

# WEIGHTED LEAST-SQUARES CRITERIA FOR SEISMIC TRAVELTIME TOMOGRAPHY

**James G. Berryman**

Lawrence Livermore National Laboratory

Electronics Engineering

P. O. Box 808 L-156

Livermore, CA 94550

and

Courant Institute of Mathematical Sciences

New York University

251 Mercer St.

New York, NY 10012

*Abstract* – Methods are developed for design of linear tomographic reconstruction algorithms with specified properties. Assuming a starting model with constant slowness, an algorithm with the following properties is found: (1) The optimum constant for the starting model is determined automatically. (2) The weighted least-squares error between the predicted and measured traveltime data is as small as possible. (3) The variance of the reconstructed slowness from the starting model is minimized. (4) Rays with the greatest length have the least influence on the reconstructed slowness. (5) Cells with most ray coverage tend to deviate least from the background value. The resulting algorithm maps the reconstruction problem into a vector space where the contribution to the inversion from the background slowness remains invariant, while the optimum contributions in orthogonal directions are found. For a starting model with nonconstant slowness, the reconstruction algorithm has analogous properties.

## Introduction

Our goal in seismic travelttime tomography is to find a slowness (reciprocal wave speed) model consistent with measured travelttime data within some specified measurement tolerance. Let  $t$  be the measured travelttime  $m$ -vector such that  $t^T = (t_1, \dots, t_m)$  where  $t_i$  is the travelttime along the  $i$ -th ray path (a  $T$  superscript implies the transpose). Although various methods of parameterizing the slowness model have been studied, we restrict our attention here to a model based on cells of constant slowness. Thus, if we are working on a two-dimensional cross section – say between two vertical boreholes – we form our model conceptually by dividing the region between our sources and receivers into rectangular cells of constant slowness. On the other hand, if our problem is three-dimensional, the cells are blocks of constant slowness. Then  $s$  is the model slowness  $n$ -vector  $s^T = (s_1, \dots, s_n)$  with  $s_j$  being the slowness of the  $j$ -th cell, satisfying

$$Ms = t \tag{1}$$

where  $M$  is an  $m \times n$  matrix whose matrix elements  $l_{i,j}$  are determined by the  $i$ -th ray path through the  $j$ -th cell of the model. In general, we do not have good *a priori* approximations to the ray paths if the medium is very inhomogeneous (*i.e.*, having contrasts of 15% or more). However, for the present application, we will assume that the ray path matrix has been fixed with a known set of (what are possibly trial) ray paths.

The point of view presented here is quite different from that presented in more typical weighted least-squares methods such as *Iterative Reweighted Least Squares* – discussed for example by Scales *et al.*[1] – where the weights are functions of the travelttime residual. Unlike Bording *et al.*[2] and others, we do not restrict our analysis to changes  $\Delta s$  in  $s$  relative to some known background. We take explicit account of the fact that the linear weighted least-squares problem we study is actually derived from a nonlinear reconstruction problem. The (possibly) substantial uncertainties in the path length matrix  $M$  due to ray bending effects are acknowledged and treated from the outset. The importance of this point of view derives from the fact that the errors introduced into the reconstruction by a poor choice of  $M$  may be far more significant than the measurement errors introduced through  $t$ . The method developed here will ultimately be used as one step in a comprehensive nonlinear inversion algorithm such as that of Lytle and Dines [3]. However, we should emphasize that the analysis of the present work is still restricted to the linear inverse problem associated

with (1). Extensions to the nonlinear inversion problem will be presented elsewhere.

Once a set of trial ray paths and corresponding values of  $l_{i,j}$  are known, the model vector  $s$  may be underdetermined if  $m < n$  or overdetermined if  $m > n$ . Thus, finding a “solution” to (1) requires the use of a generalized inverse [4-6] which in turn often implies a least-squares estimate [7] of the model slowness vector. A general objective function to be minimized might have the form suggested by Herman[8]

$$\phi_{\mu}(s) = (t - Ms)^T W_1 (t - Ms) + \mu(s - s_b)^T W_2 (s - s_b), \quad (2)$$

where  $W_1$  and  $W_2$  are (respectively)  $m \times m$  and  $n \times n$  real, symmetric weight matrices,  $\mu$  is some scalar determined by the relative importance of the second term compared to the first, and  $s_b$  is some special background value of the slowness vector to which the final result should be close. Many criteria have been proposed in order to ensure a unique solution to the reconstruction problem. For example, positivity of the slowness vector  $s$  is clearly required for any physical model, but for seismic tomography maintaining the positivity of  $s$  is seldom a problem. Therefore, we will make no special effort to control this feature through the objective function itself. Some of the most important features of the reconstructed slowness needing control are: (i) The (weighted) least-squares error between the measured and predicted traveltimes data should be as small as possible. (ii) For linear tomographic inversion with a good starting model, the variance of the reconstructed slowness from the specified background should be as small as possible for the given data [8]. (iii) The (possibly weighted) mean slowness should be the best possible in some sense. Other criteria that have been proposed include: (iv) Rays of greatest length should be weighted least in the reconstruction [9]. (v) Cells with the most rays traveling through them should be the ones with the most accurate reconstructed values. The motivation for each of these criteria and the impact they have on determining the values of the weights and other constants in (2) will be developed more fully in the discussion that follows.

The techniques developed here may be used in many currently existing least-squares tomographic inversion methods to improve their stability and convergence properties, while making relatively minor changes in the architecture of these algorithms.

## Homogeneous Background

We treat the minimization problem for homogeneous background  $s_b$  in this section, and then generalize to more complex background slowness in the next section. In a comprehensive nonlinear inversion algorithm, we often expect constant slowness to be the initial guess.

### A. Optimum homogeneous slowness

The crudest sort of reconstruction one can imagine is to take the available data and form the optimum constant slowness vector consistent with that data. By optimum constant, we mean that it uses all the data, that it produces a minimum error in some appropriate least-squares functional, and that it agrees with the exact result if the medium is actually constant. If the medium is homogeneous (constant slowness) and the trial rays are straight, the exact value for the slowness is easily seen to be given by

$$\sigma_0 = \frac{\sum_i t_i}{\sum_{ij} l_{i,j}}. \quad (3)$$

The slowness scalar  $\sigma_0$  is the total time taken by all the rays to travel over all their paths divided by the total length of all their paths. When the medium is not actually homogeneous, a least-squares error criterion can be used based on minimizing the functional

$$\psi(\sigma) = (t - M\sigma v)^T W_3 (t - M\sigma v), \quad (4)$$

where  $W_3$  is another  $m \times m$  real, symmetric weight matrix and  $v^T = (1, \dots, 1)$  is an  $n$ -vector of ones. The minimum of (4) occurs when

$$v^T M^T W_3 (t - M\sigma_0 v) = 0, \quad (5)$$

or equivalently when

$$\sigma_0 = \frac{v^T M^T W_3 t}{v^T M^T W_3 M v}. \quad (6)$$

Now Eq. (6) can be simplified by introducing some new notation. Define the row sums  $L_i$  and column sums  $C_j$

$$L_i = \sum_{j=1}^n l_{i,j}, \quad C_j = \sum_{i=1}^m l_{i,j}. \quad (7)$$

The quantity  $L_i$  is seen to be the total length of path  $i$ . The quantity  $C_j$  is the total length of all the ray-path segments that pass through cell  $j$ , so we will call this the “coverage” of cell  $j$ . Any cell with  $C_j = 0$  is uncovered and therefore lies outside the span of our data for the current choice of ray paths. We retain only the covered cells in the reduced slowness vector  $\tilde{s}$  of length  $\tilde{n} \leq n$ . The matrix  $M$  may similarly be reduced to  $\tilde{M}$  by deleting the corresponding columns of zeros. Finally, define diagonal  $m \times m$  and  $\tilde{n} \times \tilde{n}$  matrices  $L$  and  $C$  respectively whose diagonal elements are given by the nonzero sums in (7). For simplicity, we assume that  $\tilde{n} = n$  in the following discussion. Then, the diagonal matrices  $L$  and  $C$  are given by

$$L = \begin{pmatrix} L_1 & & \\ & \ddots & \\ & & L_m \end{pmatrix} \quad (8)$$

and

$$C = \begin{pmatrix} C_1 & & \\ & \ddots & \\ & & C_n \end{pmatrix}. \quad (9)$$

Now in addition to the  $n$  vector  $v$  we have already introduced, we define  $u^T = (1, \dots, 1)$ , an  $m$ -vector of ones. Then,

$$Mv = Lu \quad (10)$$

and

$$M^T u = Cv \quad (11)$$

follow easily from the definitions. Substituting (10) into (6) we obtain

$$\sigma_0 = \frac{u^T L W_3 t}{u^T L W_3 L u}. \quad (12)$$

For comparison, note that (3) may be rewritten as

$$\sigma_0 = \frac{u^T t}{u^T L u}. \quad (13)$$

Now, neglecting a possible (but clearly irrelevant) arbitrary scale factor in the weight matrix  $W_3$ , we see that (12) will agree with (13) if

$$W_3 L u = u, \quad (14)$$

*i.e.*, if  $u$  is a right eigenvector of  $W_3L$  with unit eigenvalue. Of course, (14) does not uniquely determine  $W_3L$  since we see easily that two possible choices are given by

$$W_3L = I, \tag{15}$$

where  $I$  is the  $n \times n$  identity matrix, and by

$$W_3L = L^{-1}MC^{-1}M^T. \tag{16}$$

In either case, the resulting  $W_3$  is real and symmetric as required. Although other choices clearly exist, the simplest choice for  $W_3$  appears to be the one determined by (15), *i.e.*,

$$W_3 = L^{-1}. \tag{17}$$

With this choice, (12) and (13) are identical.

The special background value of  $s$  for our analysis of (2) in the remainder of this section will be taken to be

$$s_b = \sigma_0 v, \tag{18}$$

where  $\sigma_0$  is determined by (13).

### B. Weighted least-squares error

Next we consider (2) with  $\mu = 0$ . Then, the error functional

$$\phi_0(s) = (t - Ms)^T W_1 (t - Ms) \tag{19}$$

should be minimized with respect to  $s$ . The standard result is that the minimum occurs for  $s$  satisfying

$$M^T W_1 M s = M^T W_1 t. \tag{20}$$

If the square matrix  $M^T W_1 M$  is nonsingular, *i.e.*, has no zero eigenvalues, then (20) is easily solved by inverting the matrix; this result is just the well-known solution associated with the Gauss-Markov theorem [6]. However, in virtually all cases of practical interest, this matrix is singular and some method from the theory of generalized inverses will be required [10]. One particularly simple method to analyze involves the iterative solution of (20) [8,11-13]. However, the success of this method requires that the matrix whose pseudoinverse is to be approximated in this way must have its eigenvalues  $\lambda$  restricted to

the unit interval  $0 \leq \lambda \leq 1$ . We can precondition the matrix by multiplying (20) by some positive definite, diagonal matrix  $D^{-1}$  and reassociating matrix factors so that

$$(D^{-1}M^TW_1MD^{-1})Ds = D^{-1}M^TW_1t. \quad (21)$$

The diagonal elements of  $D$  are chosen so that the eigenvalues of the matrix in parentheses lie in the unit interval. We define an  $n$ -vector  $z$

$$z = Ds \quad (22)$$

and, since the matrix in parentheses in (21) is clearly symmetric, a matrix  $A$  can always be found such that

$$A^T A = D^{-1}M^TW_1MD^{-1}. \quad (23)$$

Then, (20) may be rewritten as

$$A^T Az = D^{-1}M^TW_1t \quad (24)$$

which can be solved by iteration according to

$$z^{(0)} = D^{-1}M^TW_1t, \quad (25)$$

$$z^{(k+1)} = z^{(0)} + (1 - A^T A)z^{(k)}. \quad (26)$$

It is well-known [14,15] that the  $z^{(k)}$  sequence converges to the solution of (24) obtained from the Moore-Penrose generalized inverse as  $k \rightarrow \infty$ . We will develop this result in more detail by examining the eigenvalue (singular value) decomposition of  $A^T A$  in a manner similar to that of Aki and Richards [16].

Following Lanczos [17], we will introduce the  $(m+n) \times (m+n)$  real, symmetric matrix  $H$  determined by  $M$

$$H = \begin{pmatrix} 0 & M \\ M^T & 0 \end{pmatrix} \quad (27)$$

and the corresponding eigenvalue problem

$$\begin{pmatrix} 0 & M \\ M^T & 0 \end{pmatrix} \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix} = \lambda \begin{pmatrix} L & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix}. \quad (28)$$

In (28), the vectors  $u_\lambda$  and  $v_\lambda$  are of length  $m$  and  $n$  respectively. We see in particular that Eqs. (10) and (11) are of the form (28) so that the vector  $(u^T, v^T) = (1, \dots, 1)$  is an eigenvector

of  $H$  with eigenvalue unity. This fact is significant, and provides the motivation for studying the eigenvalue problem (28), as we shall show.

We will now proceed to manipulate (28) into a more illuminating form. First, we invert the diagonal matrix on the right side of (28) and multiply through, obtaining

$$\begin{pmatrix} 0 & L^{-1}M \\ C^{-1}M^T & 0 \end{pmatrix} \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix} = \lambda \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix}. \quad (29)$$

[Note that, if  $(u_\lambda^T, v_\lambda^T)^T$  is an eigenvector of (29), then  $(u_{-\lambda}^T, -v_{-\lambda}^T)^T$  is also an eigenvector.]

Next we square the matrix on the left side of (29) to find

$$\begin{pmatrix} L^{-1}MC^{-1}M^T & 0 \\ 0 & C^{-1}M^TL^{-1}M \end{pmatrix} \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix} = \lambda^2 \begin{pmatrix} u_\lambda \\ v_\lambda \end{pmatrix}. \quad (30)$$

Finally, we multiply (30) by the matrix

$$\begin{pmatrix} L^{\frac{1}{2}} & 0 \\ 0 & C^{\frac{1}{2}} \end{pmatrix}$$

(these square root matrices are well-defined since  $L$  and  $C$  both have only real, positive, diagonal elements) and then reassociate some terms to reach the desired result

$$\begin{pmatrix} L^{-\frac{1}{2}}MC^{-1}M^TL^{-\frac{1}{2}} & 0 \\ 0 & C^{-\frac{1}{2}}M^TL^{-1}MC^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} L^{\frac{1}{2}}u_\lambda \\ C^{\frac{1}{2}}v_\lambda \end{pmatrix} = \lambda^2 \begin{pmatrix} L^{\frac{1}{2}}u_\lambda \\ C^{\frac{1}{2}}v_\lambda \end{pmatrix}. \quad (31)$$

Comparing (31) and (23), we see that, if the following choices are made for the weight matrix

$$W_1 = L^{-1} \quad (32)$$

(note that this is the same choice that was made for  $W_3$ ) and the preconditioning matrix

$$D = C^{\frac{1}{2}}, \quad (33)$$

then the matrix  $A$  becomes

$$A = L^{-\frac{1}{2}}MC^{-\frac{1}{2}} \quad (34)$$

and (31) may be rewritten as

$$\begin{pmatrix} AA^T & 0 \\ 0 & A^TA \end{pmatrix} \begin{pmatrix} w_\lambda \\ x_\lambda \end{pmatrix} = \lambda^2 \begin{pmatrix} w_\lambda \\ x_\lambda \end{pmatrix}, \quad (35)$$



where  $w_\lambda = L^{\frac{1}{2}}u_\lambda$  and  $x_\lambda = C^{\frac{1}{2}}v_\lambda$ . That the eigenvalues of  $A$  all lie in the range  $[0,1]$  as desired follows from the fact that, by construction, the matrices  $MC^{-1}$  and  $M^TL^{-1}$  have all their column sums equal to unity. Furthermore, since  $(u^T, v^T)^T$  is a strictly positive eigenvector with eigenvalue  $\lambda = 1$  from (25) and (26), we know that this eigenvalue is both maximal and simple from the Perron-Frobenius theory of non-negative matrices [18,19]. [A technical point arising here is this: The matrix  $A^TA$  must be “primitive,” *i.e.*, some positive power of it must have all positive matrix elements. This condition will often be satisfied for problems in tomography. If it is not satisfied, we separate  $A^TA$  into submatrices for which it is satisfied and proceed with the analysis as presented for each of the submatrices.] In addition, we know that the eigenvectors  $x_{\lambda_j}, x_{\lambda_{j'}}$  for distinct eigenvalues  $\lambda_j^2, \lambda_{j'}^2$  are orthogonal, and we may assume that those eigenvectors which share a common eigenvalue have been orthogonalized.

Returning to the iterative scheme (25)-(26) now that we have an explicit representation for  $A$ , we find that

$$z^{(0)} = A^TL^{-\frac{1}{2}}t = \sum_{j=1}^n \zeta_j x_{\lambda_j}, \quad (36)$$

where

$$\zeta_j = x_{\lambda_j}^T A^TL^{-\frac{1}{2}}t \quad (37)$$

assuming that the eigenvectors have been normalized so that  $x_\lambda^T x_\lambda = 1$ . Furthermore, the eigenvalue decomposition of  $A^TA$  is given by

$$A^TA = \sum_{j=1}^n x_{\lambda_j} \lambda_j^2 x_{\lambda_j}^T. \quad (38)$$

Thus, we have in general that

$$(I - A^TA)^p z^{(0)} = \sum_{j=1}^n (1 - \lambda_j^2)^p \zeta_j x_{\lambda_j}. \quad (39)$$

It follows easily from (26) that the  $k$ -th iterate may be expressed as

$$z^{(k)} = \sum_{p=1}^k (I - A^TA)^p z^{(0)} \quad (40)$$

so from (39) we find that

$$z^{(k)} = \sum_{j=1}^n \sum_{p=0}^k (1 - \lambda_j^2)^p \zeta_j x_{\lambda_j}. \quad (41)$$

Let the eigenvalues be ordered in decreasing magnitude so that

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0, \quad (42)$$

where  $r$  is the rank of  $A^T A$ . Then, (41) may be rewritten as

$$z^{(k)} = \zeta_1 x_{\lambda_1} + \sum_{j=2}^r \sum_{p=0}^k (1 - \lambda_j^2)^p \zeta_j x_{\lambda_j}. \quad (43)$$

Using the fact that

$$\sum_{p=0}^{\infty} (1 - \lambda_j^2)^p = \frac{1}{1 - (1 - \lambda_j^2)} = \frac{1}{\lambda_j^2}, \quad (44)$$

we see that

$$z^{(\infty)} = \zeta_1 x_{\lambda_1} + \sum_{j=2}^r \frac{\zeta_j}{\lambda_j^2} x_{\lambda_j}. \quad (45)$$

Clearly, the rate of convergence of  $z^{(k)}$  to  $z^{(\infty)}$  depends only on the value of the lowest nonzero eigenvalue  $\lambda_r$  since the term  $(1 - \lambda_r^2)^k$  is the most slowly decreasing factor appearing in (43) as  $k \rightarrow \infty$ . Eq. (45) gives the eigenvector decomposition of the solution of (24) that would be obtained using the Moore-Penrose generalized inverse of  $A^T A$ .

Two important facts should now be noted about (43). First, since the right null spaces of  $A$  and  $A^T A$  are the same, (43) is orthogonal to the null space of  $A^T A$  for all  $k$ . This fact follows immediately from (37), since  $\zeta_j = 0$  for any  $j$  such that  $A x_{\lambda_j} = 0$ . This is also a very useful fact because it means that the solution  $z^{(\infty)}$  is the unique solution of minimum norm  $z^T z$  consistent with the data; thus, any other  $z$  for which (19) has a minimum will be equal to  $z^{(\infty)}$  plus some terms from the null space of  $A^T A$ . Second, the coefficient of the eigenvector  $x_{\lambda_1}$  is the same for all  $k$ .

Eq. (43) provides a solution to the problem of finding the slowness with minimum error for (19) when  $W_1 = L^{-1}$ . Interestingly, it does more than that. Consider the term  $\zeta_1 x_{\lambda_1}$  which

has constant coefficient (as  $k \rightarrow \infty$ ) starting from the initial iterate  $z^{(0)}$  of (25). We see that

$$\zeta_1 = x_{\lambda_1}^T A^T L^{-\frac{1}{2}} t = v^T M^T L^{-1} t / (v^T C v)^{\frac{1}{2}} \quad (46)$$

and, using (10) together with the fact that  $v^T C v = u^T L u$ , we therefore find that

$$\zeta_1 x_{\lambda_1} = \sigma_0 C^{\frac{1}{2}} v \quad (47)$$

where  $\sigma_0$  is given by (13). Recalling that  $\sigma_0 v$  is the optimum homogeneous background, we see that this result is quite remarkable. Eq. (47) shows that, by transforming into this particular vector space using the conditioning matrices  $L$  and  $C$ , we have found a solution of a weighted least-squares problem that automatically has minimum (weighted) norm consistent with the data – but that also has exactly the right value of the coefficient for the optimum constant background. This latter property is guaranteed to be satisfied at any stage of the iterative process for obtaining the solution of (24). Furthermore, the weighted least-squares criterion (19) with  $W_1 = L^{-1}$  is the only one with all these properties for the constant background problem – this fact follows by running through the steps of the derivation of (28)-(31) backwards assuming that  $L$  and  $C$  are arbitrary diagonal matrices and then noting that (28) can only be consistent with (10), (11), and (13) for this *unique* choice of  $L$  and  $C$ .

The author first learned of the weighting scheme (32)-(34) rederived here on reading an unpublished manuscript of Burkhard [15]. The goal of obtaining a convergent iterative method motivated Burkhard to make the choices (32) and (33). The choice for weight matrix  $W_1 = L^{-1}$  is also one of a class of such weights based on powers of the ray path lengths discussed by Frank and Balanis [9]. The arguments given for such weights are: (i) Signal to noise ratio is expected to be better on shorter paths than on longer ones, since the overall attenuation will typically be smaller. (ii) Shorter trial paths are more likely to correspond to real paths that remain completely in the image plane for two-dimensional reconstruction problems.

### C. Minimum variation from background

Now consider the second term in (2)

$$\xi(s) = (s - s_b)^T W_2 (s - s_b). \quad (48)$$

The motivation for using (48) as a criterion is that, although there may be many “solutions” to (1), the ones with least deviation from some prescribed (smooth) background are most likely to have no spurious structure. It is not desirable to have a reconstruction method that produces “interesting” features that are not real. High frequencies in the reconstruction should therefore be damped by some means. A stationary point of (48) occurs when  $s$  satisfies

$$W_2(s - s_b) = 0, \tag{49}$$

*i.e.*, when  $s - s_b$  is in the null space of  $W_2$ . If the weight matrix is positive definite, then (49) implies that

$$s = s_b. \tag{50}$$

We could choose  $W_2 = I$ , but this choice ignores the fact that some of the cells have significantly more coverage than others. The cells with most coverage ought to be given the most weight in (48) and clearly any cell with no coverage should have no influence on the final result. The simplest choice of weight matrix satisfying these requirements is

$$W_2 = C. \tag{51}$$

Then, we see that (48) becomes

$$(s - s_b)^T C (s - s_b) = (z - z_b)^T (z - z_b) \tag{52}$$

where  $z = C^{\frac{1}{2}} s$  and

$$z_b = \zeta_1 x_{\lambda_1}. \tag{53}$$

Minimum variance criteria such as (48) are appropriate only as secondary criteria [8]. Forcing  $s$  to the value  $s = s_b$  at the absolute minimum of (48) is too strong a constraint. A more reasonable constraint to impose is that the solution should have its component in the direction of  $s_b$  equal in magnitude to  $s_b$  but other components orthogonal to this direction should also be allowed, *i.e.*,  $z_b^T (z - z_b) = 0$ . In other words, the solution should be restricted to a hyperplane whose projection along the direction of  $z_b$  has the correct magnitude. We see that (43) has precisely this property. So it is possible to find a slowness vector achieving the absolute minimum of (19) consistent with the data, while also satisfying a sensible minimum variance criterion.

#### D. Minimizing the general objective function

Now consider minimizing the full objective function (2) with the particular choices of weight matrices motivated in the preceding analysis

$$\phi_\mu(s) = (t - Ms)^T L^{-1}(t - Ms) + \mu(s - s_b)^T C(s - s_b). \quad (54)$$

Using (34) and  $z = C^{\frac{1}{2}}s$ , (54) may be transformed to

$$\phi_\mu(z) = (y - Az)^T (y - Az) + \mu(z - z_b)^T (z - z_b) \quad (55)$$

where  $A$  is again given by (34) and  $y = L^{-\frac{1}{2}}t$ . The minimum of (55) occurs when  $z$  satisfies

$$A^T(Az - y) + \mu(z - z_b) = 0, \quad (56)$$

or equivalently (since  $A^T A z_b = z_b$ ),

$$(A^T A + \mu I)(z - z_b) = A^T y - z_b. \quad (57)$$

Using (36) and (53), we notice immediately that

$$z_b^T (A^T y - z_b) = 0, \quad (58)$$

which means that the right hand side of (57) is orthogonal to  $z_b$ . Notice further that the eigenvectors of  $A^T A$  are also eigenvectors of  $A^T A + \mu I$ , since

$$(A^T A + \mu I)x_{\lambda_j} = (\lambda_j^2 + \mu)x_{\lambda_j}. \quad (59)$$

Therefore, the solution of (57) with

$$\Delta z = z - z_b, \quad (60)$$

and

$$\Delta z^{(0)} = A^T y - z_b, \quad (61)$$

has the form

$$\Delta z = \sum_{j=2}^r \frac{\zeta_j}{\lambda_j^2 + \mu} x_{\lambda_j}, \quad (62)$$

where for  $2 \leq j \leq r$

$$\zeta_j = x_{\lambda_j}^T \Delta z^{(0)} \quad (63)$$

are precisely the same numbers given by (37). For finite  $\mu > 0$ , (62) is the unique solution of (57) since the matrix on the left hand side of (57) is nonsingular.

The complete “solution” of (1) is now  $s = C^{-\frac{1}{2}}z$ , with  $z$  being the solution of (57), plus some slowness vector from the null space of the path length matrix  $M$ . In the absence of additional information, the vector from the null space may be omitted.

Eq. (57) has some clear computational advantages over (24). In particular, if the smallest nonzero eigenvalue  $\lambda_r$  is very small [on the order of the round off error for computations of  $O(1)$ ], then the computation of (43) and (45) will be numerically unstable, while the presence of a small but finite  $\mu$  in (57) will stabilize the computation of (62).

Note that the result (57) has much in common with Marquardt’s [20,21] algorithm for solving such problems using damped least-squares. In the presence of severe nonlinearity – which may often be the case for tomographic reconstruction of seismic data – it may be advisable to use the damped algorithm with an optimum constant  $\mu < 1$  found through trial and error as suggested by Aki and Richards [16]. Various alternative methods for choosing  $\mu$  have been considered and many of these are summarized in the review by Titterton [22].

Finally, note that  $\mu$  may be viewed as an interpolation parameter. As  $\mu \rightarrow 0$ , (62)  $\rightarrow$  (45); as  $\mu \rightarrow \infty$ , (62)  $\rightarrow$  (50). For finite but small  $\mu$ , (62) interpolates between (45) and (50) while providing a stable approximation to (45).

## Inhomogeneous Background

We will now generalize the discussion of the preceding section to the problem of finding appropriate minimization functionals of the form (2) when  $s_b$  is not a constant slowness vector. This situation will generally arise in the later stages of a comprehensive nonlinear inversion algorithm [3].

### *A. Functional design*

If  $s_b$  is not constant, then (10) and (11) no longer play an important role in the design of the least-squares functional. The question arises as to whether some vectors other

than  $u$  and  $v$  can play a similar role. The main issue is whether a vector  $(\bar{u}^T, \bar{v}^T)^T$  can be constructed that has unit eigenvalue for some problem of the form

$$\begin{pmatrix} 0 & M \\ M^T & 0 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} = \bar{\lambda} \begin{pmatrix} \bar{L} & 0 \\ 0 & \bar{C} \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix}. \quad (64)$$

It turns out that such a vector can always be found. Furthermore, there is sufficient freedom in the choice of the vector that an additional condition may be placed on the minimization functional if desired.

Define

$$\bar{v} = s_b / \sigma \quad (65)$$

where  $\sigma$  is some positive constant slowness value chosen for convenience to make  $\bar{v}$  dimensionless [one reasonable choice would be  $\sigma = \sigma_0$  where  $\sigma_0$  is given by (13)], and let  $\bar{u} = u$  for now. Then

$$M\bar{v} = \bar{L}\bar{u} \quad (66)$$

implies that the diagonal components of  $\bar{L}$  satisfy

$$\bar{L}_i = \sum_{j=1}^n l_{i,j}(s_b)_j / \sigma = (t_b)_i / \sigma, \quad (67)$$

where  $(t_b)_i$  is the travelttime for the  $i$ -th ray. Similarly,

$$M^T \bar{u} = C v = \bar{C} \bar{v} \quad (68)$$

where the diagonal components of  $\bar{C}$  satisfy

$$\bar{C}_j = \sigma C_j / (s_b)_j. \quad (69)$$

A derivation analogous to the one leading to (35) then gives

$$\begin{pmatrix} \bar{A}\bar{A}^T & 0 \\ 0 & \bar{A}^T\bar{A} \end{pmatrix} \begin{pmatrix} \bar{w}_\lambda \\ \bar{x}_\lambda \end{pmatrix} = \bar{\lambda}^2 \begin{pmatrix} \bar{w}_\lambda \\ \bar{x}_\lambda \end{pmatrix} \quad (70)$$

where

$$\bar{A} = \bar{L}^{-\frac{1}{2}} M \bar{C}^{-\frac{1}{2}} \quad (71)$$

with  $\bar{w}_\lambda = \bar{L}^{\frac{1}{2}} \bar{u}_\lambda$  and  $\bar{x}_\lambda = \bar{C}^{\frac{1}{2}} \bar{v}_\lambda$ . Because our design of the eigenvalue problem begins by first constructing a strictly positive vector  $(\bar{u}^T, \bar{v}^T)^T$  with eigenvalue unity, it again follows from

the Perron-Frobenius theory of nonnegative matrices [18,19] that this eigenvalue is simple and that it is the maximum eigenvalue of  $\bar{A}^T \bar{A}$  and  $\bar{A} \bar{A}^T$ , and also  $\bar{A}$ . The only caveat is that  $\bar{A}^T \bar{A}$  must be assumed “primitive” as defined in Section 2. Thus, by construction, the eigenvalues  $\bar{\lambda}_j^2$  lie in the range  $[0,1]$ .

An additional degree of freedom arises in the problem with inhomogeneous background, since no constraint analogous to (13) applies. The choice  $\bar{u} = u$  is the simplest, but it is not required. In fact, we are free to specify either  $\bar{u}$  or  $\bar{L}$  in (66) – once one is specified the other is determined. The choice is constrained only by the requirements that all the components of  $\bar{u}$  must be positive, and that the diagonal components of  $\bar{L}$  must be positive (the off-diagonal components vanish). A different choice for  $\bar{u}$  also leads to a different choice of  $\bar{C}$  in (68), *i.e.*, simply replace  $C_j$  in (69) by the  $j$ -th component of the vector  $M^T \bar{u}$ . This additional degree of freedom implies that some other constraint on the final results can be forced by an appropriate choice of these weights. Although not required, the particular choices given by (67) and (69) have several features in common with the choices made in Section 2 which may be enough to recommend them generally: For example, (69) still has the desirable property that the elements of the weight matrix  $\bar{C}$  are directly proportional to the elements of the cell coverage matrix  $C$ , and therefore tend to weight the cells with the most coverage most heavily. Similarly, (67) leads to a weighting favorable to ray paths with the least traveltime – which has virtually the same logic to recommend it as that used by Frank and Balanis [9] for the shortest path length weighting.

The analysis of (70) exactly parallels that of (35). Furthermore, if we make the choices

$$W_1 = \bar{L}^{-1} \tag{72}$$

and

$$W_2 = \bar{C} \tag{73}$$

in (2), then all of the analysis of Section 2 applies to the inhomogeneous background problem with barred quantities replacing unbarred.

We conclude that the analysis of the nonconstant background problem is virtually identical to that for constant background once the weight matrices have been determined.



### *B. Nonlinear inversion*

A general nonlinear iterative inversion algorithm based on the preceding analysis may be summarized as follows: Given traveltimes data but no *a priori* knowledge of a sensible background slowness model, first assume a constant background and compute the best estimate of the model slowness using (25)-(26) with the weights (32) and (33). If the resulting least-squares error is too large, use the new model slowness as the background and repeat the calculation. Continue until the resulting least-squares error is comparable to the error in the measured traveltimes data. Note that, except for the particular choices of weight matrices and an as yet unspecified algorithm for computing the ray paths for the background model, this algorithm is the same as that used by Lytle and Dines [3].

## **Discussion**

Tomographic reconstruction of a slowness model from traveltimes data based on Eq. (1) implicitly assumes that a background slowness  $s_b$  is known: the ray paths used to determine the path length matrix  $M$  should be based on the application of Fermat's principle or Snell's law to the construction of sensible rays through such a model. If the background is homogeneous, the rays are straight. If the background is inhomogeneous, the rays are bent. In either case, the background slowness model should be expected to play a special (and explicit) role in the reconstruction process. The methods developed in this paper show that weighted least-squares criteria can always be found that map the reconstruction problem into a vector space where the contribution to the inversion from the background slowness remains invariant; meanwhile the reconstruction proceeds to find optimum contributions in orthogonal directions. The resulting algorithm produces a hybrid slowness model both consistent with the data in the least-squares sense and possessing the least possible variance from the background.

Applications and extensions of these methods to nonlinear traveltimes inversion will be presented elsewhere.

### **Acknowledgments**

I thank D. M. Goodman, D. H. Rothman, and T. J. Yorkey for helpful conversations. This work was supported in part by the Air Force Office of Scientific Research through grants AFOSR 85-0017 and AFOSR 86-0352, and also in part under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48 and supported specifically by the Engineering Research Program and the Earth Sciences Department Institutional Research and Development Program.

## References

- [1] J. A. Scales, A. Gersztenkorn, and S. Treitel, “Fast  $l_p$  solution of large, sparse, linear systems: Application to seismic travel time tomography,” *J. Comput. Phys.*, vol. 75, pp. 314-333, 1988.
- [2] R. P. Bording, A. Gersztenkorn, L. R. Lines, J. A. Scales, and S. Treitel, “Applications of seismic travel-time tomography,” *Geophys. J. R. astr. Soc.*, vol. 90, pp. 285-303, 1987.
- [3] R. J. Lytle and K. A. Dines, “Iterative ray tracing between boreholes for underground image reconstruction,” *IEEE Trans. Geosci. Remote Sensing*, vol. GE-18, pp. 234-240, 1980.
- [4] T. J. Boullion and P. L. Odell, *Generalized Inverse Matrices*, New York: Wiley, 1971.
- [5] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*, New York: Wiley, 1971.
- [6] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, New York: Academic, 1972.
- [7] R. Penrose, “On best approximate solutions of linear matrix equations,” *Proc. Cambridge Phil. Soc.*, vol. 52, pp. 17-19, 1956.
- [8] G. T. Herman, *Image Reconstruction from Projections – The Fundamentals of Computerized Tomography*, New York: Academic, pp. 103-106 and 210-221, 1980.
- [9] M. S. Frank and C. A. Balanis, “Weighted least squares techniques in electromagnetic geophysical inversions,” *Proc. IGARSS’87 Symposium, Ann Arbor, 18-21 May, 1987*, pp. 7-12, 1987.
- [10] F. Natterer, *The Mathematics of Computerized Tomography*, New York: Wiley, 1986.
- [11] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *J. Theor. Biol.*, vol. 36, pp. 105-117, 1972.
- [12] K. A. Dines and R. J. Lytle, “Computerized geophysical tomography,” *Proc. IEEE*,

- vol. 67, pp. 1065-1073, 1979.
- [13] S. Ivansson, “Remark on an earlier proposed iterative tomographic algorithm,” *Geophys. J. R. astr. Soc.*, vol. 75, pp. 855-860, 1983.
- [14] A. Ben-Israel and D. Cohen, “On iterative computation of generalized inverses and associated projections,” *SIAM J. Num. Anal.*, vol. 3, pp. 410-419, 1966.
- [15] N. R. Burkhard, “Resolution and error of the back projection technique algorithm for geophysical tomography,” Lawrence Livermore National Laboratory Rep., UCRL-52984, September, 1980.
- [16] K. Aki and P. G. Richards, *Quantitative Seismology – Theory and Methods*, Vol. II, San Francisco: Freeman, Sect. 12.3, 1980.
- [17] C. Lanczos, *Linear Differential Operators*, New York: Van Nostrand, Chapt. 3, pp. 100-162, 1961.
- [18] E. Seneta, *Non-negative Matrices – An Introduction to Theory and Applications*, New York: Wiley, pp. 1-6, 1973.
- [19] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, New York: Academic, pp. 27-32, 1979.
- [20] D. W. Marquardt, “An algorithm for least squares estimation of non-linear parameters,” *J. Soc. Indust. Appl. Math.*, vol. 11, pp. 431- 441, 1963.
- [21] D. W. Marquardt, “Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation,” *Technometrics*, vol. 12, pp. 591-612, 1970.
- [22] D. M. Titterington, “Common structure of smoothing techniques in statistics,” *Intern. Stat. Rev.*, vol. 53, pp. 141-170, 1985.