Migration velocity analysis by optimization: linear theory

Paul Fowler

ABSTRACT

A method is presented for finding the velocity function that yields the optimal prestack time migrated image from seismic data. A linear theory is developed which relates changes in the estimated migration velocities to changes in an underlying interval velocity model. This velocity analysis method thus provides not only the migration velocities, but also a corresponding interval velocity model. The algorithm is fully automatic, and requires no picking of traveltimes or horizons. It should correctly estimate velocities in the presence of complex geologic structures. Two versions of the algorithm are developed: the first requires that velocity varies only weakly laterally, whereas the second should handle much larger lateral velocity variations.

INTRODUCTION

In this paper I discuss a formulation of velocity analysis as an optimization problem. My goal is to design an efficient velocity analysis algorithm that requires neither manual picking of reflecting horizons or velocity values, nor extensive ray tracing, yet makes few restrictions on the nature of the reflectors or the velocity field. In particular, I emphasize velocity analysis in regions of complex subsurface geologic structure, where migration is required to produce an adequate image of the subsurface. Because I use migration operators as an intrinsic part of the velocity analysis algorithm, I term this method migration velocity analysis. I first discuss the particular methodological approach taken here toward extracting velocity information from seismic reflection data, and its relation to conventional stacking velocity analysis techniques. Next I consider the case in which structure varies but the velocity function is assumed to be only weakly

laterally variable, and show how the techniques developed by Toldi (1985) can be combined directly with the velocity space dip-moveout and migration techniques I have presented in previous papers (Fowler, 1984a, 1984b). I then examine an extension of these methods applicable for both complex structure and laterally varying velocity functions. The derivation of a theory for this latter case forms the largest part of this paper.

Some of the most challenging problems of velocity analysis and imaging are in those areas where both the velocity function and the reflector geometry vary significantly laterally. In such situations it is generally difficult to obtain good images of the reflector positions unless the velocity is known quite precisely, but it may be equally difficult to determine the velocities well unless the structure is already accurately known. Conventional techniques of stacking velocity analysis can fail due both to lateral velocity variation and to structural effects. Pre-stack time migration methods may be used for finding velocities in complicated structural regions if the velocity function is laterally constant or only slowly varying (Fowler, 1984a, 1984b, 1985; Shurtleff, 1984; Tieman, 1984). Toldi (1985) attacked the other side of the problem: he assumed that the reflector geometry was simple and known, and showed how to invert observed stacking velocities for lateral anomalies in the interval velocities. The principal extension I make upon his work is to use for velocity analysis pre-stack migration instead of normal moveout (NMO) and stacking. By doing so, I hope to be able to resolve laterally varying velocities in the presence of complex structure rather than have to restrict attention to simple planar beds.

Those readers familiar with Toldi's work will recognize immediately the extent to which I have drawn upon his exceptionally lucid presentation as a basis for the present work. Velocity analysis using NMO and stacking is simpler in many ways than migration velocity analysis, so for a reader baffled by the current paper I recommend consulting Toldi's dissertation (1985) as an introduction to this type of velocity analysis algorithm. However, in an attempt to keep this paper reasonably self-contained, I repeat portions of the derivations which are nearly identical for both the stacking velocity and migration velocity approaches. I shall attempt as I proceed to highlight both the similarities and the differences between Toldi's approach and mine, but the cognoscenti may find it possible merely to skim the sections concerning the general formulation of the problem and the 1-D algorithm.

The approach taken in this paper to interval velocity analysis is not the only one possible. Most other methods which have been suggested differ principally in requiring the manual or automatic picking of travel-times or reflecting horizons, or both (Hubral and Krey, 1980; Gray and Golden, 1983; Sword, 1985; Bishop, et al., 1985). The

optimization algorithm proposed here can be construed as implicitly performing an automatic picking, as well, but the "picking" is for migration velocities, which I believe, like stacking velocities, can usually be determined quite reliably.

VELOCITY ANALYSIS AS AN OPTIMIZATION PROBLEM

The approach to velocity analysis taken here is based upon specification of three functions: i) an imaging operator which, upon input of a particular velocity function, converts the seismic data into an (approximate) image of the subsurface geology, ii) an objective function which measures in some sense the quality of the resulting image, and iii) a conversion scheme which relates the imaging velocities to an underlying model of the actual acoustic velocities of the rocks. In a general sense, velocity analysis is taken to consist of trial imaging with a variety of velocity functions, and selection of that one which which maximizes the objective function. The imaging operator used most commonly in conventional velocity analysis is NMO and stack; in this paper I focus on the use of pre-stack time migration. The objective functions used are ordinarily based on some smoothed or time averaged form of energy or magnitude of signal amplitude.

This formulation of velocity analysis as an optimization problem depends upon exploiting the redundancy present in data which has been collected as multiple experiments providing overlapping information about the subsurface. It is perhaps easiest to consider the data as if it arose from a series of constant offset experiments. If the earth model were restricted to flat horizontal reflectors and the velocity were allowed to depend only on depth, then a correct choice of the normal moveout velocity would align the various offset panels better than an incorrect one. (This is in essence a definition of a correct NMO velocity.) Hence, the signal amplitude in the stack would be greater using this NMO velocity function than using any other one for which the constructive summation would be replaced in part by destructive interference due to misalignment. The same principle may be applied to imaging operators more sophisticated than NMO and stack: the best velocity function for imaging is the one which maximizes the magnitude of the amplitude of the signal in the resulting image.

In practice, one probably would not want to use simply the rectified amplitude of the signal, due to the many zeroes, but instead would want to smooth or average to remove the waveform. Two candidates for good objective functions which I intend to test are the energy in the signal, smoothed over a time window of a few periods of the dominant frequency, and the analytic envelope of the signal. Conventional stacking velocity analysis often uses semblance, a coherency measure which may be expressed as a normalized energy; I do not see any good way to incorporate the (non-linear) semblance

normalization into a pre-stack migration scheme.

Besides an objective function, one needs to specify an imaging operator appropriate to one's assumptions about allowed reflector geometry and velocity field. The combination of NMO and stack forms a kinematically accurate imaging operator under the assumptions of flat reflectors and laterally invariant (or only weakly varying) velocity. Lynn and Claerbout (1982), Loinger (1983), Rocca and Toldi (1983), and Toldi (1985) have developed theories showing how stacking velocities may also be inverted for information about laterally varying velocities. The basic principle of their approach is to find the underlying interval velocity model which best explains the observed pattern of stacking velocities; I apply a similar methodology to migration velocities. The NMO and stack operator runs into problems in the presence of complex structure, because it is not easy to unscramble the effects of structure on the observed stacking velocities from those due to lateral velocity variations. I take the viewpoint here that a better starting point in this case is provided by the velocities associated with pre-stack time migration, since this kind of migration yields an appropriate imaging operator for most complex structures until lateral velocity variations become quite large.

The third important component of a velocity analysis scheme of the nature discussed in this paper is the relation between the imaging velocity function and the underlying interval model. I have defined above what I mean by an optimal imaging velocity. The velocity which is physically most meaningful and important, however, is the interval velocity, or local acoustic velocity of the rock. The two velocities could only be expected to be equal if the earth were of a single constant velocity and one's choice of imaging operator were adequate to allow for the effects of whatever structure were present. In general, the relation between these two velocity functions can be very complex; accurate representation of this relation lies at the core of successful extension of laterally invariant velocity analysis to the laterally varying case. The interval velocity is critical for good imaging of seismic data for several reasons. First, as Toldi demonstrated clearly, even in the laterally invariant case simple-minded optimization of the imaging velocity function without reference to the corresponding interval velocity model can result in wildly unreasonable interval velocities; it is the interval velocity model on which constraints to the optimization should be formulated, not on the imaging velocities. Second, in the laterally varying case, the optimum image derived from laterally invariant imaging operators (e.g., a stack or a pre-stack time migration) may give a reasonably well focused image, yet significantly misrepresent the true subsurface structure, since reflectors may be systematically mispositioned. In order to obtain an accurate depth image of the subsurface, the effects of lateral variation in the interval velocities must be

fully allowed for. In particular, a high quality interval velocity model is a prerequisite for any depth migration, either pre- or post-stack. Hence, for migration velocity analysis just as for stacking, the objective function to be maximized should, if at all possible, be related back to an underlying model of interval velocities.

VELOCITY AS A FUNCTION OF DEPTH ALONE

I follow Toldi in referring to the problem of velocity analysis under the restriction of lateral invariance as the 1-D problem; the dimensionality is of the velocity field, not of the structure. The extension of his algorithm for this case to migration velocity analysis is straightforward. Nonetheless, I outline in some detail the 1-D migration velocity algorithm, my purpose being to summarize the formulation and implementation of the optimization algorithm in a context in which the mathematical derivations involved are minimal, before the complexities of the full two-dimensional algorithm are tackled. Before proceeding further I should point out that, although I continue to follow convention and refer to "velocity analysis", most of the theory and implementation in the rest of this paper is cast in terms of inverse velocity, or slowness.

The imaging operator I use for both the 1-D and the 2-D cases is pre-stack time migration. In implementation, the algorithm begins by imaging the data with a suite of reference slowness functions; choosing these functions to be just a range of constant slownesses is easiest both to conceptualize and to implement. This reference imaging may be done practically by pre-stack F-K migration (Shurtleff, 1984; Tieman, 1984), by velocity space dip-moveout and zero-offset migration (Fowler, 1984a, 1984b), or possibly by pre-stack phase shift migration (Fowler, 1985). I treat the results of these various methods as equivalent for the present discussion. I also assume that the slowness axis is sampled sufficiently densely to allow reconstruction of any image corresponding to a variable slowness function by interpolation between the reference panels actually generated.

To calculate the objective function I first calculate either the energy in the image, smoothed over a short time window, or the analytic envelope of the image; in either case, I use $E(w,y,\tau)$ to designate this function and refer to it generically as "energy". Given a slowness function $w(y,\tau)$, the objective function Q is then defined as

$$Q(\mathbf{w}) = \sum_i \sum_j E(w(y_i,\tau_j),y_i,\tau_j) \tag{1}$$

Evaluation of the objective function may be done either by extracting the image corresponding to a specified slowness function, and then summing up E for that image, or by calculating E for each reference panel, and interpolating and summing the E

values corresponding to the choice of slowness function.

In principle, the migration velocity analysis method so far is quite similar to Toldi's approach to stacking velocity analysis. The objective function differs in that I use an unnormalized energy measure instead of semblance, but this difference is minor. The space of reference images is similar, but note that the migration method requires imaging all midpoints at once, whereas (for the 1-D case) the stacking analysis could be done on isolated midpoints. The last step, the relation between the imaging slownesses and the interval slownesses, is the same for the stacking and migration methods: the one-dimensional algorithms both make the conventional identification of imaging velocities with root-mean-square (rms) averages of the interval velocities:

$$w_j(\mathbf{m}) = \frac{1}{[v_{rms}(\mathbf{m})]_j} = \left[\frac{\tau_j}{\sum_{i=1}^{j} \frac{(\tau_i - \tau_{i-1})}{m_i^{\,2}}} \right]^{1/2} \tag{2}$$

The interval slowness model $\mathbf{m}$ corresponding to a given migration slowness function $\mathbf{w}$ can thus be calculated from Dix's equation.

Having specified the three functions which determine the nature of the velocity analysis method, I can proceed to outline an algorithm to implement it. The following is a "generic" steepest ascent algorithm taken nearly verbatim from Toldi (1985, p.21); it can be applied for each midpoint y_i of interest.

Set $\mathbf{m}$ to starting model: $\mathbf{m} = \hat{\mathbf{m}}$
Set $\mathbf{w}$ to starting value: $\mathbf{w} = \hat{\mathbf{w}}$
Calculate $\mathbf{G}$ by taking derivatives of $\mathbf{w}$ at $\hat{\mathbf{m}}$.

Begin loop on iterations

1. Form $\nabla_{\mathbf{m}} Q$ at current model point $\mathbf{m}$:

$$(\nabla_{\mathbf{w}} Q)_j = \frac{E(w_j(\mathbf{m})+\Delta w, y_i, \tau_j) - E(w_j(\mathbf{m}), y_i, \tau_j)}{\Delta w}$$

$$\nabla_{\mathbf{m}} Q = \mathbf{G}^T \nabla_{\mathbf{w}} Q$$

2. Line search for α that maximizes $Q(\mathbf{m}+\alpha\nabla_{\mathbf{m}} Q)$

$$Q(\mathbf{m}+\alpha\nabla_{\mathbf{m}} Q) = Q[\mathbf{w}(\mathbf{m})+\alpha\mathbf{G}\nabla_{\mathbf{m}} Q]$$

3. Update model

$$\mathbf{m} = \mathbf{m}+\alpha\nabla_{\mathbf{m}} Q$$

$$\mathbf{w} = \mathbf{w}+\alpha\mathbf{G}\nabla_{\mathbf{m}} Q$$

recalculate $\mathbf{G}$ by taking derivatives at new $\mathbf{m}$.

End loop on iterations.

The notation of the above algorithm needs to be explained. The migration slowness function is represented by **w** and the interval slowness model by **m**. In this example, both are taken as functions of the zero-offset travel-time, τ, for a fixed midpoint y. The energy value calculated for a particular slowness w and τ is designated by $E(w,\tau)$, and the objective function Q is the sum of E along a given velocity function, as defined above. The matrix **G** comprises elements $G_{jp} = \dfrac{\partial w_j}{\partial m_p}$ which can be calculated by differentiating the rms formula (equation 2) relating **w** to **m** to get

$$\frac{\partial w_j}{\partial m_p} = \frac{\tau_p - \tau_{p-1}}{\tau_p} \left[\frac{w_j(\mathbf{m})}{m_p} \right]^3 \qquad \text{for } p \le j \tag{3}$$

$$= 0 \qquad \text{for } p > j$$

The two gradients $\nabla_{\mathbf{w}} Q$ and $\nabla_{\mathbf{m}} Q$ refer to the changes in Q due to changes in **w** and **m**, respectively. The first gradient is calculated by finite difference approximation from the energy data space, and the second by the chain rule. The steps of updating **w** and **m** are themselves linear approximations.

The algorithm as shown here is a simple steepest-ascent method. In practice, one would want to use more sophisticated methods which converge more rapidly, and would want to include constraints on the optimization. I briefly discuss these matters after a description of the 2-D algorithm; the reader may also consult Toldi's dissertation for more detailed discussion. Most major features of the 1-D algorithm carry over to the 2-D case: the optimization is guided by the gradient $\nabla_{\mathbf{m}} Q$, which can be calculated indirectly from $\nabla_{\mathbf{w}} Q$ if the matrix **G** is known. The 2-D algorithm differs from the 1-D principally in the way that **G** is defined and calculated.

VELOCITY AS A FUNCTION OF BOTH DEPTH AND LATERAL POSITION

In the previous section I assumed that velocity was allowed to vary in the lateral direction only weakly if at all. In this section I derive a theory which attempts to find the laterally and vertically varying interval velocity model which optimally explains the observed peaks in the migration velocity spectrum. That is, I now require only that for each y and τ there is some w for which time migration focuses the data better than for other values of w. I no longer assume that the migration velocity is simply an rms average of the overlying interval velocities; instead I derive a more complicated linearized

relation between **m** and **w**.

The 2-D algorithm uses the same data space as the 1-D, namely panels of pre-stack constant velocity migrations of the original seismic data. The objective function Q is also the same, either the energy smoothed over a short time window or the analytic envelope. The 2-D algorithm, like the 1-D, is guided by $\nabla_{\mathbf{m}} Q$, which tells how to update the model so as to increase the value of the objective function. I write the migration slowness **w** as a function of y and τ, so $w_{ij} = w(y_i, \tau_j)$. Similarly, I write the interval slowness **m** as a function of x and z, so $m_{pq} = m(x_p, z_q)$. For a given component of $\nabla_{\mathbf{m}} Q$ one has

$$(\nabla_{\mathbf{m}} Q)_{pq} = \frac{\partial Q}{\partial m_{pq}} = \sum_i \sum_j \frac{\partial Q}{\partial w_{ij}} \frac{\partial w_{ij}}{\partial m_{pq}} \tag{4}$$

As in the 1-D algorithm, $\frac{\partial Q}{\partial w_{ij}}$ can by approximated from the data by a finite difference calculation:

$$(\nabla_{\mathbf{w}} Q)_{ij} = \frac{\partial Q}{\partial w_{ij}} \approx \frac{E(w_{ij}(\mathbf{m}) + \Delta w, y_i, \tau_j) - E(w_{ij}(\mathbf{m}), y_i, \tau_j)}{\Delta w} \tag{5}$$

Thus what needs to be found is an expression for $\frac{\partial w_{ij}}{\partial m_{pq}}$, relating a change in interval slowness at a particular point (x_p, z_q) to the resulting change in observed migration slowness at some point (y_i, τ_j). To derive an approximate analytic expression for these derivatives, I follow Toldi's lead, and use the intermediary of travel-times.

Consider first a single point diffractor at (x_d, z_d) in a medium of constant slowness w. If one were to run a seismic survey passing over this point, the kinematics of the pre-stack point diffractor would be given by the pyramid equation (Claerbout, 1985)

$$t = w\sqrt{z_d^2 + (y - h - x_d)^2} + w\sqrt{z_d^2 + (y + h - x_d)^2} \tag{6}$$

Suppose now that the slowness model were perturbed. The travel-time data for the point diffractor would now be a set $\{t_{ik}, y_i, h_k\}$ which would no longer satisfy equation (6) exactly. However, if the perturbations were not too large, it should be possible to define a slowness W, a zero-offset time T, and a location Y for which an equation of the form

$$t = \sqrt{T^2 + W^2(y - h - Y)^2} + \sqrt{T^2 + W^2(y + h - Y)^2} \tag{7}$$

best fits the data points in a least-squares sense. Note that it is necessary to consider

the changes in T and Y as well as W because, for a laterally varying perturbations in the slowness model, it would not in general be true that $T = Wz_d$ and $Y = x_d$ as they would be for the constant slowness background of the starting model.

The pyramid equation (7) does not lend itself to a $t^2 - x^2$ linearizing parametrization such as Toldi used for least-squares analysis of stacking slownesses. Instead, I solve the problem of fitting a pyramid through the data points $\{t_{ik}, y_i, h_k\}$ by linearizing around an initial value of $(\hat{W}, \hat{T}, \hat{Y})$. Then

$$t_{ik} \approx t(\hat{W}, \hat{T}, \hat{Y}) + \frac{\partial t}{\partial W} \Delta W + \frac{\partial t}{\partial T} \Delta T + \frac{\partial t}{\partial Y} \Delta Y \tag{8}$$

where all the partial derivatives are evaluated at $(\hat{W}, \hat{T}, \hat{Y}, y_i, h_k)$.

Solving the least-squares system yields a solution of the form

$$\begin{bmatrix} \Delta W \\ \Delta T \\ \Delta Y \end{bmatrix} = \frac{1}{D} \begin{bmatrix} \sum_{i,k} A_{ik} \ \Delta t_{ik} \\ \sum_{i,k} B_{ik} \ \Delta t_{ik} \\ \sum_{i,k} C_{ik} \ \Delta t_{ik} \end{bmatrix} \tag{9}$$

The coefficients A_{ik}, B_{ik}, C_{ik}, and D are all functions of W, T, and Y, as well as depending on the geometry of the seismic experiment. Values for them are derived in Appendix A; the expressions become quite involved.

These equations (9) describe how the $W, T,$ and Y arising from a single point diffractor at (x, z) change when the model is perturbed, and hence also the travel-times. I intend, as the notation suggests, to identify (W, T, Y) with $(w(y, \tau), \tau, y)$. This identification deserves some discussion. The definition of migration slowness by maximizing the energy in the image is not tractable for easy analytical expression, since it is fundamentally dependent on the wave nature of the original data. Hence I have substituted a definition based on ray tracing and travel-times. The two are not truly interchangeable, as a simple example shows. Consider again a single point diffractor in a constant slowness medium. The travel-times would satisfy equation (7) exactly, so W would equal the medium slowness w. The actual data which would be gathered by a seismic experiment passing over the point diffractor could be represented to a reasonable approximation by convolution of this travel-time pyramid with a wavelet. Assume the wavelet is simple in form and of a fairly short duration. Think of pre-stack time migration as a diffraction stack, or surface integral, along a pyramid shaped surface; the optimal migration slowness would be the one which maximizes this diffraction stack integral. For the constant slowness medium, the migration slowness estimate could be expected to equal

the actual medium slowness. The two definitions of w and W would thus be expected to yield the same number. Now suppose that one trace has a large travel-time perturbation introduced, shifting it by an amount greater than the length of the wavelet. The migration based estimate of w would not be affected by the single erratic value, since the best diffraction stack would still be obtained by summing together all the remaining traces along the same pyramidal surface as before and ignoring the trace with the large shift. The least-squares estimate of W, on the other hand, might be shifted significantly by such a anomalous travel-time point. However, such a dramatic change in only a single trace could only be expected to arise from a large, very localized perturbation in the underlying interval slowness model. Smaller, more broadly distributed changes in the interval slowness would cause more gentle changes in the travel-time surface, and it is is reasonable to expect that W would be adequately close to w. This is one of several approximations I make which I expect may prevent this method from resolving high wavenumber components of the velocity field well.

A further problem remains to be dealt with before I can turn to the question of relating travel-time changes to changes in the interval slowness model, and complete the derivation of the linear operator. I have found how W, T, and Y change for a fixed diffractor point. However, what I really want to know is how w changes for a fixed (y,τ). Changing the model causes a given diffractor point to appear at a different (y,τ), so I need to figure out not just how $w(y,\tau)$ changes, but also which diffractor point it represents in the updated model. Let primes indicate the perturbed model, so the diffractor point associated with $w(y,\tau)$ is associated with $w'(y',\tau')$ for the new model. Let $\Delta w = w' - w$, $\Delta y = y' - y$, and $\Delta\tau = \tau' - \tau$. Also, let the partial derivatives of w relative to a fixed (y,τ) be indicated by the usual partial derivative notation $\frac{\partial w}{\partial m}$, etc., and let the partial derivatives relative to a fixed diffractor location (x,z) be denoted by $\frac{\delta w}{\delta m}$, etc. Then I have

$$w'(y',\tau') \approx w(y,\tau) + \sum_{p,q} \frac{\delta w}{\delta m_{pq}} \Delta m_{pq} \tag{10}$$

Also to first order, I have,

$$\left.\frac{\partial w'}{\partial y'}\right|_{(y',\tau')} \approx \left.\frac{\partial w}{\partial y}\right|_{(y,\tau)} \tag{11}$$

and

$$\left.\frac{\partial w'}{\partial \tau'}\right|_{(y',\tau')} \approx \left.\frac{\partial w}{\partial \tau}\right|_{(y,\tau)} \tag{12}$$

I can then write

$$
\begin{aligned}
w'(y,\tau) &\approx w'(y'-\Delta y,\tau'-\Delta\tau) \\
&\approx w'(y',\tau') - \left.\frac{\partial w'}{\partial y'}\right|_{(y',\tau')} \Delta y - \left.\frac{\partial w'}{\partial \tau'}\right|_{(y',\tau')} \Delta\tau \\
&\approx w(y,\tau) + \sum_{p,q} \frac{\delta w}{\delta m_{pq}} \Delta m_{pq} - \left.\frac{\partial w}{\partial y}\right|_{(y,\tau)} \Delta y - \left.\frac{\partial w}{\partial \tau}\right|_{(y,\tau)} \Delta\tau \\
&\approx w(y,\tau) + \sum_{p,q} \frac{\delta w}{\delta m_{pq}} \Delta m_{pq} - \frac{\partial w}{\partial y} \sum_{p,q} \frac{\delta y}{\delta m_{pq}} \Delta m_{pq} - \frac{\partial w}{\partial \tau} \sum_{p,q} \frac{\delta \tau}{\delta m_{pq}} \Delta m_{pq}
\end{aligned}
$$

I now make the identification of (W,T,Y) with $(w(y,\tau),\tau,y)$ and get

$$
\frac{\partial w_{ij}}{\partial m_{pq}} = \frac{\partial W_{ij}}{\partial m_{pq}} - \left(\frac{\partial w}{\partial \tau}\right)_{ij} \frac{\partial T_{ij}}{\partial m_{pq}} - \left(\frac{\partial w}{\partial y}\right)_{ij} \frac{\partial Y_{ij}}{\partial m_{pq}} \tag{13}
$$

The derivatives of w with respect to τ and y can be evaluated by finite differences. The derivatives of W, T, and Y with respect to the model have been recast previously in terms of derivatives with respect to changes in travel-times. To complete the linearization, I need a relation between Δt_{ik} and Δm_{pq}.

I follow Toldi's methodology again, tracing rays and using Fermat's principle. The major difference is that a point perturbation in m affects not just one midpoint and offset, but a family of them, since the same ray may be followed for many different combinations of midpoints and offsets. I index midpoints by the subscript i and offsets by k. The subscript a refers to the anomaly coordinates. For a given ray S_{ik} one has

$$
t_{ik} = \int_{S_{ik}} dS_{ik}\, m(x_a, z_a) \tag{14}
$$

Invoking Fermat's principle, I perturb the model and calculate the changes in travel times integrating the slowness perturbations along the *unperturbed* ray S_{ik}:

$$
\Delta t_{ik} = \int_{S_{ik}} dS_{ik}\, \Delta m(x_a, z_a) \tag{15}
$$

This last calculation is valid for quite a general model, but to apply it directly requires tracing many rays at every iteration. I choose instead to analytically evaluate these derivatives against a simple constant slowness background model for which the ray paths are straight, and use these approximate values in place of the more accurate values which would be calculated by ray tracing using an iteratively updated model. Should greater accuracy prove necessary, it might prove possible to calculate similar results for a depth variable background model; I have not yet done so.

I now calculate what the travel-time derivatives would be in a simple, constant slowness reference medium. Figure 1 shows the geometry of the ray path for a particular diffractor point, midpoint, and offset. Represent the travel-time data as (t_{ik}, y_i, h_k). Let the subscript d refer to the coordinates of a particular diffractor point, and let the subscript a refer to the location of a slowness anomaly, that is, a particular element of the model $\mathbf{m}$. Using the notation of figure 1, equation (15) becomes

$$\Delta t_{ik} = \int_{x_a} dx_a \int_{z_a} dz_a \, \Delta m(x_a, z_a) \left(\frac{\delta_1}{\cos\theta_{ik}} + \frac{\delta_2}{\cos\phi_{ik}} \right) \tag{16}$$

where

$$\delta_1 \equiv \delta[x_a - y' + \mu_{ik}(z_a)] \tag{17}$$

and

$$\delta_2 \equiv \delta[x_a - y' - \mu_{ik}(z_a)] \tag{18}$$

Rewrite equation (9) as

$$\Delta W(Y,T) = \frac{1}{D} \sum_i \sum_k A_{ik} \, \Delta t_{ik} \tag{19}$$

and substitute from eq (16) for ΔT to yield

$$\Delta W(Y,T) = \frac{1}{D} \sum_i \sum_k A_{ik} \int_{x_a} dx_a \int_{z_a} dz_a \, \Delta m(x_a, z_a) \left(\frac{\delta_1}{\cos\theta_{ik}} + \frac{\delta_2}{\cos\phi_{ik}} \right) \tag{20}$$

Pull the integrals outside of the sums and make the identification of $W(Y,T)$ with $w(y_d, \tau_d)$ to get

$$\Delta w(y_d, \tau_d) = \int_{x_a} dx_a \int_{z_a} dz_a \, G_W(y_d, \tau_d, x_a, z_a) \Delta m(x_a, z_a) \tag{21}$$

where

$$G_W(y_d, \tau_d, x_a, z_a) = \sum_{i=1}^{N_y} \sum_{k=1}^{N_h} \frac{A_{ik}}{D} \left(\frac{\delta_1}{\cos\theta_{ik}} + \frac{\delta_2}{\cos\phi_{ik}} \right) \tag{22}$$

This Green function $G_W(y_d, \tau_d, x_a, z_a)$ can be identified with $\dfrac{\partial W_{ij}}{\partial m_{pq}}$ in equation (13); they both represent the change in $w(y, \tau)$ due to a perturbation in $m(x, z)$. Evaluation of the the Green function G_W in a form suitable for implementation involves substituting for the trigonometric terms in equation (22) and using the delta functions to eliminate one of the sums; details may be found in Appendix B. The result for G_W is

$$G_W(y_d, \tau_d, x_a, z_a) = \frac{\gamma}{\tau_d D} \left[\tau_d^2 + w^2(\gamma x_a - y_d)^2 \right]^{1/2} \times \tag{23}$$

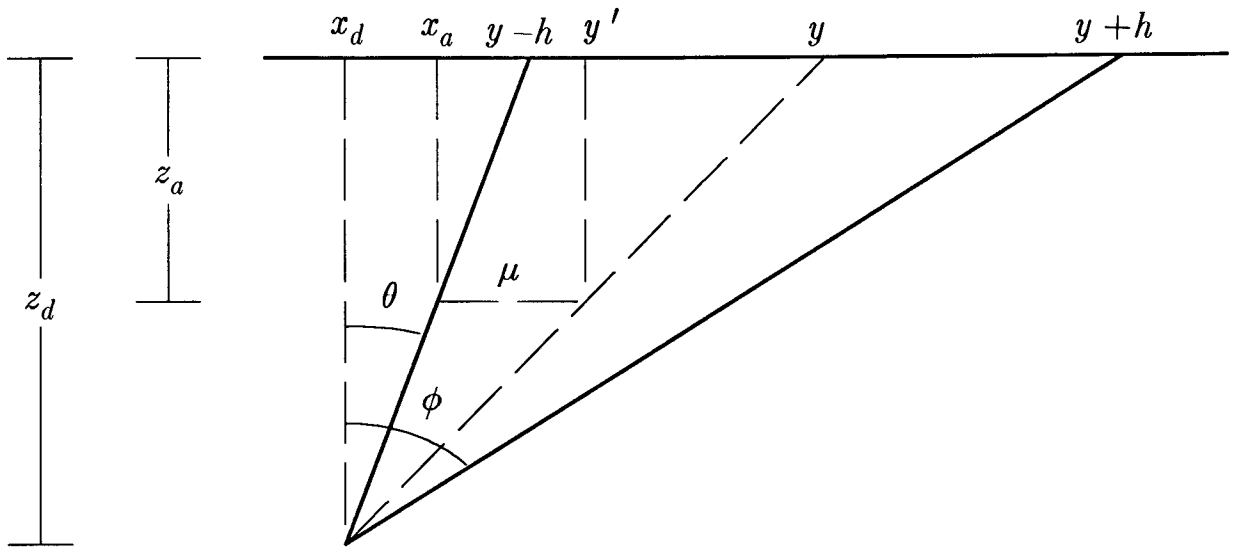

FIG. 1. Geometry of rays for a single diffractor point and a constant background slowness. The rays for a single trace with midpoint y and offset h are shown. The diffractor is at (x_d, z_d). The point in the model at which the slowness is perturbed is designated by (x_a, z_a). The quantities θ, ϕ, μ, and y' are used in calculating the effect on the travel-time of perturbing the slowness.

$$\sum_{k=1}^{N_h} \left[A\,(y=\gamma x_a + h_k, h_k) + A\,(y=\gamma x_a - h_k, h_k) \right]$$

where

$$\gamma = \frac{w\,\tau_d}{w\,\tau_d - z_a} \tag{24}$$

One can also write similar Green function representations for $G_T(y_d, \tau_d, x_a, z_a) = \dfrac{\partial T_{ij}}{\partial m_{pq}}$ and for $G_Y(y_d, \tau_d, x_a, z_a) = \dfrac{\partial Y_{ij}}{\partial m_{pq}}$ simply by substituting B_{ik} or C_{ik} in place of A_{ik} in equations (22) and (23).

From the above discussion, Equation (13) becomes

$$G_{ijpq} = (G_W)_{ijpq} - \left(\frac{\partial w}{\partial \tau}\right)_{ij} (G_T)_{ijpq} - \left(\frac{\partial w}{\partial y}\right)_{ij} (G_Y)_{ijpq} \tag{25}$$

I have thus written the migration slowness perturbation directly as a linearized function of the interval slowness perturbations and **G** is now in a form suitable to incorporate in an optimization algorithm of the same general outline as used in the 1-D example. The new 2-D algorithm has the same form as the 1-D except that now all midpoints *must* be considered, and the computation of **G** is much more involved. The steps for computing **G** as developed above can be summarized as follows:

For each (y_d, τ_d)
  Calculate $\frac{\partial w}{\partial y} \approx \frac{w(y_d+\Delta y, \tau_d) - w(y_d, \tau_d)}{\Delta y}$
  Calculate $\frac{\partial w}{\partial \tau} \approx \frac{w(y_d, \tau_d+\Delta\tau) - w(y_d, \tau_d)}{\Delta \tau}$
  Calculate $D(w(y_d, \tau_d), y_d, \tau_d)$ by eq. (A8)
  For each (x_a, z_a)
    Calculate γ by eq. (24)
    For each h_k
      Calculate $A(w(y_d, \tau_d), y_d, \tau_d, y = \gamma x_a + h_k, h_k)$ by eq. (A5)
      Calculate $A(w(y_d, \tau_d), y_d, \tau_d, y = \gamma x_a - h_k, h_k)$ by eq. (A5)
      Calculate corresponding B, C values by eqs. (A6) and (A7)
    End loop over h_k
    Calculate $G_W(y_d, \tau_d, x_a, z_a)$ by eq. (23)
    Calculate corresponding G_T and G_Y
    Calculate $G(y_d, \tau_d, x_a, z_a)$ by eq. (25)
  End loop over (x_a, z_a)
End loop over (y_d, τ_d)

Note that, despite the debauch of indices in equation (25), **G** can still be treated as a matrix, albeit a large one. Note also that many elements of **G** are actually zero. I do not consider turning rays, so the loop over z_a can end at $z_a \le \tau_d / w(y_d, \tau_d)$. The loop over x_a can also be limited by considering the maximum travel-time represented in the data, since travel-times for very wide angle rays may exceed the recording limit. As a reasonable approximation (based on straight ray paths), I expect **G** to be zero for $x_a \ge z_a \sqrt{(1-\tau^2/t_{\max}^2)}$.

DISCUSSION OF IMPLEMENTATION

Both the one and two dimensional algorithms begin by generating a suite of pre-stack time migrated sections and computing E at each point in each section; this space of reference migrated images effectively constitutes the "data space" for the inversion. The 2-D algorithm uses the data at all midpoints simultaneously during the velocity analysis; the 1-D algorithm may be applied to selected midpoints once the data space has been constructed. However, since much of the cost of the velocity analysis is computing the data space of reference migrations, I shall assume that the 1-D algorithm is also to be applied at all midpoints. The starting model $\hat{\mathbf{m}}$ for the 2-D algorithm will probably be slowness as a function of depth only, to make computation of the initial $\hat{\mathbf{w}}$ easy. The starting model for the 1-D case would be the same, or possibly just a single slowness constant with depth.

As Toldi discussed in his dissertation (Toldi, 1985), there are several major enhancements possible on the basic algorithms of the kind presented here. Smoothing the energy panels over slowness for the first few iterations can help make up for a poor starting model. Use of conjugate-gradient methods such as Partan (Luenberger, 1973) should lead to a large acceleration in convergence. The basic idea remains the same as in the simple steepest-ascent method outlined here, but the gradient information is incorporated in a more efficient manner in such methods. Toldi found that it was not necessary to recalculate $\mathbf{G}$ at each step, and that re-using the same $\mathbf{G}$ for several iterations gave entirely acceptable results. If this proves true here also, it will provide a substantial speed-up, since once the reference migration data panels are calculated, most of the work of the algorithm appears to be involved in calculating the $\mathbf{G}$ matrix. I note in passing that dropping the G_T and G_Y terms from equation (25) should still give a convergent velocity analysis algorithm; I expect that the resulting velocity model would be systematically distorted in the same manner that a time migration mispositions events relative to a depth migration, since the terms incorporating the lateral and depth corrections have been suppressed. This may provide a useful check during testing of the algorithm. Finally, it may prove possible to cut down the size of $\mathbf{G}$, and hence the run time, by considering only a fraction of the points on one or more of the y_d, τ_d, x_a, or z_a axes. Since the energy function is smoothed over time, sub-sampling τ_d may be quite acceptable. Sub-sampling x_a or z_a would result in a lower resolution velocity model, which may be good enough; it may not prove possible or desirable to attempt to resolve the highest wavenumber components of the model.

Toldi also discussed the incorporation of constraints during the optimization. In particular, I shall probably want to include smoothness constraints in x and z. I expect

the high wave number components in both x and z to be the least well determined. For the 1-D algorithm, I certainly do not want to allow the model to vary rapidly in x, since that would contradict the underlying theory, which assumes that w varies laterally only slowly if at all. One simple way to incorporate such smoothness constraints is to add penalty terms to the objective function Q:

$$Q(\mathbf{m}) = \sum_{i,j} E(w_{ij}(\mathbf{m}), y_i, \tau_j) - \mathbf{m}^{\mathbf{T}}(\beta \mathbf{D_x}^{\mathbf{T}}\mathbf{D_x} + \epsilon \mathbf{D_z}^{\mathbf{T}}\mathbf{D_z})\mathbf{m} \tag{26}$$

where β and ϵ would be adjustable damping parameters determining the degree of lateral and vertical smoothing to be used. For further discussion of how to include such constraints, see Toldi (1985), Claerbout (1976), and Menke (1984).

ACKNOWLEDGMENTS

I thank John Toldi, both for useful discussion and for writing so lucidly of his stacking velocity analysis techniques. I have learned much from him. I also thank Clement Kostov, Dan Rothman, and Michael Schlax for critical review of drafts of this paper, and Chuck Sword, Dave Hale, Bill Harlan, Rick Schult, and Francis Muir for helpful comments and suggestions along the way. Finally, I want to acknowledge the invaluable assistance of the symbolic algebra program MACSYMA in the least-squares computations of Appendix A.

REFERENCES

Bishop, T.N., Bube, K.P., Cutler, R.T., Langan, R.T., Love, P.L., Resnick, J.R., Shuey, R.T., Spindler, D.A., and Wyld, H.W., 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, 50, 903-923

Claerbout, J.F., 1976, Fundamentals of geophysical data processing: McGraw-Hill, Inc.

Claerbout, J.F., 1985, Imaging the earth's interior: Blackwell, London

Fowler, P.J., 1984a, Incorporating dip corrections in velocity analysis using constant velocity stacks: SEP-39

Fowler, P.J., 1984b, Velocity independent imaging of seismic reflectors: Paper presented at the 54th Annual SEG Meeting, in Atlanta

Fowler, P.J., 1985, Velocity space imaging: formalism, methods, and prospects: SEP-42

Gray, W.C., and Golden, J.E., 1983, Velocity determination in a complex Earth: Paper presented at the 53rd Annual SEG Meeting, in Las Vegas

Hubral, P., and Krey, T., 1980, Interval velocities from seismic reflection time measurements: Society of Exploration Geophysicists

Loinger, E., 1983, A linear model for velocity anomalies: Geophysical Prospecting, 31, 98-118

Luenberger, D.G., 1973, Introduction to linear and nonlinear programming: Addison-Wesley Publishing Company, Inc.

Lynn, W.S., and Claerbout, J.F., 1982, Velocity estimation in laterally varying media: Geophysics, 47, 884-897

Menke, W., 1984, Geophysical data analysis: discrete inverse theory: Academic Press

Rocca, F., and Toldi, J.L., 1983, Lateral velocity anomalies: Paper presented at the 53rd Annual SEG Meeting, in Las Vegas

Shurtleff, R.N., 1984, An f-k procedure for pre-stack migration and migration velocity analysis: Paper presented at the 46th annual meeting of the E.A.A.G. in London

Sword, C.H., 1985, Tomographic determination of interval velocities from CDR data - preliminary results: SEP-44

Tieman, H.J., 1984, Migration velocity analysis: Theoretical development and practical results: Paper presented at the 54th annual meeting of the S.E.G. in Atlanta

Toldi, J.L., 1985, Velocity analysis without picking: Ph.D. dissertation, Stanford University, also SEP-43

APPENDIX A

LEAST SQUARES SOLUTIONS FOR W, T, AND Y

In this appendix I derive expressions for least-squares solutions for W, T, and Y. The set of equations to solve are given by

$$t_{ik} \approx t(\hat{W},\hat{T},\hat{Y}) + \frac{\partial t}{\partial W}\Delta W + \frac{\partial t}{\partial T}\Delta T + \frac{\partial t}{\partial Y}\Delta Y \tag{A1}$$

To make the notation more compact, denote the partial derivatives by subscripts: $\frac{\partial t}{\partial W} \equiv t_W$, etc. Explicitly,

$$\begin{bmatrix} t_W \\ t_T \\ t_Y \end{bmatrix} = \frac{1}{\sqrt{T^2+W^2(y-h-Y)^2}} \begin{bmatrix} W(y-h-Y)^2 \\ T \\ W^2(Y-y+h) \end{bmatrix} + \frac{1}{\sqrt{T^2+W^2(y+h-Y)^2}} \begin{bmatrix} W(y+h-Y)^2 \\ T \\ W^2(Y-y-h) \end{bmatrix} \tag{A2}$$

I then need to solve the follow system of normal equations for $(\Delta W, \Delta T, \Delta Y)$:

$$\begin{bmatrix} \sum_{i,k}(t_W)_{ik}\Delta t_{ik} \\ \sum_{i,k}(t_T)_{ik}\Delta t_{ik} \\ \sum_{i,k}(t_Y)_{ik}\Delta t_{ik} \end{bmatrix} = \begin{bmatrix} \sum_{i,k}(t_W)_{ik}^2 & \sum_{i,k}(t_W)_{ik}(t_T)_{ik} & \sum_{i,k}(t_W)_{ik}(t_Y)_{ik} \\ \sum_{i,k}(t_W)_{ik}(t_T)_{ik} & \sum_{i,k}(t_T)_{ik}^2 & \sum_{i,k}(t_T)_{ik}(t_Y)_{ik} \\ \sum_{i,k}(t_W)_{ik}(t_Y)_{ik} & \sum_{i,k}(t_T)_{ik}(t_Y)_{ik} & \sum_{i,k}(t_Y)_{ik}^2 \end{bmatrix} \begin{bmatrix} \Delta W \\ \Delta T \\ \Delta Y \end{bmatrix} \tag{A3}$$

This matrix equation has the solution

$$\begin{bmatrix} \Delta W \\ \Delta T \\ \Delta Y \end{bmatrix} = \frac{1}{D} \begin{bmatrix} \sum_{i,k} A_{ik}\ \Delta t_{ik} \\ \sum_{i,k} B_{ik}\ \Delta t_{ik} \\ \sum_{i,k} C_{ik}\ \Delta t_{ik} \end{bmatrix} \tag{A4}$$

where

$$A_{ik} = \left[\sum_{r,s}(t_T)_{rs}^2 \sum_{r,s}(t_Y)_{rs}^2 - \left(\sum_{r,s}(t_T)_{rs}(t_Y)_{rs} \right)^2 \right] (t_W)_{ik} \tag{A5}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_Y)_{rs}^2 \right] (t_T)_{ik}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \sum_{r,s}(t_T)_{rs}^2 \right] (t_Y)_{ik}$$

$$B_{ik} = \left[\sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_Y)_{rs}^2 \right] (t_W)_{ik} \tag{A6}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}^2 \sum_{r,s}(t_Y)_{rs}^2 - \left(\sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \right)^2 \right] (t_T)_{ik}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_W)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}^2 \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} \right] (t_Y)_{ik}$$

$$C_{ik} = \left[\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \sum_{r,s}(t_T)_{rs}^2 \right] (t_W)_{ik} \tag{A7}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_W)_{rs}(t_Y)_{rs} - \sum_{r,s}(t_W)_{rs}^2 \sum_{r,s}(t_T)_{rs}(t_Y)_{rs} \right] (t_T)_{ik}$$

$$+ \left[\sum_{r,s}(t_W)_{rs}^2 \sum_{r,s}(t_T)_{rs}^2 - \left(\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \right)^2 \right] (t_Y)_{ik}$$

$$D = \sum_{r,s}(t_W)_{rs}^2 \sum_{r,s}(t_T)_{rs}^2 \sum_{r,s}(t_Y)_{rs}^2 - \sum_{r,s}(t_T)_{rs}^2 \left(\sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \right)^2 \tag{A8}$$

$$+ 2 \sum_{r,s}(t_W)_{rs}(t_T)_{rs} \sum_{r,s}(t_W)_{rs}(t_Y)_{rs} \sum_{r,s}(t_T)_{rs}(t_Y)_{rs}$$

$$- \sum_{r,s}(t_W)_{rs}^2 \left(\sum_{r,s}(t_Y)_{rs}(t_T)_{rs} \right)^2 - \sum_{r,s}(t_Y)_{rs}^2 \left(\sum_{r,s}(t_W)_{rs}(t_T)_{rs} \right)^2$$

It would be nice if closed form solutions or approximations could be found expressing these many summations as simple functions of the upper and lower bounds; none are apparent to me yet. Toldi was able to approximate his analogous, but much simpler, sums by integrals which could be solved explicitly. Unfortunately, not only do the integrals become quite complicated here, but if my computations are correct, they give rise to elliptic integrals which cannot have simple closed form solutions.

APPENDIX B

DERIVATION OF THE GREEN FUNCTIONS $\mathbf{G_W}$, $\mathbf{G_T}$, AND $\mathbf{G_Y}$

In this appendix I derive explicit expressions for the Green function matrices $\mathbf{G_W}$, $\mathbf{G_T}$, and $\mathbf{G_Y}$. Consider first equation (22)

$$G_W(y_d,\tau_d,x_a,z_a) = \sum_{i=1}^{N_y}\sum_{k=1}^{N_h} \frac{A_{ik}}{D}\left(\frac{\delta_1}{\cos\theta_{ik}} + \frac{\delta_2}{\cos\phi_{ik}}\right) \tag{B1}$$

I first look at the sum containing δ_1.

$$(G_W)_1(y_d,\tau_d,x_a,z_a) = \sum_{i=1}^{N_y}\sum_{k=1}^{N_h} \frac{A(y_i,h_k)\delta_1}{D\cos\theta} \tag{B2}$$

Substitute

$$\cos\theta = \tau_d \left[\tau_d^2 + w^2(y-h-y_d)^2\right]^{-1/2} \tag{B3}$$

$$\gamma = \frac{w\,\tau_d}{w\,\tau_d - z_a} \tag{B4}$$

$$y' = \frac{y}{\gamma} \tag{B5}$$

and

$$\mu = \frac{h}{\gamma} \tag{B6}$$

to get

$$\delta_1 = \delta[x_a - y' + \mu_{ik}(z_a)] \tag{B7}$$

$$= \delta\left[x_a - \frac{(y-h)}{\gamma}\right] \tag{B8}$$

$$= \gamma\,\delta[\gamma x_a - (y-h)] \tag{B9}$$

and then

$$(G_W)_1(y_d,\tau_d,x_a,z_a) = \sum_{i=1}^{N_y}\sum_{k=1}^{N_h} \frac{wA(y_i,h_k)\left[\tau_d^2+w^2(y-h-y_d)^2\right]^{1/2}\delta[\gamma x_a-(y_i-h_k)]}{(w\tau_d-z_a)D} \quad \text{(B10)}$$

$$= \frac{\gamma}{\tau_d D}\left[\tau_d^2+w^2(\gamma x_a-y_d)^2\right]^{1/2}\sum_{i=1}^{N_y}\sum_{k=1}^{N_h} A(y_i,h_k)\,\delta[\gamma x_a-(y_i-h_k)] \quad \text{(B11)}$$

The double sum thus contains non-zero terms only when $y_i-h_k=x_a$. N_y, the number of midpoints, is generally much larger than N_h, the number of offsets, so equation (B11) reduces to

$$(G_W)_1(y_d,\tau_d,x_a,z_a) = \frac{\gamma}{\tau_d D}\left[\tau_d^2+w^2(\gamma x_a-y_d)^2\right]^{1/2}\sum_{k=1}^{N_h} A(y=\gamma x_a+h_k,h_k) \quad \text{(B12)}$$

I note in passing that the y in the above equation will now usually not fall on a grid point y_i; the value of A can be calculated by interpolation between the values of A at neighboring grid points. By a derivation similar to that of equation (B12), the sum in equation (B1) containing δ_2 becomes

$$(G_W)_2(y_d,\tau_d,x_a,z_a) = \frac{\gamma}{\tau_d D}\left[\tau_d^2+w^2(\gamma x_a-y_d)^2\right]^{1/2}\sum_{k=1}^{N_h} A(y=\gamma x_a-h_k,h_k) \quad \text{(B13)}$$

so

$$G_W(y_d,\tau_d,x_a,z_a) = \frac{\gamma}{\tau_d D}\left[\tau_d^2+w^2(\gamma x_a-y_d)^2\right]^{1/2}\times \quad \text{(B14)}$$

$$\sum_{k=1}^{N_h}\left[A(y=\gamma x_a+h_k,h_k)+A(y=\gamma x_a-h_k,h_k)\right]$$

The expressions for G_T and G_Y are identical except for substituting B and C for A.