

## MATHEMATICAL METHODS FOR THE SEISMIC INVERSION PROBLEM

*S.J. Wright*

## 1. INTRODUCTION

The class of problems known as geophysical inversion problems is one in which measurements are made at points on or above the earth's surface, and mathematical techniques are used to infer something about the geophysical properties of the region under study from these measurements. One example is the problem of simultaneous earthquake hypocentre location and velocity structure determination. The data consists of seismic measurements taken at an array of recording stations on the surface. The "velocity structure" refers to the velocity of propagation of compressional shock waves through each point of the three-dimensional region in which the earthquakes and recording stations lie.

The problem (hereafter referred to as the seismic inversion problem) is essentially one of fitting the data obtained from hypothetical velocity structure and earthquake locations, to the observed data. Since in general it is impossible to completely specify the velocity structure using a finite number of parameters, we will always have an infinitely underdetermined problem. One approach, then, is to apply the parameter separation technique of Pavlis & Booker [9] to obtain a set of equations in which the velocity structure is the only unknown. These equations are used to parametrize a number of three-dimensional subvolumes (or "windows") within the region of interest, and to estimate the average velocity within each of these subvolumes. Parametrization of the unknown function which describes the velocity model is not directly attempted. This

approach is tried by Chou & Booker [4] , who base their work on the one-dimensional analysis of Backus & Gilbert [2,3] . While the method is very sophisticated, we are not aware of its use on practical problems, and we will not deal with it in this paper.

An alternative, and more widespread, approach is to approximate the velocity structure with a model which can be described using a finite number of parameters. The number of available data is assumed large enough that the problem is overdetermined, so that least squares methods can be used. Various techniques for modelling the velocity structure are discussed in section 2.2 .

Practical considerations make the resulting least squares problem very difficult to solve. Given a hypothetical set of earthquake locations and a hypothetical velocity structure, accurate calculation of the resultant data is very time-consuming. This is discussed in section 2.3 . Partial derivatives of the data with respect to locations and model parameters are also difficult to find accurately for most models (see section 3) . The sheer size of the problem may restrict the accuracy with which these quantities are found, and may also lead to storage difficulties.

The first least-squares formulation is due to Aki & Lee [1] , who solved a damped linear least squares problem (i.e. just one iteration). Their damping matrix assumed prior knowledge of standard errors in the data and the unknown parameters. Their approach is easily generalised to a nonlinear least squares problem, as in Hawley, Zandt, and Smith [6] , Spencer & Gubbins [12] , and Thurber [13] . Firbas [5] and Wesson [15] use nonlinear least squares to solve the simpler problem in which only the velocity model parameters are assumed unknown.

These authors obtain the step at each iteration either by solving the normal equations, by performing a decomposition of the Jacobian or by applying the parameter separation technique described in section 4.2 . We examine in particular the solution technique of Thurber's program HYP02 in section 4.3 .

Section 4.4 describes the proposed application of a new technique for large sparse nonlinear least squares, due to Wright & Holt [16] to the seismic inversion problem.

## 2. CALCULATING P-WAVE TRAVEL TIMES

### 2.1 Formulation of the Problem.

The data used in the seismic inversion problem are obtained from a network of seismic recording stations in the region of interest. Each earth tremor or "event" produces compressional waves and shear waves (p- and s- waves) which are monitored at each station. In particular, we are interested in the first time at which the p-wave from each event is detected at each station.

Since the region under consideration will most likely contain rocks of differing types, the p-waves will be travelling through an inhomogeneous medium. Thus the "fastest" ray path will not generally be a straight line. The problem of ray tracing is in general difficult to solve, and would comprise a major part of the computer time in any reasonable solution process for the seismic inversion problem. A typical set of ray paths for a single event is shown in Fig. 1 .

The first-arrival time of the p-wave from event  $k$  at station  $\ell$  can be thus written as

$$(2.1) \quad t_{k\ell}^c = t_{k\ell}^c(x_k, y_k, z_k, t_k; M)$$

where

- (1)  $(x_k, y_k, z_k)$  are the coordinates of event  $k$ ,
- (2)  $t_k$  is the time of occurrence of event  $k$ .
- (3)  $M$  denotes the p-wave velocity structure of the crustal region.

Our aim in solving the inverse problem is to choose values of the parameters, and  $M$ , so as to minimise the sum-of-squares objective function

$$(2.2) \quad \sum_{\ell=1}^m \sum_{k=1}^n (t_{k\ell}^c - t_{k\ell}^o)^2$$

where

$m$  is the number of stations

$n$  is the number of events

$t_{k\ell}^o$  is the observed arrival time of the  
p-wave from event  $k$  at station  $\ell$ .

## 2.2 Parametrizing the Velocity Structure.

If we can completely specify our velocity model  $M$  with a finite number of parameters, then our problem becomes a least-squares problem. Many different parametrization techniques have been tried. Perhaps the most obvious one is to split the region into a set of rectangular blocks, and assign a single velocity parameter to each. This is the approach used in Aki and Lee's [1] early attempt at the problem.

This approach made ray tracing impossible, since the model must be at least continuously differentiable for ray tracing techniques to be applied. Ray paths were taken to be straight lines, a generally unsatisfactory approximation.

Another approach is to consider  $M$  as a linear combination of standard basis functions on the whole space, e.g.

$$(2.3) \quad v(x, y, z) = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \sum_{k=0}^{n_3} A_{ijk} P_i(x) Q_j(y) R_k(z)$$

where  $P_i$ ,  $Q_j$ , and  $R_k$  could be (scaled) Legendre polynomials. The parameters for this model are the coefficients  $A_{ijk}$ . This type of approach has been tried by Firbas [5] for the simpler problem in which the hypocentres and occurrence times are fixed. However, the model produced is strongly dependent on the choice of basis functions, and from our limited experience, this method is not to be recommended.

A sort of compromise approach is to choose a regular grid of points over the region, and let the set of parameters be the velocity values at points of this grid. Velocities at other points in the region are found by spline interpolation. This approach is used by the seismic inversion FORTRAN program HYP02, which will be discussed further in later sections. The interpolation process involves forming three bicubic splines, one each for the  $x - y$ ,  $y - z$ , and  $x - z$  planes. The tensor product of these three splines gives the interpolation function. This model has continuous second partial derivatives, allowing the "bending" methods to be described in section 2.3 to be used for ray tracing.

Finally, an approach based on prior knowledge of the region is to model the velocity structure by some function which has a number of unknown parameters. These parameters are to be determined in the inversion. An example is found in Spencer & Gubbins [12] where a region in New Zealand is modelled by an increasing-depth velocity structure upon which a low velocity slab is superimposed. The velocity model is characterized by

- (1) parameters denoting velocity outside the slab.
- (2) Strike, dip, and width of the slab
- (3) two parameters which characterize velocity within the slab.

This method is also used by Wesson [15] who uses a 12-parameter function to characterize velocity in the Bear Valley area of California.

### 2.3 Ray Tracing.

The ray equation for a smooth (say  $C^2$ ) medium can be derived from first principles (see, for example, Lee & Stewart [7]) to be

$$(2.4) \quad \frac{d}{ds} \left[ u(\underline{\eta}) \frac{d}{ds} \underline{\eta} \right] = \nabla u$$

where

$$\underline{\eta} = (x, y, z) \text{ represents the ray and } u(\underline{\eta})$$

represents the reciprocal velocity of propagation (or "slowness") of the medium. If we apply the constraint

$$(2.5) \quad \left( \frac{dx}{ds} \right)^2 + \left( \frac{dy}{ds} \right)^2 + \left( \frac{dz}{ds} \right)^2 = 1$$

then  $s$  will denote arc length along the ray. The ray tracing problem consists of solving (2.4) subject to the constraint (2.5) and the boundary value conditions

$$(2.6) \quad \underline{\eta}(0) = P_S, \quad \underline{\eta}(S) = P_R$$

where  $P_S$  and  $P_R$  denote the source and receiver coordinates, respectively,  $S$  denotes the total arc length of the ray - also an unknown.

Following Pereyra, Lee, and Keller [11], and Pereyra [10] we can reduce (2.4) to a first-order system with six dependent variables (denoted by  $\omega_1, \dots, \omega_6$ ) in the standard way. To handle the unknown  $S$ , we use a change of independent variable

$$t = \frac{s}{S},$$

and introduce the dependent variable  $\omega_8 \equiv S$ . If we introduce a further dependent variable  $\omega_7$ , denoting the partial travel time, that is

$$\omega_7(t_1) = \int_0^{t_1 S} u \, ds = S \int_0^{t_1} u \, dt \quad (0 \leq t_1 \leq 1)$$

then clearly  $\dot{\omega}_7 = Su = \omega_8 u$ , and  $\omega_7(1)$  will be the total travel time from  $P_S$  to  $P_R$ . The system can now be written as

$$(2.7) \quad \left. \begin{aligned} \dot{\omega}_i &= \omega_8 \omega_{i+1} \\ \dot{\omega}_{i+1} &= \frac{\omega_8}{u(\tilde{\eta})} \left[ -G(\tilde{\omega}) \omega_{i+1} + \frac{\partial u}{\partial \omega_i} \right] \end{aligned} \right\} i = 1, 3, 5 .$$

$$\dot{\omega}_7 = \omega_8 u$$

$$\dot{\omega}_8 = 0$$

where  $G(\tilde{\omega}) = u_x \omega_2 + u_y \omega_4 + u_z \omega_6$

with  $[0,1]$  being the interval of integration.

If there is a discontinuity across which the ray must travel, this can be dealt with by doubling the size of the system and applying Snell's law at the discontinuity to obtain extra boundary conditions. However, the Pereyra ray path solver has only been applied to smooth models in the context of the seismic inversion problem, and hence discontinuities are not allowed for in current solution algorithms.

Pereyra uses an adaptive finite difference scheme to solve this two-point boundary value problem. However it can be seen that in any computation of hypothetical travel times a large number of ray paths must be found. For practical problems, this involves substantial computer time. It would be desirable, then, to sacrifice some accuracy in our computed travel times, if we could gain a one-or-two-order of magnitude saving in computer run time. To meet this requirement approximate ray tracing algorithms have been devised by Thurber & Ellsworth [14] .

Their first algorithm, ART, (for Approximate Ray Tracing) assumes that the grid-and-spline velocity model is used. They constrain their approximate ray path to be within the vertical plane which includes  $P_S$  and  $P_R$ , thus reducing the problem to two dimensions. They split the model into layers (see Fig. 2b) and assign each layer a velocity value according to its "average" value in a local rectangular volume. As in Fig. 2b, all possible ray paths appropriate to this layered model are



considered. Each ray path is then subdivided into  $N$  segments, each of length  $\Delta s$ , and the travel time integral

$$\int u(s) ds$$

is approximated by

$$(2.8) \quad T = \sum_i u_i \Delta s$$

where  $u_i$  is the slowness evaluated at the midpoint of the  $i$ -th segment. Travel times for all possible paths are compared, and the approximate path is chosen to be the one giving the fastest travel time.

The approximate path so derived can actually be used as the initial path for the algorithm of Pereyra. In fact, it is much better initial path than the usual straight-line approximation, because the latter is more likely to lead to convergence to a local minimum. Thurber & Ellsworth [14] give an example in which use of an initial path generated by ART produces convergence to the correct first-arrival path, while use of the straight line initial path produces convergence to a local-minimum path with substantially greater travel time.

Thurber & Ellsworth found that the travel times produced by ART were accurate to within normal reading error for  $p$ -wave arrival times. This is enough to justify their direct use in a seismic inversion algorithm. Thurber & Ellsworth found that there is about a 99% saving in computer time over the Pereyra method.

The second algorithm of Thurber & Ellsworth, ART 2, is based on the assumption that the ray path can be adequately approximated by an arc of a circle. A number of arcs with different radii of curvature are tried, and the dip of the plane containing the arcs is varied. The authors report that this method gave even better results than ART on their test models. However its appropriateness for a model with strong lateral heterogeneity is questionable.

### 3. THE JACOBIAN MATRIX

#### 3.1 Calculation of Travel Time Derivatives.

To solve the least squares problem (2.2), we need to be able to calculate derivatives of the computed arrival times  $t_{kl}^C$  with respect to the occurrence time, the hypocentre coordinates, and the model parameters. It is found that the Jacobian (matrix of partial derivatives) is sparse, with a known sparsity structure.

Since  $t_{kl}^C$  can be written as

$$t_{kl}^C = t_k + t_{kl}^T$$

where  $t_k$  is the occurrence time of event  $k$  and  $t_{kl}^T$  is the fastest travel time from event  $k$  to receiver  $l$ , then clearly

$$(3.1) \quad \frac{\partial t_{kl}^C}{\partial t_k} = 1.$$

Using an argument based on calculus of variations, it can be shown that the partial derivatives with respect to hypocentre coordinates are given by

$$(3.2) \quad \left\{ \begin{array}{l} \frac{\partial t_{kl}^C}{\partial x} = -u \frac{dx}{ds} \\ \frac{\partial t_{kl}^C}{\partial y} = -u \frac{dy}{ds} \\ \frac{\partial t_{kl}^C}{\partial z} = -u \frac{dz}{ds} \end{array} \right.$$

where all right-hand side quantities are evaluated at  $P_S$ , and so

$\frac{dx}{ds}$ ,  $\frac{dy}{ds}$ , and  $\frac{dz}{ds}$  are the direction cosines of the ray at  $P_S$ . This result

can also be seen intuitively. If Pereyra ray tracing is used, the partial

derivatives can be obtained directly from (3.2). However if some form

of approximate ray tracing is used, the take-off direction of the approximate

ray may be substantially different from that of the actual ray.

Thurber [13] has found that it is inappropriate to use (3.2) in this case.

Rather, he suggests using a "variational" estimate, that is, adding a small displacement to each coordinate in turn, calculating the new travel time, and using derivative estimates of the form

$$(3.3) \quad \frac{\partial t_{kl}^c}{\partial x} \approx \frac{\Delta t_{kl}^c}{\Delta x} .$$

This method requires the evaluation of three more ray paths, but remains practical because ART is so economical.

Velocity partial derivatives are given by the formula

$$(3.4) \quad \frac{\partial t}{\partial a_j} = \int_0^S \frac{\partial u}{\partial a_j} ds$$

where  $S$  is the arc length along the path and  $a_j$  is the  $j$ -th velocity parameter. Since approximate integration is used, as in (2.8), this becomes

$$(3.5) \quad \frac{\partial t}{\partial a_j} \approx \sum_{i=1}^n \frac{\partial u_i}{\partial a_j} \Delta s$$

where  $\frac{\partial u_i}{\partial a_j}$  is evaluated at the midpoint of the  $i$ -th line segment. For models such as that of Spencer & Gubbins [12],  $\frac{\partial u}{\partial a_j}$  can be calculated analytically and so (3.5) can be used directly. For the block model, for which  $a_j$  gives the slowness of block  $j$ ,  $\frac{\partial u_i}{\partial a_j}$  will be 1 if the midpoint of the  $i$ -th line element lies in block  $j$ , and zero if it does not. Again, (3.5) can be used directly.

While it would be possible to calculate  $\frac{\partial u}{\partial a_j}$  exactly for the grid-and-spline model of Thurber [13], it would not be practical since the slowness at any point is dependent on the slowness values of 32 surrounding grid points. Thurber finds that the "block" approach is best, that is,  $\frac{\partial u_i}{\partial a_j} = 1$  if node  $j$  of the grid is the closest node to the midpoint of the  $i$ -th line element. Otherwise  $\frac{\partial u_i}{\partial a_j} = 0$ . In testing on

artificial data, he found that average error in derivatives calculated by this method was 13% .

### 3.2 Structure of the Jacobian.

The vector of variables is generally ordered as follows :

$$(t_1, x_1, y_1, z_1, t_2, \dots, z_n, a_1, a_2, \dots, a_p)$$

where  $p$  is the number of velocity parameters, and the residuals are arranged thus

$$(r_{11}, r_{12}, \dots, r_{1m}, r_{21}, \dots, r_{2m}, \dots, r_{nm})$$

where  $r_{kl} = t_{kl}^c - t_{kl}^0$  (see, for example, Lee & Stewart [7] , Spencer & Gubbins [12] , Pavlis & Booker [9]) .

Since

$$\frac{\partial r_{kl}}{\partial t_j} = \frac{\partial r_{kl}}{\partial x_j} = \frac{\partial r_{kl}}{\partial y_j} = \frac{\partial r_{kl}}{\partial z_j} = 0$$

unless  $k = j$  , the columns of the Jacobian which correspond to hypocentre coordinates and occurrence times have a block diagonal structure

(see Fig. 3) .

The sparsity of the second part of the Jacobian will depend on the form of the velocity modelling. If the Spencer-Gubbins technique is used, then the number of parameters will be small and this part of the Jacobian will be dense. However, if the more general modelling techniques are used,  $p$  will be large, and this submatrix will be sparse. This is because each ray path will depend on a relatively small subset of the velocity parameters  $a_1, a_2, \dots, a_p$ .

It can be seen immediately that it is a trivial task to perform a QR decomposition on the "block diagonal" part of the Jacobian. This fact is exploited in the solution techniques to be discussed in the next section.

## 4. SOLVING THE INVERSE PROBLEM

### 4.1 Introduction.

It can be seen at this stage that the seismic inversion problem is unusual and difficult to solve. Firstly, if precise ray tracing is used, function evaluations are almost as expensive as Jacobian evaluations (although this is not the case when ART is used). Secondly, only approximate derivatives are usually known. This could conceivably lead to a situation in which the gradient is not a descent direction, although in practice this would only be likely to occur near the solution, where the gradient is vanishing. Thirdly, storage requirements could create difficulties. A typical run would involve about 30 events and about 100 stations, giving a problem with thousands of equations and hundreds of unknowns.

### 4.2 Parameter Separation.

The approach of parameter separation, outlined by Pavlis & Booker [9], is used by Thurber [13] in the program HYP02 to solve the linear least squares subproblem

$$(4.1) \quad \min_Y \|Jy + r\|_2^2$$

which arises at each iteration of the Gauss-Newton solution method for the nonlinear problem. The usefulness of parameter separation arises from its taking advantage of the ease with which a QR decomposition can be performed on the first part of the Jacobian.

If we write the Jacobian as

$$(4.2) \quad J = [A \ M] = \begin{pmatrix} A_1 & & & M_1 \\ & A_2 & & M_2 \\ & & \ddots & \\ 0 & & & A_n & M_n \end{pmatrix}$$

where each  $A_i$  is  $m \times 4$  and each  $M_i$  is  $m \times p$ , we can find orthogonal matrices  $U_i$  such that

$$(4.3) \quad U_i^T A_i = \begin{pmatrix} R_i \\ 0 \end{pmatrix}$$

where  $R_i$  is  $4 \times 4$  upper triangular (and, we assume for simplicity, non-singular).  $U_i$  can be partitioned accordingly as

$$(4.4) \quad U_i = \begin{pmatrix} U_i^{(1)} & \vdots & U_i^{(0)} \end{pmatrix}$$

where  $U_i^{(1)}$  is  $m \times 4$  and  $U_i^{(0)}$  is  $m \times (m-4)$ . Hence  $U_i^{(0)T}$  "annuls"  $A_i$ .

We can then form an  $(mn) \times (mn)$  orthogonal matrix  $U$  as follows :

$$(4.5) \quad U = \begin{pmatrix} U_1 & & & \\ & U_2 & & 0 \\ & & \ddots & \\ 0 & & & U_m \end{pmatrix} \Pi$$

where  $\Pi$  is a permutation matrix, chosen so that  $U$  satisfies

$$(4.6) \quad U^T J = \begin{pmatrix} R_1 & & & \vdots & \\ & R_2 & & & 0 \\ & & \ddots & & \\ 0 & & & R_n & \\ \hline & & & & M^{(1)} \\ \hline 0 & & & & M^{(0)} \end{pmatrix}$$

where  $M^{(0)} = \begin{pmatrix} U_i^{(0)T} M_i \\ \vdots \\ U_n^{(0)T} M_n \end{pmatrix}_{i=1}^n$  and similarly for  $M^{(1)}$ . If we rewrite (4.1) as

$$(4.7) \quad \begin{aligned} & \min_{\delta x, \delta a} \left\| \begin{pmatrix} A \\ \vdots \\ M \end{pmatrix} \begin{pmatrix} \delta x \\ \delta a \end{pmatrix} + r \right\|_2^2 \\ \Leftrightarrow & \min_{\delta x, \delta a} \left\| \begin{pmatrix} R_1 & & & \vdots & \\ & R_2 & & & \\ & & \ddots & & \\ 0 & & & R_n & \\ \hline & & & & M^{(1)} \\ \hline & & & & M^{(0)} \end{pmatrix} \begin{pmatrix} \delta x \\ \delta a \end{pmatrix} + \begin{pmatrix} r^{(1)} \\ r^{(0)} \end{pmatrix} \right\|_2^2 \end{aligned}$$

where  $r^{(1)} = \left( U_i^{(1)T} r_i \right)_{i=1}^n$ , and similarly for  $r^{(0)}$ .

The solution to (4.7) is found by solving the linear least squares problem

$$(4.8) \quad \min_{\delta a} \left\| M^{(0)} \delta a + r^{(0)} \right\|_2^2$$

for  $\delta a$ , and then solving the upper triangular system

$$(4.9) \quad \begin{pmatrix} R_1 & & & & 0 \\ & R_2 & & & \\ & & \ddots & & \\ & & & R_n & \\ 0 & & & & \end{pmatrix} \delta x = -r^{(1)} - M^{(1)} \delta a$$

to obtain  $\delta x$ .

#### 4.3 Solution Technique of HYP02.

The program HYP02, written by Thurber, Roecker, Ellsworth, and Comer, is a seismic inversion program in current use at the U.S. Geological Survey. It uses the grid-and-spline velocity model discussed earlier, and the linear least squares subproblem is solved using parameter separation. The actual implementation of parameter separation in this program is not sophisticated, but geared towards minimising storage requirements.

The subproblem (4.8) is solved by forming the normal equations,

$$(4.10) \quad M^{(0)T} M^{(0)} \delta a = -M^{(0)T} r^{(0)}$$

and taking an  $LL^T$  decomposition of the coefficient matrix (L is lower triangular). Since by definition,

$$(4.11) \quad M^{(0)T} M^{(0)} = \sum_{i=1}^n \left( U_i^{(0)T} M_i \right)^T \left( U_i^{(0)T} M_i \right)$$

it is possible to "build up" the matrix  $M^{(0)T} M^{(0)}$  by evaluating the  $M_i$  and  $U_i$  one at a time, rather than by forming  $M^{(0)}$  explicitly.

This is what in fact occurs. Only  $m$  rows of the Jacobian are evaluated at a time (i.e. each  $A_i$  and  $M_i$ ). The  $U_i$  are formed explicitly, as are  $U_i^T M_i$  and  $U_i^T r_i$ . The  $R_i$  are stored, as are the elements of  $r^{(0)}$  and  $r^{(1)}$  so found.

Damping is allowed for in the solution of (4.10) by adding a user-supplied constant to the diagonal elements of  $M^{(0)T} M^{(0)}$ . No damping is used in the solution of (4.9) for hypocentre adjustments.

Having obtained the steps  $\delta a$  and  $\delta x$ , restrictions are applied to the magnitude of each component. No component of  $\delta a$  or  $\delta x$  is allowed to exceed a user-supplied constant, which is meant to be chosen so that the new point is within a region of approximate linearity about the old point. For the regions which he tested, Thurber applied upper bounds of 3 km on hypocentre coordinate movement and .5 km/s on velocity parameter changes for each step.

Once the step has been made, there is no check for residual norm decrease at the new point. It is possible that the residual vector norm at the new point is larger than at the old point.

#### 4.4 Proposed Improvements to HYP02.

The approach of HYP02 may seem ad-hoc from a mathematical programming viewpoint, but we must remember that the authors wanted to minimise storage requirements, and to minimise the possibility of taking an unsuccessful step. Also, the program users would not usually be interested in iterating to convergence. Rather, they would take a fixed number of iterations and be happy if there was a substantial decrease in residual norm.

However, it is proposed to apply a new implementation of a large sparse nonlinear least squares solver to the seismic inversion problem, using the framework of HYP02. This new solver, due to Wright



& Holt [16] uses a Levenberg-Marquardt approach, in which a linear subproblem of the form

$$(4.12) \quad \min_{\delta x} \left\| \begin{pmatrix} J \\ \lambda I \end{pmatrix} \delta x + \begin{pmatrix} r \\ 0 \end{pmatrix} \right\|_2^2$$

whose solution is

$$(4.13) \quad \delta x_\lambda = - (J^T J + \lambda^2 I)^{-1} J^T r$$

is considered at each iteration. In the standard Levenberg-Marquardt method, the exact solution (4.13) is used, and the damping parameter  $\lambda$  is varied until an "acceptable" step  $\delta x_\lambda$  (one which produces a sufficient decrease in residual norm) is found. For small-and-medium-sized problems, the step  $\delta x_\lambda$  is computed by finding a decomposition of the coefficient matrix in (4.12). For large sparse problems, decomposition would not be practical, and so we could use either a direct method or an iterative method to solve the linear least squares problem (4.12).

The Inexact Levenberg-Marquardt method of Wright & Holt applies the iterative algorithm LSQR of Paige & Saunders [8] to (4.12). However, instead of finding  $\delta x_\lambda$  accurately, we only solve the subproblem to within a certain tolerance. In other words, if  $t$  is defined by

$$(4.14) \quad t = (J^T J + \lambda^2 I) \delta x + J^T r$$

then we accept  $\delta x$  as our "inexact" solution if

$$(4.15) \quad \frac{\|t\|}{\|J^T r\|} \leq \eta < 1$$

where  $\eta$  is a parameter which is reset at the start of each iteration. Clearly if  $\delta x$  in (4.14) is the exact solution from (4.13),  $t$  will be the zero vector.

It is shown (in Wright & Holt [16]) that a decrease in function value can be obtained at each iteration if the condition (4.15) is applied, and the damping parameter  $\lambda$  is chosen sufficiently large. Further, it is proved that the sequence of iterates will converge to a stationary point of the sum-of-squares objective function.

The foregoing results only require that  $\eta$  be less than 1 at each iteration. However, in the case where the sum-of-squares at the optimal point is zero, superlinear convergence can be obtained by ensuring that the sequence of values of  $\eta$  approaches zero. In fact, if the condition

$$(4.16) \quad \eta \leq 2 \|J^T r\|$$

is enforced at each step (where  $(2J^T r)$  is the gradient of the objective function  $\|r\|_2^2$ ), the convergence will be quadratic. For small-residual problems (including, hopefully, the seismic inversion problem), a good strategy would be to apply the condition (4.16) as the optimal point is approached, in the hope of accelerating the convergence.

The iterative method LSQR of Paige and Saunders requires matrix-vector products of the form  $Jp$  and  $J^T p$  to be calculated at each iteration. The larger the value of  $\eta$ , the fewer iteration of LSQR are required for the inexact solution  $\delta x$  to satisfy (4.14) and (4.15). In addition, if  $\lambda$  is sufficiently large, these conditions can be satisfied after only one iteration of LSQR ( $\delta x$  will then be a negative multiple of the gradient).

Another advantage of LSQR is that it allows (4.12) to be solved for a number of different values of  $\lambda$  simultaneously, at the cost of very little extra computation and storage.

It will be necessary to store the Jacobian  $J$  explicitly, but this can be done economically in the seismic inversion problem, since from (4.2), the first submatrix  $A$  can be stored as

$$\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix}$$
 in  $(mn)$  rows and only 4 columns. In fact, only 3 columns need be used since the first column of each  $A_i$  (containing the derivatives with respect to occurrence time) is the vector whose elements

are all 1. The other submatrix  $M$  can be stored using standard sparse-matrix storage techniques.

Our initial choice of the damping parameter  $\lambda$  will be important, because in practical terms, we are met with two conflicting requirements. Firstly, if  $\lambda$  is given too small an initial value, function evaluations will be wasted in raising it to an appropriate value. Secondly, if  $\lambda$  is chosen too large, the first few steps will be too small, and may not result in a substantial decrease in the residual sum-of-squares.

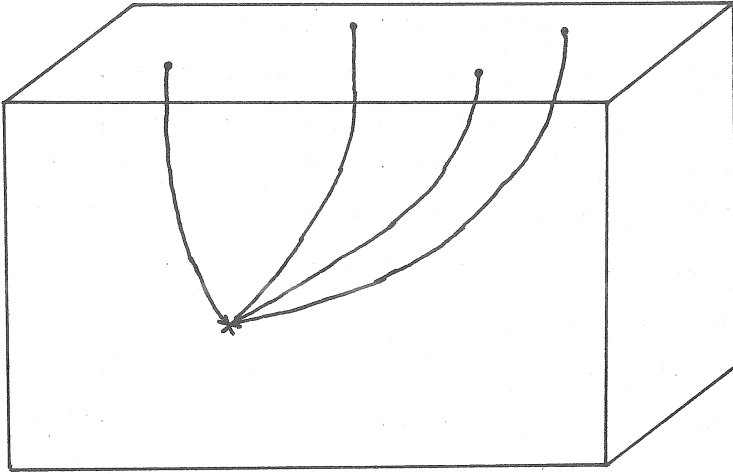


Figure 1 : Typical set of ray paths for a single event and a set of four recording stations.

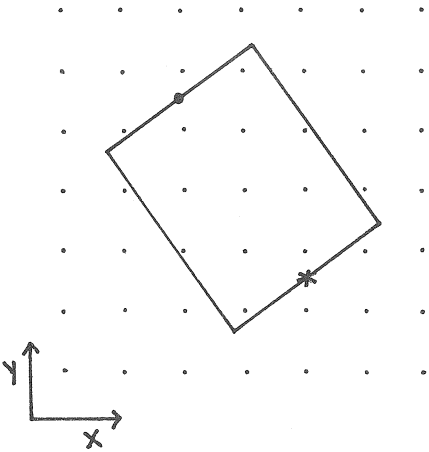


Figure 2a : ART : Velocity for each layer is averaged over a rectangular region between source and receiver.

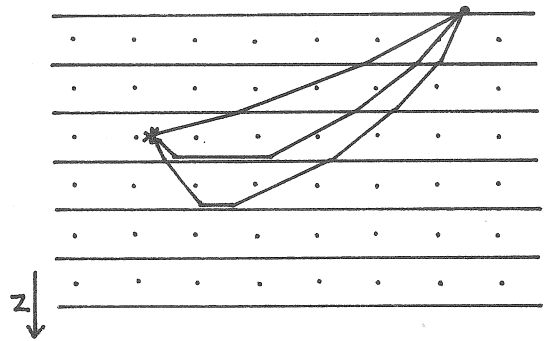


Figure 2b : ART : All ray paths appropriate to the layered model are considered.

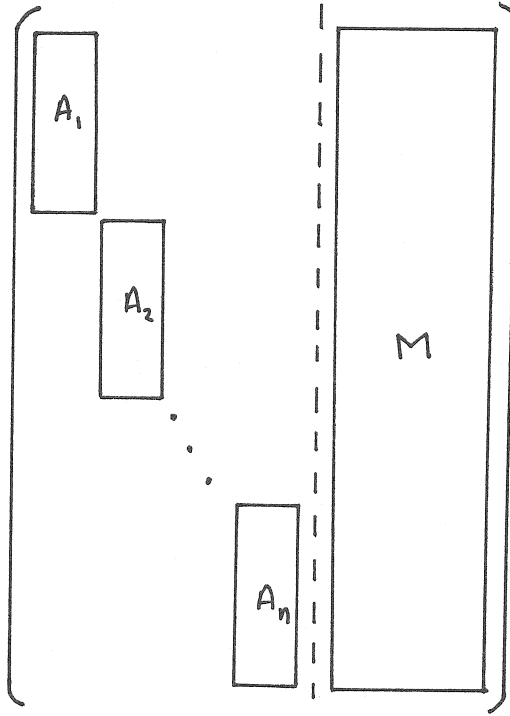


Figure 3 : Structure of the Jacobian for standard ordering of equations and variables. The submatrix  $M$  is sparse ; each of the  $A_i$ 's is  $m \times 4$  and dense.

## 5. CONCLUSION

We have attempted to describe the current modelling and numerical techniques used for an important geophysical inversion problem. Insofar as our own contribution to the least squares minimization is concerned, this paper is clearly a progress report. The new algorithm of Wright and Holt has been successfully run on small, synthetic problems, but not yet on real data. In the near future, we will have actual data available from the U.S. Geophysical Survey's seismic network in the San Francisco Bay area, by courtesy of Dr. W. Lee. A comparative study of the existing HYP02 and a version incorporating our least squares solver will then be undertaken.

## REFERENCES

- [1] K. Aki and W. Lee, "Determination of Three-Dimensional Velocity Anomalies under a Seismic Array using First p-arrival Times from Local Earthquakes", *J. Geophys. Res.* 81 (1976), 4381 - 4399.
- [2] G. Backus & F. Gilbert, "The Resolving Power of Gross Earth Data", *Geophys. J.R.Astr. Soc.* 16 (1968), 169 - 205.
- [3] G. Backus & F. Gilbert, "Constructing P-velocity models to fit Restricted Sets of Travel Time Data", *Bull. Seism. Soc. Am.* 59 (1969), 1407 - 1414.
- [4] C. Chou & J. Booker, "A Backus-Gilbert approach to Inversion of Travel Time Data for Three-Dimensional Velocity Structure", *Geophys. J.R. Astr. Soc.* 59 (1979), 325 - 344.
- [5] P. Firbas, "Inversion of Travel Time Data for Laterally Heterogeneous Velocity Structure - Linearization Approach", *Geophys. J.R. Astr. Soc.* 67 (1981), 189 - 198.

- [6] B. Hawley, A. Zandt, and R. Smith, "Simultaneous Inversion for Hypocentres and Lateral Velocity Variations : an Iterative Solution with a Layered Model", *J. Geophys. Res.* 36 (1981), 7073 - 7086.
- [7] W. Lee and S. Stewart, *Principles and Applications of Microearthquake Networks*, Academic Press, 1981.
- [8] C. Paige and M. Saunders, "LSQR : An Algorithm for Sparse Linear Equations and Sparse Least Squares", *ACM Trans. Math Softw.* 8 (1982), 43 - 71.
- [9] A. Pavlis and J. Booker, "The Mixed Discrete-Continuous Inverse Problem : Application to the Simultaneous Determination of Earthquake Hypocentres and Velocity Structure", *J. Geophys. Res.* 85 (1980), 4801 - 4810.
- [10] V. Pereyra, "Two-Point Ray Tracing in Heterogeneous Media and the Inversion of Travel Time Data", in *Computing Methods in Applied Sciences and Engineering* (R. Glowinski & J. Lions, eds.), North-Holland, 1980.
- [11] V. Pereyra, W. Lee, and H. Keller, "Solving two-point Seismic Ray-tracing Problems in a Heterogeneous Medium", *Bull. Seism. Soc. Am.* 70 (1980), 79 - 99.
- [12] C. Spencer and D. Gubbins, "Travel Time Inversion for Simultaneous Earthquake Location and Velocity Structure Determination in Laterally Varying Media", *Geophys. J.R. Astr. Soc.* 63 (1980), 95 - 116.
- [13] C. Thurber, *Earth Structure and Earthquake Locations in the Coyote Lake Area, Central California*, Ph.D. Thesis, Massachusetts Institute of Technology, 1981.

- [14] C. Thurber and W. Ellsworth, "Rapid Solution of Ray Tracing Problems in Heterogeneous Media", *Bull. Seism. Soc. Am.* 70 (1980), 1137 - 1148.
- [15] R. Wesson, "Travel Time Inversion for Laterally Inhomogeneous Crustal Velocity Models", *Bull. Seism. Soc. Am.* 61 (1971), 729 - 746.
- [16] S.J. Wright and J.N. Holt, "An Inexact Levenberg-Marquardt Method for Large Sparse Nonlinear Least Squares", submitted to *J. Austr. Math. Soc.* (1983).

Department of Mathematics,  
University of Queensland,  
St. Lucia Qld 4067,  
AUSTRALIA.