Seismic Borehole Tomography—Theory and Computational Methods

SVEN IVANSSON

Invited Paper

Tomographic inversion can be applied to seismic traveltime data to obtain a map of seismic velocities in a rock volume, although the field geometries will not, in general, allow a complete ray sampling. This paper deals with theoretical and computational aspects of tomographic inversion under such circumstances. It is shown that some usual measurement geometries could, in theory, be sufficient for determining the seismic velocity distribution uniquely (under the straight-ray approximation). From the computational point of view, however, the inverse problem can be poorly conditioned. The usual transform methods of reconstruction, developed for medical tomography, are not applicable. A brief discussion is, therefore, given of a few relevant geophysical inverse techniques and the CG-algorithm applied to a certain quadratic functional is shown to give reasonable images. A convergence result that contains "ART with smoothing" as a special case is also obtained. Techniques to correct for ray bending are discussed.

I. Introduction

This paper deals with theoretical and computational aspects of tomographic inversion of seismic traveltime data. Basically, the idea is to reconstruct a function, the seismic slowness (reciprocal velocity) distribution within a rock volume, from its integrals (the traveltimes) over many crossing paths (the rays). The specific application we have in mind concerns investigation of a rock volume, situated between boreholes and the surface of the earth, using seismic sources and receivers in the boreholes and on the surface. Field test results are presented in the accompanying paper [18].

The reconstruction problem is threefold. To what extent the slowness function is determined by the data must be found out, suitable inversion algorithms must be devised, and resolution as well as confidence properties of a computed solution must be assessed. We start by introducing some appropriate notation.

Let $x = (x_1, x_2, x_3)$ denote coordinates within the rock volume and let $m(x)$ be the unknown slowness function. Suppose there are N data values, i.e., traveltimes, which we denote by d_i, $i = 1, \cdots, N$. Each d_i may be written (high-frequency approximation)

$$d_i = g_i(m) \overset{\text{def}}{=} \int_{T_i(m)} m(x) \cdot ds$$

where g_i is a nonlinear functional, ds is arc length, and $T_i(m)$ denotes the curve, connecting source, and receiver, which yields the least possible traveltime value (cf. Fermat's principle). The nonlinearity is due to the complex dependence of $T_i(m)$ upon m. With the obvious vector notation we will also write $d = g(m)$.

In certain cases it may be necessary to allow for an anisotropic structure ([30], [28]) but except for Example 8 we will not discuss that complication.

With a suitable choice of space for the slowness models m, the functionals g_i will become Fréchet differentiable. This follows, using a physical argument, from Fermat's principle. We obtain

$$g_i(m + \delta m) - g_i(m) = \int_{T_i(m)} \delta m(x) \cdot ds$$

when δm is small (in the appropriate norm). It follows that when m is approximately proportional to $\overline{m}$ where $\overline{m}$ is an *a priori* model, the system $d = g(m)$ may be linearized by

$$d_i \approx \int_{T_i(\overline{m})} m(x) \cdot ds.$$

For a homogeneous $\overline{m}$ the paths $T_i = T_i(\overline{m})$ will be straight.

Section II contains some simple but interesting corollaries to uniqueness theorems for the "X-ray transform," whereas Sections III–VI are devoted to the computational methods.

II. Uniqueness Properties of the Inverse Problem Assuming an Infinite Data Set

It is of interest to know if a proposed basic measurement geometry could, in principle with an infinite data set, allow a unique determination of $m(x)$ by the corresponding traveltimes. In this section such questions will be discussed (we will here denote the unknown function by f instead of m).

Manuscript received March 1, 1985; revised July 27, 1985. This work was supported by the International Stripa Project.

The author is with FOA (National Defense Research Institute), Division 29, S-102 54 Stockholm, Sweden.

0018-9219/86/0200-0328$01.00 ©1986 IEEE

At first we consider the linearized problem with straight paths, thus the integrals of the sought-for function $f(x)$ along certain straight-line segments are known. More generally, one can consider the transformation which maps a function on R^n to its integrals over k-dimensional hyperplanes. For $k = 1$ ($k = n - 1$) one speaks of the X-ray transform (Radon transform). Note that in R^2 the X-ray and Radon transforms coincide. In the following we will only consider $k = 1$ and $n = 2$ or 3 although the results are more generally valid.

For $\omega \in S^{n-1}$ (the unit sphere in R^n) the parallel-beam transform $P_\omega f$ is defined by

$$P_\omega f(x) = \int_{-\infty}^{+\infty} f(x + t \cdot \omega) \cdot dt$$

for $x \in \omega^\perp$ (the vectors orthogonal to ω).

For $a \in R^n$, the divergent-beam transform $D_a f$ is defined by

$$D_a f(\omega) = \int_0^{+\infty} f(a + t \cdot \omega) \cdot dt, \qquad \text{for } \omega \in S^{n-1}.$$

At first, we collect a few theorems from the literature which are of interest. (The support of a function f, $\operatorname{supp}(f)$, is the closure of the set where it is nonzero.)

Theorem 1: Let $f \in L_1(R^n)$. Then:

a) For each $\omega \in S^{n-1}$, $P_\omega f(x)$ is well-defined for almost all $x \in \omega^\perp$ and $P_\omega f$ is integrable over $\omega^\perp$.
b) For each $a \in R^n - \operatorname{supp}(f)$, $D_a f$ is well-defined for almost all $\omega \in S^{n-1}$ and $D_a f$ is integrable over S^{n-1}. If f has compact support, the conclusion is valid for almost all $a \in R^n$ as well.

The proof is easy using Fubini's theorem and polar variables (e.g., [20]).

Theorem 2: An integrable function f on R^n is uniquely determined by

$$\{ P_\omega f(x) : \omega \in S^{n-1} \cap V, x \in \omega^\perp \}$$

where V is any two-dimensional subspace of R^n.

Proof: See [41, Cor. 1.6]. One uses the projection-slice theorem (e.g., [31]) to reduce to well-known uniqueness results for the Fourier transform.

For the two-dimensional (2D) case ($n = 2$), it is easy to see that line-integral data must be available for essentially all lines in order for an integrable f to be determined uniquely [27, Appendix B]. When f is known to have compact support, however, much less information is needed.

Theorem 3: Let $f \in L_1(R^n)$ have compact support. Then:

a) f is uniquely determined by $\{P_\omega f(x): \omega \in V, x \in \omega^\perp\}$ where V is any infinite subset of S^{n-1}.
b) f is uniquely determined by $\{D_a f(\omega): a \in A, \omega \in S^{n-1}\}$ where A is any infinite subset of R^n known *a priori* to have positive distance to the convex hull of the support of f.

Proof: See [41, Theorem 1.7] and [20, Theorem 5.1].

For a uniqueness result to hold it is crucial here that an infinite number of directions ω (sources a) are present [40, Theorem 4.3] ([20, Theorem 5.13]). A theorem of Logan [32] implies, however, that data for a finite number of directions can constrain f significantly if f is known *a priori* to be reasonably smooth.

In the seismic applications we have in mind, the sources and the receivers are, in general, located on part of the boundary of the rock volume under study. The slowness distribution outside is then immaterial (in the linearized problem), it can be considered as being identically zero. We are then indeed dealing with a function of compact support and the integrals do not change when the integration lines are (artificially) extended to infinity. However, it should be noted that the uniqueness theorems given so far are not directly applicable to the usual borehole geometries (even with infinite data sets) since all the hypotheses are not fulfilled.

Theorem 4 (The Hole Theorem): Let K be a compact and convex subset of R^n. Let $f \in L_1(R^n)$ have compact support and suppose that the line integral of f is zero when taken over any straight line which does not intersect K. It follows that f vanishes (almost everywhere) outside K.

Proof: A very simple proof is given in [42].

The hole theorem immediately yields results of interest for some possible seismic geometries.

Corollary 5 (2D VSP geometry, VSP means "vertical seismic profiling"): Consider Fig. 1. If traveltime data are

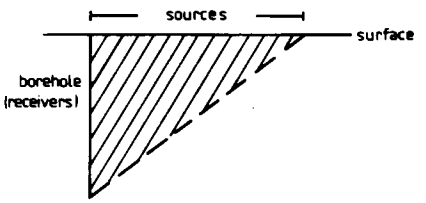

Fig. 1. 2D VSP geometry.

available for all source/receiver combinations (idealized straight paths) when the source is located on the shown part of the surface and the receiver is located in the borehole, then the (integrable) slowness distribution is uniquely determined within the shaded triangle.

Proof: Apply Theorem 4 with K equal to the dashed part of the border of the triangle.

By rotating the triangle around the borehole we easily obtain a corresponding uniqueness result for the three-dimensional (3D) VSP geometry. The reconstructed region is now a cone, Fig. 2.

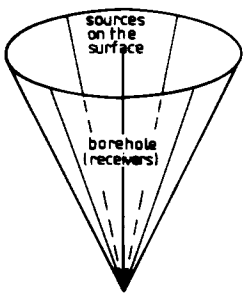

Fig. 2. 3D VSP geometry.

The argument used by Strichartz for the hole theorem can actually be used to prove more. The following result, for example, is easily obtained.

Proposition 6 ("one-sided" 2D crosshole geometry): Consider Fig. 3. If traveltime data are available for all

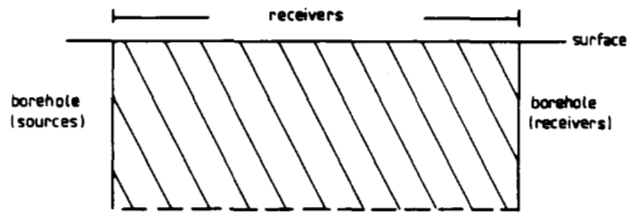

Fig. 3. "One-sided" 2D crosshole geometry.

source/receiver combinations (idealized straight paths) when the source is located in the left borehole and the receiver is located in the right borehole or on the shown part of the surface, then the (integrable) slowness distribution is uniquely determined within the shaded crosshole region.

In practice it is not enough that there exists a unique solution for an inverse problem to be manageable. The solution must also be a smooth function of the data so that the impact of measurement errors is not too large. In this respect, the borehole geometries discussed above are not particularly favorable. It would be much better if the lower border of the region could be made accessible which would allow a more complete coverage by ray paths in different directions. Iterative solution methods would also converge more rapidly in that case. These aspects will be considered in more detail in Sections IV and V. For the moment, we give an example to illustrate that the conclusion in Proposition 6 is not valid if the hypothesis is weakened slightly.

Example 7: Suppose that $x^0 \in R^2$ and that θ_1, θ_2 are two real numbers (angles) such that $\theta_1 < \theta_2 < \theta_1 + \pi$. Let φ be an $L_1(R)$ function with support in $[\theta_1, \theta_2]$ such that

$$\int \varphi(\theta) \cdot d\theta = 0.$$

Choose "arg" (the polar angle) in the interval $[\theta_1, \theta_1 + 2\pi)$ and introduce a function f on R^2 by

$$f(x) = \frac{\varphi(\arg(x - x^0))}{|x - x^0|^2}, \qquad \text{for } x \neq x^0.$$

Then as is easily verified

$$\int_L f(x) \cdot ds = 0$$

for any line L which intersects both the legs of the convex angular region formed by the apex x^0 and the directions θ_1, θ_2 (Fig. 4). Note that f has a nonintegrable singularity at x^0. We realize the kinds of nonuniqueness that appear when data are missing for all integration paths that pass close to a certain point (x^0) between the boreholes.

In the presence of anisotropy, severe nonuniqueness problems may appear. This is indicated by the following example (in reality the angular dependence is more complicated than elliptical, e.g., [30]):

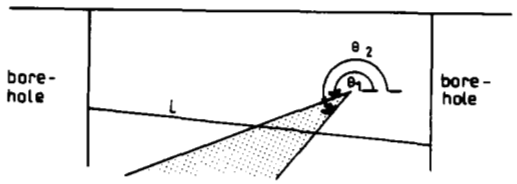

Fig. 4. Illustration for Example 7.

Example 8: Let $s(x, \theta) = f(x) \cdot \cos^2(\theta) + g(x) \cdot \sin^2(\theta)$ where f and g are two functions on R^2. Think of $s(x, \theta)$ as residual slowness at x in direction θ. If $f = \delta^2 h / \delta x_2^2$ and $g = -\delta^2 h / \delta x_1^2$ where h is a compactly supported and twice continuously differentiable function on R^2, then the line integral of $s(\cdot, \theta)$ along any line of direction θ vanishes. The verification is most easily performed in the Fourier domain using the projection-slice theorem.

Let us return to the isotropic case. For pure crosshole geometries, i.e., as Fig. 3 but without receivers on the surface, a high degree of nonuniqueness is present. This is easily seen already from Example 7 (choose the apex x^0 outside the crosshole region). Related to this is the following apparent result, which implies that thin steeply dipping low-velocity strips cannot be discriminated from very thick ones with a very small velocity contrast if the boreholes are vertical. (A vertical strip will then be completely invisible.)

Observation 9: Suppose that $\varphi \in L_1(R)$ vanishes outside the compact interval $[a, b]$ and that

$$\int \varphi(t) \cdot dt = 0.$$

Let $\mathcal{L}$ be an oriented line in R^2 and let $\text{dist}(x, \mathcal{L})$ denote the distance with sign from an arbitrary point x in R^2 to $\mathcal{L}$. Let A and B, the "source and receiver sets," be two bounded subsets of R^2 such that $\text{dist}(x, \mathcal{L}) \leqslant a$ for all $x \in A$ and $\text{dist}(x, \mathcal{L}) \geqslant b$ for all $x \in B$. Introduce a function f on R^2 such that

$$f(x) = \varphi(\text{dist}(x, \mathcal{L})), \quad \text{if } x \in \text{"the convex hull of } A \cup B\text{."}$$

Then

$$\int_L f(x) \cdot ds = 0$$

for any straight-line segment L connecting a point of A to a point of B.

As is seen from synthetic test examples, pure crosshole geometries can, in fact, be quite successful for mapping local velocity anomalies in the interior of the crosshole region. Consider the geometry in Fig. 3 but with the surface receivers removed. If traveltime data are available for all straight borehole-to-borehole rays and if the slowness function is known *a priori* close to the surface and close to the bottom of the crosshole rectangle, then a unique reconstruction is possible. This follows easily from Theorem 3.

So far we have only considered the uniqueness properties of the linearized problem with straight paths. Consider next the generalization to fixed but curved paths. For certain families of integration curves it is possible to formulate a uniqueness theorem ([39], [11], [6], [4A]).

For our original problem, that of solving the nonlinear system $d = g(m)$, the uniquness questions are, of course, still more difficult. In this case, the integration paths depend upon the unknown slowness function $f(x)$. Some

results, although of an implicit character, have been obtained by Russian mathematicians (e.g., [37], [4A]). Here we only make the following two immediate observations.

First, suppose there exists an open subset of the rock volume through which no minimum traveltime ray passes with the chosen source/receiver geometry. Then $f(x)$ is certainly not determined uniquely on this open set, since it may (for example) be increased there without affecting the traveltimes; see Fig. 5.

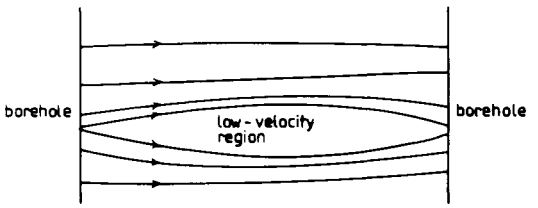

Fig. 5. Example of a low-velocity region that is not sampled with rays.

Second, consider the case of pure crosshole measurement. It is easily realized that a parallel translation of a constant-velocity strip dividing an otherwise homogeneous crosshole region, performed so that the boreholes are never intersected by the strip, will not affect the measured traveltimes; see Fig. 6. Thus the location of a strip of this kind cannot be determined by pure crosshole measurements.

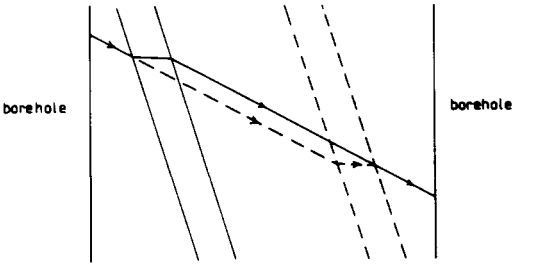

Fig. 6. Pure crosshole reconstruction is nonunique also when Fermat's principle is taken into account.

III. Techniques to Pick a Preferred Solution in the Presence of Nonuniqueness

From the outset we may fix some model space $\mathcal{M}$ to which the unknown slowness function m is assumed to belong. $\mathcal{M}$ will be some kind of infinite-dimensional manifold, for example the space of positive-continuous functions on the rock volume under study. (Since the bandwidth of the seismic signals is not infinite in practice, it is natural to impose smoothness constraints on $\mathcal{M}$.) The inverse problem is to use the data equations $d = g(m)$ to determine those $m \in \mathcal{M}$ which are compatible with certain measured traveltimes. Since the number of data equations is finite, the solutions will actually span an infinite-dimensional submanifold of $\mathcal{M}$ although the results in Section II indicate that for certain situations uniqueness will "appear in the limit" when more and more measurements are made.

As stated, the inverse problem is formidable. In practice, one usually aims at determining one particular "typical" model $\hat{m}$ which satisfies the data equations at least approximately and giving some quantitative measures to indicate how much other possible models may deviate from $\hat{m}$ in different parts of the rock volume. Thus one needs a mapping $F: U \to \mathcal{M}$ where U is some open subset of R^N such that $F(d) = \hat{m}$. Note that the range of F will be some finite-dimensional submanifold of $\mathcal{M}$, thus certainly not all of $\mathcal{M}$.

We will briefly discuss four types of methods to determine a suitable mapping F: transform methods, Backus–Gilbert solutions, the approach of Tarantola and Valette, and series-expansion methods. In the last case, the finite-dimensional range of F is specified directly whereas in the first three cases this choice is implicit ([27]).

Transform methods of reconstruction are very popular in medical X-ray tomography. The idea is to derive an inversion formula for the "continuous case" which is finally discretized (integrals are replaced by sums and so forth) in order to be applicable to real data, see, e.g., [31]. The inversion formulas are derived by transform theory. The standard formulas require a dense and uniform coverage, both angular and lateral, by integration paths. Although some papers dealing with other cases have appeared as well, for example, [25], [33], and [17], it seems as if transform methods are still difficult to use for the irregular and sparse ray-path patterns that often appear in seismic applications.

Backus and Gilbert ([3]) asked what information about the solution manifold can be obtained by combining the data in different ways. They considered linear combinations

$$\sum_{i=1}^{N} a_i \cdot d_i$$

where a_i are coefficients to be varied. Note that (in our application)

$$\sum a_i \cdot d_i = \sum a_i \cdot g_i(m) = \sum a_i \cdot \int_{T_i(m)} m(x) \cdot ds.$$

Thus if we impose the condition $\sum a_i \cdot$ (arc length of $T_i(m)$) $= 1$ then $\sum a_i \cdot d_i$ will be a certain average of $m(x)$. Despite the fact that the solution space is infinite-dimensional, averages of this type are uniquely derivable from the data. Of course, certain averages are particularly interesting. For each point x^0 in the rock volume, one would like to choose a_i so that values of $m(x)$ with x close to x^0 become dominant. Some criterion is needed to achieve this. Backus and Gilbert introduced "δ-ness criteria" to build up as δ_{x^0}-like averages as possible. They considered model functions depending on only one continuous parameter ($n = 1$), however. Since the data functionals involve integration over one-parameter curves, complications will arise when the data are used to build up δ_{x^0}-like functions in higher dimensions. One can, however, make suitable modifications: in [10] "quelling" (by integration) is used, while in [36] the integration curves are replaced by thin strips (for the 2D case). Note that one large equation system must be solved for each point x^0, fortunately it may be possible to use a common system matrix.

Another possibility is to impose additional constraints such as *a priori information* or guesses. This approach was followed in two general papers by Tarantola and Valette ([45], [46]). The particular problem of 3D velocity inversion was treated in [44] and [38]. In this approach, data and model are treated on an equal footing. Let us temporarily change the previous notation and terminology slightly to fit into the framework of Tarantola and Valette. We say that we have *a priori* information about the "ideal data" d as

well as the model m. Our information about d consists of measured values d^0, whereas the information about m consists of some reasonable *a priori* model m^0. Suppose also that uncertainties about d^0 and m^0 can be expressed in terms of a covariance matrix C_{d^0} and a covariance function $C_{m^0}(x, y)$, respectively. Just as C_{d^0} may also be viewed as an operator from R^N to R^N with inverse $C_{d^0}^{-1}$, C_{m^0} may also be viewed as an operator from $\mathcal{M}$ to $\mathcal{M}$ with inverse $C_{m^0}^{-1}$. Using a least squares criterion the problem of estimating the true d and m becomes:

Find the pair (d, m) which minimizes

$$(d^0 - d)^T \cdot C_{d^0}^{-1} \cdot (d^0 - d) + \int\int (m^0(x) - m(x)) \cdot C_{m^0}^{-1}(x, y) \cdot (m^0(y) - m(y)) \cdot dx \cdot dy$$

under the constraint $d = g(m)$.

Tarantola and Valette showed that the solution for m satisfies

$$m = m^0 + C_{m^0} \cdot G^* \cdot (C_{d^0} + G \cdot C_{m^0} \cdot G^*)^{-1} \cdot (d^0 - g(m) + G(m - m^0))$$

where G is the Fréchet derivative of g at m and G^* is the Hilbert space adjoint of G. They also suggested a fixed-point algorithm for solving this equation iteratively. For a linear system we have $g(m) = G(m)$ and the solution is even obtained directly without any iterations. The demands on computing power can, however, be considerable.

In the examples of tomographic calculations presented in this paper and in [18], the series-expansion method (e.g., [8], [22]) is used. Introduce M linearly independent basis functions $\varphi_j(x), j = 1, 2, \cdots, M$. The idea is to seek a slowness model of the form

$$m(x) = \sum_{j=1}^{M} b_j \cdot \varphi_j(x) = b^T \cdot \varphi(x)$$

which is compatible with the data $d = g(m)$. Here b_j are coefficients to be determined and the obvious vector notation has been introduced. Note that the selected model is restricted to the linear span of the φ_j's. It is essential that the basis functions are chosen flexible enough to allow an accurate representation of the actual slowness function.

Using the expansion of m in terms of φ_j, we may rewrite the *linearized* data functionals

$$\begin{aligned} d_i = g_i(m) &\approx \int_{T_i} m(x) \cdot ds \\ &= \sum_{j=1}^{M} b_j \cdot \int_{T_i} \varphi_j(x) \cdot ds \\ &= \sum_{j=1}^{M} G_{ij} \cdot b_j \end{aligned}$$

if the matrix G is defined by

$$G_{ij} = \int_{T_i} \varphi_j(x) \cdot ds.$$

Thus our data simply become N linear equations, $d = G \cdot b$, for M unknown parameters b_j.

In general this system of equations is both overdetermined and underdetermined. To pick a stable preferred b some criterion is needed, e.g., the least squares criterion. Furthermore, additional constraints will in general have to be imposed. It is natural to take the preferred b, $\hat{b}$, as the vector which minimizes $|d - G \cdot b|^2 + |C \cdot b|^2$ where C is a suitable matrix taking *a priori* "information" into account. This implies that

$$\begin{aligned} \hat{b} &= (G^T \cdot G + C^T \cdot C)^{-1} \cdot G^T \cdot d \\ &= \bar{b} + (G^T \cdot G + C^T \cdot C)^{-1} \cdot G^T \cdot (d - G \cdot \bar{b}) \end{aligned}$$

for any $\bar{b}$ which satisfies $C \cdot \bar{b} = 0$. Thus it often makes no difference whether one works with absolute or residual traveltimes. The strength of "damping" in different parts of the rock volume can be regulated by changing C.

The most common basis functions are the box-wise constant ones. The rock volume is divided into a number of rectangles (or cubes in the 3D case) and the basis functions are chosen as the characteristic functions of these rectangles. Another possibility is to use bilinear (trilinear) elements. The division into rectangles is kept, but only functions obtained by bilinear interpolation in the grid formed by the midpoints of the rectangles are considered. Bilinear elements will be used in the examples to follow, the coefficients $\hat{b}_j$ will be the estimates of the slowness function at the midpoints of the rectangles.

In Sections IV and V we will discuss properties of series-expansion solutions and how they can be computed; the linearization with fixed integration paths T_i is used.

IV. Confidence and Resolution

Our preferred coefficient vector $\hat{b}$ can be written $\hat{b} = H \cdot d$ where $H = (G^T \cdot G + C^T \cdot C)^{-1} \cdot G^T$. Supposing the measurements are contaminated with uncorrelated random noise so that $\mathrm{cov}(d) = \sigma^2 \cdot I$ (where I is the identity matrix), we see that $\hat{b}$ is a random vector with $\mathrm{cov}(\hat{b}) = \sigma^2 \cdot H \cdot H^T$. The impact of measurement errors on the estimate $\hat{m}(x) = \hat{b}^T \cdot \varphi(x)$ can thus be assessed.

Supposing the true slowness model $m(x)$ may be written $m(x) = b^T \cdot \varphi(x)$ and that $d = G \cdot b$, we obtain $\hat{b} = (H \cdot G) \cdot b$. The matrix $R = H \cdot G$ Is called the resolution matrix since it indicates the possibilities of resolving models of type $b^T \cdot \varphi(x)$. (Resolving power for other types of true slowness models can be similarly assessed [27].) Ideally, we would prefer to have $R = I$ but in general $\hat{b}$ will be a blurred version of b.

If M, the number of basis functions, if not too large, then $G^T \cdot G + C^T \cdot C$ can be inverted by Gaussian elimination and $\hat{b}$, $H \cdot H^T$, and R can easily be computed. An example, based on a 2D geometry of the type shown in Fig. 3, is given in Fig. 7. Twenty source points were used, all located in the 300-m-deep borehole to the left, and fifty receivers, thirty on the surface and twenty in the right borehole. This made up for a total of $20 \cdot (30 + 20) = 1000$ rays traversing the region. A division into $12 \times 20 = 240$ squares was made. Straight rays were used and, as is then natural, C was chosen as a suitable matrix with row sums zero thus favoring smooth slowness models. Fig. 7(a) shows the expected errors in $\hat{b}$ (i.e., $\hat{m}(x)$ at the centers of the squares) in percent for $\sigma = 0.2$ ms at the typical velocity of 5.0 km/s.

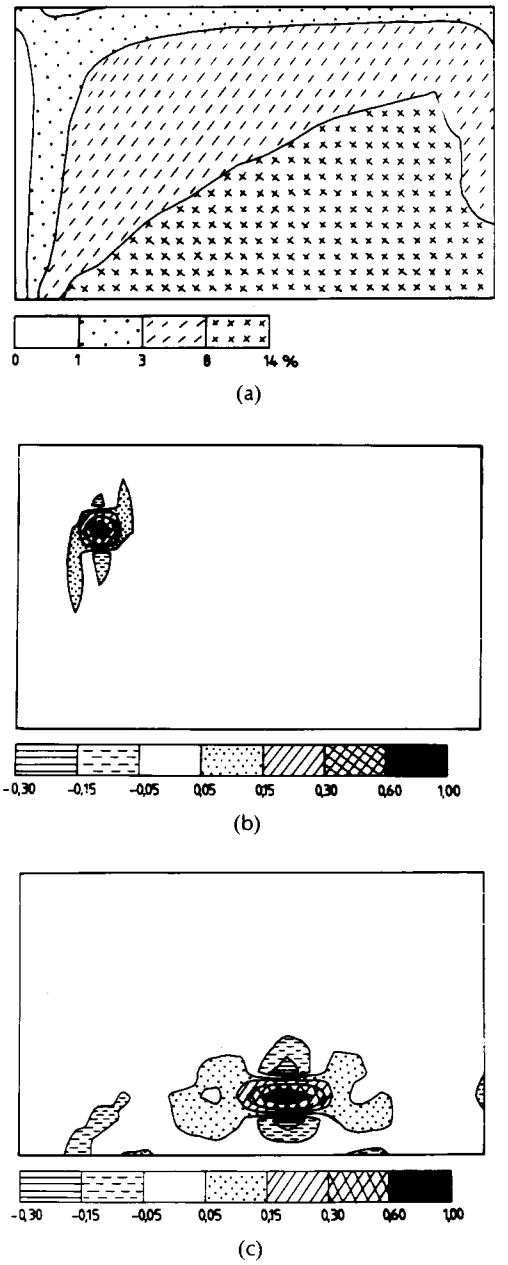

Fig. 7. Confidence and resolution for a typical "one-sided" crosshole coverage. (a) Noise-induced errors. (b) Resolving power at a point in the upper left corner. (c) Resolving power at a point in the middle of the bottom.

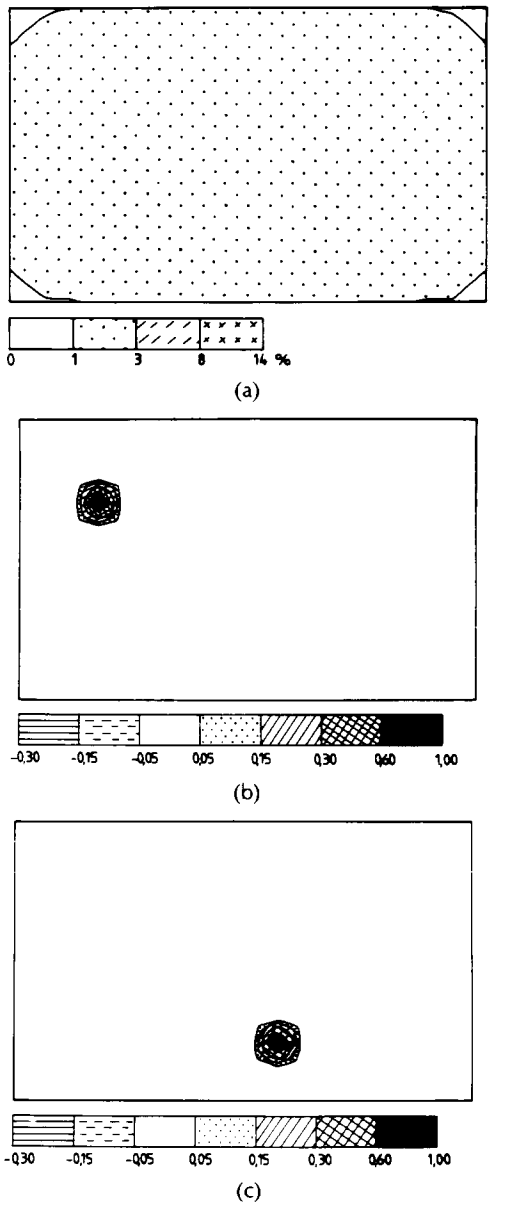

Fig. 8. (a)–(c) As in Fig. 7 but for a uniform ray coverage.

As expected, these errors will be small in the densely sampled upper left corner whereas they will be more significant downwards in the middle where the angular coverage is sparse. In parts (b) and (c), the resolving power at two different locations is shown. What is shown is actually the values, put out at the appropriate places in the crosshole region, in two rows of the matrix R. For each panel, a normalization to the maximum value 1.0 was done. Note the comparatively sharp resolution that is obtained in the upper left corner. In the middle of the lower part there is a smearing in the horizontal direction, the main direction of the rays there.

In Fig. 8 only 607 rays were used but these rays were uniformly distributed, both angularly and laterally. The lower border of the region was thus assumed to be accessible. As can be seen, this change implied a much better conditioned inverse problem.

V. SIRT, ART, AND CG

When the number of basis functions is large, as it will be particularly in 3D applications, the computation of $\hat{b}$ by Gaussian elimination will be too demanding. The original (inconsistent) system can be written

$$\begin{pmatrix} G \\ C \end{pmatrix} \cdot b = \begin{pmatrix} d \\ 0 \end{pmatrix}$$

where G and C are sparse matrices with M columns. Iterative storage-efficient "row-action" methods for computing "solutions" to such systems have been developed (e.g., [7]). Let g_i^T denote the ith row of G. Each data equation $g_i^T \cdot b = d_i$ defines a hyperplane in R^M. Let P_i be

the orthogonal-projection operator for this hyperplane, i.e.,

$$P_i(b) = b + \frac{d_i - g_i^T \cdot b}{g_i^T \cdot g_i} \cdot g_i.$$

Two famous types of methods for "solving" $G \cdot b = d$ are SIRT (Simultaneous Iterative Reconstruction Technique) and ART (Algebraic Reconstruction Technique).

Typical SIRT:

$$\hat{b}^{(k+1)} = \frac{1}{N} \cdot \sum_{i=1}^{N} P_i(\hat{b}^{(k)}).$$

Typical ART:

$$\hat{b}^{(k+1)} = P_N \cdot P_{N-1} \cdot \cdots \cdot P_2 \cdot P_1(\hat{b}^{(k)}).$$

In the SIRT case, $\hat{b}^{(k)}$ converges towards the unique solution of the normal equations $G^T \cdot G \cdot b = G^T \cdot d$ which lies closest to $\hat{b}^{(0)}$ if all g_i have length unity. The ART procedure is also convergent. Generalizations of these results are proved below.

The limit vector of ART can also be characterized [43]. In the consistent case, ART in an infinite-dimensional Hilbert space setting also converges [21, ch. I.7].

Consider Fig. 9 for an illustrative example. Some characteristics of SIRT and ART are realized already from this simple case. Usually ART converges more rapidly than SIRT, when $g_1 \perp g_2$ and $N = M = 2$ it will, in fact, converge in one iteration. In general, the angles between the successive hyperplanes $g_i^T \cdot b = d_i$, which (for the tomographic applications) are related to the angles between the rays, will affect the rate of convergence for ART [19]. Again there is reason to regret the incomplete angular ray coverage that is available in typical crosshole applications. When the equations $G \cdot b = d$ are inconsistent, even the limit vector of ART is dependent on the order of the equations (note that it always satisfies the last equation). The smoothing equations $C \cdot b = 0$ are more naturally included in a SIRT scheme.

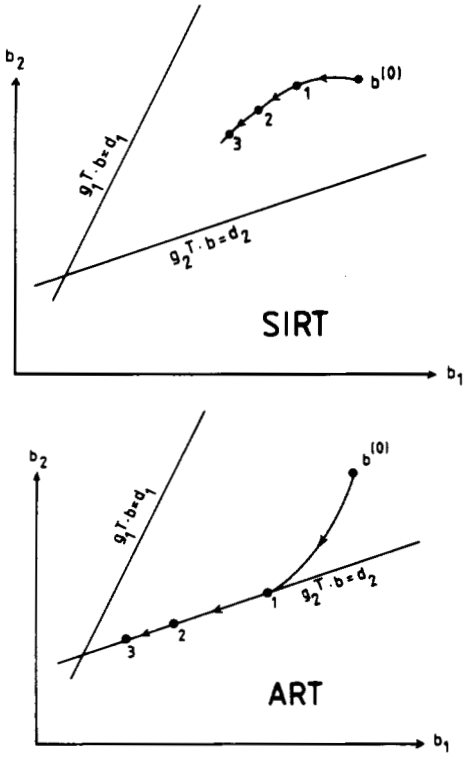

Fig. 9. The first iterates for SIRT and ART in a case with two equations and two unknowns ($N = M = 2$).

In order to improve the behavior of ART it has been suggested that additional changes to the b vector should be incorporated between the iterations; for example, constraining and selective smoothing ([24], [22]); and that underrelaxation could be used. Underrelaxation for ART will slow down the rate of convergence but will reduce the influence of the order of the equations upon the limit vector [9]. Another possibility could be to mix ART and SIRT steps in the iteration sequence. This case is covered by the following convergence result.

Theorem 10: Let G be an $N \times M$ matrix with (nonzero) rows g_i^T, C an $N_1 \times M$ matrix, d a column vector with N elements, D and S symmetric matrices which are positive semidefinite of type $N \times N$ and positive definite of type $M \times M$, respectively; b, $\hat{b}^{(k)}$, and $\hat{b}$ denote column vectors with M elements.

Let $0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_M$ be the eigenvalues of the matrix $S^{1/2} \cdot (G^T \cdot D \cdot G + C^T \cdot C) \cdot S^{1/2}$. The relaxation parameter ω is a positive-real number strictly less than 2.

An ART step is defined by

$$\hat{b}^{(k+1)} = Q_N \cdot Q_{N-1} \cdot \cdots \cdot Q_2 \cdot Q_1(\hat{b}^{(k)})$$

where

$$Q_i(b) = b + \omega \cdot \frac{d_i - g_i^T \cdot b}{g_i^T \cdot S \cdot g_i} \cdot S \cdot g_i.$$

A SIRT step is defined by

$$\hat{b}^{(k+1)} = \hat{b}^{(k)} + S \cdot \left(G^T \cdot D \cdot (d - G \cdot \hat{b}^{(k)}) - C^T \cdot C \cdot \hat{b}^{(k)}\right).$$

Let K_1 and K_2 be nonnegative integers such that $K = K_1 + K_2 > 0$. Apply ART and SIRT steps repeatedly in cycles, each cycle consisting of K_1 ART steps followed by K_2 SIRT steps.

a) If $\lambda_M < 2$, then

$$\lim_{\nu \to \infty} \hat{b}^{(\nu \cdot K)}$$

exists for an arbitrary start vector $\hat{b}^{(0)}$.

b) For "pure SIRT" ($K_1 = 0$), the limit vector $\hat{b}$ will minimize $|D^{1/2} \cdot (d - G \cdot b)|^2 + |C \cdot b|^2$. In the presence of many such minimizing vectors, $\hat{b}$ will be the one for which $(b - \hat{b}^{(0)})^T \cdot S^{-1} \cdot (b - \hat{b}^{(0)})$ is least. Also, at least for even k, $\hat{b}^{(k)}$ will be a kind of interpolation between $\hat{b}^{(0)}$ and $\hat{b}$. To obtain a rapid convergence, the numbers $|1 - \lambda_\nu|$ should be small for all ν such that $\lambda_\nu > 0$.

c) For "pure ART" ($K_2 = 0$) and consistent equations $d = G \cdot b$, the limit vector $\hat{b}$ will be equal to the orthogonal, relative to the inner product $\langle u, v \rangle = u^T \cdot S^{-1} \cdot v$, projection of $\hat{b}^{(0)}$ on the plane defined by $d = G \cdot b$.

Proof: Consider first "pure ART" with $S =$ the identity matrix I. We can give a simple proof for this case as follows (cf., [43]): Define the linear mappings R_i by

$$R_i(b) = b - \omega \cdot \frac{g_i^T \cdot b}{g_i^T \cdot g_i} \cdot g_i$$

and set $R = R_N \cdot R_{N-1} \cdot \cdots \cdot R_2 \cdot R_1$. For each i, let $H_i = g_i^{\perp}$ be the fixed-point space of R_i and choose a vector

$h_i \in H^\perp$, where $H = H_1 \cap H_2 \cap \cdots \cap H_N$ = the null space of G, such that $g_i^T \cdot h_i = d_i$. For consistent equations all h_i can be taken equal and z below will be the zero vector. We see that $Q_i(b) = h_i + R_i(b - h_i)$ for each b and by induction it follows that

$$\hat{b}^{(k)} = h_1 + (I + R + \cdots + R^{k-1})(z) + R^k(\hat{b}^{(0)} - h_1)$$

where

$$\begin{aligned} z = {} & h_N - h_1 + R_N(h_{N-1} - h_N) \\ & + R_N \cdot R_{N-1}(h_{N-2} - h_{N-1}) + \cdots \\ & + R_N \cdot R_{N-1} \cdot \cdots \cdot R_2(h_1 - h_2). \end{aligned}$$

Obviously $\|R\| \leqslant 1$ and $R(b) = b \Leftrightarrow |R(b) = b| \Leftrightarrow b \in H$. Since $H^\perp$ (= the linear span of the $g_i's$) is finite-dimensional, it follows that the restriction of R to $H^\perp$ is an operator (into $H^\perp$) with norm strictly less than one. We note that $z \in H^\perp$ and conclude that $(I + R + \cdots + R^{k-1})(z)$ must converge as k tends to infinity. Decomposing $\hat{b}^{(0)}$ with one term in H and one in $H^\perp$, $R^k(\hat{b}^{(0)} - h_1)$ is also seen to converge. This completes the proof for "pure ART" with $S = I$.

The general case can actually be reduced to the previous one by a change of variables. The idea is to diagonalize the matrix $S^{1/2} \cdot (C^T \cdot D \cdot C + C^T \cdot C) \cdot S^{1/2}$. In the new variables, the SIRT step splits into M independent equations, for details see [26], which may be interpreted as M simple ART mappings. The original ART-step equations are automatically transformed so that the "new $S'' = I$. Reinterpreting G as an augmented matrix (with at most $K_1 \cdot N + K_2 \cdot M$ rows), the reduction is completed and the proof is finished.

It should be noted that SIRT is very much related, not to say identical, to Richardson's algorithm for solving a linear equation system ([23], [29], [22]). Other mixtures of SIRT and ART than the case treated in Theorem 10 are also possible. Convergence proofs for "block-type" formulations were given in [14]; see also [1].

Proposition 11: If D and S in Theorem 10 are diagonal matrices with

$$S_{jj} \leqslant \left(\sum_{i:\, G_{ij} \neq 0} D_{ii} \cdot \left(\sum_v G_{iv}^2 \right) + \sum_{i:\, C_{ij} \neq 0} \left(\sum_v C_{iv}^2 \right) \right)^{-1}$$

then $\lambda_M \leqslant 1$.

Proof: This is a simple application of Cauchy's inequality, see [26].

Corollary 12 (ART with smoothing): Let G, C, d, S, ω, and all the Q_i's be as in Theorem 10. The iterative procedure defined by

$$\hat{b}^{(k+1)} = (I - S \cdot C^T \cdot C) \cdot Q_N \cdot Q_{N-1} \cdot \cdots \cdot Q_1(\hat{b}^{(k)})$$

will be convergent (for an arbitrary start vector) if all the eigenvalues of the matrix $S^{1/2} \cdot C^T \cdot C \cdot S^{1/2}$ are strictly less than 2.

Proof: Just take $D = 0$ in Theorem 10. Note that although $K_1 = K_2 = 1$, it is now most natural to let one iteration denote one cycle (i.e., one ART step followed by one smoothing step).

Recall that any symmetric and positive semidefinite matrix B may be written $B = C^T \cdot C$ for some positive semidefinite C. Gerschgorin's theorem (e.g., [12]) is often convenient for proving that a given matrix is positive-semidefinite and also for obtaining an upper bound to the spectral radius. Given S, it is now easy to choose a variety of $S \cdot B = S \cdot C^T \cdot C$ which fulfill the conditions of Corollary 12 and are suitable for smoothing in connection with ART. For example, any symmetric and diagonally dominant $S \cdot B$ such that the ℓ_1 norm of each row sum is strictly less than 2 will do with $S = I$. Some of the smoothing matrices in [22] are easily incorporated.

Dines and Lytle ([13]) proposed a whole "ℓ_p-family" of SIRT-like methods, with crosshole applications in mind, using backprojection of residual traveltimes to slownesses in the intersected rectangles. For the ℓ_2 case, when the Euclidean norm is used in the minimizing criterion of the backprojection, the results above provide suitable convergence proofs [28]. It is also seen that a certain amount of "overrelaxation" will improve the rate of convergence, without affecting the limit vector. The initial vector $\hat{b}^{(0)}$ is usually chosen in agreement with some very smooth (possibly homogeneous) slowness model.

For "pure SIRT" we have seen (Theorem 10) that the distribution of the eigenvalues λ_j will affect the rate of convergence. Fig. 10 shows the resulting eigenvalue distributions when a SIRT algorithm was applied to the two ray geometries used in Figs. 7 and 8. As expected, the case with the uniform ray sampling will yield the fastest convergence.

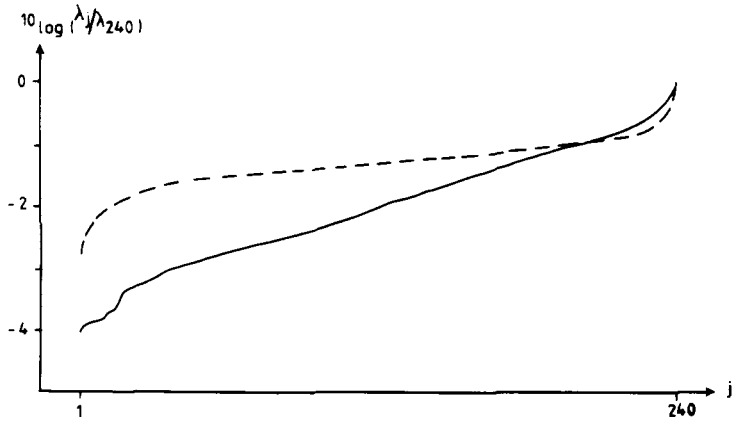

Fig. 10. Eigenvalue distributions for a SIRT algorithm applied to two geometries. The unbroken curve is for the "one-sided" geometry used in Fig. 7 while the dashed curve is for the uniform geometry used in Fig. 8.

By choosing a stronger damping (e.g., blowing up the smoothing matrix C), the rate of convergence is improved. Computations of a similar kind for "limited-angle tomography" can be found in [16].

A well-known optimization method that does not seem to have been applied very much to reconstruction problems is the conjugate gradients (CG) method. This is a certain type of iterative line-search method (e.g., [15]). When CG is used to optimize a quadratic function in M variables, the limit would actually be reached in M iterations if exact arithmetic could be used. The influence of roundoff will, however, destroy this property and M is in general too large a number of iterations to perform anyway. More interesting is the following result. When CG is used to minimize $|D^{1/2} \cdot (d - G \cdot b)|^2 + |C \cdot b|^2$, a functional from Theorem 10, the number of iterations necessary to reach a relative accuracy of ϵ (measured in a certain norm) is at most $(1/2) \cdot \ln(2/\epsilon) \cdot \sqrt{\mathcal{K}}$ where $\mathcal{K}$ is the condition number of the matrix $G^T \cdot D \cdot G + C^T \cdot C$. A proof can be found in [2]. A corresponding result for the SIRT method of Theorem 10 involves $\mathcal{K}$ rather than $\sqrt{\mathcal{K}}$. Since $\mathcal{K}$ is in general a large number, the difference can be significant [27]. CG is a good iterative method to use when the number of basis functions is too large to allow direct inversion by Gaussian elimination. Minor deficiencies are that "tricks" like selective smoothing cannot be naturally incorporated and that, compared to SIRT and ART, somewhat more computer storage is needed.

A comparison of SIRT ("overrelaxed"), ART (underrelaxed and with smoothing as described in Corollary 12), and CG (preconditioned) is given in Fig. 11. A velocity model, for which first-arrival traveltimes were computed using the same ray-path pattern as for Fig. 7, is shown in part (a). Reconstructions after 20 iterations are given in the other panels, 960 basis functions corresponding to a division into squares of side 12.5 m were used. Part (b) is for SIRT, (c) for ART, and (d) for CG. Note that the image of the low-velocity strip is less sharp in the lower part where the angular spread of the covering rays is bad. The CG-method provides the fastest convergence (the SIRT algorithm needs a few thousands iterations to achieve the same result as CG after 100 iterations). As the limit is approached, the image of the strip is improved but in the right part a high-velocity artifact develops (it begins to emerge in part (d)). It is present because the "slow" straight paths were used in the inversion. Note that nonconverged images may be of interest in themselves.

VI. Taking Account of Ray Bending

In the previous two sections, the data functionals $d = g(m)$ were linearized by using fixed (straight) integration curves T_i. This may result in certain artifacts when the velocity structure is not homogeneous. Iterative techniques to take the nonlinearity into account have been suggested. One approach is to linearize the data functionals around the current solution candidate by determining its implied ray paths and then compute suitable corrections ([5], [34]). Another possibility is to estimate the straight-line integrals of the true slowness model by adding to the measured data the appropriate time differences computed for the current model candidate ([4], [28]). In either case, extensive two-point ray tracing must be performed. This will increase the computing time most significantly.

If the initial solution is sufficiently good, it is often possible to obtain improvements by the methods described above. Unfortunately, however, the difficulties arise when an initial solution (computed by the straight-ray linearization, for example) is bad and the corrections are really needed. The rays may get trapped by the artifacts, the procedure may diverge, or it may converge to a wrong solution. In [35] the suitability of trying to correct for ray

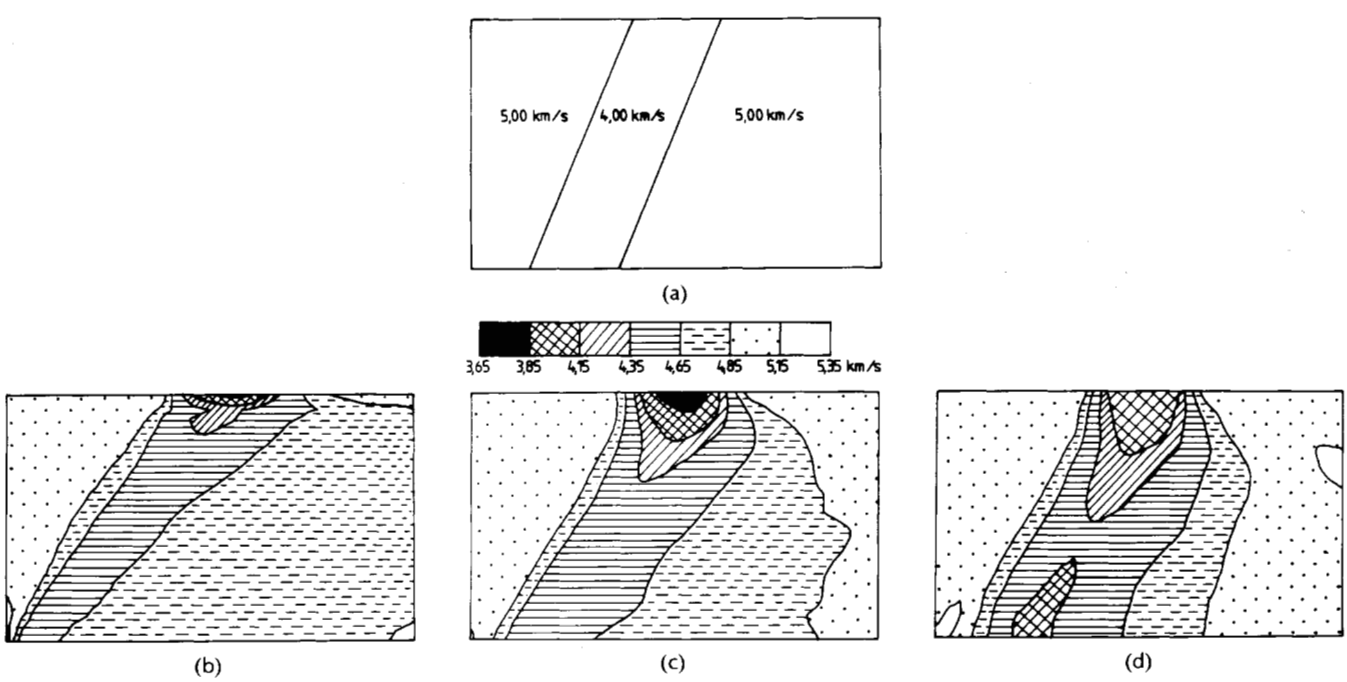

Fig. 11. A comparison of SIRT, ART, and CG. (a) The velocity model ("the phantom"). (b)–(d) Reconstructions after 20 iterations with SIRT, ART, and CG, respectively.

bending in the ways described above is questioned. Furthermore, it is the experience of several workers that straight-line tomography will do reasonably well when the velocity never departs from the average by more than 10 percent (say) (e.g., [13]).

A less ambitious way to "correct" for ray-bending errors is the following ([28]). Choose a number of suitable velocity models and compute the corresponding first-arrival traveltimes by ray tracing. Invert these traveltimes by some straight-line tomographic method and consider the images. By learning how different types of structures are imaged, where artifacts are present, and so forth, it may then be easier to interpret results from real data. The interpretation process will be one of trial-and-error comparisons with forward modeling as an essential ingredient. Recall Fig. 11 where we learned how a low-velocity strip was mapped in a certain situation. By studying synthetic examples of this kind, one also gets an appreciation of the impact of incomplete ray coverage in certain regions, which is very important. We must expect that the images will never be perfect and that ambiguities will often remain. Additional information and ideas about what is geologically plausible may be needed in order to resolve these ambiguities.

VII. Conclusions

With the possible exception of complex situations where ray bending is significant or varying anisotropy is present, seismic borehole tomography can be used for determining the velocity structure of a rock volume. With ideal infinite data sets, the usual measurement geometries could, in principle, allow a unique determination of the spatial velocity variations. In practice, however, the confidence and resolution properties of the solution would improve substantially if a more complete angular sampling of the rock volume could be obtained. Although the basic principles of a few other inversion techniques were described, this paper has mainly dealt with the series-expansion method. Particularly for 3D applications it is then often necessary to use some iterative reconstruction procedure. The CG algorithm was shown to be a suitable method. A convergence result for SIRT combined with ART was derived. Concerning ray-bending corrections it was concluded that a good initial solution is required.

Acknowledgment

The author wishes to thank J. Boman and T. Elfving for drawing his attention to the mathematical X-ray theorems and the CG method, respectively.

References

[1] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrason. Imag.*, vol. 6, pp. 81–94, 1984.
[2] O. Axelsson, "A class of iterative methods for finite element equations," *Comp. Meth. Appl. Mech. Eng.*, vol. 9, pp. 123–137, 1976.
[3] G. Backus and F. Gilbert, "The resolving power of gross Earth data," *Geophys. J. Roy. Astr. Soc.*, vol. 16, pp. 169–205, 1968.
[4] R. H. J. Bates and G. C. McKinnon, "Towards improving images in ultrasonic transmission tomography," *Australian Phys. Sci. Med.*, vol. 2, pp. 134–140, 1979.
[4A] I. N. Bernshtein and M. L. Gerver, "A condition for distinguishing matrices from hodographs," *Vychislitelnaya Seismologiya*, vol. 13, pp. 50–73, 1980.
[5] P. Bois, M. La Porte, M. Lavergne, and G. Thomas, "Well-to-well seismic measurements," *Geophys.*, vol. 37, pp. 471–480, 1972.
[6] J. Boman, "Uniqueness theorems for generalized Radon transforms," unpublished manuscript, University of Stockholm, Stockholm, Sweden, 1984.
[7] Y. Censor, "Row-action methods for huge and sparse systems and their applications," *SIAM Rev.*, vol. 23, pp. 444–466, 1981.
[8] ——, "Finite series-expansion reconstruction methods," *Proc. IEEE*, vol. 71, pp. 409–419, 1983.
[9] Y. Censor, P. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's method for inconsistent systems," *Numer. Math.*, vol. 41, pp. 83–92, 1983.
[10] C. Chou and J. Booker, "A Backus-Gilbert approach to inversion of travel-time data for three-dimensional velocity structure," *Geophys. J. Roy. Astr. Soc.*, vol. 59, pp. 325–344, 1979.
[11] A. Cormack, "The Radon transform on a family of curves in the plane," *Proc. Amer. Math. Soc.*, vol. 83, pp. 325–330, 1981.
[12] G. Dahlquist and Å. Björck, *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[13] K. Dines and J. Lytle, "Computerized geophysical tomography," *Proc. IEEE*, vol. 67, pp. 1065–1073, 1979.
[14] P. Eggermont, G. Herman, and A. Lent, "Iterative algorithms for large partitioned linear systems with applications to image reconstruction," *Lin. Alg. Applic.*, vol. 40, pp. 37–67, 1981.
[15] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. San Francisco, CA: Academic Press, 1981.
[16] A. Grünbaum, "A study of Fourier space methods for limited angle image reconstruction," *Numer. Funct. Anal. Optimiz.*, vol. 2, pp. 31–42, 1980.
[17] ——, "The limited angle reconstruction problem," in *Computed Tomography*, L. Shepp, Ed. (AMS Short Course Lecture Notes, vol. 27), 1983.
[18] M. Gustavsson, S. Ivansson, P. Morén, and J. Pihl, "Seismic borehole tomography—Measurement system and field studies," this issue, pp. 339–346.
[19] C. Hamaker and D. Solmon, "The angles between the null spaces of X-rays," *J. Math. Anal. Appl.*, vol. 62, pp. 1–23, 1978.
[20] C. Hamaker, K. Smith, D. Solmon, and S. Wagner, "The divergent beam X-ray transform," *Rocky Mount. J. Math.*, vol. 10, pp. 253–283, 1980.
[21] S. Helgason, *The Radon Transform*. Boston, MA: Birkhäuser, 1980.
[22] G. Herman, *Image Reconstruction from Projections*. New York: Academic Press, 1980.
[23] G. Herman and A. Lent, "Quadratic optimization for image reconstruction I," *Comput. Graph. Imag. Proc.*, vol. 5, pp. 319–332, 1976.
[24] ——, "Iterative reconstruction algorithms," *Comput. Biol. Med.*, vol. 6, pp. 273–294, 1976.
[25] B. Horn, "Density reconstruction using arbitrary ray-sampling schemes," *Proc. IEEE*, vol. 66, pp. 551–562, 1978.
[26] S. Ivansson, "Remark on an earlier proposed iterative tomographic algorithm," *Geophys. J. Roy. Astr. Soc.*, vol. 75, pp. 855–860, 1983.
[27] ——, "Crosshole investigations—Tomography and its application to crosshole seismic measurements," Stripa Project Internal Rep. 84-08, Stockholm, Sweden SKB/KBS, 1984.
[28] ——, "A study of methods for tomographic velocity estimation in the presence of low-velocity zones," *Geophys.*, vol. 50, pp. 969–988, 1985.
[29] A. Lakshminarayanan and A. Lent, "Methods of least-squares and SIRT in reconstruction," *J. Theor. Biol.*, vol. 76, pp. 267–295, 1979.
[30] P. C. Leary and T. L. Henyey, "Anisotropy and fracture zones about a geothermal well from P-wave velocity profiles," *Geophys.*, vol. 50, pp. 25–36, 1985.
[31] R. Lewitt, "Reconstruction algorithms, transform methods," *Proc. IEEE*, vol. 71, pp. 390–408, 1983.

[32] B. F. Logan, "The uncertainty principle in reconstructing functions from projections," *Duke Math. J.*, vol. 42, pp. 661–706, 1975.
[33] A. Louis, "Approximation of the Radon transform from samples in limited range," in *Mathematical Aspects of Computerized Tomography*, G. Herman and F. Natterer, Eds. (Lecture Notes in Medical Informatics, vol. 8), 1981.
[34] J. Lytle and K. Dines, "Iterative ray-tracing between boreholes for underground image reconstruction," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-18, pp. 234–240, 1980.
[35] G. C. McKinnon and R. H. J. Bates, "A limitation of ultrasonic transmission tomography," *Ultrason. Imag.*, vol. 2, pp. 48–54, 1980.
[36] W. Menke, "The resolving power of cross-borehole tomography," *Geophys. Res. Lett.*, vol. 11, pp. 105–108, 1984.
[37] R. G. Mukhometov, "A problem of reconstructing a Riemannian metric," *Sib. Mat. Zh.*, vol. 22, pp. 119–135, 1981.
[38] A. Nercessian, A. Hirn, and A. Tarantola, "Three dimensional seismic transmission prospecting of the Mont Dore volcano, France," *Geophys. J. Roy. Astr. Soc.*, vol. 76, pp. 307–315, 1984.
[39] V. G. Romanov, *Integral Geometry and Inverse Problems for Equations of Hyperbolic Type*. New York: Springer, 1974.
[40] K. Smith, D. Solmon, and S. Wagner, "Practical and mathematical aspects of the problem of reconstructing objects from radiographs," *Bull. Amer. Math. Soc.*, vol. 83, pp. 1227–1270, 1977.
[41] D. Solmon, "The X-ray transform," *J. Math. Anal. Appl.*, vol. 56, pp. 61–83, 1976.
[42] R. Strichartz, "Radon inversion—Variations on a theme," *Amer. Math. Monthly*, vol. 89, pp. 377–384, 1982.
[43] K. Tanabe, "Projection method for solving a singular system of linear equations and its applications," *Numer. Math.*, vol. 17, pp. 203–214, 1971.
[44] A. Tarantola and A. Nercessian, "Three-dimensional inversion without blocks," *Geophys. J. Roy. Astr. Soc.*, vol. 76, pp. 299–306, 1984.
[45] A. Tarantola and B. Valette, "Generalized nonlinear inverse problems solved using the least-squares criterion," *Rev. Geophys. Space. Phys.*, vol. 20, pp. 219–232, 1982.
[46] ——, "Inverse problems—Quest of information," *J. Geophys.*, vol. 50, pp. 159–170, 1982.