SAND2004-xxx
Unlimited Release
Printed August 2004

Seismic Event Location Using Levenberg-Marquardt Least Squares Inversion, 2nd Edition

Sanford Ballard
Next Generation Monitoring Systems Department
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185

Abstract

The most widely used algorithm for estimating seismic event hypocenters and origin times is iterative linear least squares inversion. In this paper we review the mathematical basis of the algorithm and discuss the major assumptions made during its derivation. We go on to explore the utility of using Levenberg-Marquardt damping to improve the performance of the algorithm in cases where some of these assumptions are violated. We also describe how location parameter uncertainties are calculated. A technique to estimate an initial seismic event location is described in an appendix.

Acknowledgement

The author thanks Chris Young and Paul Reeves of the Data Exploitation Department for thoughtful discussions related to the development of the material described in this report. Special thanks goes to Neill Symons for pointing out the utility of the Levenberg-Marquardt algorithm and for his help with the related mathematics. Chris Young, Neill Symons and David Gallegos reviewed the manuscript.

Contents

1. Introduction

The iterative linear least squares inversion algorithm is arguably the most widely used algorithm for locating seismic events. This technique was originally proposed by Geiger (1910) and is described in detail by Jordan and Sverdrup (1981) and Lay and Wallace (1995). The core algorithm is the basis for the United States National Data Center (USNDC) location applications and a next-generation application called LocOO currently under development at Sandia National Laboratories. The purpose of this report is to illuminate the basis of the algorithm so that users of the codes based on it will have a greater appreciation of its capabilities, limitations and idiosyncrasies. In Section 2 we review the mathematical basis of the algorithm and discuss the major assumptions made during its derivation. This section describes material that applies to both the USNDC code and to LocOO. The convergence criteria implemented by LocOO to determine when to terminate iteration is described in Section 3. Sections 4 and 5 describe how poorly constrained events and events which involve violations of model assumptions are handled. The details presented in these sections apply to LocOO but not the USNDC codes. The final section describes the calculation of uncertainties on the event location parameters that is implemented by both LocOO and the USNDC codes. A technique to estimate an initial seismic event location, described in the appendix, is unique to LocOO.

2. Formulation of the Inverse Problem

The seismic event location problem can be described as follows: We are presented with N observations of seismic arrival times, station-to-event azimuth, and horizontal slowness, which we denote as a vector **d** of length N, and we wish to determine the location in time and space of the seismic event which produced the observations. The location is described by a vector **m**, of length $M = 4$, which contains the event latitude, longitude, depth and origin time. In order to find **m**, we need an operator **F**, which relates **m** and **d**:

$$\mathbf{F}(\mathbf{m}) \cong \mathbf{d} \tag{2.1}$$

For the seismic event location problem, **F** is an earth model as defined by the geometry of the Earth and a set of travel time tables with associated corrections.

The left and right sides of Equation 2.1 are only approximately equal since the observations, **d**, are corrupted by measurement errors, and the Earth model, **F**, is only an imperfect representation of the true Earth. The uncertainties are captured in the analysis by associating with each observation d_i and prediction F_i, data and model uncertainty estimates $\sigma_{d,i}$ and $\sigma_{F,i}$, respectively, which have a combined uncertainty $\sigma_{r,i}$ given by

$$\sigma_{r,i} = \sqrt{\sigma_{d,i}^2 + \sigma_{F,i}^2} \tag{2.2}$$

The $\sigma_{r,i}$ are guaranteed to be greater than zero since all observations have finite uncertainty (models should have uncertainties as well). For convenience, we define an $N \times N$ diagonal matrix $\boldsymbol{\sigma}$, containing the values $\sigma_{r,i}$. It may often be true that errors associated with some observations are correlated with the errors of other observations, and it is necessary to account for this so that correlated observations do not exert undue influence on the final location. Following the lead of Chang et al. (1983) we define an $N \times N$, symmetric correlation matrix **R** that contains the correlation coefficients for each pair of observations. The diagonal elements of **R** are all 1, while the off-diagonal elements range between –1 and 1. The data covariance matrix, $\sum$, is given by

$$\sum = \boldsymbol{\sigma} \mathbf{R} \boldsymbol{\sigma} \qquad (2.3)$$

We wish to assign weights to the observations in order to reduce the influence of observations that have large uncertainties and which are highly correlated with other observations. To that end we seek a matrix of weights, **w**, such that

$$\mathbf{w}^{\mathrm{T}} \mathbf{w} = \sum^{-1} \qquad (2.4)$$

From a computational perspective, **w** is the Cholesky decomposition of the inverse of $\sum$. Note that if the observations are uncorrelated, then **R** will be the identity matrix and **w** will be equal to $\boldsymbol{\sigma}^{-1}$. For any observation that has non-zero correlation coefficients with one or more other observations, the weight that will be applied to that observation will depend not only on the observation's own uncertainty, but on the properties of the other observations to which it is correlated as well.

Returning now to the seismic event location algorithm, Equation 2.1 is a statement of the forward problem: given an Earth model, **F**, and an arbitrary seismic event location, $\mathbf{m}_0$, we can calculate a set of predictions, $\mathbf{F}(\mathbf{m}_0)$, of our observations, **d**. The problem we wish to solve is the inverse problem: given an Earth model, **F**, find **m**, the location of the seismic event that produces a set of predictions, **F**(**m**), which agree as well as possible with the observations, **d**.

We must specify precisely what we mean by 'agree as well as possible'. To that end we define a vector **r** of weighted residuals

$$\mathbf{r} = \mathbf{w} \cdot \left[\mathbf{d} - \mathbf{F}(\mathbf{m}_0)\right] \qquad (2.5)$$

We seek the event location **m** that minimizes the sum of the squared weighted residuals, $\left\|\mathbf{r}\right\|^2$, defined as

$$\left\|\mathbf{r}\right\|^2 = \sum_{i=1}^{N} r_i^2 \qquad (2.6)$$

Since **F** is non-linear, we cannot solve the inverse problem directly. The standard approach is to start with an initial estimate of the event location, $\mathbf{m}_0$, and try to deduce an appropriate perturbation to that location, $\delta\mathbf{m}$, such that

$$\mathbf{m}_0 + \delta\mathbf{m} = \mathbf{m} \qquad (2.7)$$

In other words, find the change in event location $\delta\mathbf{m}$, which will move the event location from the initial estimate of the event location $\mathbf{m}_0$, to the event location $\mathbf{m}$, that, when operated on by the Earth model, will produce predictions that agree as well as possible with the observations. Combining Equations 2.1 and 2.7 we have

$$\mathbf{F}(\mathbf{m}_0 + \delta\mathbf{m}) \cong \mathbf{d} \tag{2.8}$$

To solve Equation 2.8 for $\delta\mathbf{m}$, we first linearize it by expanding the left hand side into a Taylor series

$$\mathbf{F}(\mathbf{m}_0 + \delta\mathbf{m}) \cong \mathbf{F}(\mathbf{m}_0) + \left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0}\right)\delta\mathbf{m} + \left(\frac{\partial^2\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0^2}\right)\frac{\delta\mathbf{m}^2}{2!} + \left(\frac{\partial^3\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0^3}\right)\frac{\delta\mathbf{m}^3}{3!} + \ldots \tag{2.9}$$

Ignoring all terms of order 2 and greater and substituting the remainder into Equation 2.8 yields

$$\left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0}\right)\delta\mathbf{m} \cong \mathbf{d} - \mathbf{F}(\mathbf{m}_0) \tag{2.10}$$

Multiplying both sides of Equation 2.10 by $\mathbf{w}$ yields

$$\mathbf{w}\cdot\left(\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0}\right)\delta\mathbf{m} \cong \mathbf{w}\cdot\left[\mathbf{d} - \mathbf{F}(\mathbf{m}_0)\right] \tag{2.11}$$

If we define a matrix $\mathbf{A}$ of weighted partial derivatives

$$\mathbf{A} = \mathbf{w}\cdot\frac{\partial\mathbf{F}(\mathbf{m}_0)}{\partial\mathbf{m}_0} \tag{2.12}$$

Equation 2.11 becomes

$$\mathbf{A}\cdot\delta\mathbf{m} \cong \mathbf{r} \tag{2.13}$$

Let us examine the terms of this equation. The right hand side is a vector of N weighted residuals, defined as the differences between the observations, $\mathbf{d}$, and the predictions of those observations derived from the Earth model, $\mathbf{F}$, applied to the estimate of the event location, $\mathbf{m}_0$. The weighting factors are the combined measurement and model uncertainties. Their inclusion gives greater weight in the analysis to uncorrelated residuals that have smaller uncertainties, and it renders all of the residuals unitless.

The left hand side of Equation 2.11 consists of three terms. The first is the inverse of the weighting matrix, which we have already used in Equation 2.5 to weight the vector of residuals. The second term in Equation 2.11 describes an $N \times M$ matrix of partial derivatives of the N model predictions with respect to the M components of the estimated event location. Each

Table 1 – Partial derivatives of predicted travel time, *T*, station to event azimuth, *Az*, and horizontal slowness, *Sh*, with respect to event hypocenter and origin time. The partial derivatives with respect to the event hypocenter are expressed in Cartesian coordinates: *x* is east, *y* is north and *z* is depth. *t* is origin time, Δ is the distance from the event to the station (measured as an angular distance at the center of the earth), and α is the azimuth from the event to the station measured clockwise from north. All the derivative terms in the interior of Table 1 are determined by interpolation of travel time tables.

	$\frac{\partial}{\partial x}$	$\frac{\partial}{\partial y}$	$\frac{\partial}{\partial z}$	$\frac{\partial}{\partial t}$
T	$-\frac{\partial T}{\partial \Delta}\sin\alpha$	$-\frac{\partial T}{\partial \Delta}\cos\alpha$	$\frac{\partial T}{\partial z}$	1
Az	$-\frac{\cos\alpha}{\sin\Delta}$	$\frac{\sin\alpha}{\sin\Delta}$	0	0
Sh	$-\frac{\partial^2 T}{\partial \Delta^2}\sin\alpha$	$-\frac{\partial^2 T}{\partial \Delta^2}\cos\alpha$	$\frac{\partial^2 T}{\partial z \partial \Delta}$	0

element of this matrix describes the amount by which one of the predictions will increase as a result of a positive change in one of the *M* components of the estimated event location. When the predictions increase, the residuals decrease so the left hand side is a recipe for decreasing the residuals. The partial derivatives are taken in Cartesian coordinates: *x* (easterly), *y* (northerly) and *z* (depth). Coordinate axis *y* corresponds to the direction of increasing latitude while *x* corresponds to the direction along a great circle path that is perpendicular to the meridian that passes through the epicenter. The partial derivatives of travel time, station-to-event azimuth, and horizontal slowness, with respect to event hypocenter and origin time are presented in Table 1. Azimuth derivatives are from Bratt and Bache (1988). The third term on the left side of Equation 2.11, $\delta\mathbf{m}$, is a vector of length *M* containing finite changes in each of the *M* components of the event location.

The right and left hand sides of Equation 2.13 are related by an "approximately equal" symbol for two reasons. The first, as discussed earlier, is because our measurements are corrupted by errors. The second reason that Equation 2.13 is only approximately correct is because the high order terms in Equation 2.9 were omitted. Equation 2.13 will be correct only to the extent that, in regions of the model parameter space near $\mathbf{m}_0$, residuals are directly proportional to changes in model parameters. In this context, "near" means within a distance characterized by $\delta\mathbf{m}$.

Because of the linearity assumption made in the derivation of Equation 2.13, $\mathbf{m}_0 + \delta\mathbf{m}$ will not, in general, be located at the point in parameter space where $\|\mathbf{r}\|^2$ is a minimum. If the linear approximation is a reasonably good one, then $\delta\mathbf{m}$ will move us to a point in parameter space, $\mathbf{m}_0 + \delta\mathbf{m}$, where $\|\mathbf{r}\|^2$ is smaller than it was at $\mathbf{m}_0$. If we accept the improved location as the initial estimate for another application of Equation 2.13, then the solution will be improved further still. If we continue to iteratively apply Equation 2.13 in this manner, we will ultimately reach a situation where the residuals are very close to the minimum we seek and the solution ceases to change significantly. At that point we can conclude that we have reached convergence and accept the solution as the final solution. In essence, we are breaking the larger problem of finding the event location **m** where $\|\mathbf{r}\|^2 = \|\mathbf{r}\|^2_{\min}$ into a sequence of sub-problems each of which is concerned with finding a change in location, $\delta\mathbf{m}$, that moves us closer to **m**. The remainder of this section is concerned with solving just one of the sub-problems.

Returning to Equation 2.13 we note that it tells us only how changes in a given event location will affect residuals between observed and predicted seismic quantities. What we seek is the particular value of $\delta\mathbf{m}$ that leads to $\|\mathbf{r}\|^2_{\min}$. To find that value of $\delta\mathbf{m}$, we must seek the least squares solution to Equation 2.13 which is found by minimizing E^2, given by

$$E^2 = \sum_{i=1}^{N}\left(r_i - \sum_{j=1}^{M} A_{ij}\delta m_j \right)^2 \qquad (2.14)$$

Minimizing E^2 will yield the maximum likelihood estimate of $\delta\mathbf{m}$ (Press et al., 2002).

Note that E^2 is not the same as $\|\mathbf{r}\|^2$. E^2 relates to the solution of one iteration of the overall problem, not the overall problem itself. However, E^2 and $\|\mathbf{r}\|^2$ will be equal at the end of the last iteration when we have reached the solution to the overall problem of finding **m**. At that point, $\delta\mathbf{m}$ will be zero (or very small), so $E^2_{\min} = \|\mathbf{r}\|^2_{\min}$.

$E^2_{\min}$ is found by taking the derivative of E^2 with respect to each of the model parameters and setting them equal to 0:

$$\frac{\partial E^2}{\partial(\delta m_k)} = -2\sum_{i=1}^{N}\left(r_i - \sum_{j=1}^{M} A_{ij}\delta m_j \right) A_{ik} = 0\,; \quad k = 1, M \qquad (2.15)$$

Reorganizing terms, this becomes

$$\sum_{i=1}^{N}\left(\sum_{j=1}^{M} A_{ij}\delta m_j \right) A_{ik} = \sum_{i=1}^{N} r_i A_{ik}\,; \quad k = 1, M \qquad (2.16)$$

or equivalently

$$\mathbf{A}^T \cdot \mathbf{A} \cdot \delta\mathbf{m} \cong \mathbf{A}^T \cdot \mathbf{r} \qquad (2.17)$$

One way to solve Equation 2.17 is to replace matrix **A** by its singular value decomposition (SVD; Menke, 1989; Lay and Wallace, 1995). SVD decomposes **A** into a set of matrices **U**, **W** and $\mathbf{V^T}$ where **U** is an $N \times M$ column-orthogonal matrix and $\mathbf{V^T}$ is the transpose of an $M \times M$ orthogonal matrix. **U** and **V** are orthogonal in the sense that their columns are orthonormal. The columns of **V** are M-dimensional vectors that describe the principal axes of the error ellipsoid of the solution vector $\delta\mathbf{m}$. **W** is an $M \times M$ diagonal matrix containing the so-called singular values of matrix **A**. Their squared inverses describe the dimensions of the principle axes of the error ellipsoid. The diagonal elements of **W** are all zero or positive (Press et al., 2002).

Replacing **A** in Equation 2.17 with its singular value decomposition, $\mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T$, we find

$$\mathbf{V} \cdot \mathbf{W}^2 \cdot \mathbf{V}^T \cdot \delta\mathbf{m} = \mathbf{V} \cdot \mathbf{W} \cdot \mathbf{U}^T \cdot \mathbf{r} \qquad (2.18)$$

Noting that since **U** and **V** are orthonormal their transposes are equal to their inverses, it is straightforward to solve for $\delta\mathbf{m}$

$$\delta\mathbf{m} = \mathbf{V} \cdot \mathbf{W}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{r} \qquad (2.19)$$

or equivalently

$$\delta m_k \cong \sum_{i=1}^{M} \left(\frac{\mathbf{U}_i \cdot \mathbf{r}}{w_{ii}} \right) \mathbf{V}_{k,i} \ ; k = 1, M \qquad (2.20)$$

where the subscripts i on matrices **U** and **V** indicate column numbers and w_i indicates the i'th diagonal element of **W**. Equations 2.17 and 2.18 indicate that each element of $\delta\mathbf{m}$ is a linear combination of the columns of **V**, with coefficients obtained by forming the dot products of the columns of **U** with the vector of residuals, **r**, scaled by the singular values.

After finding $\delta\mathbf{m}$ we add it to $\mathbf{m}_0$ and test for convergence as described in Section 3. If the convergence criteria are not satisfied, we replace $\mathbf{m}_0$ with $\mathbf{m}_0 + \delta\mathbf{m}$ and start another iteration.

There is a minor complication that arises as a result of the fact that epicentral components of $\delta\mathbf{m}$ are in Cartesian coordinates (δx, δy), while those of $\mathbf{m}_0$ are in spherical coordinates (latitude and longitude). To add $\delta\mathbf{m}$ to $\mathbf{m}_0$, we first compute the distance d, that the event is to be moved according to

$$d = \sqrt{\delta x^2 + \delta y^2} \qquad (2.21)$$

and the direction in which the event is to be moved according to

$$\varphi = \tan^{-1}\left(\frac{\delta x}{\delta y} \right) \qquad (2.22)$$

and then move the epicentral components of $\mathbf{m}_0$ distance d in direction φ using spherical trigonometry. The depth and origin time components of $\delta\mathbf{m}$ and $\mathbf{m}_0$ can be simply added together in a straightforward manner.

3. Convergence Criteria

Iteration is terminated when either some maximum number of iterations have been exhausted or when $\|\mathbf{r}\|^2$ ceases to change significantly. More specifically, convergence is declared when

$$\left| \frac{\|\mathbf{r}\|_i^2}{\|\mathbf{r}\|_{i-1}^2} - 1 \right| < tolerance \qquad (3.1)$$

Where subscripts i and i–1 indicate the current and previous iterations, respectively. A reasonable value for tolerance might be 1×10^{-3}.

Convergence can also be declared when damping is being applied and small changes in location fail to reduce $\|\mathbf{r}\|^2$ relative to its value at the conclusion of the previous iteration. See Section 5 for detailed discussion.

4. Poorly Constrained Events

Up to this point we have assumed that the solution to our problem was well constrained by the available data. What if this is not the case? The degree to which **A** is ill-conditioned is characterized by the ratio of the largest to the smallest singular values, a quantity that is referred to as the condition number.

One of the significant advantages of the SVD algorithm is that after computing the SVD of **A**, we can examine **W** to assess potential difficulties and take steps to alleviate them before finally computing $\delta\mathbf{m}$ using Equation 2.20 (Press et al., 2002).

If **A** is singular, then one or more of the singular values (diagonal elements of **W**) will be equal to zero and the condition number will be infinite. The difficulty will be encountered when we attempt to compute Equation 2.20 since we will be attempting to divide by zero. In this case we must take action before computing Equation 2.20 or our algorithm will generate an exception and fail.

In cases where the condition number is extremely large, round off errors in the computer may come in to play, yielding a solution with wildly large components that send our event location off into regions that make no physical sense. Fortunately, this situation appears to be rare in overdetermined seismic event location problems. More common is for the condition number to be moderately large, in which case it is possible to solve for all of the location parameters, but it may not be particularly useful to do so. The uncertainties on the combination of parameters corresponding to the very small singular values may be extremely large (larger than the dimensions of the Earth, for example) rendering the results meaningless.

When we encounter a situation with a large condition number (greater than 10^6, say) our best option is to change one of the very small singular values to zero (Press et al., 2002). As will be described in the next section, zero singular values will be rendered ineffectual, which causes the algorithm to simply not solve for the parameter that correspond to the manipulated singular value. Parameters with zero singular values will not deviate from their initial estimates.

5. Non-linear Effects

It is possible for our solution to diverge, meaning that at the conclusion of a given iteration we find that $\|\mathbf{r}\|^2$ has increased over its value at the beginning of the iteration. This indicates failure of our algorithm and can be attributed to violation of the linearity assumption inherent in Equation 2.11.

Consider the $\|\mathbf{r}\|^2$ hypersurface over our model parameter space. When the linearity assumption is strictly valid, this hypersurface will be ellipsoidal in shape and may be highly elongated in some directions relative to others (Figure 1a). For highly elongated ellipsoids, $\|\mathbf{r}\|^2_{\text{min}}$ will reside at the lowest point of a long, linear, trough-shaped surface (the tip of the long arrow in Figure 1a). From any point in model parameter space, a single step described by the $\delta\mathbf{m}$ vector

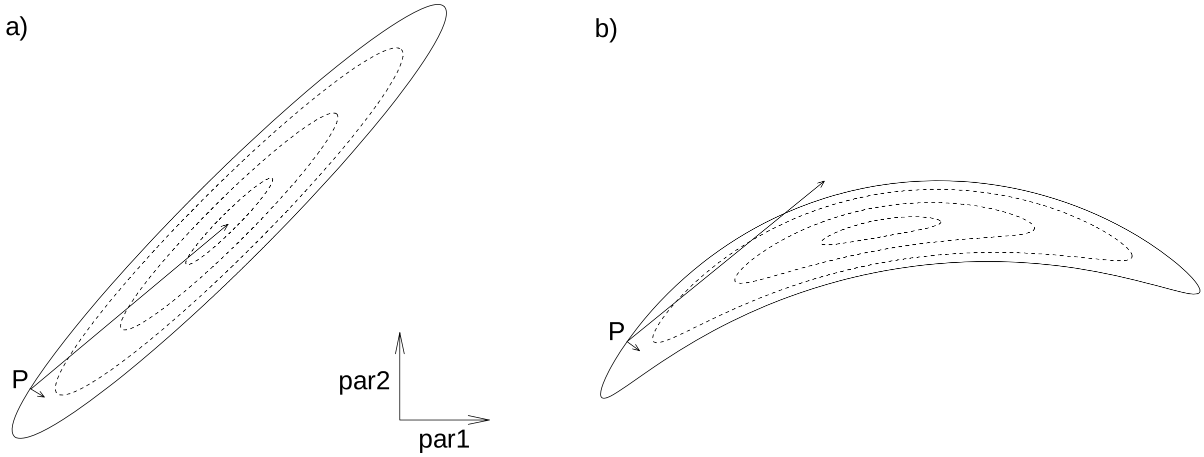

Figure 1 – Contours of constant $\|\mathbf{r}\|^2$ over a 2-dimensional parameter space defined by parameters *par1* and *par2*. a) represents a case where the linearity assumption is strictly valid while b) represents the case where the linearity assumption is invalid. Point P represents the location of the solution vector at the beginning of an arbitrary iteration. The long arrows represents the direction and length of the step calculated by the linear least squares method while the short arrows represent the direction of steepest descent.

will take us to the position in model parameter space where $\|\mathbf{r}\|^2 = \|\mathbf{r}\|^2_{\min}$. Because of the ellipsoidal shape of the $\|\mathbf{r}\|^2$ contours, however, the direction of $\delta\mathbf{m}$ will not, in general, correspond to the direction of steepest descent (the direction that is directly down the gradient of $\|\mathbf{r}\|^2$ with respect to the model parameters). In fact, for extremely elongated ellipsoids, the angle between the direction of steepest descent and the direction of $\delta\mathbf{m}$ may approach 90°. As an alternative to the linear least squares algorithm, we could take many small steps in the direction of steepest descent, but such algorithms tend to be very slow compared to least squares.

When the linearity assumption is not valid, $\|\mathbf{r}\|^2_{\min}$ will reside at the lowest point in a non-linear trough-like feature of potentially complex topology, and the position pointed to by $\delta\mathbf{m}$ may be characterized by a value of $\|\mathbf{r}\|^2$ which is higher than the value of $\|\mathbf{r}\|^2$ at the current position (Figure 1b). In this case, we would be better off just taking a small step in the direction of steepest descent.

Best of all would be if we could interpolate smoothly between taking the step specified by $\delta\mathbf{m}$ and taking a small step in the direction of steepest descent. A method to accomplish this was originally proposed by Levenberg (1944) and Marquardt (1963) and is described in Press et al. (2002). To derive their algorithm we start with the linear least squares solution, Equation 2.17.

$$\mathbf{A}^T \cdot \mathbf{A} \cdot \delta\mathbf{m} = \mathbf{A}^T \cdot \mathbf{r} \tag{5.1}$$

Marquardt's insight was to add a positive constant term, λ, to the diagonal elements of the $M \times M$ matrix $\mathbf{A}^T \cdot \mathbf{A}$

$$(\mathbf{A}^T \cdot \mathbf{A} + \lambda\mathbf{I}) \cdot \delta\mathbf{m} = \mathbf{A}^T \cdot \mathbf{r} \tag{5.2}$$

Replacing $\mathbf{A}$ with its singular value decomposition, $\mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T$, we find

$$\left(\mathbf{V} \cdot \mathbf{W}^2 \cdot \mathbf{V}^T + \lambda\mathbf{I}\right) \cdot \delta\mathbf{m} = \mathbf{V} \cdot \mathbf{W} \cdot \mathbf{U}^T \cdot \mathbf{r} \tag{5.3}$$

Solving for $\delta\mathbf{m}$ yields

$$\delta\mathbf{m} = \mathbf{V} \cdot \left(\mathbf{W}^2 + \lambda\mathbf{I}\right)^{-1} \cdot \mathbf{W} \cdot \mathbf{U}^T \cdot \mathbf{r} \tag{5.4}$$

or

$$\delta m_k \cong \sum_{i=1}^{M} \left(\frac{w_{ii}}{w_{ii}^2 + \lambda} \right) \left(\mathbf{U}_i \cdot \mathbf{r}\right) \mathbf{V}_{k,i} \; ; k = 1, M \tag{5.5}$$

Equations 5.4 and 5.5 differ from Equations 2.17 and 2.18 in that

$$\mathbf{W}^{-1} \to \left(\mathbf{W}^2 + \lambda\mathbf{I}\right)^{-1} \cdot \mathbf{W} \tag{5.6}$$

or, equivalently

$$\frac{1}{w_{ii}} \to \frac{w_{ii}}{w_{ii}^2 + \lambda} \tag{5.7}$$

This has several desirable effects. First of all, when $w_i = 0$, equation 2.20 returns ∞ while Equation 5.7 returns 0 (for $\lambda \neq 0$). This is highly desirable since it means that parameters with zero singular values (parameters that are unconstrained by the data) will contribute nothing to the step rather than contributing infinitely to it. Furthermore, small singular values are affected more than larger ones. This acts to reduce the elongation of highly elongated ellipsoids, making them more spherical and hence reducing the angle between the solution vector, $\delta\mathbf{m}$, and the direction of steepest descent (see Figure 1).

Ultimately, the Levenberg-Marquardt transformation has two effects on $\delta\mathbf{m}$: $\delta\mathbf{m}$ is simultaneously reduced in magnitude and rotated in the direction of the direction of steepest descent. When $\lambda = 0$, we have our unmodified least squares solution. As $\lambda \to \infty$, the direction of $\delta\mathbf{m}$ will approach the direction of steepest descent and its length will approach zero. It remains only to select the appropriate value of λ such that we take as large a step as possible that still moves our solution down the $\|\mathbf{r}\|^2$ gradient (ie., such that $\|\mathbf{r}\|_{i+1}^2 < \|\mathbf{r}\|_i^2$).

Our algorithm proceeds in the following manner:

1) Begin by initializing the applied damping factor, λ, to a negligible initial value, λ_0, such as 10^{-8}, at the start of the first iteration.
2) Start each iteration i by calculating $\delta\mathbf{m}_i$, $\mathbf{m}_{i+1}$ and $\|\mathbf{r}\|_{i+1}^2$.
3) If $\|\mathbf{r}\|_{i+1}^2 \geq \|\mathbf{r}\|_i^2$:
   - discard $\mathbf{m}_{i+1}$
   - increase λ by some factor λ_x (10 for example)
   - return to Step 2.
4) If $\|\mathbf{r}\|_{i+1}^2 < \|\mathbf{r}\|_i^2$:
   - replace $\mathbf{m}_i$ with $\mathbf{m}_{i+1}$
   - if $\lambda < \lambda_0$, divide it once by λ_x
   - return to Step 2 to start the next iteration.

This algorithm seeks to keep λ as small as possible by initializing it to a neglible value and by reducing it by factor λ_x at the conclusion of each successful iteration. Keeping λ small keeps the step size as large as possible. The algorithm is willing to increase λ as much as necessary, however, to ensure that $\|\mathbf{r}\|^2$ is never allowed to increase from one iteration to the next.

It is possible that the event location $\mathbf{m}_i$ at the start of some iteration i is characterized by a value of $\|\mathbf{r}\|_i^2$ which is very close to $\|\mathbf{r}\|_{\min}^2$ and a value of $\|\mathbf{r}\|_{i+1}^2$ that is significantly lower than $\|\mathbf{r}\|_i^2$ will never be found. This situation is handled by monitoring the length of $\delta\mathbf{m}$, an approximation of which is given by

$$\|\delta\mathbf{m}\| = \sqrt{(\Delta \cdot R)^2 + dz^2 + (dT \cdot v)^2} \tag{5.8}$$

where

Δ is the great circle distance the epicenter moved, in radians
R is the radius of the earth in km
dz is the change in hypocenter depth, in km
dT is the change in origin time, in seconds
v is an estimate of the seismic velocity of the medium near the hypocenter (assumed to be constant 8 km/second).

As λ increases, $\|\delta\mathbf{m}\|$ will decrease. If $\|\delta\mathbf{m}\|$ decreases below some threshold value (0.01 km for example), then the algorithm is terminated. At this point, not only is the current iteration concluded but solution convergence can be declared. The reason this conclusion can be reached is that repeated applications of our algorithm have failed to cause $\delta\mathbf{m}$ to point down the $\|\mathbf{r}\|^2$ gradient. Further applications may ultimately achieve that end but only after having reduced the magnitude of $\delta\mathbf{m}$ so severely that it results in only a negligible change in $\mathbf{m}$. We must therefore be at, or very near, the value of $\mathbf{m}$ where $\|\mathbf{r}\|^2 = \|\mathbf{r}\|_{\min}^2$.

6. Uncertainty Intervals

In this section, the method used to calculate uncertainty intervals surrounding the event location and origin time is described. It is critical to note that the technique described here assumes that the uncertainties in the residuals ($\boldsymbol{\sigma}$ in Equations 2.3 and 2.10) are normally distributed. If the residual uncertainties are not normally distributed, then the method outlined in this section is invalid. The analysis also assumes that Earth model, **F** in Equation 2.1, is locally linear with respect to the seismic event location (see discussion in Section 1).

After iterating to convergence we require estimates of the uncertainty in the computed model parameters. We would like these uncertainties to be expressed in terms of an *M*-dimensional confidence ellipsoid surrounding the computed event location in time and space. This ellipsoid defines a region in the *M*-dimensional model parameter space that contains some percentage of the total probability distribution. For a given event, one can point to the associated confidence ellipsoid and say for example, “there is a 95% probability that the true event location falls within this region”. The larger the size of the uncertainty interval, the larger the probability that the true event location falls within it.

Recall that our location problem was over-determined in the sense that we had more observations than unknown parameters. Our solution algorithm solved this system in a least-squares sense. In other words, we found the set of model parameters **m** that minimized the sum of the squared weighted residuals, $\|\mathbf{r}\|^2$ (Equation 2.6)

One can think of $\|\mathbf{r}\|^2$ as a surface over the model parameter space. It represents the sum of the squared weighted residuals corresponding to all sets of model parameters. Our solution, **m**, is the set of model parameters that resides at the point in model parameter space where $\|\mathbf{r}\|^2$ is a minimum. As one moves away from **m** in any direction in model parameter space, $\|\mathbf{r}\|^2$ increases by some amount $\Delta\|\mathbf{r}\|^2$. For a model that is linear in its parameters, a contour of constant $\Delta\|\mathbf{r}\|^2$ describes an *M*-dimensional ellipsoid centered on the best-fit model parameters, **m**. The larger the ellipsoid, the more of the model parameter probability space is incorporated within the ellipsoid. The trick is to select an ellipsoid of just the right size, ie., choose the right value of $\Delta\|\mathbf{r}\|^2$, so that the ellipsoid encompasses the specified percentage of the total model probability space.

6.1 Shape of the Parameter Uncertainty Intervals

The shape of the *M*-dimensional ellipsoid that describes a $\Delta\|\mathbf{r}\|^2 = 1$ contour is returned by the SVD algorithm. The columns of the $M \times M$ matrix **V** contain unit vectors along the principal axes of the ellipsoid and the lengths of the axes are inversely proportional to the squared singular values $\mathbf{W}^2$. Here we use the **W** that has had small singular values set to zero, as described in Section 4, but has not been modified by λ, as described in Section 5.

Typically, we do not wish to express the uncertainties in our model parameters as the joint uncertainties, as represented by **V** and **W**. Generally, we are interested in the uncertainty in a single parameter (the origin time for example), independent of the uncertainties in the other parameters. Or, we might like to express the joint uncertainty in two or three of the parameters (the event epicenter or hypocenter, for example). To accomplish this, we must first project the full *M*-dimensional probability space onto the lower dimensional subspace of interest.

First, we convert **V** and **W** into the $M \times M$ covariance matrix, **C**, as follows (Press et al., 2002):

$$\mathbf{C} = \mathbf{V} \cdot \mathbf{W}^{-2} \cdot \mathbf{V}^T \qquad (6.1)$$

or, equivalently

$$C_{jk} = \sum_{i=1}^{M} \frac{V_{ji} V_{ki}}{w_i^2} \quad j = 1, M; \;\; k = 1, M \qquad (6.2)$$

The uncertainty in a single model parameter, m_i, independent of the others, is simply

$$\sigma_{m,i} = \sqrt{C_{ii}} \tag{6.3}$$

To determine the joint uncertainty in two parameters, for example, the 2 dimensional ellipse surrounding the epicenter of the event, take the intersection of the two rows and columns of **C** corresponding to the components of interest. Copy these into a 2 × 2 matrix and calculate it's inverse, **c.** The equation of the ellipse of interest is

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1 \tag{6.4}$$

where x is east, y is north and $c_{12} = c_{21}$. Multiplying this out yields

$$c_{11}x^2 + 2c_{12}xy + c_{22}y^2 = 1 \tag{6.5}$$

which is the equation for an ellipse in the x, y plane. To remove the xy term (Beyer, 1981) rotate the coordinate axes about the origin through an acute angle $\theta = \tan^{-1}\gamma$ where γ is the positive root of

$$c_{11}\gamma^2 + (c_{11} + c_{22})\gamma - c_{12} = 0 \tag{6.6}$$

Applying this algorithm we find that

$$\theta = \tan^{-1}\left(-\varepsilon + \sqrt{\varepsilon^2 + 1}\right) \tag{6.7}$$

where $\varepsilon = (c_{11} - c_{22})/2c_{12}$. This is the orientation of the major axis of the ellipse relative to the x axis. The strike of the major axis of the confidence ellipse, measured from north, is 90° - θ.

To find the unscaled lengths of the semi-major and semi-minor axes of the confidence ellipse we first rotate the ellipse such that its major and minor axes align with the x and y coordinate axes, respectively. This is accomplished by replacing x and y in Equation 6.5 as follows (Beyer, 1981):

$$\begin{aligned} x &= x'\cos\theta - y'\sin\theta \\ y &= x'\sin\theta + y'\cos\theta \end{aligned} \tag{6.8}$$

If we perform this substitution, set $y' = 0$, and solve for x' we find the length of the semi-major axis of the ellipse

$$x' = \left(c_{11}\cos^2\theta + 2c_{12}\cos\theta\sin\theta + c_{22}\sin^2\theta\right)^{-\frac{1}{2}} \tag{6.9}$$

Performing the substitution, setting *x*' = 0, and solving for *y*' we find the length of the semi-minor axis of the ellipse:

$$y' = \left(c_{11}\sin^2\theta - 2c_{12}\cos\theta\sin\theta + c_{22}\cos^2\theta\right)^{-\frac{1}{2}} \tag{6.10}$$

6.2 Sizes of the Parameter Uncertainty Intervals

At this point, the uncertainty interval, whether it has one, two, three or four dimensions, encloses a space defined by $\Delta\|\mathbf{r}\|^2$ = 1, which yield uncertainty intervals that contain 68.3% of the total probability space. We wish to scale the magnitudes of the uncertainties such that they include some other, user specified, proportion of the total probability space, *p* (90 or 95% for example). Further, the uncertainty estimates assume that the uncertainties in the residuals that were used to weight the residuals (**σ** in Equations 2.3 and 2.10) were exactly equal to the actual uncertainties. This may not be the case so we wish to introduce a data variance scale factor, s_σ^2, to account for this possibility.

To accomplish these objectives, we scale the magnitudes of the uncertainties by a factor κ_p^2 which consists of two terms: a statistical term that relates the value of $\Delta\|\mathbf{r}\|^2$ to the probability that the true event location falls within the uncertainty ellipse defined by $\Delta\|\mathbf{r}\|^2$, and s_σ^2, defined above. There are two end-member methods of estimating the appropriate value for κ_p^2, and a continuum in between. After determining κ_p^2, we simply multiply the model uncertainties (standard errors) by κ_p in order to scale them so that they enclose the desired percentage of the total probability space.

6.2.1 Confidence Uncertainty Intervals

The first possibility relies entirely on *a posteriori* information (Flinn, 1965). It is based on the premise that the residuals that remain after we have obtained the best fit event location accurately reflect the actual data variances. A plausible estimate of the data variance scale factor, given the best fit solution just obtained, is the mean squared weighted residual that resulted from the inversion

$$s_{aposteriori}^2 = \frac{\|\mathbf{r}\|^2}{N-M} \tag{6.11}$$

Recall that **r** is the vector of observed – predicted values, divided by the uncertainty in the residuals. If each observed value differed from the prediction by exactly the value of the uncertainty, then each element of **r** would be 1, $\|\mathbf{r}\|^2$ would equal N^2 and $s_{aposteriori}^2 \cong 1$ for

$N >> M$. Values of $s^2_{aposteriori}$ significantly less than or greater than 1 might suggest that the original estimates of the uncertainties in the residuals were too large or small, respectively.

These considerations suggest that $s^2_{aposteriori}$ is a reasonable estimate of s^2_{σ}, so long as $N >> M$. When N is small, $s^2_{aposteriori}$ may be too large since the denominator is $N – M$, not N. $s^2_{aposteriori}$ can also be a poor estimate of s^2_{σ} by coincidence when N is small since there are few samples.

For confidence uncertainty intervals, we set $s^2_{\sigma} = s^2_{aposteriori}$.

When using only *a posteriori* information, the appropriate statistical parameter is the F-Distribution probability function and κ^2_p becomes

$$\kappa^2_p = s^2_{\sigma} M' F_p\left[M', N - M\right] \tag{6.12}$$

where M' is the number of components included in the joint uncertainty calculation ($M' \le M$) and $F_p\left[M', N - M\right]$ is the F-statistic at the p confidence level with M' and N-M degrees of freedom.

6.2.2 Coverage Uncertainty Intervals

The next possibility is that the residual variances were, in fact, known perfectly *a priori* (Evernden, 1969). Perhaps we have high confidence that the data and model uncertainties that were introduced into the analysis when the residual vector and the derivative matrix were weighted in Equations 2.3 and 2.10 were actually correct. In this case set $s^2_{apriori} = 1$. Alternatively, if many estimates of $s^2_{aposteriori}$ had been accumulated in a series of prior experiments in which events in the same region had been located then we could define $s^2_{apriori}$ to be the mean normalized data variance observed during some number of prior experiments, J:

$$s^2_{apriori} = \frac{1}{J}\sum_{j=1}^{J}(s^2_{aposteriori})_j \tag{6.13}$$

For coverage uncertainty intervals, we set $s^2_{\sigma} = s^2_{apriori}$.

When using only *a priori* information, the appropriate statistical parameter is the Chi-Square probability function, $\chi^2_p\left[M'\right]$, and κ^2_p becomes

$$\kappa^2_p = s^2_{\sigma}\chi^2_p\left[M'\right] \tag{6.14}$$

6.2.3 K-weighted Uncertainty Intervals

These two possibilities can be considered end-members of a continuous spectrum in which both *a priori* and *a posteriori* information about the data variances are used (Jordan and Sverdrup, 1981). To incorporate both types of information, we redefine

$$\kappa_p^2 = s_\sigma^2 M' F_p\left[M', K+N-M\right] \tag{6.15}$$

and define the data variance scale factor

$$s_\sigma^2 = \frac{K s_{apriori}^2 + \left\|\mathbf{r}\right\|^2}{K+N-M} \tag{6.16}$$

where K is an integer that can range from 0 to ∞. When $K = \infty$, $s_\sigma^2 = s_{apriori}^2$, $M' F_p[M', \infty] = \chi_p^2[M']$, and we obtain coverage uncertainty intervals. When $K = 0$, $s_\sigma^2 = s_{aposteriori}^2$, and we obtain confidence uncertainty intervals. For intermediate values of K, the uncertainty intervals reflect both *a priori* and *a posteriori* information about the data variances. Jordan and Sverdrup (1981) recommend setting $K = 8$.

We use algorithms supplied by Press et al. (2002) to compute $\chi_p{}^2$ and F_p for any desired confidence level.

7. Conclusion

In this paper, we have reviewed the mathematical basis of the linear least squares inversion algorithm, which is the most widely used algorithm for estimating seismic event hypocenters and origin times. We have placed particular attention on the assumption of linearity of the predictions of the observations with respect to the seismic event location, showing how the assumption arises in the derivation and discussing its implications. We also outlined the damping technique used by LocOO to improve its performance in situations where the linearity assumption is violated. Finally, we described the methods implemented to calculate the uncertainties on the calculated seismic event location.

References

Ballard, S. (2002), Improved Seismic Event Location Using a Damped Least Squares Algorithm, Proceedings of the 24th Annual Seismic Research Review Symposium, Ponte Vedra Beach, Florida, USA.

Beyer, W. H., ed., (1981), *CRC Standard Mathematical Tables, 26th Edition*, CRC Press, Boca Raton, FL.

Bratt, S. R. and T. C. Bache (1988), Locating Events with a Sparse Network of Regional Arrays, Bull. Seis. Soc. Am., v. 78, pp. 780-798.

Chang, A. C., R. H. Shumway, R. R. Blandford and B. W. Barker, Two methods to improve location estimates – preliminary results, Bull. Seis. Soc. Am., v. 73, pp. 281-295.

Evernden, J. F. (1969), Precision of epicenters obtained by small numbers of world-wide stations, Bull. Seism. Soc. Am., v. 59, pp. 1365-1398.

Flinn, E. A. (1965), Confidence regions and error determinations from seismic event location, Rev. Geophys., v. 3, pp. 157-185.

Geiger, L. (1910), Herdbestimmung bei erdbeden ans den ankunftzeiten, K. Gessel. Wiss. Goett. v. 4, pp. 331-349.

Jordan, T. H. and K. A. Sverdrup (1981), Teleseismic Location Techniques and their Application to Earthquake Clusters in the South-Central Pacific, Bull. Seis. Soc. Am., v. 71, pp. 1105-1130.

Lay, T., and T. C. Wallace (1995), *Modern Global Seismology*, Academic Press.

Levenberg, K. (1944), A method for the solution of certain non-linear problems in least squares, Quart. Appl. Math., v. 2, pp. 164-168.

Marquardt, D. W. (1963), Journal of the Society for Industrial and Applied Mathematics, v. 11, pp. 431-441.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (2002), *Numerical Recipes in C++, The Art of Scientific Computing*, 2nd Edition, Cambridge University Press.

Appendix
Initial estimate of event location

Most seismic event location algorithms require an initial estimate of the event hypocenter and origin time. This appendix describes a technique for obtaining such an estimate from a set of seismic observations. The event location is estimated in the order presented below with subsequent estimates superceding previous estimates.

Travel time

If there are no travel time observations, then the location algorithm is aborted. Otherwise, the station which observed the earliest arrival time (which corresponds to the shortest travel time) is selected. The initial estimate of the event location epicenter is set to the latitude and longitude of this station, the depth is set to zero, and the initial estimate of the event origin time is set to 100 seconds earlier than the arrival time observed at this station.

Azimuth observations

If there are azimuth observations, then it is very important that the initial estimate of the event location be within the correct quadrant for most of the azimuth observations (Bratt and Bache, 1988).

A single azimuth observation

If there is only a single azimuth observation, then LocOO places the initial estimate of the event location 10 degrees away from the station that made the azimuth observation, in the direction of the azimuth observation.

Two azimuth observations

If there are two azimuth observations, then we want to set the event location to the great circle intersection defined by the two station locations/azimuth observations. This location is calculated as follows:

1. For each of the two stations, determine the latitude and longitude of a point 90° away from the station in the direction of the observed azimuth.
2. For each of the two stations, convert both the station location and the location of the 90° point computed in step 1 to 3-component unit vectors centered at the center of the earth.
3. For both stations, compute the elements of the unit vector that is normal to the plane defined by the station vector and the 90° point. This is the normalized cross product of the station vector and the 90° point vector.
4. Compute the cross product of the two normal vectors. This vector will point to one of the two great circle intersections defined by the two planes containing the station locations and the azimuth directions. Note that if the two normal vectors are parallel to each other, the

length of the cross product will be zero and it is not possible to obtain a valid estimate of the event location using this technique.

5. In general, two planes through the center of the Earth intersect at two points on the surface of the earth which are on exactly opposite sides of the globe. The cross product of the two normal vectors calculated in step 4 will point to one of these points but it may not be the one that is pointed to by the two observed azimuth observations. To determine if we are pointing to the correct great circle intersection, compute the dot product of the great circle intersection vector and the 90° point vector associated with either of the two stations. If the dot product is less than zero, then we are pointing at the wrong great circle intersection and the great circle intersection vector should be flipped 180° to point to the correct intersection.

More than two azimuth observations

If there are more than two azimuth observations, compute the great circle intersection vectors of each pair of observations and calculate the sum of all the vectors. Use this location as the initial estimate of the event epicenter. Pairs of azimuth observations which lie along the same great circle path will generate great circle intersection vectors with zero length and hence will not contribute to the final location estimate. Non-parallel pairs of azimuth observations will have contributions that are weighted by the angle at which they intersect, with perpendicular observations contributing the most.

DISTRIBUTION:

20	1138	S. Ballard, 6533
1	1138	D. Carr, 6533
1	0975	E. Chael, 5736
1	0750	G. Elbring, 6116
1	1138	L. Ellis, 6500
1	1138	D. Gallegos, 6533
1	1138	J. Gauthier, 6533
1	0975	J. M. Harris, 5736
1	0975	P. Herrington, 5736
1	1138	J. Hipp, 6533
1	1138	L. Jensen, 6533
1	1138	R. Keyser, 6533
1	1138	J. Merchant, 6533
1	1138	P. C. Reeves, 6533
1	1138	R. Simons, 6533
1	0750	N. Symons, 6116
1	0750	M. Walck, 6116
1	1138	C. J. Young, 6533
1	9018	Central Technical Files, 8945-1
2	0899	Technical Library, 9616
1	0619	Review & Approval Desk, 9612 For DOE/OSTI