

TOMOGRAPHIC INVERSION IN REFLECTION SEISMOLOGY

P. R. Williamson

1. Reflection Seismology

In reflection seismic exploration we attempt to determine as much as possible about the nature of a region of the earth (near the surface) from the propagation of energy from some artificial source, e.g. an explosion located at or near the surface, to an array of receivers. At each receiver a record of displacement or velocity against time after source activation is produced for analysis. In most such work the source and receivers are colinear and a "seismic line" is shot by interspacing the shotpoint source location and receivers along the line (with a typical receiver separation being 25-50m.), and repeating the shooting process many times until shots have been made all along some desired length. For such a line the problem is assumed to be two-dimensional, with no significant variation perpendicular to the vertical plane containing the line in its vicinity (the interpretation being subject to correction for large systematic variations revealed by cross-lines). A typical length of seismic line in exploration work would be several km. long, made up of repeated use of an array containing 24-48 receivers with length up to ~1 km.

Energy reaches the receivers either directly (surface waves) or by reflections from or refractions by changes in the earth's properties. The direct energy is not generally very informative, but in regions of the crust where it is stratified, i.e. composed of a series of

geologically distinct layers, the reflected arrivals may in principle permit estimation of the depth and/or shape of layer boundaries, the reflection coefficients at these boundaries, and the propagation velocities within layers. The former are of particular interest since we may hope to delineate possible stratigraphic traps, e.g. for hydrocarbons, but the velocities may also be of direct interest in geological discrimination, as well as being necessary for depth calculations.

In typical conventional seismic data processing the receiver records are arranged into "common midpoint gathers" i.e. all traces which were generated with the same central point between shot and receiver are grouped together and ordered by shot-receiver offset; the traces in each gather are then effectively added together, after a time-varying correction has been made for offset, to produce a "stack section". This correction requires an estimate of the velocity field, and this is achieved assuming a locally one-dimensional structure (variation with depth only). The stack section, which is still essentially a time section, is finally converted into a representation of the earth's stratification (depth section) by "migration"; this includes correction for the effects of dipping (non-horizontal) reflectors. All this amounts to a crude but very robust inversion. However the search for increasingly subtle traps now requires more accurate, higher resolution inversion procedures that make full use of the large amounts of available data at only reasonable computational cost. It is to this end that the work described here has been directed.

2. Forward Modelling

Mathematical modelling of the above process is usually based upon the elastic wave equation, generally under the assumption that the earth is isotropic. However, in our attempt to produce a more detailed velocity field and thus improve the imaging of the reflectors, we work with a subset of the available data, the traveltimes of the particular arrivals from reflector(s) of interest. We therefore simplify the model by assuming that a ray approximation is valid, i.e. that the dominant wavelength is less than the scale lengths of structures of interest. We further reduce the scale of the problem by considering only a single layer extending from the surface down to the one reflector of interest. Initially, while developing the inversion procedure, we assume that the reflector lies parallel to the surface at a known depth, so that we concentrate only on velocity recovery. Now, because traveltimes are line integrals of the reciprocal of the velocity ("slowness"), we may adopt a tomographic approach to the inversion. [Tomography literally means the production of pictures of slices through 3-D objects, so to that extent all 2-D reflection seismology may be considered to be tomographic. However with the advent of medical X-ray tomography it has acquired a more specific connotation in high-resolution imaging.]

Initially the region under consideration is modelled as a collection of square cells of uniform properties; shot/receiver points (S/R's) are spaced uniformly along the surface at 25 m. intervals, and this (somewhat arbitrary) separation fixes the minimum cell side. Thus a model 500m. long by 250m. deep divides into a 20x10 cell grid. We shall consider rays from each S/R to itself and all others, but as there

may be more than one possible ray path for a given shot-receiver pair, only the path corresponding to the shortest travel time is considered.

The forward problem is to trace a ray from source to receiver via the reflector, and evaluate its traveltime (in fact the forward problem does not actually require the actual raypath but it is necessary for the inversion). For a general velocity field this is not easy to accomplish directly, leaving aside the fact that we do not really wish to attempt to implement the reflection process directly by the rule of equal angles of incidence and reflection on a potentially rather approximately-specified interface especially when it becomes more complicated than the flat, level surface considered here. Therefore we propose a scheme in which Fermat's Principle is directly applied. (Fermat's principle states that a physical raypath between fixed endpoints is such as to extremise (usually minimise) the traveltime.)

A series of rays is timed one-way from each S/R down to the reflector so that their arrival points more or less span its length at intervals rather less than the cell side, and the traveltimes, arrival positions and take-off angles stored. A cubic spline routine is then used to produce continuous functions with continuous first derivatives for the traveltime and take-off angle in terms of the arrival position's horizontal coordinate, x . Then to trace the first arrival ray between the i^{th} and j^{th} S/R's we add their two time functions, $t_i(x)$ and $t_j(x)$, and find that x which this sum is minimum. The spline functions now yield a traveltime and the take-off angles from which the ray may be traced (in two parts).

We should note at this point that the square-cell parameterisation proposed gives rise to the possibility of rather unphysical effects, which may cause problems, due to abrupt changes in properties at cell boundaries. These problems are to some extent reduced by the interpolation procedure in conjunction with the exercising of some care in selection of one-way rays for the interpolation, but in severe cases we adopt a more sophisticated triangular-cell discretisation, with velocities fixed at the vertices and determined elsewhere by linear interpolations over each cell. Thus velocity is continuous at cell boundaries and the effects of division into cells are reduced (but still apparent due to gradient discontinuities).

3. Inversion: Preliminaries

We may formulate the forward-modelling process by the equation

$$\mathbf{d} = \mathbf{g}(\mathbf{s}) ,$$

\mathbf{g} being the vector functional generating the traveltimes for the velocity model \mathbf{s} . Specifically

$$g_i = \int_{R_i} s \, d\ell$$

where R_i denotes the integration to running over the i^{th} raypath, itself dependent upon \mathbf{s} . In the case of the square-cell parameterisation the integration becomes a sum :

$$g_i(\mathbf{s}) = \sum_j \ell_{ij} s_j$$

where ℓ_{ij} is the length of the element of the i^{th} ray in the j^{th} cell, and s_j is the velocity in the cell. Since ℓ_{ij} depends on s_j , the problem is non-linear (and in fact the form of the dependence means that g is only once differentiable) but it may be locally linearized. In particular we approximate the equation

$$\mathbf{d} = \mathbf{g}(\mathbf{s})$$

where \mathbf{d} is the set of observed times, by expanding $\mathbf{g}(\mathbf{s})$ to first order about some starting model \mathbf{s}_0 as

$$\mathbf{g}(\mathbf{s}) = \mathbf{g}(\mathbf{s}_0) + \mathbf{G}(\mathbf{s}-\mathbf{s}_0) + O(|\mathbf{s}-\mathbf{s}_0|^2)$$

where $G_{ij} = \partial g_i / \partial s_j$ is a component of the Fréchet derivative of g_i . This gives the linear system

$$\mathbf{d}_0 - \mathbf{g}(\mathbf{s}_0) = \mathbf{G}(\mathbf{s}-\mathbf{s}_0) \quad (1)$$

$$\text{so} \quad \mathbf{s} = \mathbf{s}_0 + \mathbf{G}^\dagger(\mathbf{d}_0 - \mathbf{g}(\mathbf{s}_0))$$

where \mathbf{G}^\dagger is some suitable generalised inverse or damped least squares inverse of \mathbf{G} . [Note that in the square-cell case it can be shown, using Fermat's principle, that $G_{ij} = \ell_{ij}$.]

Unfortunately \mathbf{G} quickly becomes very large as the size of the model increases, and so for this application direct inversion is impractical (as well as inaccurate if the non-linearity is significant and \mathbf{s}_0 is not close to \mathbf{s}). We therefore favour a minimisation

approach and, following many authors (e.g. Tarantola and Valette [4], consider a least square statistic in which "data" and "model" are placed on an equal footing. In particular we seek to minimize the statistic

$$S = \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_0)$$

over the parameter vector $\mathbf{x} = \begin{pmatrix} \mathbf{d} \\ \mathbf{s} \end{pmatrix}$. Here \mathbf{C}_x is an a priori covariance matrix of the form

$$\mathbf{C}_x = \begin{bmatrix} \mathbf{C}_{dd} & \mathbf{C}_{ds} \\ \mathbf{C}_{sd} & \mathbf{C}_{ss} \end{bmatrix}$$

\mathbf{C}_{dd} is the expected covariance matrix for the observed travel times; \mathbf{C}_{ss} is that for the velocity parameters, reflecting the scale length of expected variations. As there appears to be little justification for assuming \mathbf{C}_{ds} ($= \mathbf{C}_{sd}^T$) to be non-zero, these blocks are discarded, giving the familiar form

$$(2) \quad S = \frac{1}{2} \left\{ (\mathbf{d} - \mathbf{d}_0)^T \mathbf{C}_{dd}^{-1} (\mathbf{d} - \mathbf{d}_0) + (\mathbf{s} - \mathbf{s}_0)^T \mathbf{C}_{ss}^{-1} (\mathbf{s} - \mathbf{s}_0) \right\}$$

The above formalism also places \mathbf{d}_0 , the observed times, and \mathbf{s}_0 , the starting or prior model, on an equal footing, so that we may expect the minimum to locate a compromise between \mathbf{s}_0 and the model implied by \mathbf{d}_0 (just as a generalized inverse would normally compromise between apparently inconsistent elements of \mathbf{d}_0).

Minimisation of the S statistic given above subject to the constraints $\mathbf{d} = \mathbf{g}(\mathbf{s})$ may be achieved using Lagrange multipliers, giving

$$(3) \quad \mathbf{s} = \mathbf{s}_0 + \mathbf{C}_{ss} \mathbf{G}^T \mathbf{C}_{dd}^{-1} (\mathbf{d}_0 - \mathbf{g}(\mathbf{s})) .$$

Typically we simplify by taking $\mathbf{C}_{dd} = \sigma_d^2 \mathbf{I}$ and $\mathbf{C}_{ss} = \sigma_s^2 \mathbf{R}_s$, where \mathbf{R}_s is a "smoothing matrix". Furthermore if, as is generally the case, \mathbf{s}_0 merely represents a starting model based on vague prejudice rather than definite a priori information, we use the following iterative algorithm :

$$\mathbf{s}_{n+1} = \mathbf{s}_n + \mu \mathbf{R}_s \mathbf{G}^T (\mathbf{d}_0 - \mathbf{g}(\mathbf{s}_n)) \quad (\mu = \sigma_s^2 / \sigma_d^2)$$

i.e. we update the \mathbf{s}_0 used in \mathbf{S} after each step. This is effectively a non-linear generalisation of Landweber iteration so the model term in \mathbf{S} acts merely to limit the step length in the direction given by the solution of the linear problem. In the purely linear case when $\mathbf{g}(\mathbf{s}) = \mathbf{G}\mathbf{s}$, and taking $\mathbf{R}_s = \mathbf{I}$ (this formalism can always be achieved by appropriate 'coordinate transformations' if \mathbf{R}_m is non-singular), it can be shown that provided $\mu < 2/\sigma_{\max}^2$, where σ_{\max} is the largest singular value of \mathbf{G} , such an iterative process will eventually converge to a solution such as that in Eq.1. above. Indeed, if we write

$$\mathbf{d}_0 - \mathbf{g}(\mathbf{s}_0) = \mathbf{d}_0 - \mathbf{G}\mathbf{s}_0 = \sum_{i=1}^N a_i \mathbf{u}^{(i)}$$

with N the number of data, and the $\mathbf{u}^{(i)}$ being the data-space singular vectors of \mathbf{G} with corresponding model-space singular vectors $\mathbf{v}^{(i)}$ and singular values σ_i , then after n iterations we get

$$\mathbf{s}_n = \mathbf{s}_0 + \sum_{i=1}^p a_i (1 - (1 - \mu \sigma_i^2)^n) / \sigma_i \mathbf{v}^{(i)}$$

Here p is the rank of G .

The factor $(1-(1-\mu\sigma_1^2)^n)/\sigma_1$ is a damped approximation to the factor $1/\sigma_1$ appearing in the generalised inverse G^\dagger ; it prevents excessive noise amplification in small- σ components. After a reasonable number of iterations with a sensible choice of μ (left open by the uncertainty of σ_s^2), the solution is similar in form to the damped least squares inverse of Marquardt [2]. Thus the number of iterations directly parameterises the trade-off between resolution and parameter variance (see e.g. Backus and Gilbert [1]).

Of course in our non-linear problem this result will not carry over quantitatively, but we may expect that if the non-linearity is mild enough then the behaviour is qualitatively similar. Therefore we have an inversion scheme which is computationally facile and as robust as we choose to make it. Tests on synthetic data show that it is successful in dealing with the problem under consideration, i.e. the imaging of vertically-varying velocity fields above a reflector of known shape and position, and will also give some idea of lateral variations if a sufficiently wide range of offsets is used. A similar algorithm also succeeds with the complementary problem of ascertaining reflector depth and shape beneath a known (uniform) velocity field despite the increased non-linearity and enhanced effects of the non-differentiability mentioned earlier. The reflector is characterised by depth points with a 25m. horizontal separation (i.e. between each column of cells in the square-cell velocity model) joined by straight line segments; the spline interpolation is relied upon to smooth out the effects of the gradient discontinuities of this discretisation.

4. Inversion: Full Problem

Unfortunately naive application of the above inverse scheme to the problem of simultaneous recovery of the velocity field and the depth/shape of the underlying reflector proves less successful even after we have attempted to deal with the problem of having two sorts of parameters which are dimensionally inhomogeneous by dividing each type by a reference value. (This renders all parameters dimensionless and of roughly equal amplitude if we choose actual reference values and is the simplest, but not necessarily the best, way of dealing with what I shall refer to as the "unit problem" in further discussions below.) In fact we observe in a typical attempted reconstruction that the correct systematic variation of the velocity field is only recovered above roughly half-way up the model, the recovery point is higher at the ends and deeper in the middle, and the range of variation is generally somewhat reduced. Below half-way there appears to be little or no recovery of the velocity field at all. Furthermore the estimated reflector position corresponds to the best data fit obtainable with the reconstructed velocity field, i.e. the estimated reflector attempts to compensate for failings in the estimated velocity field.

The reason for this appears to be that velocities can only be determined for those cells transversed by rays whose reflection points are fairly widely spread over the reflector, and thus that are not sensitive to the local behaviour of the interface. On the other hand a cell near the reflector will typically be transversed by rays which only meet a small region of the reflector in the neighbourhood of the cell, with the result that errors in a reflector depth point may be more or less cancelled by appropriate 'errors' in the estimates of velocity in

cells immediately above it. To put this another way, the fact that similar changes in the model-generated data $g(s)$ can be produced by changes in either the velocity or depth parameters raises the possibility that in the parameter space the set of minima for S is not a single point but a connected region. In the linearization this nonuniqueness manifests as the appearance of nearly zero singular values that represent local depth/velocity field trade-offs. The resultant indeterminacy is more serious than that encountered in problems that estimate the velocity field alone since the usual methods of dealing with it (choice of starting model, smoothing, etc.) do not resolve it. Also the actual effect on the solution model of incorrect estimation of such a component is more significant than the effects induced by such errors during modelling of the velocity field alone; the latter usually consist only of high frequency noise or "edge effects".

The method of dealing with this indeterminacy problem which we propose is to run the inversion in several stages. In the first stage the parameterisation is one large cell with a straight reflector defined by its depth at each end. At each successive stage the parameterisation is made finer (with cell sides decreased by a factor typically in the range 2-4). until the original cell size is reached or until no further significant improvement in fitting the data is obtained. This approach offers the hope that, even if there are near zero singular components in the final model linearization, the previous, better-determined, stages will at least have produced good starting models. Thus we define progressively expanding subspaces of the final-parameterisation model space within which the inversions are sought.

However we now run up against the chief problem of descent methods: their slow convergence rates. Previously, in the velocity-only problem, we were not bothered by this since:

(a) Although it is slow in terms of numbers of iterations in achieving (near-)complete convergence, as each iteration is itself so cheap it is still quicker than direct inversion in achieving a reasonable level of convergence for large problems.

(b) Its failure to recover small singular components quickly is perceived as an advantage considering the likely effects of noise and non-linearity on such components.

However, in the velocity/depth problem the data misfit in early and intermediate stages is likely to be more strongly affected by misparameterisation than by noise and so we look for the maximum rate of convergence that is reasonably possible. Therefore we seek something quicker than our basic descent scheme which retains, as far as possible, its advantages (stability and overall computational facility). The result is the following family of subspace search methods.

Suppose that \mathbf{s} is the vector of velocity parameters and \mathbf{b} is the vector of reflector parameters, so that

$$\mathbf{m} = \begin{bmatrix} \mathbf{b} \\ \mathbf{s} \end{bmatrix}$$

is the full vector of model parameters. We first consider a local quadratic approximation S^Q to S expanded about \mathbf{m}_0

$$S^Q(\mathbf{m}_0 + \delta\mathbf{m}) = S(\mathbf{m}_0) + \delta\mathbf{m}^T \boldsymbol{\theta} + \frac{1}{2} \delta\mathbf{m}^T H \delta\mathbf{m}$$

where $\boldsymbol{\theta} = \nabla S(\mathbf{m}_0)$ and $H = \nabla^2 S(\mathbf{m}_0)$. Straightforward minimisation of S^Q gives.

$$H\delta\mathbf{m} = -\boldsymbol{\theta}$$

yielding the standard least squares inverse.

In the expanding subspace algorithm we restrict $\delta\mathbf{m}$ to an n -dimensional parameter subspace defined by the n vectors, $\mathbf{a}^{(i)}$, where $n \ll M$ the number of model parameters. Thus

$$\delta\mathbf{m} = \sum_{i=1}^n \alpha_i \mathbf{a}^{(i)}.$$

Minimisation of S^Q is now carried out with respect to the α_i , and writing A for the matrix whose i^{th} column is $\mathbf{a}^{(i)}$, (i.e. $A \equiv [\mathbf{a}^{(1)}, \mathbf{a}^{(2)} \dots \mathbf{a}^{(n)}]$), and $\boldsymbol{\alpha}$ for the vector whose j^{th} component is α_j , we get

$$\delta\mathbf{m} = A\boldsymbol{\alpha} \quad \text{where} \quad A^T H A \boldsymbol{\alpha} = -A^T \boldsymbol{\theta}$$

i.e. a "projected" set of equations. This has the advantage that the matrix $A^T H A$ is much smaller and, if the $\mathbf{a}^{(i)}$ are well-chosen, better-conditioned than H . The system of $\mathbf{a}^{(i)}$ suggested by Shilling [3] seems as good as, if not better than, any other

1. $\mathbf{a}^{(1)} = \beta_1 \boldsymbol{\theta}$
2. $\mathbf{a}^{(i)} = \beta_i \mathbf{H}^i \boldsymbol{\theta}$, $i \geq 1$.

where the β_i are normalizing constants. If $n = 1$ then this method reduces to steepest descent. If n is chosen so that $n > 1$ iterative application of this higher-order approximation scheme results in more rapid (in terms of both number of iteration and CPU time) convergence towards an acceptable solution.

We now redefine our statistic S , simplifying slightly and allowing for both a priori knowledge and further step-length limitation, as:

$$S = \frac{1}{2} \{ |\mathbf{d}_0 - \mathbf{g}(\mathbf{m})|^2 + \zeta^2 |\mathbf{m} - \mathbf{m}_c|^2 + \sigma^2 |\mathbf{m} - \mathbf{m}_0|^2 \}$$

where \mathbf{m}_c is the current model and, \mathbf{m}_0 the a priori one. Now σ^2 controls damping via the a priori information and ζ^2 is an additional step-length restraint. We have temporarily neglected the possibility of smoothing since: (a) it would not really be appropriate for the larger cells in all but the last one or two stages, and (b) it requires \mathbf{R}_m^{-1} for strict implementation. (To avoid inversion of \mathbf{R}_m we could apply ad hoc smoothing near the end of each step.) Now

$$\begin{aligned} \boldsymbol{\theta}(\mathbf{m}) &= -\mathbf{G}^T(\mathbf{d}_0 - \mathbf{g}(\mathbf{m})) + \sigma^2(\mathbf{m} - \mathbf{m}_0) \\ \mathbf{H}(\mathbf{m}) &= \mathbf{G}^T \mathbf{G} - \nabla \cdot \mathbf{G}^T(\mathbf{d}_0 - \mathbf{g}(\mathbf{m})) + (\rho^2 + \sigma^2) \mathbf{I} . \end{aligned}$$

Unfortunately $\nabla \cdot G^T$ is non-existent in many places (since G varies discontinuously irrespective of the parameterisation), and even when it does exist it is well-nigh impossible to calculate exactly. Therefore we have approximated those parts of it that we can estimate and discarded the rest. Note that discarding it altogether would amount to merely using the linear approximation to $g(\mathbf{m})$.

In developing our subspace approach we have temporarily ignored the problem of choice of units mentioned above, i.e. that the direction of θ is dependent upon the effective choice of relative unit magnitudes of the two types of parameter :

$$\theta = \begin{bmatrix} \theta_s \\ \theta_b \end{bmatrix} = \begin{bmatrix} \nabla_s S \\ \nabla_b S \end{bmatrix}$$

so if we change units as follows : $s'_i = \alpha s_i$, $b'_j = \beta b_j$ then in the new units

$$\theta'(s', b') = \begin{bmatrix} \frac{1}{\alpha} \theta_s(s', b') \\ \frac{1}{\beta} \theta_b(s', b') \end{bmatrix}$$

so that in general θ' is not colinear with θ . Similarly

$$H = \begin{bmatrix} H_s & H_{sb} \\ H_{bs} & H_{bb} \end{bmatrix}$$

so the relative magnitudes of the blocks will vary with unit choice, and therefore the recommended step. However, since the various possible

θ 's all lie in a plane defined by the model vectors

$$\frac{1}{|\theta_s|} \begin{bmatrix} \theta_s \\ 0^s \end{bmatrix} \quad \text{and} \quad \frac{1}{|\theta_b|} \begin{bmatrix} 0 \\ \theta_b \end{bmatrix},$$

the problem is solved for the descent method by connecting it into a two-dimensional search in this plane, using these as the search vectors. The result finds the minimum of S^Q in this plane, i.e. effectively chooses units so that the contours of S^Q in this plane are circular and the descent direction points to the minimum. Similarly we may conduct a unit-independent second order search by adding four directions corresponding to $H\theta$:

$$\begin{bmatrix} H_{ss} & \theta_s \\ 0 & \end{bmatrix}, \begin{bmatrix} 0 \\ H_{bb} & \theta_b \end{bmatrix}, \begin{bmatrix} H_{sb} & \theta_b \\ 0 & \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ H_{bs} & \theta_s \end{bmatrix} \quad (\text{appropriately orthonormalised})$$

The scheme outlined above has shown considerable potential for improvement over the fixed-cell descent scheme in synthetic tests, but the solution shows some dependence on the actual configuration of the cells in the intermediate stages, which is clearly undesirable. Work is in progress to attempt to overcome or avoid this problem: one possible idea is to use some kind of variable scale-length smoothing, but as yet no tests have been carried out on this.

REFERENCES

- [1] Backus, G. and Gilbert F. *Uniqueness in the inversion of inaccurate gross data*. Phil. Trans. R. Soc. Lond. A, 266, 123-192, 1970.
- [2] Marquardt, D.W. *An algorithm for least-squares estimation of non-linear parameters*. J. Soc. Ind. Appl. Maths., 11, 431-441, 1963.

- [3] Shilling, J. *Topics in Data Analysis*. Postgraduate Lecture Course at the Dept. of Appl. Maths. and Theor. Phys., Univ. of Cambridge. 1982.
- [4] Tarantola A. and Valette B. *Generalised nonlinear inverse problems solved using the least-squares criterion*. Rev. Geophys. and Space Phys., **20**, 219-232, 1982.

Research School of Earth Sciences
Institute of Advanced Studies
Australian National University
G.P.O. Box 4, Canberra, 2601
Australia