

Traveltime tomography in anisotropic media—II. Application.

R. G. Pratt and C. H. Chapman*

Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada

Accepted 1991 August 28. Received 1991 August 28; in original form 1991 January 4

SUMMARY

Cross-borehole seismic data have traditionally been analysed by inverting the arrival times for velocity structure (traveltime tomography). The presence of anisotropy requires that tomographic methods be generalized to account for anisotropy. This generalization allows geological structure to be correctly imaged and allows the anisotropy to be evaluated. In a companion paper we developed linear systems for 2-D traveltime tomography in anisotropic media. In this paper we analyse the properties of the linear system for quasi-compressional waves and invert both synthetic and real data. Solutions to the linear systems consist of estimates of the spatial distributions of five parameters, each corresponding to a linear combination of a small subset of the 21 elastic, anisotropic velocity parameters. The parameters describe the arrival times in the presence of weak anisotropy with arbitrary symmetries. However, these parameters do not, in general, describe the full nature of the anisotropy. The parameters must be further interpreted using additional information on the symmetry system. In the examples in this paper we assume transverse isotropy (TI) in order to interpret our inversions, but it should be noted that this final interpretation could be reformulated in more general terms.

The singular value decomposition of the linear system for traveltime tomography in anisotropic media reveals the (expected) ill-conditioning of these systems. As in isotropic tomography, ill-conditioning arises due to the limited directional coverage that can be achieved when sources and receivers are located in vertical boreholes. In contrast to isotropic tomography, the scalelength of the parametrization controls the nature of the parameter space eigenvectors: with a coarse grid all five parameters are required to model the data; with a fine grid some of the parameters appear only in the null space.

The linear systems must be regularized using external, *a priori* information. An important regularization is the expectation that the elastic properties vary smoothly (an ad hoc recognition of the insensitivity of the arrival times to the fine-grained properties of the medium). The expectation of smoothness is incorporated by using a regularization matrix that penalizes rough solutions using finite difference penalty terms. The roughness penalty sufficiently constrains the solutions to allow the smooth eigenvectors in the null space of the unconstrained problem to contribute to the solutions. Hence, the spatial distribution of all five parameters is recovered. The level of regularization required is difficult to estimate; we advocate the analysis of a suite of solutions. Plots of the solution roughness against the data residuals can be used to find ‘knee points’, but for the fine tuning of the regularization one has little recourse but to examine a suite of images and use geological plausibility as an additional criterion.

The application of the regularized numerical scheme to the synthetic data reveals that the roughness penalty should include terms that penalize high gradients in addition to penalizing high second derivatives. Only when this constraint was included were the features of the original model recovered. The inversions of the

*Now at: Schlumberger Cambridge Research, PO Box 153, Cambridge, CB3 0HG, UK.

field data yield good images of the expected stratigraphy and confirm previous estimates of the magnitude of the anisotropy and the orientation of the symmetry axis. The solutions further indicate an increase in anisotropy from the top to the bottom of the survey region that was not previously detected.

Key words: anisotropy, regularization, travelt ime tomography.

1 INTRODUCTION

Since at least 1971, cross-borehole seismic surveying has been used as a remote sensing tool in the exploration and exploitation of hydrocarbon reserves (Bois *et al.* 1971, 1972). The same geophysical technique has also been used extensively in mining applications (Mason 1981; Leung & Downey 1988), in waste disposal site evaluation (Wong, Hurley & West 1983; Paulsson, Cook & McEvilly 1985; Bregman, Bailey & Chapman 1989a) and in civil engineering site surveying (LaPorte *et al.* 1973; Saito *et al.* 1988; Wright *et al.* 1988). The objectives of crosshole seismic surveys include mapping the distributions of seismic velocities, locating and mapping fracture distributions, stratigraphic and structural imaging, and mapping of stress regimes. Further objectives include the mapping of velocity alterations induced by industrial processes, such as steam injection or *in situ* combustion (for example, Bregman, Hurley & West 1989c). Geophysical tomography, in which arrival times are inverted in order to provide an estimate of the distribution of seismic velocities, has emerged as a promising data processing technique capable of meeting these objectives.

There are three complicating factors that make geophysical tomography more difficult than its algorithmic predecessor, medical tomography. The first two factors are the awkward geometric restrictions that borehole geometries impose and the large variations in seismic velocities that are encountered in the Earth. These two factors have been dealt with extensively in the literature. The geometric restrictions imply a coverage of the subsurface by ray paths that is deficient in ray angles. This aperture limitation results in data that are not sufficient to uniquely specify the velocities. Regularization of the equation systems using appropriate constraints (for example, damped least squares) is essential. The second factor, large velocity variations, introduces non-linearities into the equation system, since the solution is dependent upon the ray paths, which are themselves determined by the velocities (i.e. the unknowns). The non-linearity of the equation system must be approached by relaxation, solving iteratively in turn for the ray paths and for the velocities (Cottin *et al.* 1986; Bregman *et al.* 1989a). However, in some cases the velocity variations may be so large that the geometric ray approximation itself is inadequate. In these cases a more complete modelling technique, such as the method of finite differences, is required to generate the partial derivatives required (Luo & Schuster 1989).

Anisotropy is the third complication. Surprisingly, little research in geophysical tomography exists that admits this complication or that tries to address the problem of travelt ime inversion in the presence of anisotropy. Yet an extensive literature exists that documents the presence of

seismic anisotropy in the Earth's crust (see Crampin 1987, for a review). The major evidence of crustal anisotropy is the observation of shear wave splitting. In an exploration context, shear wave splitting has been observed recently in a multitude of case studies using multicomponent surface recorded data (for example, Lynn & Thomsen 1990; Frasier & Winterstein 1990). Other direct evidence of anisotropy is the measurement of velocity anisotropy in laboratory measurements. Thomsen (1986) has summarized a large number of laboratory core sample measurements.

Shear wave splitting can be observed in the presence of very weak anisotropy, as the splitting depends on velocity differences along the same propagation directions. The directional dependence of velocities in conventional (non-tomographic) studies is difficult to separate from the effects of heterogeneity. In tomographic studies, the wide directional coverage allows quasi-compressional (*qP*) or quasi-shear (*qS*) anisotropy to be observed directly, rather than inferred from shear wave splitting and interference effects.

The geometric ray (high frequency) approximation is usually used in tomography to model arrival times. Since the rays will always inadequately sample the continuous distribution of the elastic properties of the Earth, one could expect severe alias errors. Aliasing is not generally considered to be a problem in isotropic tomography because seismic waves are band-limited, and arrival times are often insensitive to the fine-grained properties of the medium. However, when fine-grained properties are spatially correlated (for example in finely layered media), the long-wavelength properties are anisotropic. We can only safely assume that the arrival times are unaffected by the fine-grained properties of the medium if we include terms in the inversions that model the long-wavelength anisotropy.

Anisotropy appears to be ubiquitous, yet tomography has been applied widely (and often successfully) without accounting for anisotropy. To some extent this can be understood in terms of scalelengths, since the Earth is heterogeneous over a wide range of characteristic lengths. At seismic wavelengths, although layered sediments, fractured rocks and sediments containing amounts of oriented anisotropic mineral constituents will behave anisotropically, disturbances of the symmetries will weaken the anisotropy (Kerner 1989). Formations have different amounts of anisotropies at different scalelengths and, therefore, the assumption of isotropy may be justified when the anisotropy is negligible at the relevant scale. Nonetheless, several tomography data sets exist, gathered by various groups, in which anisotropy does play a significant role. However, until now, no rigorously developed formalism for tomography existed that could be used to account for the anisotropy. Instead, the approach taken has been to use crude methods to estimate the effect

of the anisotropy and remove it from the data before proceeding under the assumption of isotropy. McCann *et al.* (1989) have published an example that shows the efficacy of this approach. Others have presented similar approaches (for example, Peterson *et al.* 1990).

This paper is the second of two papers that present and apply a formalism for the application of traveltimes tomography in anisotropic media. In our first paper (Chapman & Pratt 1992, hereafter referred to as Paper I) we derived analytical expressions for traveltimes perturbations due to anisotropy. By assuming weak anisotropy and by assuming that sources and receivers were confined to a single imaging plane, we showed that the traveltimes were sensitive only to a limited subset of the elastic tensor components. We gave explicit expressions for the partial derivatives using both straight and curved rays. Using the partial derivatives, appropriate linear systems for traveltimes inversion in the presence of anisotropy can be formulated. The solution of the resultant linear systems is an estimate of the distribution (throughout the interwell region) of the elastic tensor components.

An important aspect of the approach used in Paper I is that the linear systems do not assume any particular symmetries. However, in order to extract geological parameters from the solutions (i.e. to interpret the results), it is necessary to assume a symmetry system. In an appendix of Paper I we provided a transformation for mapping transverse isotropy (TI) parameters to the tensor components (the 'global parameters'). For *qP*-waves, there are the same number of global parameters as TI parameters (five). However, the two parameter sets do not span the same space. The inverse transformation, obtaining the TI parameters from the global parameters, requires some additional information. For example, where the symmetry axis is nearly vertical, additional information on the azimuth of the symmetry axis must be provided. Both layered media and media with a single set of oriented cracks have TI symmetries, so that the TI to global transformations will be appropriate for these special cases. For the purposes of the examples in this paper we assume TI symmetries; the symmetry axes are assumed to be nearly vertical (but need not be exactly vertical). For more general symmetry systems, appropriate transformations could also be formulated; more *a priori* information on the symmetry axes would then be required.

In this paper we investigate the application of the results of Paper I to *qP*-wave traveltimes tomography. We analyse the numerical properties of these systems, and show that, as in isotropic problems, a crosshole geometry leads to ill-conditioned linear equation systems. Just as regularization is required for isotropic tomography, we require adequate regularization to solve the anisotropy problem, which is even less well posed. We show that damped least squares is inadequate, and that it is essential to apply smoothing constraints [we use constraints similar to those applied in the 1-D electromagnetic problem by Constable, Parker & Constable (1987)].

In order to demonstrate the application of the numerical scheme, we investigate data from a simulation problem (synthetic data) and data from a field crosshole experiment. The field data were obtained in near-surface sediments known to be anisotropic [the data are those used in the

study by McCann *et al.* (1989), described above]. In near-surface sedimentary environments, the intrinsic anisotropy of clay materials is combined with the anisotropic behaviour of finely layered sediments. This leads to a symmetry in the velocities about approximately vertical axes. The field data we examine support this conjecture about the nature of the anisotropy in these sediments.

2 METHODOLOGY

2.1 *P*-wave perturbation equations

In this section we briefly review the formulation given in Paper I for *qP*-wave traveltimes tomography in anisotropic media. In the majority of published case studies, tomography has been carried out using the first arrivals, generally assumed to be *P*- or *qP*-waves. First arrivals are far easier to use than the later arriving shear waves. In anisotropic media, shear wave tomography requires that the shear waves be correctly identified (as fast and slow split shear waves) and correctly picked. It is difficult to pick any arrival, but shear waves are far more difficult to accurately separate from other arrivals, with which they interfere. Here we do not consider the shear wave tomography problem. This allows us to use a simpler (although less general) notation than in Paper I.

In Paper I we gave perturbation methods for the approximate computation of arrival times in anisotropic media. These methods, first developed by Červený & Jech (1982) and by Jech & Pšenčík (1989), relate traveltimes perturbations to perturbations of the components of the elastic tensor. This formulation allows the anisotropy to be perfectly general. The perturbation approach relies on the validity of Fermat's principle, which allows the perturbation in the traveltimes due to small perturbations in the ray trajectory to be ignored.

It is extremely convenient to use isotropic unperturbed media, and to consider anisotropic effects to be due to weak perturbations of isotropic systems. This assumption allows one to proceed without developing new methods for ray tracing in anisotropic media—all rays are traced in unperturbed isotropic media. In view of the past success in assuming isotropy, we consider this to be a satisfactory approximation. Iterative inversions yield estimates of the corrections to the elastic tensor, some of which are required due to errors in the isotropic velocities, and some of which are required due to ignoring the anisotropy.

The wave propagation properties of the medium are represented by a spatial distribution of the elastic tensor, a_{ijkl} . The tensor a_{ijkl} contains the 81 density normalized media stiffnesses, $a_{ijkl} = c_{ijkl}/\rho$, of which only 18 are independent (Federov 1968). We proceeded in Paper I by introducing small perturbations, δa_{ijkl} . For the case in which the original elastic tensor distribution is everywhere isotropic, we obtained the expression

$$\delta T = -\frac{1}{2} \int_{\mathcal{L}} \alpha^2 p_i p_l p_j p_k \delta a_{ijkl} dT, \quad (1)$$

[equation (32) in Paper I], which relates the traveltimes perturbation for *qP*-waves, δT , to an integration over the ray path, \mathcal{L} , of a weighted sum of the components of δa_{ijkl} . The weighting in the terms of the integrand is determined

by the components of the slowness vector, \mathbf{p} , in the unperturbed, isotropic media. In equation (1) the integration measure, dT , is an incremental measure of time in the original (unperturbed) medium and α is the isotropic velocity distribution in the unperturbed medium.

In the original, isotropic media the slowness vector can be expressed as $\mathbf{p} = \hat{\mathbf{p}}/\alpha$, where $\hat{\mathbf{p}}$ is the unit vector tangent to the ray trajectory. Furthermore, along the ray path,

$$\frac{dT}{dl} = \frac{1}{\alpha}, \quad (2)$$

where dl is an increment of length along the ray path. Equation (1) can be rewritten as

$$\delta T = -\frac{1}{2} \int_{\mathcal{P}} \frac{1}{\alpha^3} \hat{p}_i \hat{p}_j \hat{p}_k \delta a_{ijkl} dl, \quad (3)$$

which shows clearly how the weighting terms in the integrand are determined by the ray direction, $\hat{\mathbf{p}}$.

In Paper I we made the further assumption that data are collected for source–receiver pairs that are all co-planar. Small deviations from the plane can be corrected for by projection onto the plane with appropriate travelt ime corrections. Once this is assumed, the coordinate system can be oriented so that the x_2 coordinate is normal to the plane. In isotropic 2-D media rays are confined to the (x_1, x_3) plane and, hence, \hat{p}_2 is always zero. This reduces the number of independent terms contributing to the integral in equation (3) to 5:

$$\begin{aligned} \delta T = & -\frac{1}{2} \int \frac{1}{\alpha^3} \hat{p}_1^4 \delta q_1 dl - \frac{1}{2} \int \frac{1}{\alpha^3} \hat{p}_1^3 \hat{p}_3 \delta q_2 dl \\ & - \frac{1}{2} \int \frac{1}{\alpha^3} \hat{p}_1^2 \hat{p}_3^2 \delta q_3 dl - \frac{1}{2} \int \frac{1}{\alpha^3} \hat{p}_1 \hat{p}_3^3 \delta q_4 dl \\ & - \frac{1}{2} \int \frac{1}{\alpha^3} \hat{p}_3^4 \delta q_5 dl, \end{aligned} \quad (4)$$

where

$$\begin{aligned} \delta q_1 = \delta a_{1111}, \quad \delta q_2 = 4\delta a_{1131}, \quad \delta q_3 = 2\delta a_{1133} + 4\delta a_{3131}, \\ \delta q_4 = 4\delta a_{3331}, \quad \delta q_5 = \delta a_{3333}. \end{aligned} \quad (5)$$

To first order, varying any of the other elastic parameters other than the six that appear above will not affect the qP arrival times. Of the six parameters, a_{1133} and a_{3131} appear only in combination. To first order, 2-D tomographic imaging can ignore variations in the other parameters. If one is primarily interested in correcting tomographic velocity estimates for anisotropy, this result is encouraging since only five parameters need be included. However, if one is interested in characterizing the symmetry system or otherwise quantifying the anisotropy and its distribution, this result indicates (as one would expect) that there is an insufficient amount of information in 2-D qP -wave arrival times to characterize fully the elastic tensor.

The five q parameters describe the intersection of the wavefront with the plane of the survey exactly, but there exists an infinite set of media that have the same wavefronts in the plane of the survey. The result of an inversion for these five parameters can be interpreted only in terms of other *a priori* information. Knowledge of the symmetry

system of the rock properties can provide this information. We return to this aspect of the methodology in Section 2.3.

2.2 Piecewise homogeneous media and bent ray tomography

In this section, equations for compressional wave tomography in anisotropic media are formulated in terms of media that are piecewise homogeneous. The assumption of piecewise homogeneity implies that rays consist of straight line segments and bend only at interfaces (the ‘bent ray’ method). It is clear from the literature (Bording *et al.* 1987; Dyer & Worthington 1988a, b; Worthington *et al.* 1989) that bent ray tomography is of considerable utility. The advantage of the bent ray formulation is the ease with which it can be programmed and the speed at which it will run. The interfaces at which rays are to be bent are determined by interpretation of straight ray tomograms. For the purposes of ray tracing, the interfaces should approximately define geological units, or regions. This allows a skilled interpreter to incorporate external knowledge (for example, geological plausibility) and thus to constrain the method. Below, we give a brief review of the bent ray method in isotropic media in order to introduce its use in anisotropic media.

2.2.1 Bent ray tomography in isotropic media

In isotropic media the linear equations for tomography relate travelt ime perturbations to perturbations in the media slownesses. The linear equation system is

$$\delta T_k = F_{ki} \delta U_i, \quad \text{or} \quad \delta \mathbf{T} = \mathbf{F} \delta \mathbf{U}, \quad (6)$$

where T_k is the k th arrival time and U_i is the isotropic slowness at the i th location in the model. The δT and δU variables are the perturbations to T and U . F_{ki} is a large, sparse matrix (the Fréchet matrix) containing the partial derivatives, $\partial T_k / \partial U_i$. In order to compute the partial derivatives, rays must be traced through the starting model of the geological regions. In piecewise homogeneous models the partial derivative of the k th time with respect to changes in the i th slowness parameter is simply the length of the ray segment, Δl_{ki} , of the k th ray in the i th region. Following the computation of the partial derivatives, one can proceed to solve equation (6) for the velocity perturbations using an appropriate numerical technique.

In the computation of the partial derivatives required in equation (6), one can choose to keep the existing model parametrization and proceed to solve the tomography equation for the regional velocity perturbations only. However, more commonly, the model is reparametrized in terms of homogeneous rectangular cells on a 2-D grid. Equation (6) can then be used to solve for travelt ime perturbations in each cell.

The next step in bent ray tomography is to interpret the resultant image. If the interpretation indicates that the regional model used to generate the ray paths is inconsistent with the result, the procedure may be repeated with a new regional model. This iteration is required in order to take into account the non-linear nature of the tomographic problem due to the dependencies of the ray paths on the velocity model. This step is interpretive, and it allows the

possible models to be constrained by an understanding of the geological setting. The most satisfactory implementation of this step is to superimpose new interfaces on top of the current tomographic image using a graphical input device.

2.2.2 Bent ray tomography in anisotropic media

We now proceed to discretize equation (4). By introducing homogeneous regions (or cells), we have for the k th ray path,

$$\delta T_k = F_{kij} \delta q_{ij} \quad (7)$$

[see equation (38) in Paper I] where δq_{ij} is the j th δq_j parameter ($j = 1, 2, \dots, 5$) in the i th region. F_{kij} is the sensitivity (or partial derivative) of the arrival time of the k th ray to changes in δq_{ij} . Replacing the integration in equation (4) by a summation, and replacing dl by Δl_{ki} (the distance traversed by the k th ray path in the i th cell), we obtained [equation (40) in Paper I]

$$F_{kij} = \frac{\partial T_k}{\partial q_{ij}} = -\frac{1}{2\alpha^3} \Delta l_{ki} \hat{p}_1^{5-j} \hat{p}_3^{j-1}, \quad \text{for } j = 1, \dots, 5 \quad (8)$$

(no summation over j). The quantities above are simple to compute the ray path in the original (isotropic) medium and the geometry of the regions (Fig. 1).

Equations (7) and (8) form a linear system for tomography in anisotropic media. The perturbation for the k th ray is a sum of five contributions from each cell. Given a set of measurements, δT_k , one can proceed to choose an appropriate numerical scheme to solve for δq_{ij} . Before proceeding, it is useful to recast the system in a matrix form. Introducing the new subscript, $l = 5(i - 1) + j$, we have

$$\delta T_k = F_{kl} \delta q_l, \quad \text{or} \quad \delta \mathbf{T} = \mathbf{F} \delta \mathbf{q}. \quad (9)$$

In order to facilitate the constraints we use in a later section, we introduce an additional model parameter, $\delta \bar{q}_0 = \delta \alpha^2$ (in contrast to Paper I here we use the perturbation of the square of the isotropic velocity). The remaining parameters are modified accordingly [see equation (36) in Paper I]. These parameters, $\delta \bar{q}_j$, $j = 1, \dots, 5$ represent purely anisotropic components. We then obtain

$$\delta T_k = F_{kl} \delta \bar{q}_l, \quad \text{or} \quad \delta \mathbf{T} = \mathbf{F} \delta \bar{\mathbf{q}}, \quad (10)$$

[the subscript l is changed so that $l = 6(i - 1) + j$], with

$$F_{ki0} = -\frac{1}{2\alpha^3} \Delta l_{ki}. \quad (11)$$

The remaining partial derivatives for ($j = 1, 2, 3, 4, 5$) are unchanged.

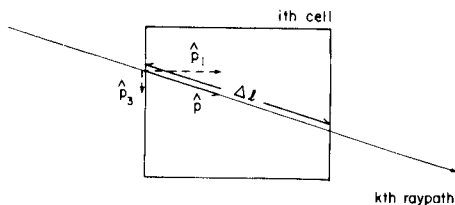


Figure 1. The partial derivatives of the traveltime for the k th ray path in the i th cell of a piecewise homogeneous medium can be computed from Δl , the length of the ray segment and the components of $\hat{\mathbf{p}}$, the tangent to the ray trajectory.

It should be noted that equations (9) and (10) lead to non-linearities in the inverse problem, even when ray curvature due to velocity structure can be ignored. The non-linearity arises because a solution for $\delta \mathbf{q}$ can contain an isotropic component, indicating that the initial choice of the distribution of α was inappropriate. This is overcome by relaxation: recomputing the distribution of α , relinearizing and resolving the equations. How the α 's are computed from the solutions, $\delta \mathbf{q}$, is dealt with in Section 2.4.

2.3 The global to TI transformation

The solution of the linear equation systems in equations (9) and (10) yields an estimate of the perturbations, or errors, in the five parameters, δq_j . Using anisotropic tomography and given a starting model, one thus recovers the spatial distribution of the five parameters, \mathbf{q} (in contrast to our earlier notation, here \mathbf{q} is a five-component vector representing the solution at only a single location). The non-uniqueness of five parameters to fully describe the elastic tensor was noted in Section 2.1. Knowledge of the symmetry system of the rock properties can provide the information required to be able to interpret the five parameters geologically. In general, the lower the symmetry is, the more parameters are required to describe the elastic properties and the more external information is required.

Often, a suitable anisotropic system is TI. As described in the Introduction to this paper, this is a suitable system in the examples that follow in Sections 4 and 5. For TI systems, there are also five parameters that describe the qP -wave velocities. It should be noted that these five parameters and the five elastic constants, \mathbf{q} , are not the same—they do not even span the same space! If the rock properties are known to exhibit TI symmetries, this information constrains the solutions almost sufficiently to allow the TI parameters to be estimated from \mathbf{q} . Actually, as we show below, this problem is itself singular and an additional degree of freedom needs to be removed before the TI can be obtained.

The mapping of TI parameters to \mathbf{q} (the global parameters) is the subject of appendix E in Paper I. Here we consider the inverse problem: mapping the five global parameters to the TI parameters, following the application of anisotropic tomography. This mapping must be carried out at each spatial location. The five global parameters can be expressed as

$$\mathbf{q} = \mathbf{H}(\theta, \phi) \begin{pmatrix} A \\ B \\ C \end{pmatrix} \quad (12)$$

[equation (E4) in Paper I], where the polar angle, θ , and the azimuthal angle, ϕ , are the Euler angles specifying the symmetry axis. The matrix \mathbf{H} is a 5×3 matrix whose elements are non-linear functions of θ and ϕ (the elements are given explicitly in Paper I).

The TI parameters A , B and C in equation (12) are given by

$$A = a_{1111}, \quad B = a_{1133} + 2a_{3131}, \quad C = a_{3333}. \quad (13)$$

In equation (13) only, the subscripts $ijkl$ refer to a coordinate system congruent with the TI axis.

Equation (12) introduces non-linearities into the global to TI transformation. The transformation can be achieved by iteration from a starting guess of the values of A , B , C , θ and ϕ . At each iteration equation (12) is used to compute the misfit between the global and TI parameters, partial derivatives are computed and new hexagonal parameters are generated. In practice we have found that the transformation is nearly always singular, or very nearly singular. This can be overcome by eliminating the degree of freedom associated with the lowest singular value, so that only four TI parameters are recovered. The degree of freedom eliminated is usually some linear combination of the two Euler angles. For example, for near symmetry axes ($\theta = 0$), ϕ is indeterminate. In the case of near vertical axes we always proceed by adding the constraint $\phi = 0.0$. It may be helpful in the future to incorporate external information on the orientation of the symmetry axes (for example from analysis of the polarization of split shear waves).

Thomsen (1986) has introduced an alternative parameter set for TI media. The Thomsen parameters simplify the directional velocity dependencies and facilitate a geological interpretation. The five Thomsen parameters are α_0 (the velocity along the axis of symmetry), ϵ (the non-dimensional 'amount' of anisotropy), δ (a second non-dimensional parameter that describes the anellipticity of the slowness sheets) and two direction cosines that describe the orientation of the axis of symmetry. Given A , B and C it is straightforward to compute α_0 , ϵ and δ [see Paper I, equation (E9)]; specifically,

$$\alpha_0 = C^{1/2}, \quad \epsilon = (A - C)/2C, \quad \delta = (B - C)/C. \quad (14)$$

The last relationship is valid only within the weak anisotropy assumption.

Due to the non-linearities certain non-uniquenesses arise. For example, for a given set of global parameters, both $(\alpha_0, \epsilon, \delta, \theta, \phi)$ and $(\alpha_0, -\epsilon, \delta, \theta + \pi/2, \phi)$ fit the same q_j parameters. Both sets of parameters yield the same curve of slownesses in the plane of the survey, but they represent very different media (the slowness sheet of one of these is prolate, the other is oblate). The approach we have taken to avoid this particular non-uniqueness is to constrain ϵ to be positive. Other problems arise due to local minima in the fitting when the slowness sheets are not ellipsoids (when $\delta \neq \epsilon$). We have found that a global search over several hundred values of θ is a useful way to begin this transformation.

2.4 Isotropic components of velocities

It was noted at the end of Section 2.2 that equations (9) and (10) are non-linear in the isotropic velocity, α , since α appears both in the partial derivatives and in the solution \mathbf{q} . The set of five δq_j 's may contain an isotropic component, indicating the original isotropic velocity model was inadequate. This can be approached numerically by relaxation. A new isotropic model is chosen after each inversion and the equations are relinearized and re-solved. In tests we have carried out the solutions stabilize after several iterations (three or four is typical); the perturbations are then purely anisotropic.

The question remains as to how the isotropic velocities may be chosen from the five global parameters, \mathbf{q} . One

obvious choice would be always to choose the horizontal phase velocities, $q_1^{1/2}$. A second choice would be the vertical velocities, $q_5^{1/2}$. These are straightforward, but they may not yield interpretable tomographic images, especially in complex regions in which the axis of symmetry may vary. In TI media other possibilities offer themselves. One could use the velocities along the symmetry axes, $C^{1/2}$. Using $C^{1/2}$ has the advantage of being independent of the orientation of the symmetry axis. However, these estimates are all biased: in TI systems with a vertical symmetry axis (TIV systems) systems the horizontal velocities are often fastest and the vertical velocities are slowest. This choice of an isotropic component of the velocity may not be a good choice in terms of the subsequent relinearization. Instead, a velocity that represents an average over all directions is required.

An appropriate isotropic velocity is supplied by the harmonic decomposition of the elastic tensor given by Backus (1970). Backus showed that a general fourth-order, 3-D elastic tensor could be decomposed into five 'harmonic' tensors, one of order 4, two of order 2 and two of order 0. The higher order tensors contain information on the orientations of the symmetry axes and on the magnitude of the anisotropy. In isotropic media the higher order tensors vanish and the two zeroth-order tensors (scalars) are directly related to the isotropic Lamé parameters. The zeroth-order tensors define the isotropic component of both the compressional and the shear velocities.

The isotropic compressional velocity component of a general elastic tensor was given by Backus (1970) as

$$\alpha^2 = (21/15)Q_{jj}, \quad (15)$$

where Q_{jj} is the trace of the matrix,

$$Q_{ij} = (a_{ijmm} + 2a_{imjm})/21 \quad (16)$$

(this notation differs slightly from that used by Backus, since we use a density normalized tensor). Equations (15) and (16) reveal the fundamental inability of 2-D, qP tomography to resolve even the zeroth-order components of the elastic tensor. Explicitly,

$$\begin{aligned} Q_{11} &= (3a_{1111} + a_{1122} + a_{1133} + 2a_{1313} + 2a_{1212})/21, \\ Q_{22} &= (a_{1122} + 3a_{2222} + a_{2233} + 2a_{2323} + 2a_{1212})/21, \\ Q_{33} &= (a_{1133} + a_{2233} + 3a_{3333} + 2a_{2323} + 2a_{1313})/21. \end{aligned} \quad (17)$$

Thus,

$$\begin{aligned} Q_{jj} &= (3q_1 + 3q_5 + q_3 + 3a_{2222} + 2a_{1122} \\ &\quad + 4a_{1212} + 2a_{2233} + 4a_{2323})/21, \end{aligned} \quad (18)$$

in which only the first three terms are determined by qP tomography in a 2-D plane.

To remove the non-uniqueness, again it is necessary to assume some symmetry property. If we assume a TIV system, then $a_{2222} = a_{1111} = q_1$, $a_{1122} + 2a_{1212} = a_{1111} = q_1$ and $2a_{2233} + 4a_{2323} = 2a_{1133} + 4a_{1313} = q_3$. Making these substitutions into equation (18) yields, for TIV systems

$$Q_{jj} = (8q_1 + 2q_3 + 3q_5)/21. \quad (19)$$

Similarly, for an arbitrary orientation of the axis of symmetry,

$$Q_{jj} = (8A + 4B + 3C)/21. \quad (20)$$

In the degenerate case of isotropy, where $A = B = C$, we obtain $\alpha^2 = A$, as expected. The higher weighting of A in equation (20) reflects the fact that most of the 3-D slowness sheet lies close to the transverse slowness.

In order to extract a suitable isotropic component of velocity from the five \mathbf{q} parameters, given a TI media, we first apply the global to TI transformation to obtain A , B and C , and then use equation (20) followed by equation (15) to compute α .

3 THE LINEAR SYSTEM AND REGULARIZATION

3.1 The singular value decomposition of the linear system

In equations (6), (9) and (10), we formulated linear systems for tomography in isotropic and in anisotropic media. Each of these systems is characterized by a Fréchet matrix \mathbf{F} . The $M \times N$ Fréchet matrix is a linear mapping from the N -dimensional parameter space to the M -dimensional data space. For isotropic tomography N is the number of cells chosen to discretely represent the velocity field M is the number of ray paths. For anisotropic tomography using equation (9) (or equation 10), N is the number of cells multiplied by five (or by six). Because each ray crosses only

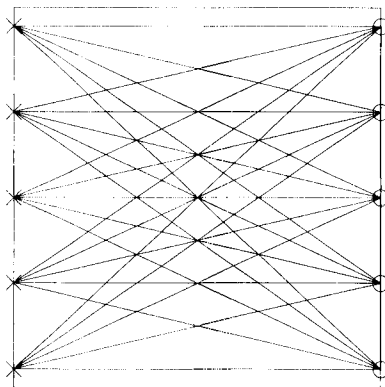


Figure 2. The ray geometry used to generate the linear systems analysed in Figs 3, 4 and 5.

a small number of cells, \mathbf{F} is extremely sparse. It is important to take advantage of this sparseness in the computer solution of tomographic problems.

Linear systems such as equations (6), (9) and (10) are, for most exploration geometries, extremely ill-conditioned. The ill-conditioning is a result of the source–receiver geometries and often cannot be overcome. The singular value decomposition (SVD) of the Fréchet matrices for crosshole tomography has been used effectively by Bregman, Bailey & Chapman (1989b) to analyse the crosshole tomography problem. Here, we review these results for isotropic tomography and continue the analysis for traveltimes tomography in anisotropic media.

For any matrix \mathbf{F} there exists the SVD,

$$\mathbf{F} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{21}$$

(Lanczos 1961). The matrix \mathbf{U} is an $M \times M$ orthogonal matrix of eigenvectors that span the data space, the matrix \mathbf{V} is an $N \times N$ orthogonal matrix of eigenvectors that span the parameter space and $\mathbf{\Lambda}$ is an $M \times N$ diagonal matrix whose diagonal elements are known as the singular value of \mathbf{F} . The singular values are always non-negative and the number of non-zero singular values is always equal to or less than the minimum of M and N . Zero-valued singular values cancel the contributions of the corresponding eigenvectors. The matrices in equation (21) can be partitioned, so that

$$\mathbf{F} = \mathbf{U}_p \mathbf{\Lambda}_p \mathbf{V}_p^T, \tag{22}$$

where the subscript p indicates that only the columns of \mathbf{U} and \mathbf{V} that are multiplied by non-zero singular values are retained. The remaining columns of \mathbf{V} span the null space of \mathbf{F} . Parameter configurations that are linear combinations of vectors lying completely within the null space have no effect on the modelled traveltimes.

Figure 2 shows the ray paths we consider in the following model study. The rays originate at five source locations located at the right-hand edge and connect each source with the five receiver locations at the left-hand edge. The results of the SVD of the isotropic tomography Fréchet matrix for this configuration were discussed by Bregman *et al.* (1989b). In Fig. 3 we reproduce some of these results. The velocity

SVD (25 data, 25 cells, 25 parameters)

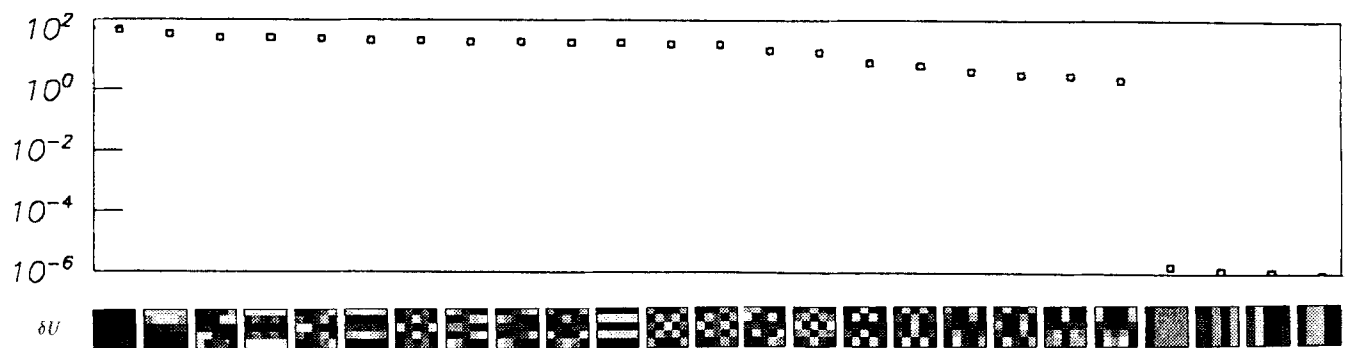


Figure 3. Eigenvectors and associated singular values for isotropic tomography using the model shown in Fig. 2. The cell geometry is $5 \times 5 = 25$ cells.

field is discretely represented by a 5×5 grid of points, leading to a total of 25 parameters which are used to attempt to fit the 25 data. There are thus 25 singular values; these are plotted in Fig. 3 above the corresponding eigenvectors of the parameter space (i.e. the columns of \mathbf{V}). There are four zero-valued singular values for this configuration; these four are associated with four eigenvectors that define the purely vertical structures in the model. Apart from the four zero singular values, the lowest singular values are associated with patterns centred over cell locations close to the top and bottom edges of the survey (the poorly sampled cells). Bregman showed that the ill-conditioning could be alleviated if additional data from sources or receivers located on the surface were included. Unfortunately, in many applications the distance to the surface is much larger than the distance between the wells, so that this approach is often not possible. Furthermore, the near surface is often more heterogeneous, anisotropic and attenuating than everything else, leading to additional imaging problems.

Figure 4 shows the SVD when the additional degrees of freedom required to describe the anisotropy are included (using equation 9), without increasing the data volume. Since five parameters are required at each cell location (δq_j , $j = 1, \dots, 5$), there are now 125 eigenvectors required to span the parameter space, each of which can be partitioned into five spatial distributions, one for each δq_j . The first 25 of these eigenvectors are plotted below their corresponding singular values. All remaining singular values are identically zero, so that only the eigenvectors shown can influence the modelled travelt ime data. These eigenvectors contain primarily components of δq_1 and, for the smaller singular values, components of δq_2 . Almost all the contributions for the remaining parameters lie within the null space of the problem as posed (with some exceptions at the centre of the model). The survey resolves the δq_j in roughly numerical

order; the δq_j with larger values of j have less influence on the data. This phenomenon can be understood in terms of the formula for the partial derivatives, equation (8). For small values of j , the magnitudes of the partial derivatives are larger for near-horizontal rays (with large \hat{p}_1 components) and smaller for near-vertical rays. The spectrum of singular values and eigenvectors is a reflection of the distribution of ray angles, which is biased toward the near-horizontal rays. Again, if sources or receivers were located along the top edge of the survey, this would alleviate the ill-conditioning and change the distribution of the eigenvectors.

Figure 5 shows the SVD obtained when the total number of cells is reduced. The grid is now 3×2 , so that there are only six cells (30 parameters) altogether. The pattern of the spectrum we described in the previous paragraph is clear. In this case, however, the δq_3 , δq_4 and δq_5 parameters have a greater influence on the data. Because the cell size is increased, the cells contain more ray segments. In order to fit the data using the parametrization it is necessary to vary all five parameters within the cells. For this parametrization there are 23 non-zero singular values, although these values fall off drastically beyond the 20th singular value. The eigenvectors for the zero singular values contain linear combinations of vertical structures in several parameters. The low singular values are associated with specific linear combinations of the δq_4 and δq_5 parameters.

We conclude from the previous two paragraphs, and from Figs 3 and 4, that crosshole arrival times can be modelled in very different ways, depending on how the model is parametrized. If a fine grid is used in the discretization, the arrival times can be modelled using only δq_1 and δq_2 . If a coarser grid is used, then one needs to include the remaining parameters in order to fit the data. The only way in which one can decide on the validity of either of these models is to incorporate other criteria.

SVD (25 data, 25 cells, 125 parameters)

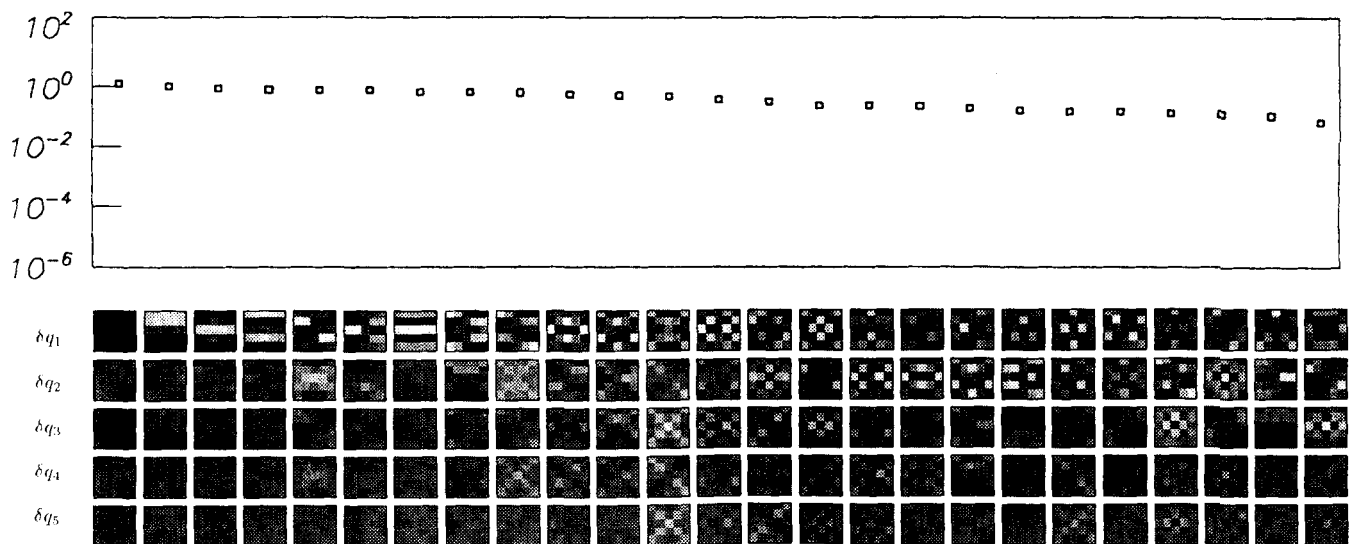


Figure 4. Eigenvectors and associated singular values for anisotropic tomography using the model shown in Fig. 2. The cell geometry is as in Fig. 3, but here five parameters per cell are used.

SVD (25 data, 6 cells, 30 parameters)

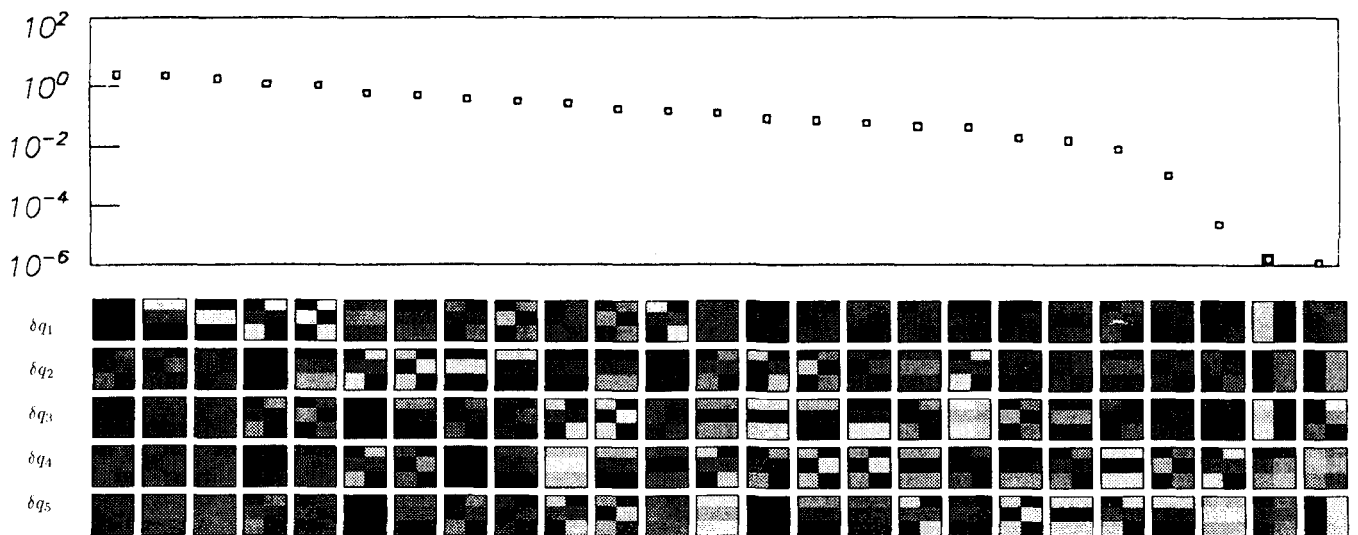


Figure 5. Eigenvectors and associated singular values for anisotropic tomography using the model shown in Fig. 2. The cell geometry is $2 \times 3 = 6$ cells.

For the multiparameter case, the scalelength of the parametrization is more important than that in isotropic tomography. In isotropic tomography, when a fine grid is used, the data can usually be modelled by varying only the long-wavelength components, since decreasing singular values are generally associated with decreasing wavelengths in the eigenvectors. In the fine-grid multiparameter case, the spectrum does not arrange itself according to wavelength. For example, in Fig. 3 the short-wavelength components of δq_1 have more influence than the long-wavelength components of the remaining δq_i 's. This difficulty in modelling data using multiple parameters has also been noted (in a different context) by Scales, Docherty & Gersztenkorn (1990). Even in the single parameter case, when too fine a parametrization is used, the eigenvector spectrum is no longer arranged in decreasing wavelength order. This can lead to artifacts in isotropic tomographic images.

3.2 The inverse problem and the nature of *a priori* information

The condition numbers for the matrices in the examples in the previous section (Figs 3, 4 and 5) are effectively infinite. In solving equations (6), (9) and (10) this ill-conditioning needs to be addressed. The eigenvectors associated with the zero singular values (i.e. the vectors that span the null space) are unconstrained by the data. Any combination of these eigenvectors can be added to the solution without affecting the result. The inverse method must include some criteria for selecting the components of the null space that appear in the solution. Techniques for including external information serve to regularize the inverse problem (Tikhonov & Arsenin 1977).

Possible forms of external information that might be

incorporated are:

- (1) a starting model (often of unknown certainty);
- (2) an estimate of the magnitude of the anisotropy (for example, isotropy is expected);
- (3) an expectation that the elastic properties vary smoothly; and
- (4) some combination of the above.

The third item requires comment. In actuality, elastic parameters vary at all scalelengths down to microscopic levels. If seismic waves propagated as rays, the elastic parameters would necessarily be always undersampled. Therefore, parameter estimates would be subject to severe aliasing errors. If the data are assumed to contain information on only the smooth components of the parameter distribution, this places a limit on the grain of the geology that one expects will influence the data. This approach, in ad hoc sense, is a recognition of the fact that, because the seismic energy is band-limited, the arrival times of transmitted waves are insensitive to the fine-grained properties of the medium. In anisotropic media the media must be adequately sampled, not only to prevent spatial aliasing, but also to provide the required parameter resolution.

There are a number of different philosophical approaches that can be taken to incorporate external (*a priori*) information, although most yield similar numerical schemes. Tarantola (1987) presents a stochastic formulation for the inverse problem that lends itself to the incorporation of any of the items enumerated above. In the stochastic inverse the probability distribution is described for a range of models. The probability of a given model, $\delta \mathbf{q}$, is proportional to

$$\exp\left(-\frac{1}{2} S(\delta \mathbf{q})\right), \quad (23)$$

where the least-squares objective functional, $S(\delta \mathbf{q})$ is

defined as

$$S(\delta\mathbf{q}) = (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T})^T \mathbf{C}_d^{-1} (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T}) + \delta\mathbf{q}^T \mathbf{C}_p^{-1} \delta\mathbf{q}. \quad (24)$$

In equation (24) the data space and parameter space *a priori* covariance matrices, \mathbf{C}_d and \mathbf{C}_p , are used to weight the relative contributions to the objective function of the data residuals (the differences between modelled and observed data) and the parameter differences (the differences between *a priori* and estimated parameters). This version of the probability distribution is dependent on the assumption of Gaussian probability distributions both for the data and for the model parameters. Neither assumption can really be justified. However, there is little choice but to proceed with the assumption of Gaussian model statistics, as one is unlikely in any case to be able to determine a more appropriate statistical model. Because Scales, Gersztenkorn & Treitel (1988) have shown that the least-squares formulation can be used to treat non-Gaussian data statistics by iteratively reweighting the data following each least-squares inversion, it is appropriate to continue to use the least-squares formulation.

Using equations (23) and (24), any solution to the tomography problem, $\delta\mathbf{q}$, can be assigned a unique probability value. The most probable solution is the one that minimizes the functional, $S(\delta\mathbf{q})$. Normally, the observations, $\delta\mathbf{T}$, are the differences between the *a priori*, isotropic model and the observed data. If the *a priori* model is anisotropic this can equally well be incorporated (nonetheless, partial derivatives are always computed using an isotropic model).

Since the data are often assumed to be of uniform variance and uncorrelated, replace \mathbf{C}_d with $\mathbf{C}_d = \sigma_d^2 \mathbf{I}$ in equation (24):

$$S(\delta\mathbf{q}) = \sigma_d^{-2} (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T})^T (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T}) + \delta\mathbf{q}^T \mathbf{C}_p^{-1} \delta\mathbf{q}. \quad (25)$$

In order to obtain the required solution, differentiate $S(\delta\mathbf{q})$ with respect to the model parameters, and the result equal to zero. This leads to the solution

$$(\mathbf{F}^T \mathbf{F} + \sigma_d^2 \mathbf{C}_p^{-1}) \delta\mathbf{q} = \mathbf{F}^T \delta\mathbf{T}. \quad (26)$$

In the case of infinite *a priori* model variances this equation reduces to the well-known normal equation for least squares. For singular \mathbf{F} , the normal equation cannot be solved ($\mathbf{F}^T \mathbf{F}$ has no inverse); the presence of a finite \mathbf{C}_p matrix allows a solution to be obtained.

We now explore some of the possible forms of the matrix, \mathbf{C}_p . If the *a priori* information is limited to a set of model parameters that are expected to be uncorrelated and have uncertain variance (constraint 1), then

$$\mathbf{C}_p = \sigma_p^2 \mathbf{I}, \quad (27)$$

where σ_p is chosen to be of the order of the magnitude of the parameters. This leads to the damped least-squares solution

$$(\mathbf{F}^T \mathbf{F} + \lambda^2 \mathbf{I}) \delta\mathbf{q} = \mathbf{F}^T \delta\mathbf{T}, \quad (28)$$

with $\lambda^2 = \sigma_d^2 / \sigma_p^2$.

In order to be able to constrain solutions to confirm with an expected level of anisotropy (i.e. constraint 2), a modification of the covariance matrix given in equation (27) is required. This constraint implies that the *a priori*

variances of the anisotropy parameters are smaller than the variance of the squared velocities. We formalize this by invoking the alternative parametrization given in equation (10), in which $\delta\bar{q}_j$ with $j = 0$ correspond to the velocities and in which $\delta\bar{q}_j$ with $j = 1, \dots, 5$ represent purely anisotropic parameters. This allows the correct variances to be assigned to each parameter, specifically

$$[\mathbf{C}_p]_{ll} = \sigma_p^2 \quad \text{for } j = 0 \\ = \sigma_p^2 \frac{\lambda^2}{\lambda^2 + \kappa^2} \quad \text{for } j = 1, \dots, 5 \quad (29)$$

(no summation over l). The relationship between the subscripts l and j has been given at the end of Section 2.2.2. κ is thus a measure of the decrease in the *a priori* variances of the anisotropy parameters relative to the variance of the velocity.

The form of the *a priori* parameter covariance matrix given in equation (29) leads to the damped least-squares solution

$$(\mathbf{F}^T \mathbf{F} + \lambda^2 \mathbf{I} + \kappa^2 \mathbf{I}') \delta\bar{\mathbf{q}} = \mathbf{F}^T \delta\mathbf{T}, \quad (30)$$

where \mathbf{I}' is a modified identity matrix in which the diagonal elements that correspond to $j = 0$ parameters are set equal to zero. Both this form and equation (28) are special cases of damped least-squares solutions with zero *a priori* parameter covariances and non-uniform parameter variances.

The expectation of smoothness (constraint 3) can also be incorporated into the stochastic inverse, at least in theory. This information can be implemented using a covariance between parameters spatially separated by a distance, Δr , of the form,

$$\sigma_p^2 \exp \left[-\frac{1}{2} \left(\frac{\Delta r^2}{L^2} \right) \right], \quad (31)$$

where L represents the distance over which one expects the solution to be smooth. Since \mathbf{C}_p is a smoothing operator, the effect of \mathbf{C}_p^{-1} in equation (25) is to penalize roughness in the solution.

In practice, the use of a covariance operator designed to penalize roughness is costly to implement, since the matrix \mathbf{C}_p becomes dense (destroying the sparseness of the Fréchet matrices). Instead, it is more effective to add a derivative of the solution to the objective function (Constable *et al.* 1987):

$$S(\delta\mathbf{q}) = \sigma_d^{-2} (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T})^T (\mathbf{F}\delta\mathbf{q} - \delta\mathbf{T}) + \hat{\varepsilon}^2 \delta\mathbf{q}^T \mathbf{R}^T \mathbf{R} \delta\mathbf{q}, \quad (32)$$

where \mathbf{R} is a discrete, sparse, difference operator (the regularization operator). The actual form of \mathbf{R} can vary. Either the first difference or second difference forms can be used; the former is used to penalize solutions with large first derivatives (large spatial gradients) in the solution, the latter penalizes solutions with large second derivatives (large values of the spatial Laplacian operator). If $\hat{\varepsilon}^2 \mathbf{R}^T \mathbf{R} = \lambda^2 \mathbf{I}$ (i.e. the zeroth difference operator) then the regularization operator penalizes the magnitude of the solution, as in damped least squares. In Section 4.2 we give explicit formulae for first and second difference operators.

Equation (32) leads to a solution of the form

$$(\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{R}^T \mathbf{R}) \delta\mathbf{q} = \mathbf{F}^T \delta\mathbf{T}, \quad (33)$$

with $\varepsilon^2 = \hat{\varepsilon}^2 \sigma_a^2$. The operator $\hat{\varepsilon}^2 \mathbf{R}^T \mathbf{R}$ replaces the role of \mathbf{C}_p^{-1} in equation (26) (it is also a roughness operator), but it is much easier to compute and does not destroy the sparseness of the formulation. Equation (33) can also be written in a generic form as

$$\delta \mathbf{q} = \mathbf{F}^\dagger \delta \mathbf{T}, \quad (34)$$

where \mathbf{F}^\dagger is the pseudo-inverse of \mathbf{F} , given by

$$\mathbf{F}^\dagger = (\mathbf{F}^T \mathbf{F} + \varepsilon^2 \mathbf{R}^T \mathbf{R})^{-1} \mathbf{F}^T. \quad (35)$$

3.3 The effect of constraints on the SVD

In order to gain insight into the operation of the constraints discussed in the previous section it is useful to consider the SVD of the pseudo-inverse, \mathbf{F}^\dagger (equation 35). Substitution of the SVD of \mathbf{F} (equation 22) yields

$$\mathbf{F}^\dagger (\mathbf{V}_p \mathbf{\Lambda}_p^2 \mathbf{V}_p^T + \varepsilon^2 \mathbf{R}^T \mathbf{R})^{-1} \mathbf{V}_p \mathbf{\Lambda}_p \mathbf{U}_p^T. \quad (36)$$

For the case of infinite *a priori* variances ($\varepsilon \rightarrow 0$) equation (36) reduces to the generalized inverse,

$$\mathbf{F}^\dagger = \mathbf{V}_p \mathbf{\Lambda}_p^{-1} \mathbf{U}_p^T. \quad (37)$$

The generalized inverse simply sets the components of the null space equal to zero in the solution. Problems occur in applying a generalized inverse when, as in Figs 2 and 4 there are singular values within $\mathbf{\Lambda}_p$ that lie close to zero. Multiplying the data eigenvectors, \mathbf{U}_p by $\mathbf{\Lambda}_p^{-1}$ leads to unacceptable amplification of noise in the data. The data fitting will be too successful and the solutions will be dominated by those eigenvectors that actually have little influence on the data. Furthermore, as we have shown in Section 3.1, the scalelength of the parametrization will determine the form of the solution that the generalized inverse will yield: parametrization on a fine grid will yield fine-grain solutions, with only δq_1 and δq_2 terms, whereas a coarse parametrization will yield coarse-grain solutions with estimates of all δq_j parameters.

If $\varepsilon^2 \mathbf{R}^T \mathbf{R} = \lambda \mathbf{I}$ (damped least squares), then some of the deficiencies of the generalized inverse are rectified. Under this regularization, equation (36) reduces to

$$\mathbf{F}^\dagger = \mathbf{V}_p \mathbf{\Lambda}_p^{-1} \mathbf{D} \mathbf{U}_p^T, \quad (38)$$

with the diagonal damping matrix

$$\mathbf{D} = \frac{\mathbf{\Lambda}_p^2}{\mathbf{\Lambda}_p^2 + \lambda^2 \mathbf{I}}. \quad (39)$$

The damping matrix contains diagonal elements ranging in values from 0 to 1. As the singular values become small (less than λ in magnitude) the damping acts to reduce their effect in the pseudo-inverse. A judicious choice of damping can avoid the overfitting of noisy data. However, the eigenvectors in the null space do not contribute to the solution and the scalelength of the parametrization will, therefore, as in the generalized inverse, determine the nature of the solution.

For more general regularization operators, \mathbf{R} , (for example, the first and second difference forms) equation (36) can only be reduced to

$$\mathbf{F}^\dagger = \mathbf{V} [\mathbf{\Lambda}^T \mathbf{\Lambda} + \varepsilon^2 (\mathbf{R} \mathbf{V})^T \mathbf{R} \mathbf{V}]^{-1} \mathbf{\Lambda}^T \mathbf{U}^T. \quad (40)$$

The partitioned form of the SVD is no longer used. To the extent to which $(\mathbf{R} \mathbf{V})^T \mathbf{R} \mathbf{V}$ is not diagonal, this pseudo-inverse contains contributions from eigenvectors within the null space of the unconstrained problem. In contrast to the generalized inverse and damped least squares, the additional requirement of smoothness allows components from the null space of the unconstrained problem to contribute to the solution. This is desirable, since the scalelength of the parametrization should not determine the nature of the solution. It is preferable to use the parameter, ε , to control the regularization. Reconstructing the solution using equation (40) inhibits the contribution of the rough eigenvectors, but allows the smooth eigenvectors in the null space of the unconstrained problem to contribute to the solution.

3.4 The numerical solution of the inverse problem

The solution of the traveltime tomography problem in the presence of anisotropy has been defined, generically, by use of the pseudo-inverse, \mathbf{F}^\dagger (equation 35). Normally, one would like to avoid actually forming the pseudo-inverse (which is costly for most realistic problems). There are many equation solvers that solve the normal equation for least squares without explicitly forming a pseudo-inverse. The regularized solution, equation (33), can be manipulated to yield the normal equation,

$$(\hat{\mathbf{F}}^T \hat{\mathbf{F}}) \delta \mathbf{q} = \hat{\mathbf{F}}^T \delta \hat{\mathbf{T}}. \quad (41)$$

This form, exactly equivalent to equation (33), is obtained by making use of the augmented Frechét matrix,

$$\hat{\mathbf{F}} = \begin{pmatrix} \mathbf{F} \\ \varepsilon \mathbf{R} \end{pmatrix} \quad (42)$$

(which still has N columns, but now has $M + N$ rows) and the augmented data vector,

$$\delta \hat{\mathbf{T}} = \begin{pmatrix} \delta \mathbf{T} \\ 0 \end{pmatrix}. \quad (43)$$

To solve the inverse problem we use the LSQR solver for damped least squares. LSQR is a modified conjugate gradient technique for damped least squares, originally proposed by Lanczos (1950), but not numerically tractable until implemented by Paige & Saunders (1982). An important feature of the LSQR solver is that it allows full advantage to be taken of the sparseness of the matrices, \mathbf{F} and \mathbf{R} . The LSQR solver incorporates only standard damping of the form given in equation (28). When using the more general forms of regularization operators the LSQR algorithm can still be used, provided the augmented matrix and data vectors, $\hat{\mathbf{F}}$ and $\delta \hat{\mathbf{T}}$, are provided.

4 APPLICATION TO COMPUTER GENERATED DATA

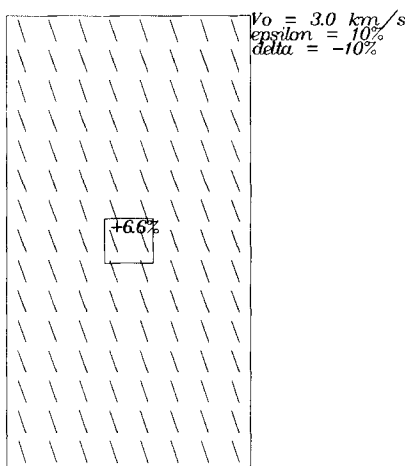
In the following sections we describe the application of the numerical scheme, equation (41). We generated data by computer simulation, we inverted the data to obtain the spatial distributions of the global parameters \mathbf{q} and, finally, we used the transformation described in Section 2.3 to map these to the isotropic velocities, α , the Thomsen

parameters, ϵ and δ and the two Euler angles, θ and ϕ . As described in Section 2.3, in order to apply the final global to TI transformation, it was necessary to add two further constraints: that the azimuthal angle, ϕ be zero and that ϵ be greater than zero.

4.1 Model description and data generation

Figure 6 depicts the anisotropic model and the survey geometry used in the following study. The two ‘boreholes’ are separated by 100 m, and the length of the source and receiver arrays is 200 m. The model is uniformly TI, and includes a central velocity anomaly as indicated by the box. The TI parameters for this model are $\epsilon = 10$ per cent and $\delta = -10$ per cent. The symmetry axis is contained within the survey plane and is tilted 20° from the vertical. The velocity α_0 along the symmetry axis is 3.0 km s^{-1} except in the central anomaly. The central anomaly is characterized by the same anisotropy parameters, but contains an increase in velocity ($\alpha_0 = 3.2 \text{ km s}^{-1}$). The monopole terms for these two regions are $\alpha = 3.113 \text{ km s}^{-1}$, for the surrounding region and $\alpha = 3.321 \text{ km s}^{-1}$, for the central anomaly.

11 source locations (along the left-hand edge of the model) and 11 receiver locations (along the right-hand edge of the model) were used in the synthetic data generation. Arrival times for all 121 source–receiver pairs were generated using an exact 2-D ray theory. Because the symmetry axis is contained within the plane 2-D ray tracing is sufficient. The anisotropies are weak (10 per cent) and we therefore expect the perturbation approach to be valid; this was confirmed by comparing the modelled times with times modelled using equation (4)—the differences were of the order of the machine precision. We also deliberately chose the velocity anomaly to be small (6.67 per cent) in order to be able to ignore non-linearities. In this study of travelt ime inversion we treat only the linear aspects of the problem; we do no relinearize and we use only straight rays. The tomography problem is often only weakly non-linear and we



Schematic of synthetic study

Figure 6. Model for synthetic study. The model is uniformly transversely isotropic (TI), with $\epsilon = 10$ per cent, $\delta = -10$ per cent. The symmetry axis is contained in the survey plane and tilted 20° from the vertical. The central anomaly is a 6.67 per cent increase in velocity.

expect the non-linear aspects of the problem to be soluble in a straightforward manner by relaxation.

4.2 Regularization

There are 121 travelt ime data available from the computer simulation. We parametrized the model using a grid 8 cells wide and 15 cells high. We used the parameter set $\bar{\mathbf{q}}$, so that the total number of parameters was $6 \times 8 \times 15 = 720$. From the arguments presented in Section 3.3, the use of a generalized inverse or damped least squares would imply that one does not expect that the parameters δq_3 , δq_4 , δq_5 will appear in the solution. To extract information about these parameters, it is important to sacrifice some of the spatial resolution of the fine parametrization grid. This was done by constraining the parameters to be spatially smooth, in the manner described in Section 3.2.

In Section 3 we related the constraints on anisotropic tomography to *a priori* parameter covariance matrices. Although ideally these matrices should be specifically computed from available data (well logs, for example) in a statistical fashion, in reality the relevant information is often not available. Instead, we now show the results we obtained for a range of regularization levels, and develop methods for selecting appropriate regularization levels from a suite of results. The two parameters we varied were ϵ , the roughness penalty and the κ , the anisotropy penalty.

To apply both roughness and anisotropy penalties to the tomography problem, we used an augmented Fréchet matrix of the form

$$\hat{\mathbf{F}} = \begin{pmatrix} \mathbf{F} \\ \epsilon \mathbf{R}_x \\ \epsilon \mathbf{R}_z \\ \epsilon \mathbf{R}_{\nabla^2} \\ \kappa \mathbf{I} \end{pmatrix} \quad (44)$$

and an augmented data matrix

$$\hat{\mathbf{T}} = \begin{pmatrix} \mathbf{T} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (45)$$

where \mathbf{R}_x , \mathbf{R}_z and \mathbf{R}_{∇^2} are the roughness matrices for the x and z gradient terms (the first difference operators) and the Laplacian term (the second difference operator). The Fréchet matrix, \mathbf{F} , contains $M = 121$ rows and $N = 720$ columns. The augmented Fréchet matrix, $\hat{\mathbf{F}}$, contains $M + 4N = 2901$ rows and $N = 720$ columns.

To compute the first difference operators \mathbf{R}_x and \mathbf{R}_z we used centred finite difference approximations to the first derivatives, $\partial/\partial x$ and $\partial/\partial z$. These operators can be represented by the finite difference ‘stars’:

$$\frac{\partial}{\partial x} \approx \frac{1}{\Delta} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \frac{\partial}{\partial z} \approx \frac{1}{\Delta} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}. \quad (46)$$

To compute the second difference operator we used a centred finite difference approximation of the Laplacian, ∇^2 . The operator can be represented by the 9 point finite

difference 'star':

$$\nabla^2 \approx \frac{1}{\Delta^2} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (47)$$

In equations (46) and (47) Δ is the width of a cell. If the cells are not square the equations must be slightly modified. The finite difference stars in equations (46) and (47) cannot be used at the edges of the survey. Instead, for the first difference operators we use forward or backward differencing schemes. The second differences are left unconstrained at the edges.

Each row of \mathbf{R}_x , \mathbf{R}_z and \mathbf{R}_{∇^2} corresponds to a particular parameter, $\delta\bar{q}_l$, where each l implicitly defines both i and j in $\delta\bar{q}_{ij}$ (the j th $\delta\bar{q}_j$ parameter in the i th cell). The non-zero elements of the l th row of each matrix are computed by overlaying the star with its centre at the i th cell, and entering the values of the star into the corresponding column locations. This leads to a block tridiagonal form for each matrix. The diagonal blocks contain tridiagonal submatrices and the off-diagonal blocks contain diagonal submatrices.

4.3 Effect of the regularization parameters

We have not yet treated the problem of selecting the appropriate levels of constraints, ϵ and κ . In general, as the magnitudes of ϵ or κ are decreased the data fit improves. However, using values that are too small will result in

overfitting of the data and amplification of the data errors by the small eigenvalues of the problem. Constable *et al.* (1987) and Scales *et al.* (1990) recommend choosing the roughness penalty, ϵ , by selecting a value at which the data are fit to within exactly one standard deviation. Two related problems arise when attempting to implement this criterion. First, the standard deviation of the data is not in practice known accurately. Second, the criterion is based on the assumption of Gaussian data statistics. In traveltime tomography the data errors are often non-Gaussian, since the errors are dominated by picking errors. Furthermore, in regions of complex velocity structure, the ray theory used to model the traveltimes is inadequate, leading to modelling errors. Modelling errors, if not accounted for, reveal themselves as data errors with decidedly non-Gaussian statistics. An alternative approach is to carefully examine a suite of solutions and use a number of indicators to estimate optimal regularization parameters.

Using the synthetic data from the model in Fig. 6, equation (41) was solved for a suite of values of the regularization parameters, ϵ and κ . All solutions were generated using a constant, isotropic velocity, $\alpha = 3.09$, as a starting model (the average of all the average ray velocities). In order to demonstrate the dependence of the solutions on the regularization parameters, in Fig. 7 we display the RMS values of the data residuals,

$$\left(\frac{1}{M} (\mathbf{F} \delta\mathbf{q} - \delta\mathbf{T})^T (\mathbf{F} \delta\mathbf{q} - \delta\mathbf{T}) \right)^{1/2}; \quad (48)$$

and the roughnesses of the solutions as measured by the

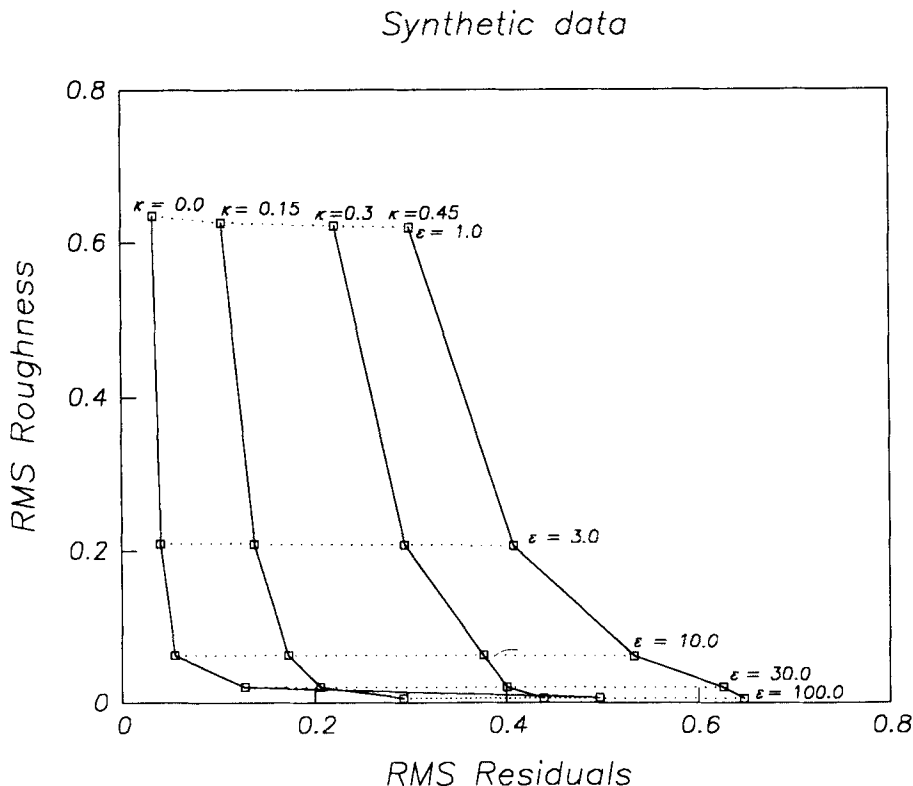


Figure 7. Solution roughness trade-off curves for the synthetic data set. A suite of solutions for varying values of roughness penalty (ϵ) and anisotropy penalty (κ) were analysed.

RMS value,

$$\left(\frac{1}{N}(\mathbf{R} \delta \mathbf{q})^T(\mathbf{R} \delta \mathbf{q})\right)^{1/2}, \quad (49)$$

where

$$\mathbf{R} = \begin{pmatrix} \varepsilon \mathbf{R}_x \\ \varepsilon \mathbf{R}_z \\ \varepsilon \mathbf{R}_{\nabla^2} \end{pmatrix}. \quad (50)$$

An examination of Fig. 7 reveals that as ε is decreased (i.e. as the constraints on the roughnesses are relaxed), as expected the solutions are rougher and the residuals are smaller. Typically, the curves exhibit ‘knee points’ beyond which, if the constraints are relaxed further, the RMS roughnesses rapidly become very large. This rapid increase in roughness is an indication of an insufficient level of constraint. Conversely, increases in ε lead to large increases in RMS residuals. A rapid increase in residuals is an indication of too high a level of constraint. The value of ε at the knee point is thus a good estimate of the optimal level of constraint.

We advocate the use of the trade-off curves such as those in Fig. 7 in conjunction with an examination of the actual parameter estimates in order to select the parameter, ε . For the model problem, the trade-off curves show that, for low values of κ , there is a knee point in the range $3.0 < \varepsilon < 30.0$. In order to choose a value from this region, we examined the actual images obtained from the solutions. The display shown in Fig. 8(a) depicts 12 velocity tomograms [the monopole velocity (α) images] obtained from 12 of our test solutions. The unacceptable roughness in the low ε solutions is evident. This roughness shows up in part as artifacts in the regions close to the source and receiver locations (the ‘aliasing artifacts’). The low ε solutions also display lobes of high velocity in the top right-hand and bottom left-hand corners. These lobes are required to account for the traveltimes if the anisotropy parameters are not resolved (the ‘unresolved parameters artifacts’). For any value of κ , as ε is increased, first the aliasing artifacts disappear, then the unresolved parameters artifacts disappear. We selected the value $\varepsilon = 7.0$ based on the appearance of the top row of tomograms in Fig. 8(a).

For a given value of ε changing κ does not change the solution roughness, but results in an increased level of residuals. Nonetheless, a given level of residuals can always be obtained for any value of κ by relaxing the roughness constraint. Fig. 8(b) shows the 12 anisotropy tomograms (the images of ϵ) obtained from the same 12 solutions. The anisotropy of the solution at ($\varepsilon = 10.0$, $\kappa = 0.0$) is reasonably homogeneous and close to the true value of 10 per cent. As the roughness constraint is relaxed, the solutions display incorrect and inhomogeneous anisotropies. As κ is increased, as expected, the solutions show less and less anisotropy. The corresponding velocity tomograms on Fig. 8(a) for increasing κ values display artifacts in the form of incorrect high-velocity lobes. These lobes can often be diagnostic of an incorrect anisotropy in the solution.

In order to select an appropriate value for κ , we show one final set of plots. Fig. 8(c) is a display of all the data residuals for the 12 solutions. Each pixel in each of these 12 panels is coloured according to the traveltimes residual for a single source–receiver pair. The pixels are arranged so that

the source depth decreases from left to right along the horizontal axis, and the receiver depth decreases from top to bottom along the vertical axis. This format of arranging source–receiver locations is commonly known as a ‘stacking chart’ and is extensively used in surface reflection data processing. For all solutions in which a non-zero value of κ was used, there is a systematic appearance to these residuals: there are bands of positive and negative residuals that correspond to source–receiver pairs that have a common ray angle. For example, source–receiver pairs of equal depths appear along the diagonal—these pairs are all linked by horizontal rays. Because these data contain only computation errors due to machine rounding, a valid inversion should yield residuals with no such systematic patterns. The presence of residuals that are correlated with a ray angle is indicative of an incorrect anisotropy in the solution. Only at the $\kappa = 0.0$ level do the residuals show the expected random behaviour. Displaying all the traveltimes residuals in plots such as these is far more diagnostic than simply computing the RMS level of the residuals.

In this section we have advocated selecting the regularization parameters partly according to the appearance of a suite of tomograms. This may strike mathematical purists as lacking in rigour. Nonetheless, this approach is consistent with the philosophy of the stochastic inverse, in which an interpreter would not be given a single result, but a suite of possible results each with a given probability value. We have not been able to compute quantitative probability estimates, due to an inadequate knowledge of either the data statistics or the parameter statistics. However, we have shown how several indicators can be used together to qualitatively estimate the most probable solution from a suite of results. The interpreter is free to pursue the interpretation based on any or all of these results. The result that is chosen will reflect the interpreter’s *a priori* understanding of the geological setting. In terms of the larger objective of creating a geological model based on all available information, this approach is entirely satisfactory.

4.4 Results

Figure 9 displays the complete solution we obtained from the synthetic data, with the regularization parameters $\kappa = 0$ and $\varepsilon = 7.0$. The velocity estimates (Fig. 9a) correspond to the true values of $\alpha = 3.113$ for the isotropic component and 3.19 for the horizontal velocity. The location but not the magnitude of the velocity anomaly has been recovered. The absence of the rough components in the solution has resulted in a smoothing of the anomaly and an error in the peak value of the velocity anomaly. The magnitudes of the anisotropy parameters (Fig. 9b) also correspond to the true values of $\epsilon = 10$ per cent and $\delta = -10$ per cent. Fig. 9(b) shows how remarkably accurate the orientations of the symmetry axis within the source receiver plane (the polar angle, θ) have been estimated.

5 APPLICATION TO FIELD DATA

5.1 Geological setting and survey specification

In 1988 the British Geological Survey (BGS) carried out a series of borehole experiments at Purton, in the UK

(McCann *et al.* 1989). The boreholes penetrated the Oxford Clay, an Upper Jurassic clay formation outcropping in the south of England. At the location of the survey, the clay is about 100 m thick and overlies a limestone layer. The Oxford clay is largely homogeneous and plane layered, but there are stringers of variable sand content that influence its elastic properties significantly. The clay has been determined by a number of experiments to exhibit transverse isotropy with near-vertical symmetry axes (King 1983; Barnes 1984; Kerner, Dyer & Worthington 1989). The anisotropy factor is of the order of 10 per cent.

Although the BGS sought to investigate attenuation, they also carried out a traveltime experiment using two boreholes 60 m deep and 20 m apart. The tomography data consisted of 210 arrival times obtained using 15 sources in one of the boreholes and recorded at 14 receivers in the other borehole. The source, known as a 'sparker', was a downhole electrode assembly across which a capacitor bank was discharged (see McCann, Grainger & McCann 1975, for a description of the source). The sparker generates frequencies between 0.5 and 2 kHz. The receivers used in this experiment were hydrophones.

5.2 Straight ray isotropic tomography results

In the study by McCann *et al.* (1989), the Purton crosshole data were used to form velocity tomograms. McCann *et al.* found that their initial image, formed without taking anisotropy into account, bore little relationship to the known stratigraphy. In Fig. 12(a) we show a similar image, also formed using isotropic tomography. Fig. 12(a) bears similarities to some of the underconstrained images shown in the bottom row of Fig. 8(a). In this case the high-velocity lobes are distributed symmetrically, leading us to expect a nearly vertical symmetry axis.

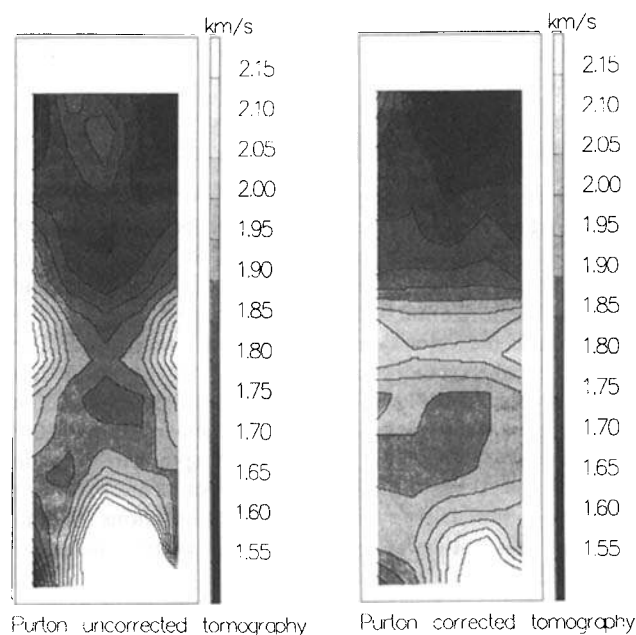


Figure 12. Isotropic tomographic images from the Purton data. (a) The velocity tomogram obtained without accounting for anisotropy. (b) The velocity tomogram obtained by correcting arrival times under the assumption of a 10 per cent, homogeneous, elliptical anisotropy.

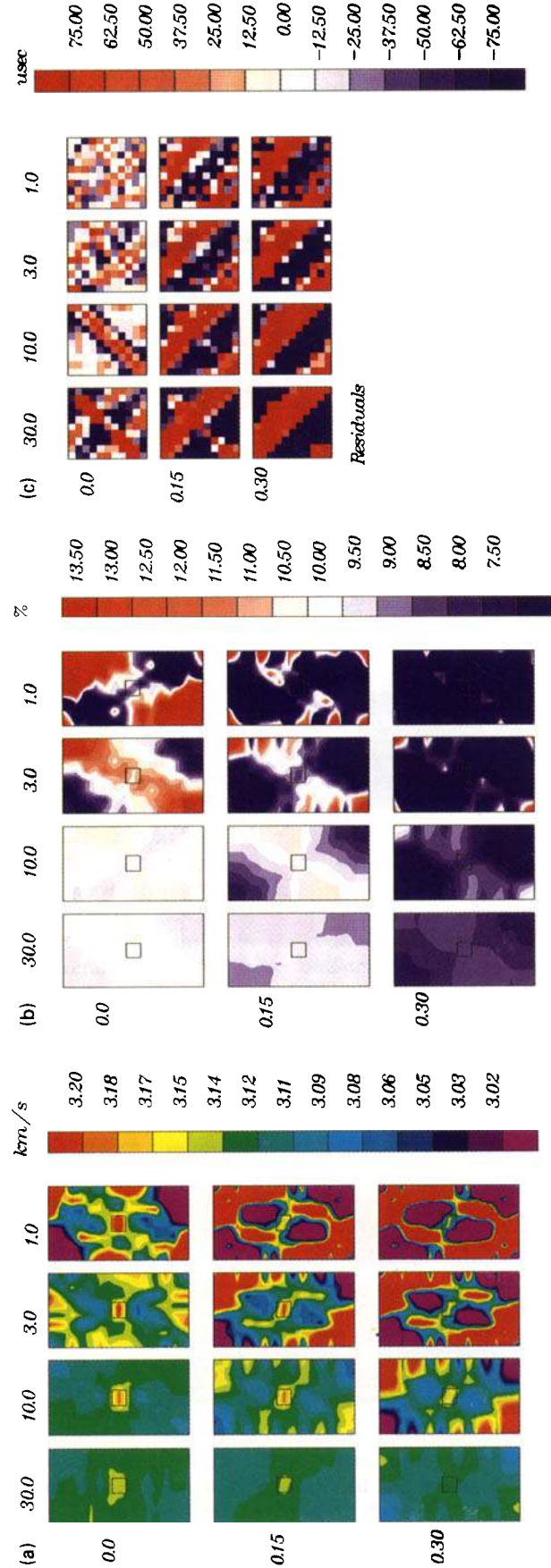
Following this demonstration of the serious artifacts that result when isotropy was assumed, McCann *et al.* then showed that, by assuming a uniform, elliptical ($\epsilon = 10$ per cent, $\delta = 10$ per cent) anisotropy (with a vertical axis of symmetry), they could correct the measured times according to assumed ray paths. Using these corrected data a more plausible image was obtained. We also tried this approach. The result is shown in Fig. 12(b). The image shows the layered stratigraphy that is expected in the Oxford clay. The high-velocity zone in the centre of the images corresponds lithologically to a zone of higher sand content. The tomogram in Fig. 12(b) effectively demonstrates the improvement in stratigraphic imaging obtained using a very crude assumption about the anisotropy. There is an indication, especially at the top and bottom of the image, that the anisotropy assumption is not completely correct. In the next section we show that the data support a more rigorous approach.

5.3 Parametrization and regularization

For the purposes of the anisotropic tomography inversion of the Purton data, the model was parametrized using a grid 8 cells wide and 25 cells high, yielding a total of 200 cells. A constant velocity, isotropic model with $\alpha = 1.81 \text{ km s}^{-1}$ was the initial model for the computation of the partial derivatives \mathbf{F} and the data differences $\delta\mathbf{T}$. We again used six $\delta\bar{q}_i$ parameters at each cell location in order to separately constrain the isotropic and anisotropic parts of the solution. The total number of parameters used was therefore 1200. In selecting the regularization parameters for the data set, ϵ and κ , we followed the methodology established in Section 4.3 with the synthetic data; i.e., use a number of indicators derived from a suite of solutions with different regularization levels.

Figure 13 shows the trade-off curves obtained for all combinations of ϵ and κ that were tested. The knee points of the solution roughness curves occur in the vicinity of $3.0 < \epsilon < 7.0$. The appropriate level of the anisotropy penalty must be estimated by evaluating the images and the data residuals, shown in Fig. 10. It is evident from Fig. 10(a) (the velocity images) that, as the anisotropy penalty is increased, the velocity structure becomes less plausible geologically. The ϵ images, Fig. 10(b), show that, for $\kappa = 0.01$ (the top row), high values of ϵ (up to 20 per cent) are obtained near the bottom of the survey. As κ is increased, the anisotropy at the bottom decreases and a compensating, high-velocity lobe appears at the bottom of the velocity images. As the $\epsilon = 0.7$, $\kappa = 1.0$ level, the velocity tomogram is very similar to the isotropic tomography result, Fig. 12(a).

The sensitivity of the inversions to the choice of the regularization parameters can be judged qualitatively by examining Figs 10(a) and (b). The plane layered nature of the solutions is quite robust and is lost only when the anisotropy penalty terms become large. The magnitudes of the anisotropies are also insensitive to ϵ and κ at the centre of the survey region. At the top and bottom of the survey, where the ray coverage is poor, the anisotropy images are sensitive to the penalty terms. However, the pattern of increasing anisotropy from top to bottom is present on almost all of these images, as is the sharp increase in anisotropy below the high-velocity layer.



Synthetic data: parameter tests (velocities)

Figure 8. 12 solutions to the synthetic data problem, corresponding to decreasing ϵ (left to right) and increasing κ (top to bottom). (a) The velocity images. (b) The anisotropy (ϵ) images. (c) The data residuals for each source receiver pair plotted in a 'stacking chart' format (see text).

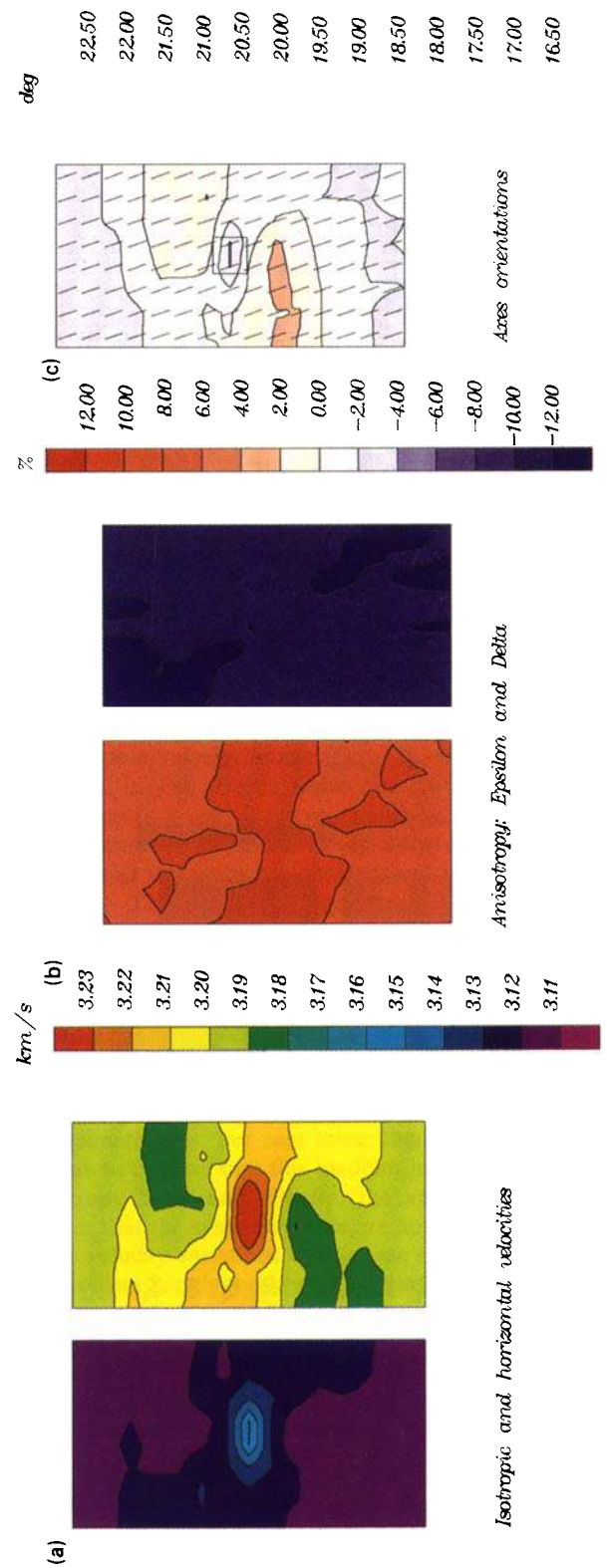


Figure 9. Result obtained from the synthetic data, computed with regularization parameters $\epsilon = 7.0$ and $\kappa = 0.0$. (a) The isotropic velocity and horizontal images. (b) The ϵ and δ images and (c) the estimated symmetry axis orientations (θ). In (c) the length of the markers is proportional to the magnitude of the anisotropy factor, ϵ .

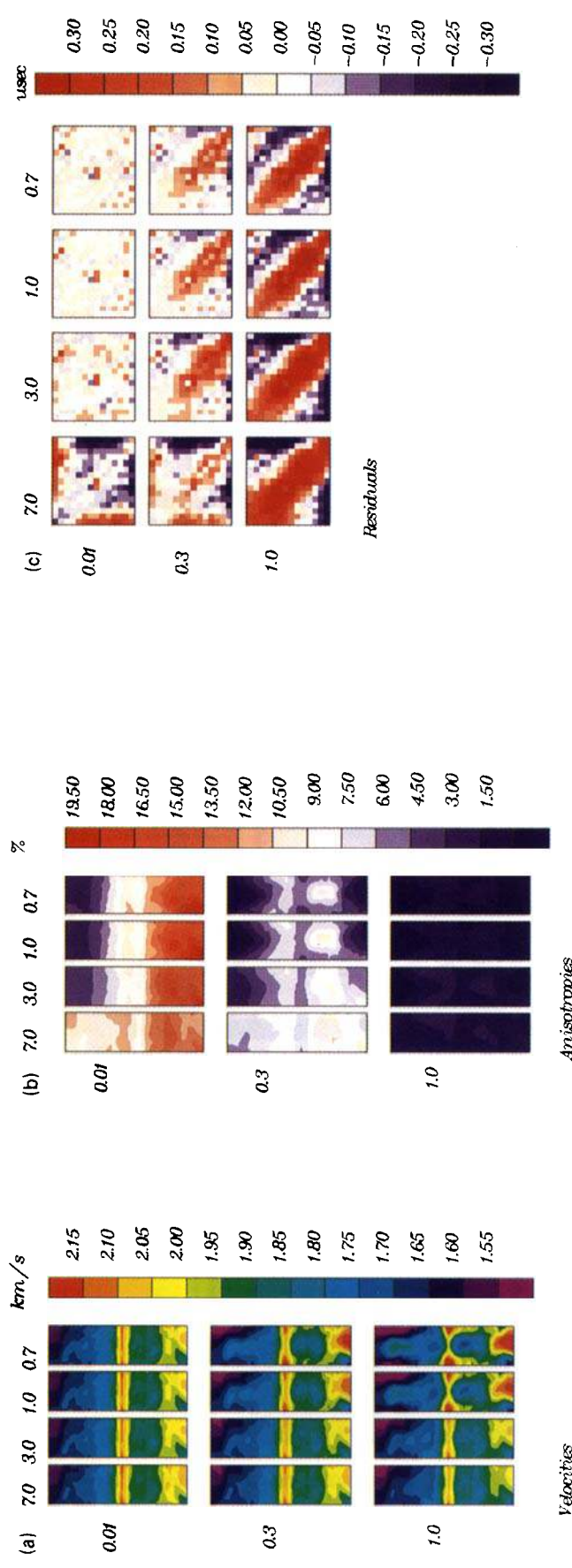


Figure 10. 12 solutions to the traveltimes inversion of the Purton field data, corresponding to decreasing ϵ (left to right) and increasing κ (to to bottom). (a) The velocity images. (b) The anisotropy (ϵ) images. (c) The data residuals for each source receiver pair plotted in a 'stacking chart' format.

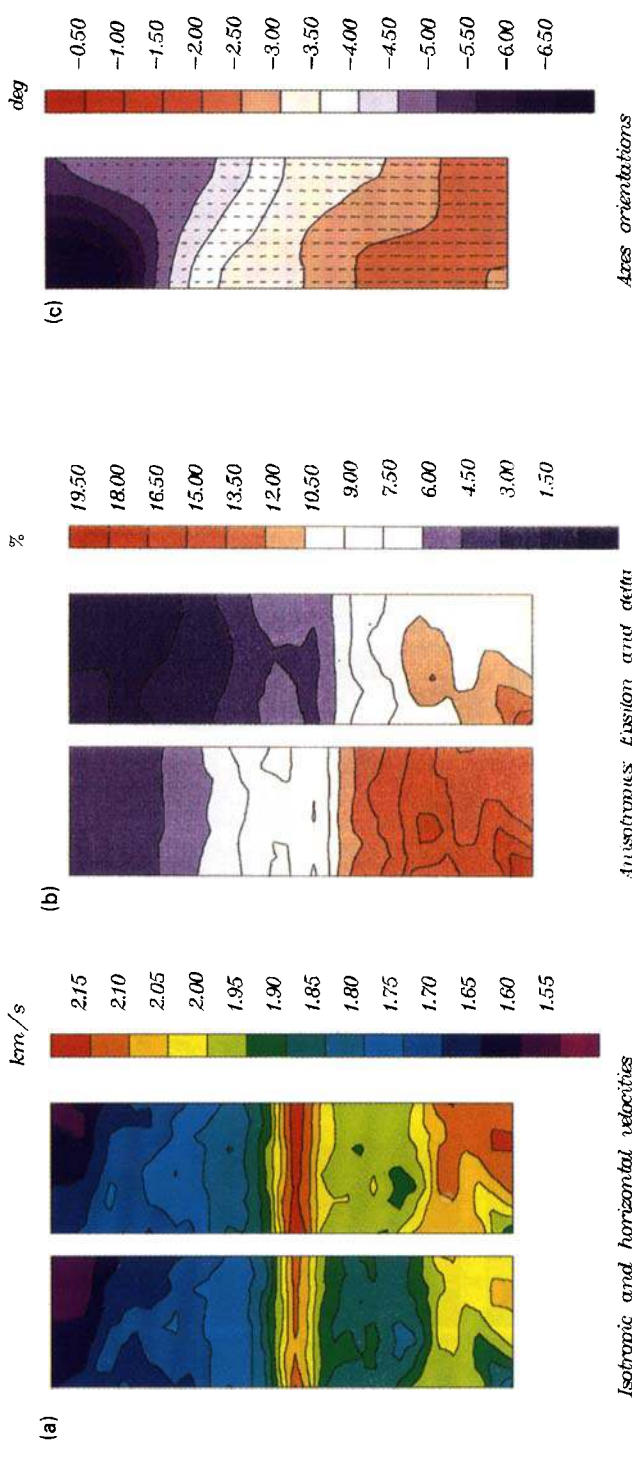


Figure 11. Anisotropic tomography results using the Purton data. The solutions were computed using regularization parameters $\epsilon = 3.0$, $\kappa = 0.01$. (a) The isotropic velocity and horizontal velocity images. (b) The ϵ and δ images and (c) the estimated symmetry axis orientations (θ). In (c) the length of the markers is proportional to the magnitude of the anisotropy factor, ϵ .

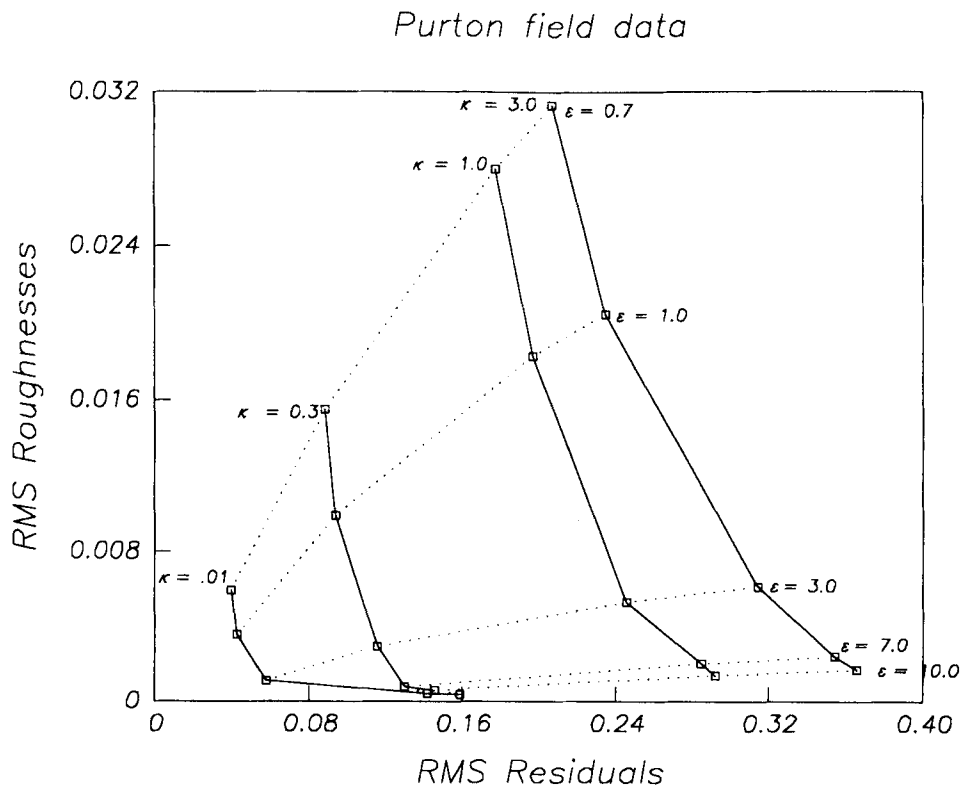


Figure 13. Solution roughness trade-off curves for the Purton field data. A suite of solutions for varying values of roughness penalty (ϵ) and anisotropy penalty (κ) were analysed.

All the traveltimes residuals for these 12 solutions for the Purton data are shown in Fig. 10(c). The display is exactly equivalent to Fig. 8(c). As in Fig. 8, for all solutions in which κ was greater than 0.01, there is a systematic appearance to the residuals: there are bands of positive and negative residuals that correspond to source–receiver pairs that have a common ray angle. In the vicinity of the knee point $\epsilon = 3.0$, $\kappa = 0.01$ the residuals can be seen to contain several ‘outliers’. These outliers are due to traveltimes picking errors. Most of these are positive, which is consistent with the expectation that traveltimes pick underestimate the true times in difficult data areas. The least-squares formulation places too much weight on the outliers, which do not conform to the assumption of Gaussian data statistics. Rather than using a lower value of ϵ to reduce the outliers, we prefer to avoid fitting these. This is an ad hoc approach; a more rigorous procedure is to reweight the residuals according to their fit, and resolve the equations. This latter procedure, known as iteratively reweighted least squares (IRLS), has been shown (Schlossmacher 1973; Scales *et al.* 1988) to be equivalent to using L_p norms in the objective function.

5.4 Anisotropic tomography results

Using the arguments presented in the previous section, we decided to proceed with the regularization parameters $\epsilon = 3.0$ and $\kappa = 0.01$. The system was relinearized, the partial derivatives were recomputed and the linear equation system was resolved. A total of three relinearizing steps were used, after which the solutions were essentially stable (i.e. the solutions, $\delta\bar{q}$, contained no further isotropic component).

The results for straight rays are shown in Fig. 11. The monopole velocity image (Fig. 11a) is similar to the image shown in Fig. 12(b), which was obtained by assuming a uniform ($\epsilon = \delta = 10$ per cent) anisotropy with a vertical axis of symmetry. The image in Fig. 11(a) no longer contains the low- and high-velocity lobes present at the top and bottom of Fig. 12(b). The ϵ images show low anisotropies in the near-surface sediments, with increasing values of ϵ at the bottom of the model. Thus, it would appear that the image in Fig. 12(b) was overcorrected for anisotropy in the shallow region and undercorrected at depth. It is interesting to note that the anisotropy appears to increase sharply below the high-velocity (sandier) layer within the Oxford Clay. The values of δ obtained from the inversion are everywhere approximately 4–5 per cent lower than the values of ϵ ; this appears to confirm that this medium has almost an elliptical anisotropy behaviour. Finally, the values of θ recovered confirm that the symmetry axes are approximately vertical.

The velocity variations in this case study are limited, and the problem is only weakly non-linear. Although not shown here, tests were carried out using a curved ray approach. These tests yielded images very similar to those shown in Fig. 11. In areas with more dramatic variations in velocities a curved ray approach will be required to account for the non-linear dependency of the ray paths on the solution.

6 DISCUSSION

In Paper I we proposed theories for 2-D tomography in anisotropic media using both qP - and qS -waves. In this paper we have investigated the application of the theory for qP -wave tomography. As expected, the linear systems are extremely ill-conditioned, and appropriate constraints are

required in order to extract meaningful solutions. Using singular value decompositions (SVD) of the linear systems, we have identified the expected spatial heterogeneity, or roughness, of the medium as an important constraint. The geometry of the ray paths defines an upper bound of heterogeneity that can be resolved; media with finer grained heterogeneity than this upper bound are undersampled, and the solutions are subject to alias errors. In anisotropic tomography the scalelength at which aliasing occurs is correspondingly longer than in isotropic tomography, given the same set of rays. If this scalelength is larger than the Fresnel zone width of the wave propagation, alias errors will be significant. We believe that this consideration should be incorporated into the specification of crosshole surveys.

Following the SVD analysis of the linear systems, we gave a regularization scheme that included a penalty term for solutions with large roughnesses. Using a synthetic example, we showed how important this constraint is in allowing features of the true structure to be recovered. We found that penalizing the second derivative of the solution was insufficient: we also required a term in the regularization that penalized large gradients in the solutions. In selecting the regularization parameters (i.e. the level of the smoothness constraints) for the synthetic data inversion, we found it essential to analyse a suite of solutions. A band of acceptable regularization levels can be determined by examining plots of solution roughness against the data fit and searching for knee points. For fine tuning one has little recourse but to examine a suite of images and select a result on the basis of geological plausibility. In determining a level of constraint for anisotropy penalty terms, the examination of plots of all traveltimes residuals are invaluable.

In the final section of this paper we applied our numerical scheme to crosshole arrival times measured in near-surface clay sediments known to be anisotropic. We found that we could obtain useful estimates of the spatial distributions of the velocity and of the anisotropy. The images indicate that the anisotropy of the sediments increases from the top of the geological column to the bottom, with a sharp increase in anisotropy below a high-velocity layer associated with a higher sand content. This description of the heterogeneity of the anisotropy was not available previously.

In this paper we have limited our investigations to TI media and to problems in which the nonlinear dependence of the ray paths on the solutions can be ignored. We consider that the nonlinear aspects of the problem will be soluble in a straightforward manner using techniques already well established in the isotropic tomography problem (for example, Bregman *et al.* 1989a). Our present research is directed toward incorporating these 'curved ray' methods and accounting for more general symmetry systems, with a view to solving more general tomography problems. The application of the theory for shear wave tomography (given in Paper I) will also be an important area for future research, although, as we indicated in the Introduction, this may have to await further progress in the technology of borehole seismic measurements.

ACKNOWLEDGMENTS

The research in this paper was partially supported by a National Science and Engineering Research Council (Canada) Operating Grant OGP0009130. The authors also

wish to thank Amoco Production Company, Tulsa and Amoco Canada Petroleum, Canada for the financial support provided. We also wish to thank Schlumberger Cambridge Research for helping to defray the costs of the colour pages in this paper. Dr John Scales of Amoco provided useful comments to us on regularization schemes for multiparameter problems. The help of Dr D. M. McCann of the British Geological Survey (BGS) and the kind permission of the BGS to use the Purton field data is gratefully acknowledged. One of us (RGP) holds an NSERC Postdoctoral Fellowship at the University of Toronto. We gratefully acknowledge the facilities provided by Bullard Laboratories, Department of Earth Sciences, University of Cambridge while one of us (CHC) was on leave.

REFERENCES

- Backus, G. E., 1970. A geometrical picture of anisotropic elastic tensors, *Rev. Geophys. Space Phys.*, **8**, 633–671.
- Barnes, S. R., 1983. Seismic anisotropy in the Oxford Clay at Purton, near Swindon, *MSc thesis*, University of Birmingham.
- Bois, P., La Porte, M., Lavergne, M. & Thomas, G., 1971. Essai de détermination automatique des vitesses sismiques par mesures entre puits, *Geophys. Prosp.*, **19**, 42–83.
- Bois, P., La Porte, M., Lavergne, M. & Thomas, G., 1972. Well-to-well seismic measurements, *Geophysics*, **37**, 471–490.
- Bording, R. P., Gersztenkorn, A., Lines, L. R., Scales, J. A. & Treitel, S., 1987. Applications of seismic traveltimes tomography, *Geophys. J. R. astr. Soc.*, **90**, 285–303.
- Bregman, N. D., Bailey, R. C. & Chapman, C. H., 1989a. Crosshole seismic tomography, *Geophysics*, **54**, 200–215.
- Bregman, N. D., Bailey, R. C. & Chapman, C. H., 1989b. Ghosts in tomography: the effects of poor angular coverage in 2-D seismic traveltimes inversion, *Can. J. Expl. Geophys.*, **25**, 7–27.
- Bregman, N. D., Hurley, P. A. & West, G. F., 1989c. Seismic tomography at a fire flood site, *Geophysics*, **54**, 1082–1090.
- Červený, V., 1972. Seismic rays and ray intensities in inhomogeneous anisotropic media, *Geophys. J. R. astr. Soc.*, **29**, 1–13.
- Červený, V. & Jech, J., 1982. Linearized solutions of kinematic problems of seismic body waves in inhomogeneous slightly anisotropic media, *J. Geophys.*, **51**, 96–104.
- Chapman, C. H. & Pratt, R. G., 1992. Traveltimes tomography in anisotropic media—I. Theory, *Geophys. J. Int.*, this issue.
- Constable, S. C., Parker, R. L. & Constable, C. G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**, 289–300.
- Cottin, J.-F., Deletie, P., Jaquet-Francillon, H., Lakshman, J., Lemoine, Y. & Sanchez, M., 1986. Curved ray seismic tomography: Application to the Grand Etang Dam (Reunion Island), *First Break*, **4**, 25–30.
- Crampin, S., 1987. The geological and industrial implications of extensive dilatancy anisotropy, *Nature*, **328**, 491–496.
- Dyer, D. & Worthington, M. H., 1988a. Some sources of distortion in tomographic velocity images, *Geophys. Prosp.*, **36**, 209–222.
- Dyer, B. C. & Worthington, M. H., 1988b. Seismic reflection tomography: a case study, *First Break*, **6**, 354–366.
- Federov, F. I., 1968. *Theory of Elastic Waves in Crystals*, Plenum Press, New York.
- Frasier, C. & Winterstein, D., 1990. Analysis of conventional and converted mode reflections at Putah Sink, California using three-component data, *Geophysics*, **55**, 646–659.
- Jech, J. & Pšenčík, I., 1989. First-order perturbation method for anisotropic media, *Geophys. J. Int.*, **99**, 369–376.
- Kerner, C., 1989. Anisotropy and scattering, *Expanded Abstracts of the 58th SEG International Meeting*, pp. 1034–1038.
- Kerner, C., Dyer, B. & Worthington, M., 1989. Wave propagation

- in a vertical transversely isotropic medium: field experiment and model study, *Geophys. J. Int.*, **97**, 292–309.
- King, R. F., 1983. Seismic anisotropy in the Oxford Clay at Purton/Swindon, *Internal Report*, University of Birmingham.
- Lanczos, C., 1950. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. NBS*, **45**, 225–282.
- Lanczos, C., 1961. *Linear Differential Operators*, Van Nostrand, Princeton, NJ.
- LaPorte, M., Lakshmanan, J., Lavergne, M. & Willm, C., 1973. Seismic measurements by transmission—Application to civil engineering, *Geophys. Prosp.*, **21**, 146–158.
- Leung, L. & Downey, M., 1988. Cross-hole seismic tomography for mineral exploration and mine planning, *Expanded Abstracts of the 58th SEG International Meeting*, pp. 328–330.
- Luo, Y. & Schuster, G. T., 1991. Wave equation travel time inversion, *Geophysics*, **56**, 645–653.
- Lynn, H. B. & Thomsen, L. A., 1990. Reflection shear-wave data collected near the principal axes of azimuthal anisotropy, *Geophysics*, **55**, 147–156.
- Mason, I. M., 1981. Algebraic reconstruction of a two-dimensional velocity inhomogeneity in the High Hazles seam of Thoresby colliery, *Geophysics*, **46**, 298–308.
- McCann, D. M., Grainger, P. & McCann, C., 1975. Inter-borehole methods acoustic methods and their use in engineering geology, *Geophys. Prosp.*, **23**, 49–65.
- McCann, C., Assefa, S., Sothcott, J., McCann, D. M. & Jackson, P. D., 1989. In-situ borehole measurements of compressional and shear wave attenuation in Oxford clay, *Sci. Drilling*, **1**, 11–20.
- Paige, C. C. & Saunders, M. A., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software*, **8**, 43–71, 195–209.
- Paulsson, B. N. P., Cook, N. G. W. & McEvilly, T. V., 1985. Elastic-wave velocities and attenuation in an underground granitic repository for nuclear waste, *Geophysics*, **50**, 551–570.
- Peterson, J. E., Tura, A., Maer, E. L. & Davey, A., 1990. Practical aspects of crosswell tomographic surveys, presented at the *International Symposium on Borehole Geophysics: Petroleum, Hydrogeology, Mining and Engineering Applications*, Soc. Expl. Geophys.
- Saito, H., Shima, H., Toshioka, T. & Ohtomo, H., 1988. Application of geotomography techniques to site investigations for civil engineering purposes, *Expanded Abstracts of the 58th SEG International Meeting*, pp. 324–327.
- Scales, J. A., Gersztenkorn, A. & Treitel, S., 1988. Fast l_p solution of large, sparse, linear systems: application to seismic travel time tomography, *J. Comp. Phys.*, **75**, 314–333.
- Scales, J. A., Docherty, P. & Gersztenkorn, A., 1990. Regularisation of nonlinear inverse problems: imaging the near-surface weathering layer, *Inv. Prob.*, **6**, 115–131.
- Schlossmacher, E., 1973. An iterative technique for least absolute deviations curve fitting, *J. Am. Stat. Assoc.*, **68**, 857–865.
- Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*, Elsevier, Amsterdam.
- Thomsen, L., 1986. Weak elastic anisotropy, *Geophysics*, **51**, 1954–1966.
- Tikhonov, A. & Arsenin, V., 1977. *Solution of Ill-Posed Problems*, Winston, Washington, DC.
- Wong, J., Hurley, P. & West, G. F., 1983. Crosshole seismology and seismic imaging in crystalline rocks, *Geophys. Res. Lett.*, **10**, 686–689.
- Worthington, M. H., East, R. J. H., Kerner, C. K. & O'Donovan, A. R., 1989. Limitations of ray theoretical tomographic imaging of crosshole seismic data for monitoring EOR flooding, *Sci. Drilling*, **1**, 47–53.
- Wright, J., Harrell, C., Wong, J., Nelson, J. S. & Gibbs, D., 1988. Crosshole seismic tomography for engineering evaluation at Theodore Roosevelt Dam, Arizona, *Expanded Abstracts of the 58th SEG International Meeting*, pp. 320–323.