solutions are usually approached within a few iterations if the associated systems are stable.

A shortcoming of derivative methods is divergence. While the cause of divergence is complex and a detailed discussion of the problem is beyond the scope of this article, it is important to know that divergence, in general, reflects the problem of instability, which, in turn, is largely governed by the sensor array geometry.

The Simplex method is the most important source location method emerged in recent years. Because of its unique iteration approach, divergence is impossible. This has given the method a huge advantage over derivative methods. The other major advantage of the method is its flexibility to deal with complicated velocity models. Unlike derivative methods, with which arrival time functions have to be established prior to the analysis, arrival time functions used in the Simplex method can be established during the data processing, which is a necessary condition in order to handle many sophisticated problems.

Finally, we would like to emphasize that source location is a subject affected by many factors and the location algorithm is only one of them. In order to use a location algorithm efficiently, we need not only a good understanding of the algorithm itself, but also a perspective view on how the algorithm relates to other factors. The two most important factors are sensor array geometry and errors associated with input data.

Sensor array geometry and system stability

The importance of the sensor array geometry in AE source location lies on the fact that we would never be able to eliminate input errors completely and the impact of these initial errors on the final location accuracy is controlled by the sensor array geometry (Ge, 1988). Good array geometry can effectively limit the impact of initial errors while a poor one will maximize the impact. In other words, the stability of the associated mathematical system is determined by the sensor array geometry. Therefore, the sensor array geometry is fundamental for an accurate and reliable AE source location.

Understanding the relation between sensor array geometry and system stability is important from two perspectives. First, the stability of an event location is independent of the location algorithm being used; that is, we can not change the degree of the sensitivity of a solution to initial errors by switching the location algorithm. If we want to improve the reliability of event locations, the only means is to improve the sensor array geometry. There is no other way around.

Second, a phenomenon that is closely associated with an unstable system is divergence. It is more difficult to approach the solution numerically when the associated system is unstable. It is, however, important to note that the convergence character does differ from method to method. We can reduce the risk of divergence by choosing an algorithm with the better convergence character, and Simplex method is such an example.

Errors associated with input data

There are a number of error sources for AE source location data. The obvious ones are timing, velocity model and sensor coordinates. The one that is often not recognized but may cause most damages is misidentification of arrival types.

An assumption that is frequently made in AE source location is P-wave triggering. In reality, many arrivals are due to S-waves or even not related signals called outlier. An outlier may be caused by either culture noises or the interference of other events. The analysis of AE data shows that S-wave arrivals may account for as much as more than 50% of total arrivals (Ge and Kaiser, 1990). As S-wave velocity is typically half of the P-wave velocity, mixing of P- and S-wave arrivals immediately introduces a large and systematical error to the location system. This has been the primary reason responsible for the poor AE source location accuracy in the past.

Although there are means to reduce the impact of input errors, such as optimization of the sensor array geometry and statistical analysis, one should not expect that any of these methods would be able to deal with large and systematical errors. These errors have to be detected and eliminated before the location process starts.

Acknowledgments

I am grateful to Dr. Kanji Ono for his encouragement to write my research experience in the area, and his detailed review of the manuscript and comments. I thank Dr. Hardy for his thorough review and the anonymous reviewer for his comments and suggestions to improve the manuscript.

References

Anderson, K. R., (1982). Robust earthquake location using M-estimates, *Phys. Earth and Planet Int.,* **30**, 119-130.

Billings, S. D., B. L. N. Kennett, and M. S. Sambridge, (1994). Hypocentre location: genetic algorithms incorporating pro lem-specific information, *Geophys. J. Int.* **118**, 693-700.

Buland, R., (1976). The mechanics of locating earthquakes, *Bull. Seism. Soc. Am.* **66**, 173-187.

Burden, R. L., J. D. Faires and A. C. Beynolds (1981). *Numerical analysis*, Prindle, Weber & Schmidt, Boston, Massachusetts.

Caceci, M. S. and W. P. Cacheris (1984). Fitting curves to data (the Simplex algorithms is the answer), *Byte* **9**, 340-362.

Ge, M., (1988). Optimization of Transducer array geometry for acoustic emission/microseismic source location. Ph.D. Thesis, The Pennsylvania State University, Department of Mineral Engineering, 237 p.

Ge, M. and P. K. Kaiser (1990). Interpretation of physical status of arrival picks for microseismic source location. *Bull. Seism. Soc. Am.* **80**, pp. 1643-1660.

Ge, M. (1995). Comment on "Microearthquake location: a non-linear approach that makes use of a Simplex stepping procedure" by A. Prugger and D. Gendzwell, *Bull. Seism. Soc. Am.* **85**, 375-377.Geiger, L. (1910). Herbsetimmung bei Erdbeben aus den Ankunfzeiten, *K. Gessell. Will. Goett.* **4**, 331-349.

Geiger, L. (1912). Probability method for the determination of earthquake epicentres from the arrival time only, *Bull. St. Louis. Univ.* **8**, 60-71.

Gendzwill, D. and A. Prugger (1989). Algorithms for micro-earthquake location, in *Proc. 4th Conf. Acoustic emission/Microseismic Activity in Geologic Structures*, Trans Tech. Publications, Clausthal-Zellerfeld, Germany, 601-605.

Holland, J. H. (1975). *Adaption in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor.

Kennett, B. L. N. and M. S. Sambridge, (1992). Earthquake location – genetic algorithms for teleseisms, *Phys. Earth and Planet Int.,* **75**, 103 -110.

Lee, W. H. K. and S. W. Stewart (1981). Principles and applications of microearthquake networks, *Adv. Geophys. Suppl.* 2.

Lienert, B. R. and L. N. Frazer (1983). An improved earthquake location algorithm, *EOS, Trans. Am. Geophys. Union*, **64**, 267.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization, *Computer J.* **7**, 308-313.

Prugger, A. and D. Gendzwill (1989). Microearthquake location: a non-linear approach that makes use of a Simplex stepping procedure, *Bull. Seism. Soc. Am.* **78**, 799-815.

Sambridge, M., K. Gallagher, (1993). Earthquake hypocenter location using genetic algorithms, *Bull. Seism. Soc. Am.* **83**, 1467-1491.

Smith, E. G. C., (1976). Scaling the equations of condition to improve conditioning, *Bull. Seism. Soc. Am.* **66**, 2075-2076.

Strang, G. (1980). Linear algerbra and its applications, Academic Press Inc., New York, New York.

Thurber, C. H. (1985). Nonlinear earthquake location: theory and examples, *Bull. Seism. Soc. Am.* **75**, 779-790.

Xie, Z, T. W. Spencer, P. D. Rabinowitz, and D. A. Fahlquist, (1996). A new regional hypocenter location method, *Bull. Seism. Soc. Am.* **86**, 946-958.