A visual relationship between Kirchhoff migration and seismic inversion

John C. Bancroft

INTRODUCTION

The exploding reflector model and the finite-difference methods automatically take care of the amplitudes when using the downward continuation method. Similarly, the FK method of migration applies a defined amplitude scaling when moving the data in the FK space. Estimation of the diffraction stack amplitudes proved more of a challenge until the Kirchhoff integral solution to the wave equation provided a theoretical foundation.

Assumptions used in the design of geological models are reviewed in preparation for evaluating the design of migration programs that are derived from the wave-equation. A review of Kirchhoff migration is then presented that begins as a diffraction stack process, and then proceeds to matched-filtering concepts and the integral solution to the wave-equation. One-dimensional (1D) convolution modelling and deconvolution are then used to introduce inversion concepts that lead to "transpose" processes and matched filtering. These concepts are then expanded for two-dimensional (2D) data to illustrate that Kirchhoff migration is a "transpose" process or matched filter that approximates seismic inversion.

MODELS

The kinematic solutions used in seismic migration are relatively straightforward to derive or estimate, and are independent of the type of models used. However, the amplitudes of the wavefield on wavefronts, or along raypaths, are not straightforward, and do depend on the type of model used. Consider for example, an expanding wavefield for a source (shot) record that is acquired along a straight line. We assume that the data represents a 2D vertical plane below the acquisition line. A synthetic 2D seismic source record could be created that is based on a 2D solution to the wave-equation. This convenient solution will contain concentrically expanding, circular wavefronts with amplitudes that vary proportionally to the inverse of the square-root of the distance travelled. A more accurate solution would be based on a 3D model with spherically expanding wavefronts, giving amplitudes that vary proportionally to the inverse of the distance travelled.

Consider Figure 1 that contains illustrations of 1D, 2D, and 3D models, all with the same constant velocity and lossless medium. A source at any point in the 1D model will propagate along the line and the amplitude will not attenuate with distance. In the 2D plane, energy from a source point will propagate outward, away from the point in concentric circles. The total sum of energy in one concentric circle will be the same in another concentric circle. The energy density along the perimeter of each circle will decrease proportionally to the reciprocal of the circumference or radius. Since the amplitude is proportional to the square root of the energy density, the amplitude will decrease proportionally to the square root of the radius or distance travelled. In the 3D volume, the energy will propagate outward in concentric spheres with an energy density

that decreases proportionally to the inverse of the surface area of the sphere, or radius squared. The amplitude (again the square-root of the energy density) will therefore decrease proportionally to the inverse of the radius. In summary then, the amplitude radiating away from a point source is:

- 1D constant
- 2D $\propto \frac{1}{\sqrt{r}}$
- 3D $\propto \frac{1}{r}$

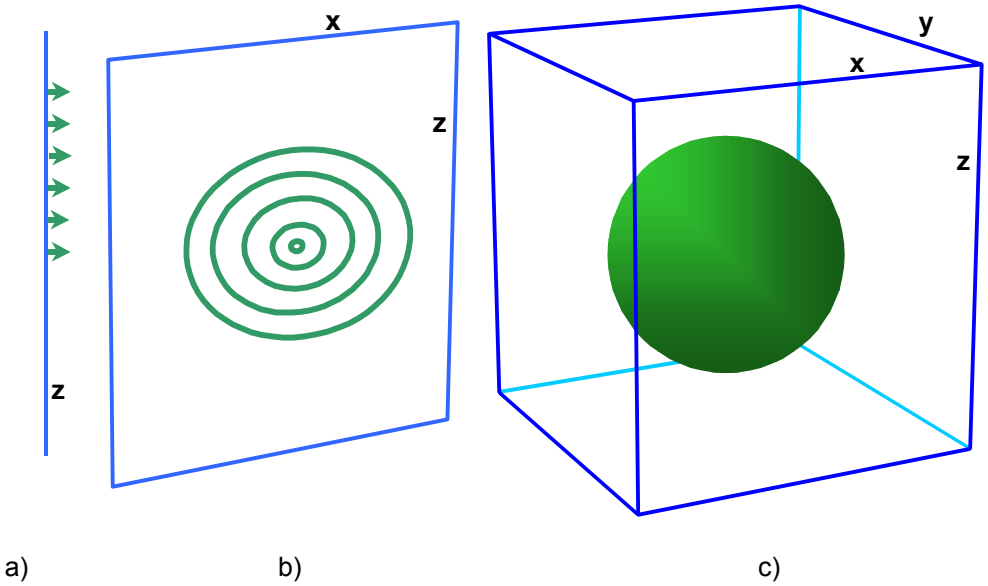


FIG. 1: Models for a) one dimension (z), b) two dimensions (x, z), and c) three dimensions (x, y, z).

Often models are defined in 2D and extended to 3D with assumptions that the geology is independent of the third dimension, a typical assumption for a 2D seismic line that is acquired normal to the dipping structure. The 2D model is defined in Figure 2a, with the corresponding 3D in part (b).

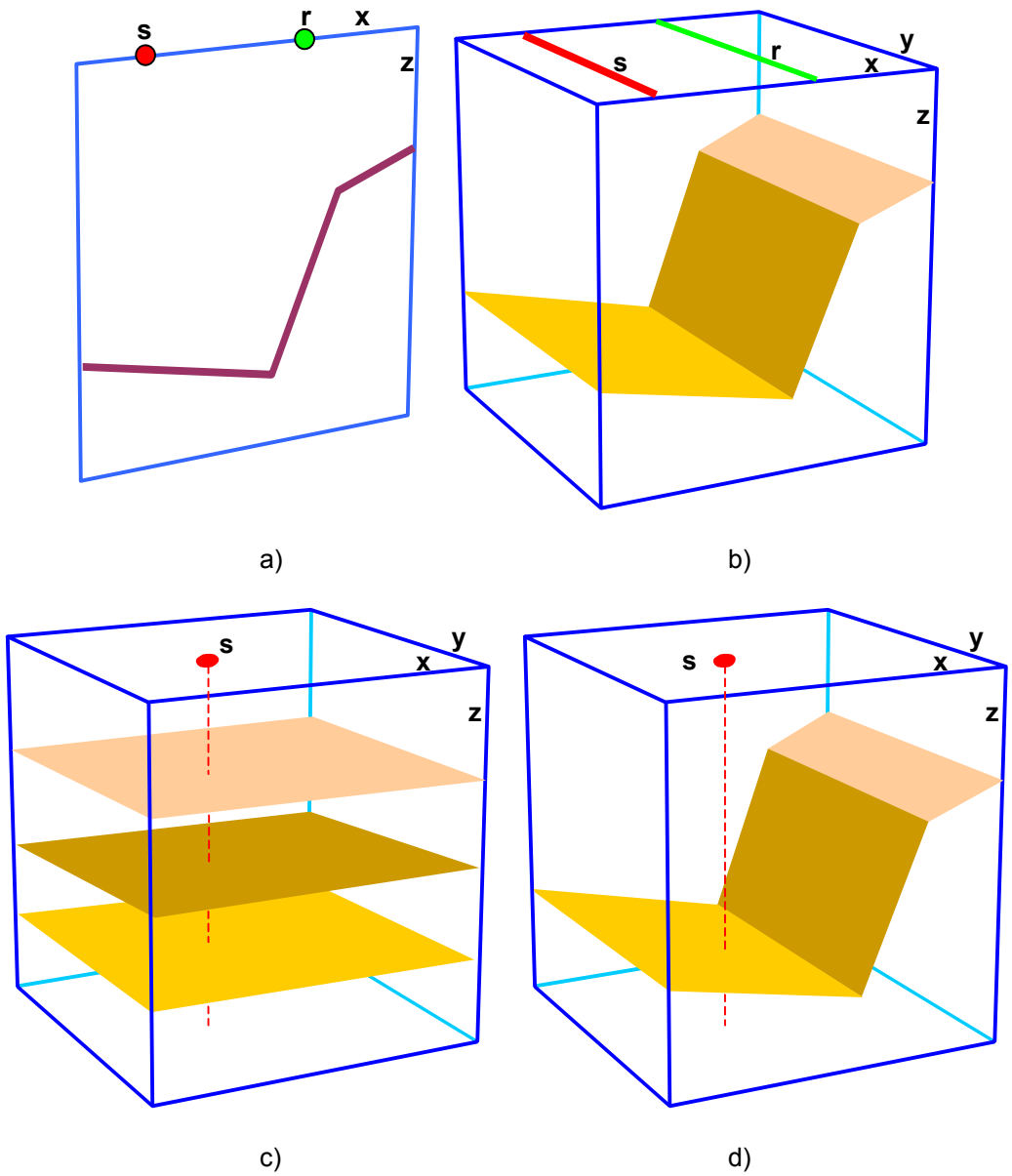


FIG. 2: Seismic models, a) 2D, b) 3D model that is equivalent to the 2D model in (a), c) a 1¼D model, and d) a 2½D model.

One might assume that a 2D seismic line modelled in the 2D space of Figure 2a would be the same as one modelled from a seismic line that crosses the surface of the 3D geology in Figure 2b. However, for the two models to be the equivalent, the 3D volume requires a line source and possibly a line receiver across the surface of the volume in the y direction. Actually, one point-receiver, anywhere along the line, should give the same result.

It becomes more complex with the definition of 1¼D for a source point in a 3D medium that varies only with z as in Figure 2c, or 2½D in which a source point is considered in a 3D volume with the geology that varies only in the (x, z) plane as illustrated in Figure 2d.

The complexity increases (or possibly simplifies) when models are created, or migration algorithms assume, that spherical divergence has been corrected with some form of "gain recovery". Fortunately, amplitudes computed using a 2D algorithm can be corrected for 3D amplitudes with simple correction factors described by Geiger (2001).

EVOLUTION OF AMPLITUDE IN KIRCHHOFF MIGRATIONS

Diffraction stacking or Kirchhoff migration produces one migrated sample at a time by: first, computing a diffraction shape for a scatterpoint at that location; second, summing and weighting the input energy along a diffraction path; and third, placing the summed energy at the scatterpoint location on the migrated section. The process is repeated for all migrated samples. During summation, the amplitudes of the input data are weighted, and it is this weighting of the input data that we are investigating, and which is the dominant objective of many inversions.

Seismic traces contain wavelets that represent different properties, depending on the assumed model. For example, with flat data, the peak amplitude of the wavelet may be assumed to represent the amplitude of a reflecting boundary, or the same wavelet may be considered part of a wavefield. The amplitude will be handled differently when combining all the traces to form an image of the subsurface. Amplitudes may be computed by a number of processes such as:

- stacking
- diffraction stacking, and matched filtering
- solutions to the wave-equation
- inversion principles

all of which are based on a specific type of model.

Stacking

Consider the preparation of traces in a common midpoint (CMP) gather where gain recovery has been applied to each trace. We now assume that the amplitudes of the wavelets represent the reflection coefficients from the subsurface geology. Normal moveout (NMO) correction has been applied to match the traveltimes of offset traces with those at zero-offset. A mute is then applied to ensure that all the contributing wavelets look similar. These wavelets are summed, and then divided by the number of contributing traces, to produce an average of the wavelets. This averaging process maintains the amplitude of the wavelet while attenuating the amplitude of noise. The result is a zero-offset trace with an improved signal to noise ratio (SNR).

This simple procedure has been used by the industry for many years and is a very powerful process. It automatically handles uneven acquisition geometries and produces a section of the subsurface that has balanced amplitudes that are ready for a zero-offset migration.

The same procedure can be applied to a prestack migration where all traces within the migration aperture are assumed to contain wavelets that represent reflectivity. For one

migrated trace location, all input traces within the migration aperture will have moveout correction defined by the double-square-root equation, followed by a mute and stacking to complete the prestack migration. Note that the stacking includes the division of the total number of contributing prestack samples for each migrated sample. It is that simple. Prestack-migrated sections that use this brute force approach will have reasonable amplitude balancing in areas with uneven acquisition geometry (without creating artificial data to fill in the missing shots).

This stacking method ignores amplitude variations with offset (AVO) that are associated with the Zoeppritz equations. However, most migration algorithms also ignore these effects, some deliberately, so that they may be observed in the migrated offset data and lead to additional parameter estimation.

Although the stacking method of controlling amplitudes is considered brute force, I do want to emphasize its robustness. Some data examples appear to produce better subsurface images with post-stack migration that with prestack migration (Cary, 2001). These examples usually occur in areas that have unbalanced acquisition geometries and the conventional stacking of common midpoint (CMP) gathers accurately balances the amplitudes. A prestack migration that is based on wavefield assumptions may not be able to handle the unbalanced geometry and produce an inferior result. In these areas, brute force stacking in the prestack migration may produce a superior result.

There are a number of weaknesses in this stacking model, especially when we consider the prestack energy, recorded at the surface, to be part of a wavefield. Prestack algorithms that consider these additional insights may have an improved image with a better SNR.

Diffraction stacking

Diffraction stack migration was an early form of the Kirchhoff method and was a relatively simple concept when considering the kinematics of the method. The seismic data was assumed to be composed of diffractions where the amplitudes decreased down the flanks by a ratio of T_0/T, as illustrated by the arrows in Figure 3.

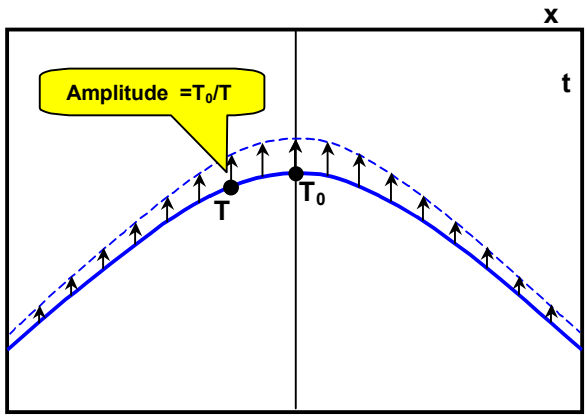
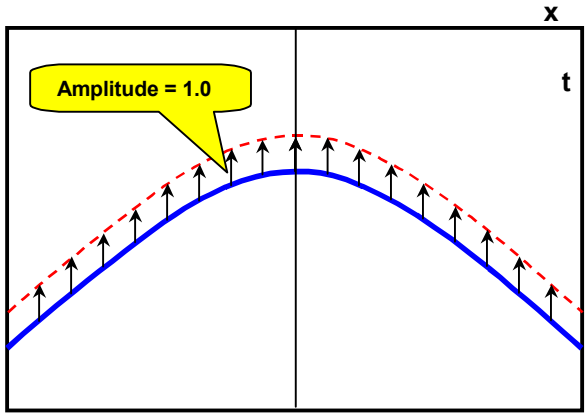
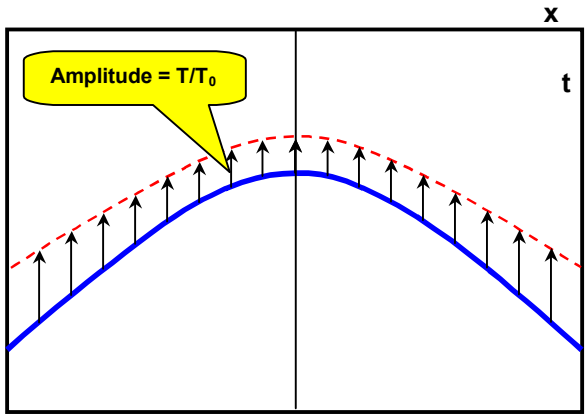


FIG. 3: Assumed amplitudes along a diffraction.

Could the migration be improved if a weighting factor was included when the energy in the diffraction was summed? If so, what would that weighting factor be? Two possible weightings are illustrated in Figures 4a and 4b. The first weighting scheme of Figure 4a does nothing at all (i.e. it is a stacking process if the sum is divided by the number of samples), but allows the program to run faster. The second weighting method assumed that the weaker parts of the diffraction should be amplified to compensate for their decreasing amplitude. The answer came with the matched-filter concept.



a)



b)

FIG. 4: Possible weighting schemes for diffraction stacking.

Matched filtering

Matched filtering is an engineering concept (see Ziemer, 1995) that extracts a known signal from a noisy channel. The noisy signal is cross-correlated with the signal that is being detected. This cross-correlation produces a zero-phase wavelet (autocorrelation) with its central peak at the zero-time of the wavelet, and with a maximum signal-to-noise ratio (SNR). This is the same concept used with vibroseis data in which the source sweep is cross-correlated with the recorded seismic signal. The cross-correlation places a zero-phase wavelet (or possibly a mixed phase wavelet), at the time of a reflection event.

In the migration problem, the signal to be detected is the diffraction. Diffraction stack migration then becomes a cross-correlation of the data with an estimate of the diffraction. According to matched-filter theory, the amplitudes on the estimated diffraction must therefore match those on the original diffraction, as defined in Figure 3. This method of migration will produce an optimum SNR on the migrated signal. Claerbout (1992, page 108) does provide a caution when using the method of matched filters when estimating amplitude verses offset; but more on that later.

The matched filter has provided a basis for choosing the amplitude weighting factors along the diffraction to provide an optimum SNR. Even if wrong amplitude weighting factors were used, the Kirchhoff method would still give a reasonable image of the subsurface. Consequently, even if the wrong model was chosen when deriving the amplitude weighting factors, reasonable results are obtained with the migration algorithms. However, there are other factors that must also be considered as described below.

Additional amplitude considerations for diffraction stacking

Figure 5a shows two diffractions with an aperture that extends to the maximum time where a shallower diffraction contains a larger summation path relative to the deeper diffraction. If the input section contained evenly distributed noise, the shallower migrated data would contain more noise as it has a larger summation path (see the examples published in Larner 1990, but first distributed privately in 1976). Early diffraction stack and Kirchhoff migrations exhibited this shallow noise, and it was considered a negative feature of the methods.

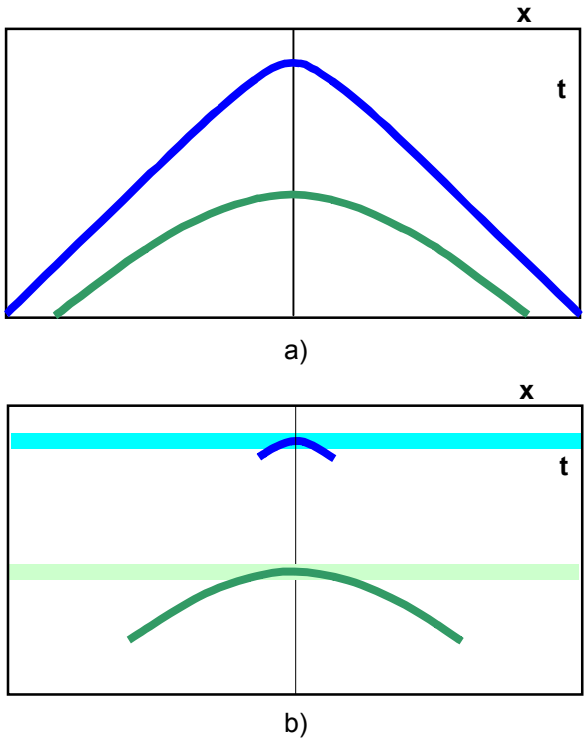


FIG. 5: The extent of diffractions, a) full aperture and b) dip-limited to 45 degrees of dip.

The shallow noise problem was rectified by limiting the aperture on a diffraction to a maximum dip limit, as illustrated in Figure 5b. Note that the shallow diffraction has a much smaller summation path and will have less noise.

Now assume that the input section contains only horizontal reflectors, illustrated by the two horizontal events in Figure 5b. The events have the same wavelet that is represented by the width of the two lines. The tangential area of the deeper diffraction is larger and will produce a larger amplitude than that summed by the shallow diffraction. The actual size of the summation window is proportional to the size of the Fresnel zone, which, for a given wavelet will increase in size proportionally to the square-root of the traveltime. (It is not proportional to T as one might expect, as the frequency of the wavelet remains constant as T changes.) We will therefore expect an additional amplitude term that compensates for the depth of a diffraction given by $1/\sqrt{T}$.

Another factor requiring attention is the effect that velocity has on the size of the summation window. The asymptotes of a low velocity are steeper with a narrow aperture, while a larger velocity has shallower asymptotes with a larger aperture. This effect, combined with the size of the Fresnel zone, produces an additional scale factor that compensates for variable velocities and is given by $1/\sqrt{V}$.

The three factors of matched filtering, the size of the Fresnel zone, and variable velocity, give a combined scale factor A_{2D} for 2D data given by

$$A_{2D} = \frac{T_0}{T}\frac{1}{\sqrt{T}}\frac{1}{\sqrt{V}} = \frac{T_0}{T\sqrt{TV}} \quad . \tag{1}$$

Note that there is also the question of spherical divergence. Has it been applied prior to migration, or is it a part of the migration process itself? In addition, the unbalanced units of equation (1) may also provide a scaling problem,

Phase considerations

The methods of modelling with diffractions, diffraction stacking, or wavefront propagation using Huygen's principle, have another feature in common, and that is a phase-shift that occurs on linear events. The phase-shift is 45 degrees for 2D data, and 90 degrees for 3D data. There is also some frequency filtering that accompanies the phase-shift. These effects are correctable by comparing the wavelets from horizontal events before and after the process. They became deterministic when it was realized that the diffraction stack method was an approximation to the Kirchhoff integral solution of the wave-equation.

KIRCHHOFF INTEGRAL SOLUTION TO THE WAVE-EQUATION

The parameters used in the diffraction stack method were estimated from the physical modelling experiments. This migration process became rigorous when it was recognized (Schneider 1978) that the Kirchhoff integral solution to the wave equation, which was used in optics, gave a theoretical solution for seismic migration. This theoretical solution provided both the amplitude and phase filters that had been previously predicted by

experimentation. A 2D integral solution to the wave-equation from Gazdag (1984) is shown in equation (2):

$$p_{2\text{-}D}\left(x_1, z_1, t=0\right) = \int \frac{\cos\beta}{\sqrt{2\pi\, rc}} \frac{\partial\, p^{1/2}}{\partial\, t}\ (x, z=0, t=r/c)\, dx \tag{2}$$

where r is the radial distance from the source receiver location to the scatterpoint, $c = V/2$, and β the geological dip for the appropriate position on the diffraction. The cosine term may be replaced by T_0/T, giving a more familiar form of:

$$p_{2\text{-}D}\left(x_1, z_1\right) \quad = \quad \int \frac{T_0}{VT^{3/2}} \frac{\partial\, p^{1/2}}{\partial\, t}\ (x, t=r/c)\, dx \tag{3}$$

The square-root derivative of the input data, when evaluated in the frequency domain, produces the 45-degree phase-shift along with an increase in the amplitude that is proportional to the square-root of the frequency. For 3D data, a full derivative is required of the input data. Equation (3) shows the amplitude scaling being applied at the time of summation. Dellinger et al. (1999) have shown that the scaling can be eliminated from the summation process (thereby increasing the speed of the algorithm) by applying the "T" weightings to the input data before summing, and the "T_0" weighting (including the velocity V at T_0) after the summation has been completed.

The Kirchhoff, FK, and downward-continuation methods of seismic migration are based on wave-equation solutions. These migration algorithms produce an image of the subsurface by propagating the energy recorded on the surface back to the area of the reflector. In contrast to these wave-equation methods, seismic inversion attempts to estimate the reflectivity of a geological model from the recorded energy. Quite often, these inversions produce an algorithm that is almost identical to that of the Kirchhoff method, with only slight changes to the amplitude scaling.

INTRODUCTION TO GEOPHYSICAL INVERSION

Geophysical forward modelling involves estimating a response, such as seismic data, from a known geological model. Geophysical inversion is the "reverse" procedure and estimates a geological model from recorded seismic data, borehole data, or other observations such as gravity and/or electromagnetics. The inversion can estimate rock parameters or create some form of a subsurface image; be a single or multi-dimensional problem; broadband or band-limited to seismic data frequencies; and be formed from a combination of geophysical data and observations. In addition, mathematical inversions, such as the least-squares method, Born-WKBJ inversion, linear and non-linear methods, etc., are used with the above geophysical inversions. Understanding the specific uses of the word inversion can be a challenge; however, an excellent starting point for a literature review is the 1988 SEG Geophysical reprint series, No. 9, edited by Lines and Levin.

Early approaches to inversion by Lindseth (1979) attempted to estimate the seismic impedance by combining integral forms of the seismic traces with low-frequency information derived from stacking velocities or sonic logs of nearby wells. Other

approaches use least-squares methods to minimize errors between real data and data created from some geological model (see all papers by Lines). Other approaches by Bleistein (2001) use advanced mathematics such as integration by stationary phase and perturbation theory to estimate reflectivity from simple models. The theory is then applied to the Kirchhoff weightings for more complex models. Topics often associated with these methods involve true amplitude processing and/or seismic illumination.

It is the intent of this section to introduce a heuristic approach to the least-squares method using basic matrix theory. I will start with a one-dimensional model of convolution that uses a two-dimensional wavelet matrix, then present the inversion as a deconvolution with a known wavelet. Two-dimensional seismic modelling and migration simply become a 2D version of the 1D example.

Convolution model

Our model for a seismic data *s(t)*, will be the reflectivity *r*, convolved with the diffraction *d(t)*, and ignoring for the moment the spatial dimensions, i.e.,

$$s(t) = r * d(t) . \tag{4}$$

If equation (4) is expressed in the frequency domain, the convolution becomes a product, i.e.,

$$S(f) = R \bullet D(f) . \tag{5}$$

This linear form of the equation enables a simple estimation of the reflectivity at each frequency from

$$R = \frac{S(f)}{D(f)} . \tag{6}$$

The actual reflectivity could then be found by summing all the frequency components using the Fourier transform. The problem with this procedure is that *D(f)* may go to zero at some frequency, and that we can't perform the division in equation (6) when there are zeros in *D(f)*. As in deconvolution, the procedure becomes an estimation problem that is solved with a variety of methods. In addition, what is the "inverse" of *D(f)*? Are there any other ways to linearize the problem? What could be done to *D(f)* that would allow the division to take place?

Linearization of convolution using linear algebra

We had linearized the convolution process above by using the frequency domain. We can also linearize convolution using matrix theory of linear algebra. I will start with a one-dimensional model where the reflectivity, wavelet, and trace are functions of time as defined by the convolution equation,

$$s(t) = r(t) * w(t) . \tag{7}$$

To many, this equation implies that we place a number of wavelets on the seismic trace, each with a starting time defined by the time of the reflector, and with a wavelet amplitude that is proportional to the amplitude of the reflectivity, as implied in Figure 6. The amplitudes on the trace will be defined when all the wavelets have been inserted.

An alternate equation that defines the convolution process is given by

$$s(t) = \int_{-\infty}^{\infty} r(\tau) \bullet w(t-\tau) d\tau \quad , \tag{8}$$

where we assume that the time is constant at t to compute one seismic trace sample at $s(t)$. We can then use any time, t, to compute all values on the trace. These two approaches to convolution, (adding an entire wavelet to a trace, or completely computing one sample at a time), produce the same result, and we will use both these concepts.

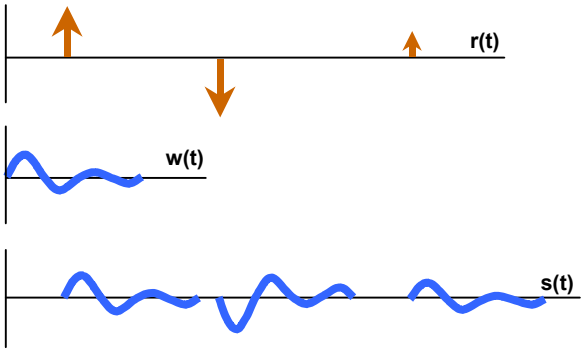


FIG. 6: Illustration of the convolution problem in a conventional manner.

The functions used in equation (7) and (8) may be sampled with the same time interval and defined in digital space as a reflectivity vector, r, with M data elements, wavelet, w, with K data elements, and a seismic vector, s, in which each element is defined the summation equation

$$s_n = \sum_m r_m w_{n-m} \quad , \tag{9}$$

in which the maximum number of samples in s is N, defined by

$$N = M + K - 1 \quad . \tag{10}$$

In equations (8) and (9), the wavelet is reversed in time and delayed as illustrated by w_r in Figure 7. One sample in the seismic vector s is computed by first, multiplying the corresponding elements in r and w_r, i.e. the dot product, and second, summing these elements. Finally, all samples in s are computed by varying n, ($1 \leq n \leq N$). Figure 7 shows the reversed wavelet, w_r, (for one output sample, n) in a vector that matches the size of the reflectivity vector, where the summation parameters in equation (9) would be

(1 ℂ *m* ℂ*M*). This form is convenient when considering the vectorized math that follows. However, an observant reader may recognize that a more efficient summation window would be defined by the *K* samples of the wavelet, i.e., (*n*-*K*+*1* ℂ *m* ℂ *n*), where *n* is the defined output sample.

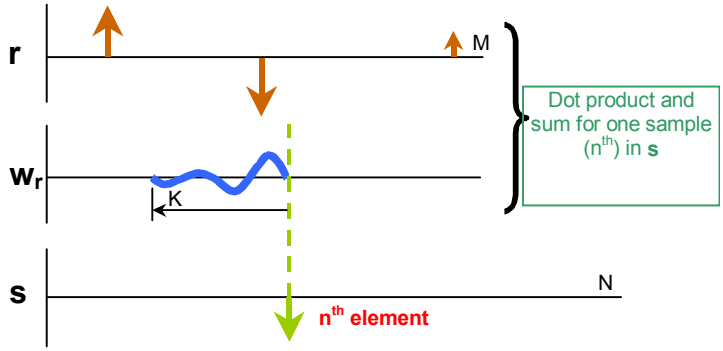


FIG. 7: Convolution defined for a single sample.

All elements of the seismic vector can be computed by repeating the process illustrated in Figure 7 for each value of *n* as illustrated in Figure 8, which shows numerous wavelets with varying delays. In this cartoon figure, and those that follow, the sample interval is much finer along the rows to define the wavelet than between the rows for illustration purposes. The intent is to illustrate the reversed wavelet moves one sample to the right when progressing to the next row, i.e. at the n^{th} row, the right side of the wavelet is at the n^{th} sample. The sample by sample product of the reflectivity vector with the vector at the n^{th} row, (containing the enlarged wavelet), is summed then stored at the n^{th} location in the seismic vector.

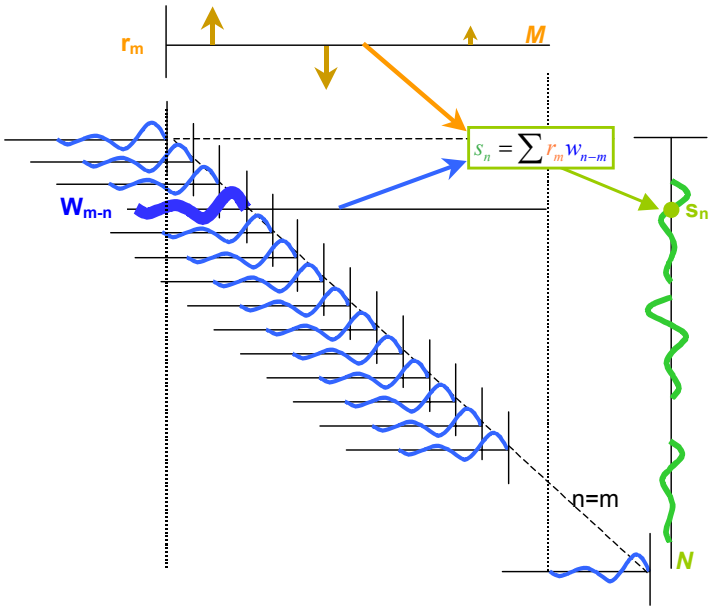


FIG. 8: Computing other samples in convolutional model.

The wavelet data illustrated in Figure 8 is in a convenient form to be defined by a two-dimensional matrix **W** that has N rows and M columns, with the number of columns defined to match the number of elements in the reflectivity vector. Each row in the matrix defines at least a portion of the delayed and time-reversed wavelet, w_{m-n}, where n represents the row number and m the column number. All the elements on a given diagonal of **W** will be the same, i.e. **W** is a Toeplitz matrix. To be consistent with matrix algebra, we define the reflectivity as a column vector **r**, and the seismic trace as a column vector **s**. The matrix equation for the convolution process then becomes a linear process,

$$\underset{[N\times M]}{\mathbf{W}}\ \underset{[M]}{\mathbf{r}} = \underset{[N]}{\mathbf{s}}, \tag{11}$$

where, for convenience, the dimensions are shown in square brackets below the matrix and vectors.

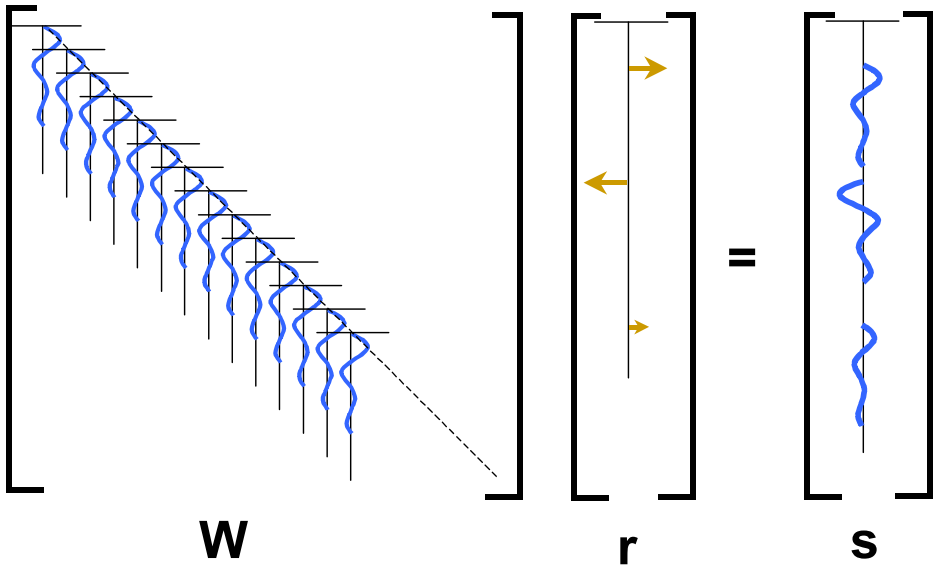


FIG. 9: Matrix form of the convolutional model

Since **W** represents two-dimensional sampled data, traces can be plotted as rows in Figure 8, or plotted as columns in Figure 9. The intent of this figure is to emphasize the impulse response form of the convolution process. Assume only one reflectivity element that progressively moves down its corresponding wavelet column. As the impulse steps down the column, the amplitude of the wavelet in **W** is added into the seismic vector, **s**. Other reflectivity impulses will also sum the corresponding wavelet to the seismic trace. This is in contrast to convolution as defined by equation (9) and illustrated in Figure 8.

An interesting feature of **W**, displayed as impulse responses in Figure 9, is that these impulse responses could vary with time and change from a high-frequency wavelet at shallow times to a low-frequency wavelet at deeper times, according to some attenuation property such as "Q". This non-stationary form of deconvolution was presented by Margrave (1998). Figure 10 contains such an image, where the wavelet increases in

length with increasing time to simulate a constant-Q. It is this time-varying property that will allow time-varying diffractions in the following section.

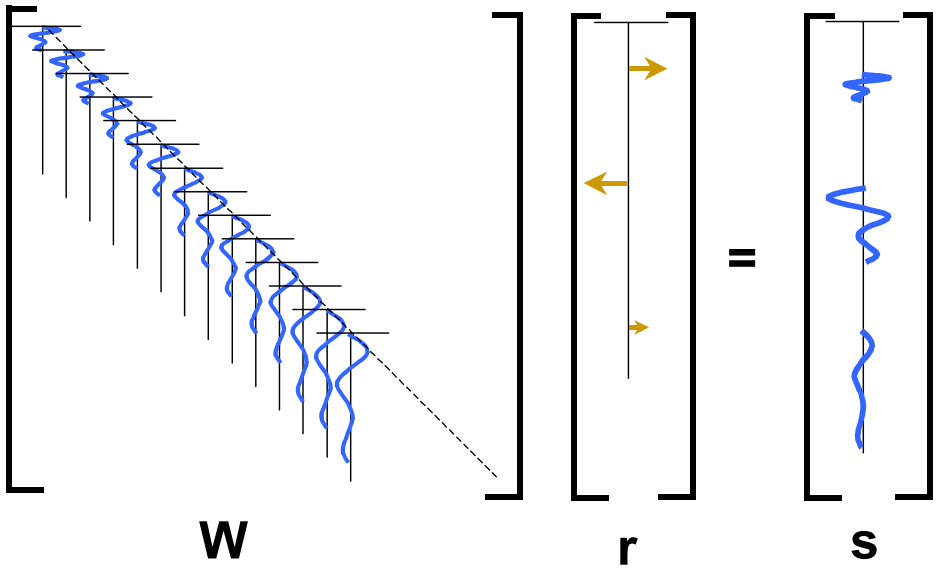


FIG. 10: Matrix form of the convolutional model with time varying wavelets.

Reversing the modelling process

It was easy to do the forward modelling where the seismic trace was computed from a reflectivity and wavelet using equation (11). We now "reverse" this modelling process to estimate the reflectivity from a given seismic trace and known wavelet. Let's start by trying to take the inverse of **W** in equation (11) to get

$$\underset{[M]}{\mathbf{r}} = ? \; \underset{[M\times N?]}{\mathbf{W}^{-1}} \underset{[N]}{\mathbf{s}}. \tag{12}$$

That probably won't work because we can't take the inverse of a matrix that is not square, and we would have to make $M = N$. That could be solved, but the biggest problem is that **W** may contain zeros, which, when inverted, would become infinite. There are a number of ways to modify **W** to make it invertible such as adding small numbers to eliminate the zeros (such as pre-whitening); however, in this paper I will only consider the linear process that leads to the "least-squares" solution.

Least-squares solution

One solution to this problem is to multiply both sides of equation (11) by the transpose of **W**,

$$\underset{[M\times N]}{\mathbf{W}^T} \underset{[N\times M]}{\mathbf{W}} \underset{[M]}{\mathbf{r}} = \underset{[M\times N]}{\mathbf{W}^T} \underset{[N]}{\mathbf{s}}, \tag{13}$$

and then inverting $\mathbf{W}^T\mathbf{W}$ giving

$$\underset{[M]}{\mathbf{r}} = \left(\underset{[M\times N][N\times M]}{\mathbf{W}^T \mathbf{W}} \right)^{-1} \underset{[M\times N][N]}{\mathbf{W}^T \mathbf{s}}. \tag{14}$$

This solution for estimating **r** is the well known "least-squares method" because is provides an optimum solution by minimizing the square of the error between the data and the estimate. The proof of the least-squares property can be found in many texts and papers (Lines 1984). The product, $\mathbf{W^T W}$, produces a square matrix referred to as **A**, which can be inverted only when there are no zeros in the spectrum of the auto-correlation wavelet.

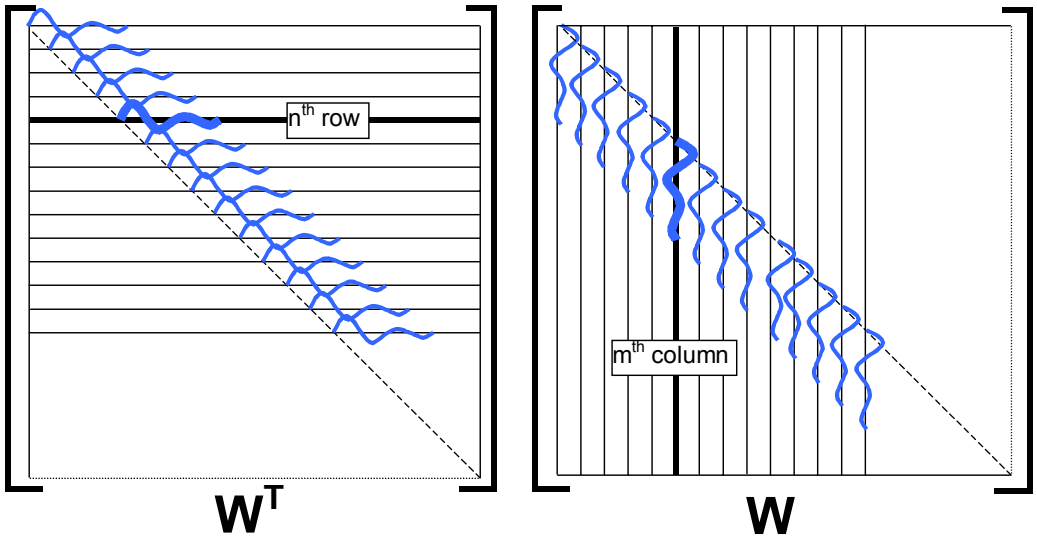


FIG. 11: Illustration of WT and W with corresponding rows and columns displayed for taking the product of the two matrices.

What does **A** look like? The two input matrix are each displayed in Figure 11 with the data plotted as rows in $\mathbf{W^T}$ and as columns in **W**, preparatory to taking the product of two matrices. For one sample in **A**, $a_{n,m}$, we take the dot product of the n^{th} row in $\mathbf{W^T}$, with a m^{th} column in **W**, as illustrated by the highlighted row and column in Figure 11. The scalar value becomes the $(n, m)^{th}$ element in the product matrix **A**, i.e.,

$$a_{n,m} = \sum_{k=1}^{N} w_{n,k}^T w_{k,n}. \tag{15}$$

This process is partially illustrated in Figure 12 where the m^{th} column of $\mathbf{W^T W}$ is plotted as a transposed vector, $\mathbf{W_m}^{\mathbf{T}}$, horizontally above **W**. Samples in the m^{th} column of **A** are computed from the dot product of the transposed vector, $\mathbf{W_m}^{\mathbf{T}}$, with the corresponding rows of **W**. This process is identical to taking the auto-correlation of the wavelet, with the result identified by the zero-phase wavelet on the column of **A**.

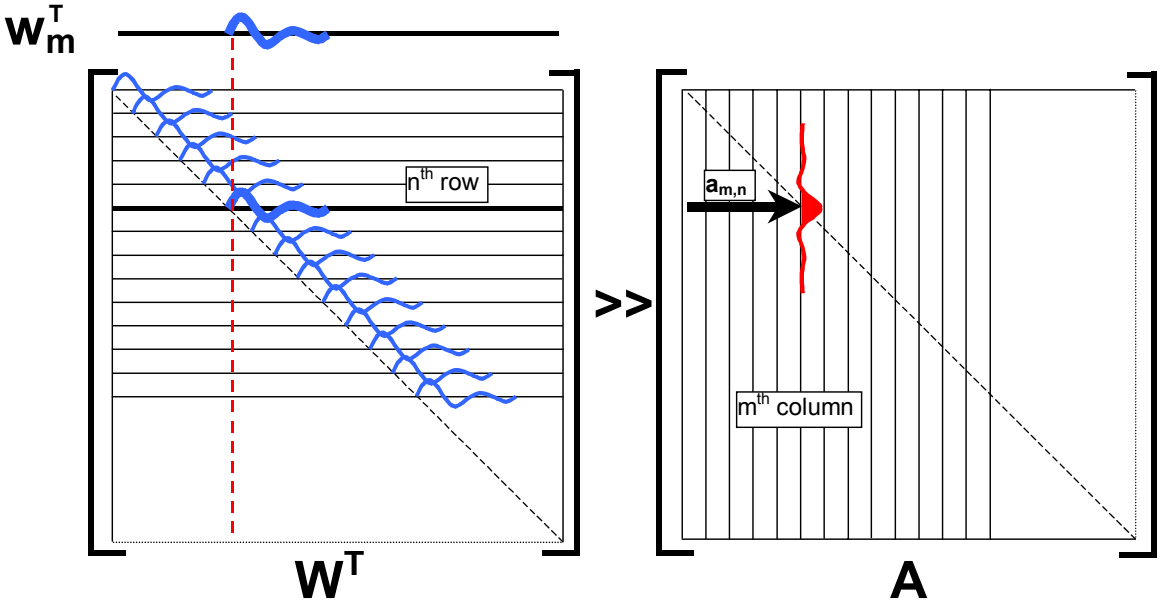


FIG. 12: One column in A is found by taking the dot product of one column in WT with all rows in W.

The square matrix, **A**, is shown in Figure 13 with vertical traces that illustrate that all the new wavelets are the delayed auto-correlation of the original wavelet. Plotting the data with horizontal traces should produce the same image (with adequate column sampling).

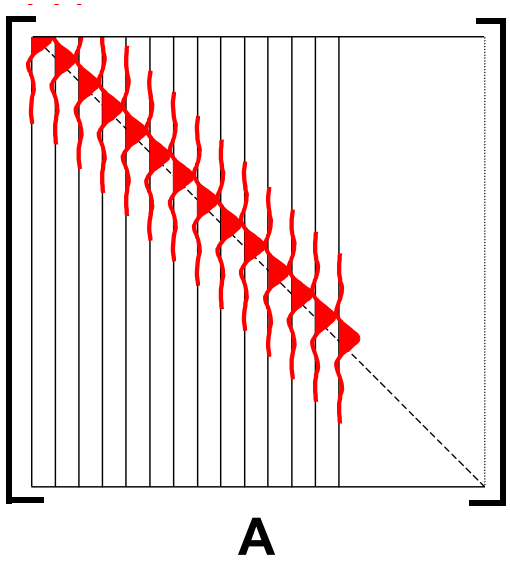


FIG. 13: The result of **A** = WTW, the product of a matrix and its transpose.

All elements on any diagonal in **A** are equal (a Toeplitz matrix) with the largest value on the main diagonal. This is very significant as **A** is similar to the identity matrix, **I**, in which the data is zero, except for unit values on the diagonal, as illustrated in Figure 14.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

FIG. 14: An identity matrix.

A feature of the identity matrix is that the inverse is also an identity matrix, i.e. $\mathbf{I^{-1} = I}$. If we assume that the auto-correlation matrix, **A**, is an approximation to the identity matrix, **I**, (except for a scale factor), we can now assume

$$\underset{[M\times N][N\times M]}{\mathbf{W}^T\ \mathbf{W}} = \underset{[M\times M]}{\mathbf{A}} \approx \underset{[M\times M]}{\mathbf{I}}, \tag{16}$$

and also that

$$\left(\underset{[M\times N][N\times M]}{\mathbf{W}^T\ \mathbf{W}}\right)^{-1} \approx \underset{[M\times M]}{\mathbf{I}}. \tag{17}$$

This simplification allows equation (14) to be reduced to a very simple form,

$$\underset{[M]}{\mathbf{r}} \approx \underset{[M\times N][N]}{\mathbf{W}^T\ \mathbf{s}}, \tag{18}$$

that allows us to define an estimate of the reflectivity, $\hat{\mathbf{r}}$, by

$$\underset{[M]}{\hat{\mathbf{r}}} = \underset{[M\times N][N]}{\mathbf{W}^T\ \mathbf{s}}. \tag{19}$$

The reflectivity can be estimated from the product of the seismic vector, **s,** with the transpose of the wavelet matrix, $\mathbf{W^T}$. A comparison of equations (12) and (19) show that our simplifications have assumed that the inverse to a matrix, $\mathbf{W^{-1}}$, can be approximated by a very simple transpose, $\mathbf{W^T}$, i.e.,

$$\underset{[M\times N]}{\mathbf{W}^T} \approx \underset{[M\times N?]}{\mathbf{W}^{-1}}. \tag{20}$$

When $\mathbf{W^T = W^{-1}}$, **W** is an orthogonal matrix with an interesting property that all vectors, composed of either columns or rows, are orthogonal, or in a three-dimensional system, the three vectors are mutually perpendicular.

By claiming the **A** matrix is equivalent to the identity matrix, **I**, we assumed the wavelet has a broad bandwidth, i.e. that the wavelet is very narrow. However, when the wavelet is broad, and the spectrum of the auto-correlation wavelet has no zeros, then the inverse of **A** will produce a spectral weighting that will shape the reflectivity to the desired spike. If we don't "divide" by this spectral weighting defined in **A**, then equation (19) will produce a reflectivity estimate that is band-limited. In summary, it is the $(\mathbf{W}^T\mathbf{W})^{-1}$ term in equation (14) that recovers the frequency content of the reflectivity. Equation (19) will only produce a band-limited form of the reflectivity.

In conventional deconvolution problems, **A** is modified to limit the smallest values of the frequency amplitudes to a preset level to eliminate zeros, which then allows **A** to be inverted. This is referred to as pre-whitening. Note however, that frequency components, beyond the bandwidth of the wavelet, will be amplified as noise and will require filtering back to the original bandwidth.

The above process has assumed that the wavelet is known. We could continue to investigate deconvolution that also estimates the wavelet, but I will leave that for the next article. We will instead proceed with a few more observations and then compare the above process with migration.

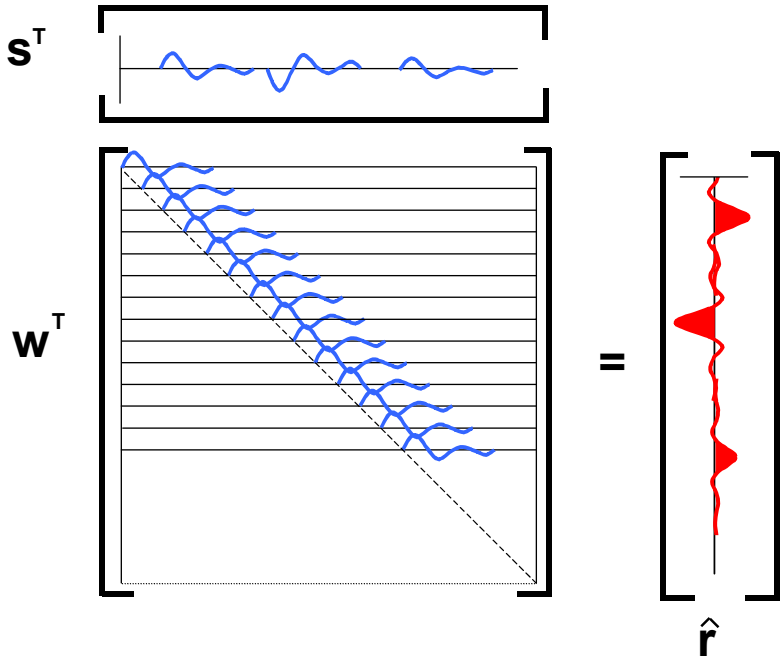


FIG. 15: Forming a reflectivity trace by cross-correlating the seismic trace with a known wavelet.

We now examine the estimated reflectivity, $\hat{\mathbf{r}}$, defined now by equation (19) and as illustrated in Figure 15. Elements in the reflectivity vector are produced by the dot product of the transpose of the seismic vector, **s**, with each row of the transposed wavelet matrix, $\mathbf{W}^T$. This cross-correlation process forms wavelets in $\hat{\mathbf{r}}$ that have a shape of the auto-correlation of the input wavelet, with the peak of the wavelet representing the location and amplitude of the reflectivity, as illustrated by the red trace in Figure 15.

We have therefore produced a band-limited form of the reflectivity. This result is identical to that obtained with a matched-filter, i.e. the cross-correlation of the seismic trace with a matching wavelet.

Using the transpose to approximate an inversion is the same as using a matched filter.

We can also therefore identify the differences between an exact inversion and an approximation obtained by a matched filter, i.e. the assumption that $\mathbf{W}^T\mathbf{W} \frown \mathbf{I}$.

KIRCHHOFF MIGRATION AS A TRANSPOSE PROCESS

There are many "reverse" processes that use the concept of approximating the inverse with a transpose. What about seismic modelling and migration? In seismic modelling we place a diffraction at every scatterpoint. When using Kirchhoff migration, we define a kinematic diffraction shape for a scatterpoint, then weight and sum the input energy defined by this shape, and insert the energy at the scatterpoint location. This is, in essence, a two-dimensional cross-correlation, and will produce a peak of energy when the model diffraction matches a diffraction in the input data, i.e. matched filtering, or by approximating the inverse process with a transpose process. We are now in a position to visualize and evaluate the limitations of our migration algorithms with true inversion.

Visualizing modelling data with diffractions

We will start by modelling a gather of reflectivity traces with the same wavelet matrix, **W**, to get a gather of traces with wavelets as illustrated in Figure 16. One reflectivity column vector, where circles now represent the amplitude of the reflectivity, produces a corresponding seismic trace vector. But, we want to be able to model a diffraction that has an additional dimension.

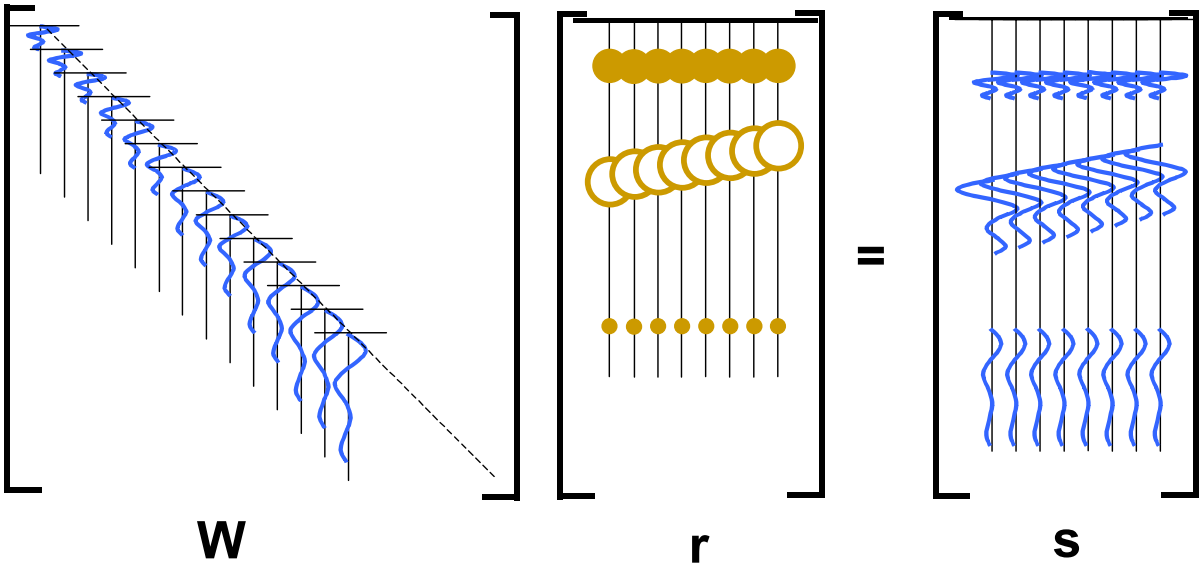
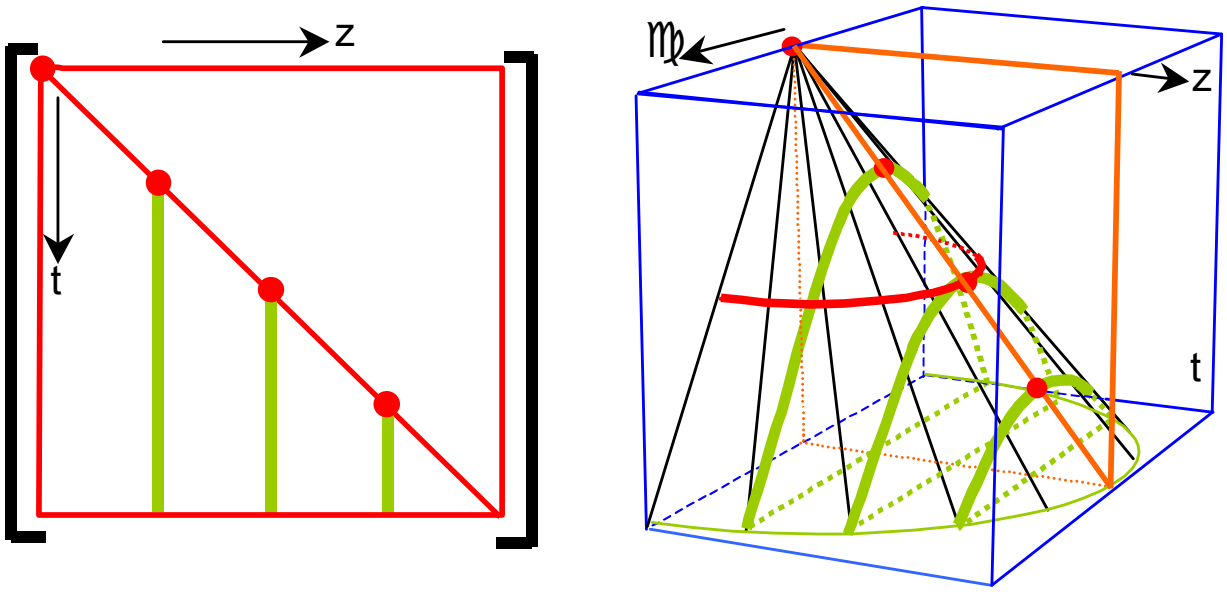


FIG. 16: Matrix form of the convolutional model with time varying wavelets.

I have defined a 3D diffraction model to accomplish this task that starts with the wavelet matrix and adds another dimension of distance out of the page as illustrated in Figure 17. The diffractions start at the surface to the left, and increase in depth to the right to form a cone. The vertical intersections of the cone form the hyperbolas.

Matrix multiplication still works because we consider one trace at a time. Rather than define the vectors and matrices with numerical values, I will indicate the dimensions on the appropriate axis.



a) b)

FIG. 17: The diffraction matrix, **D**, showing a) the (z, t) view and b) a perspective view showing the added dimension of ⇳.

Figure 17 shows two views of the diffraction matrix **D**, with dimensions (z, ⇳, t), for a constant-velocity medium. A view normal to (z, t) is shown in part (a) and is similar to the wavelet matrix **M**. The added dimension of ⇳ is visible in the perspective view in part (b). This added dimension ⇳ represents the migration aperture, which contains the diffractions, identified by the vertical lines and curves. The origin of the diffraction matrix, **D**, is at the peak of the cone where ⇳, z, and t are zero. In a constant-velocity medium, the shape of the diffractions are defined by hyperbolas, formed when a vertical planes intersect the cone.

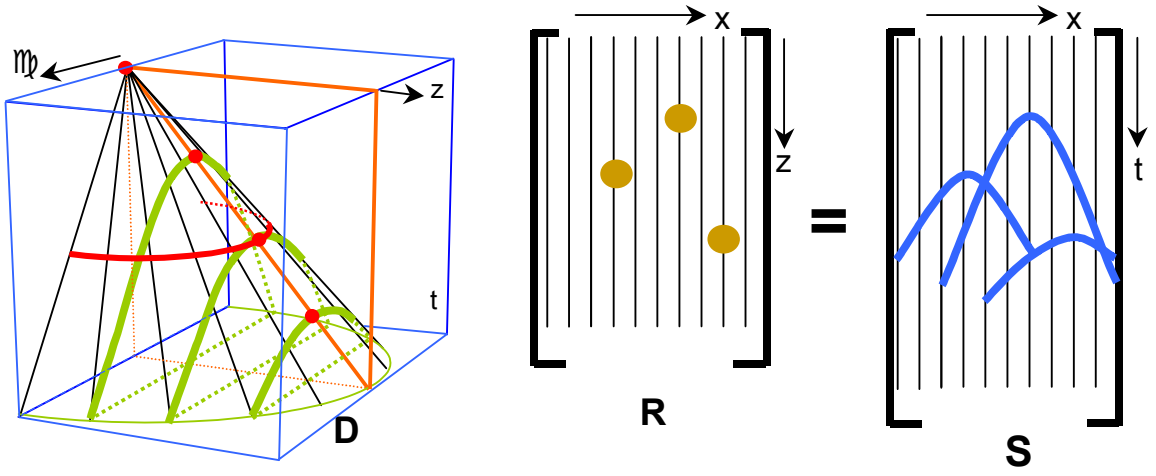


FIG. 18: Matrix view to obtain a zero-offset section, **S**, from a reflectivity matrix, **R**, and a diffraction matrix, **D**.

Seismic modelling is illustrated in Figure 18 where one sample in **S** at $s(x_p, t_q)$ is formed when the 2D matrix, **R**, is dot multiplied by one plane of the 3D diffraction matrix, **D**, at time, t. The matrix, **R**, is aligned such that the location of the migrated trace is above the location of ⇳ = *0*, as illustrated in Figure 19, which shows the **R** matrix above the diffraction matrix, **D**. As different migration traces are selected, then the matrix, **R**, above **D**, is shifted in the corresponding ⇳ direction. For example, when the red trace in **S** is being evaluated, the corresponding red trace in $\mathbf{R}^T$ is located above ⇳ = *0*.

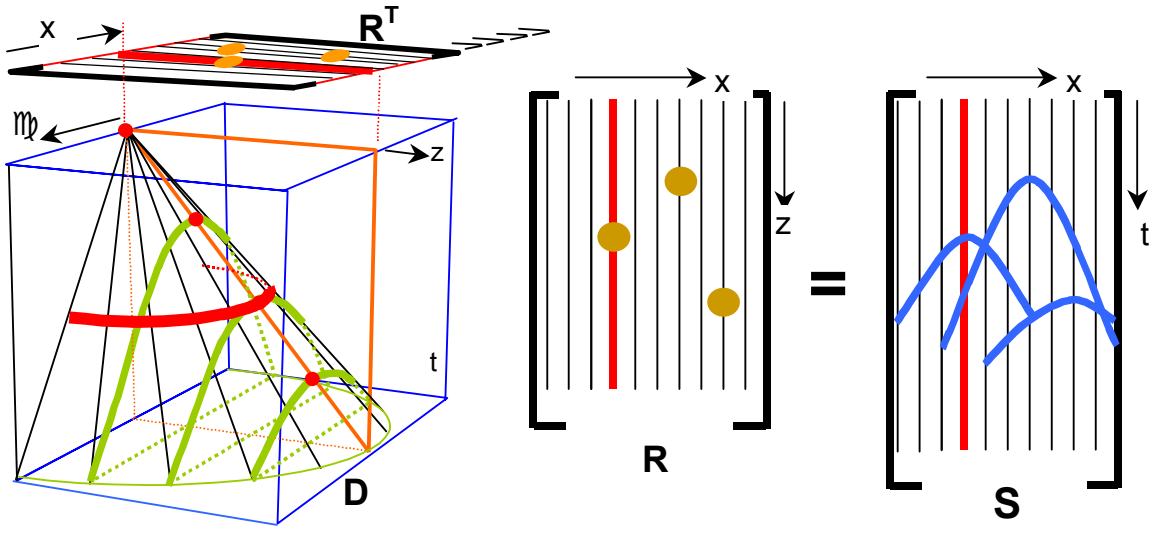


FIG. 19: Matrix view to obtain a zero-offset section, **S**, from a reflectivity matrix, **R**, and a diffraction matrix, **D**.

The time of the migrated sample t_q, is selected by the horizontal plane within **D**, and the trace location x_p , by the relative alignment of that trace above $\Updownarrow = 0$ as given by

$$s(x_p,t_q)=\sum_{\chi_i=-I}^{I}\sum_{z_j=1}^{J} d(\chi_i,z_j,t_q)\,r\left(x_p-\chi_i,z_j\right) \quad . \tag{21}$$

The **D** matrix is mainly composed of null space, with scaled values only on the surface that is defined at the cone. A constant time (slice) in **D** intersects the diffraction cone at a semi-circle, as identified by one red curve in the perspective views. This corresponds to the method of producing a seismic modelled section by summing reflector energy over a semi-circle, i.e. when creating a seismic section from modelling, we can either sum scatterpoints over a semi-circle to get one modelled sample, or spread energy from a scatterpoint along a hyperbolic diffraction.

Let's return to matrix theory for a moment, and see what happens when we take the transpose of **D** to get a migrated section from a zero-offset seismic section. The result is Figure 20 with $\mathbf{D}^T$ containing a horizontal cone. Now, the migrated section is dot-multiplied with constant-depth planes (slices) and the energy is summed where the horizontal plane intersects the cone, illustrated at one depth level by a red hyperbola. The equation for migration one sample $r(x_p, z_q)$ is

$$\hat{r}(x_p,z_q)=\sum_{\chi_i=-I}^{I}\sum_{t_j=1}^{J} d(\chi_i,z_q,t_j)\,s\left(x_p-\chi_i,t_j\right) \quad . \tag{22}$$

The dual property to hyperbolic summation in migration is that we can spread the energy along a semi-circle, defined in Figure 20 by vertical planes in $\mathbf{D}^T$ at constant time.

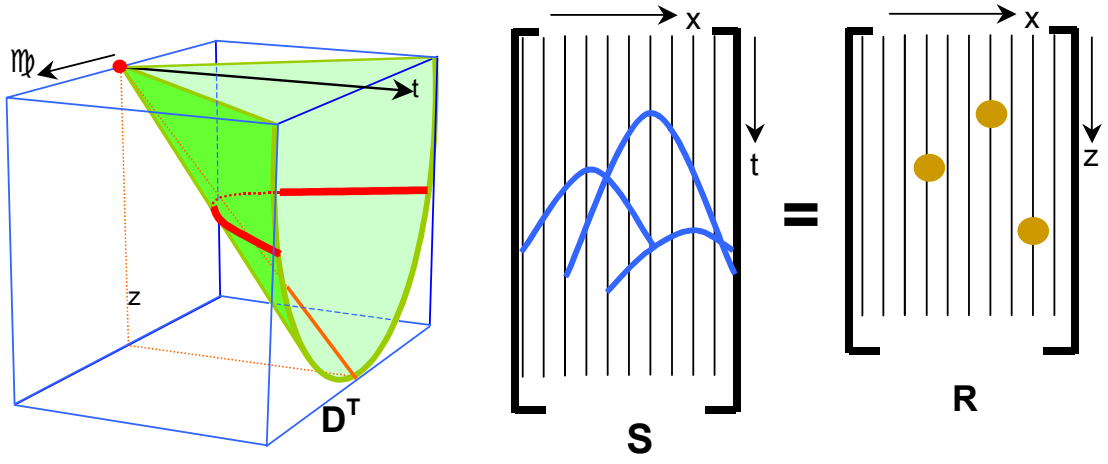


FIG. 20: Matrix view to obtain a reflectivity matrix, **R**, from a zero-offset section, **S**, and a diffraction matrix, **D**.

The diffraction matrix illustrated in Figures 17 through 19 shows the constant velocity diffractions on a cone. These diffractions can be modified to be time-varying by using RMS velocities with shapes illustrated in Figure 21a.

The diffraction shapes can also be modified to contain constant-offset diffractions as illustrated in Figure 21b. Modelling will then produce a constant-offset section. The inversion of an offset section would therefore use the transpose of the offset diffractions.

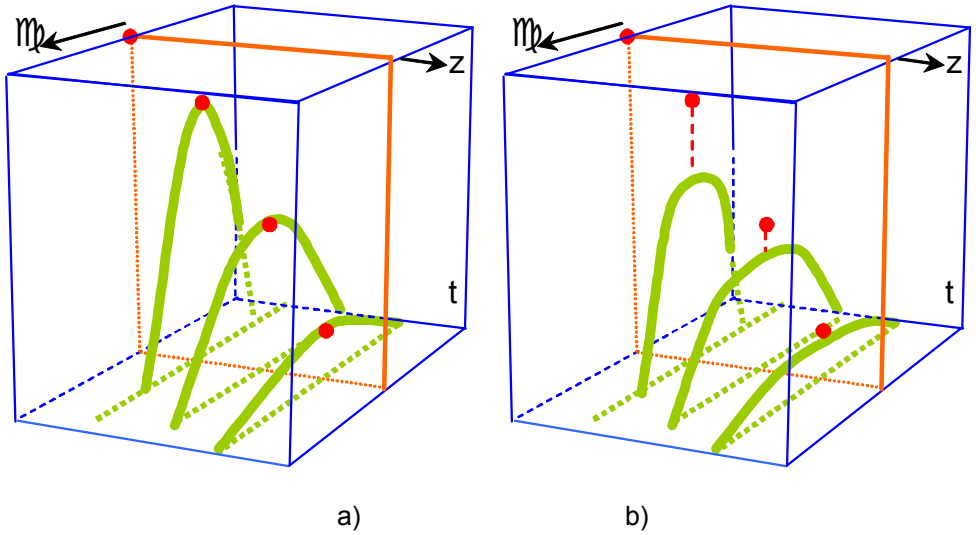


FIG. 21: a) Time-varying diffractions and b) and offset diffractions.

COMMENTS AND CONCLUSIONS

The kinematics of diffraction stack migrations are straightforward, however the amplitude weightings used when summing the diffractions are still under investigation. Migrations that are derived from solution to the wave equation include weighting schemes. The integral solution of the wave equation changed the heuristic diffraction stack process to the mathematically deterministic Kirchhoff algorithm.

Inversion techniques provide an alternative solution when imaging the subsurface and, in some cases, produce algorithms that are very similar to Kirchhoff migrations. Approximate linear algebra inversions were shown to be identical to matched filters, and the difference from an ideal inversion identified.

I have been able to represent most of the mathematics to this point with visual images to increase heuristic understanding. However, I have only touched on the principles of inversion by investigating the 1D case and extending it to 2D seismic modelling and migration.

REFERENCES

Bleistein, N., Cohen, J.K., and Stockwell, Jr., J.W., 2001, Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion, Springer

Cary, P.W. and Li, X., 2001, Some basic imaging problems with regularly-sampled seismic data, Exp. Abs., SEG Int. Nat. Exp., 981-984

Claerbout, J.F., 1992, "Earth Soundings Analysis: Processing versus inversion", Blackwell Scientific Publications

Dellinger, J., Gray, S., Murphy, G., Etgen, J., and Fei, T., 1999, Efficient two and one-half dimension true-amplitude migration, Exp. Abs., SEG Nat. Conv., 69, 1540-1543

Gazdag, J. and Squazzero, P., 1984, Migration of seismic data, Proceedings of the IEEE, 72, 1302-1315

Geiger, H.D., 2001, Relative-amplitude preserving prestack time migration by the equivalent offset method, Ph.D. Thesis, Department of Geology and Geophysics, University of Calgary

Larner, K. and Hatton, L., 1990, Wave-equation migration: two approaches (first distributed privately in 1976), First Break, 8, 443-448

Lindseth, R.O., 1979, Synthetic sonic logs - a process for stratigraphic interpretation, Geophysics, 44, 3-26

Lines, L.R. and Treitel, S., 1984, Tutorial: a review of least-squares inversion and its application to geophysical problems, Geophysical Prospecting, 32, 159-186

Lines, L.R., Schultz, A.K., and Treitel, S., 1988, Cooperative inversion of geophysical data: Geophysics, 53, 8-20.

Lines, L.R. and Levin, F.K., 1988, Inversion of Geophysical Data, Geophysical reprint series, No. 9, SEG

Margrave, G.F., 1998, Theory of non-stationary linear filtering in the Fourier domain with applications to time-variant filtering, Geophysics, 63, 255-259

Schneider, W.A., 1978, Integral formulation for migration in two and three dimensions, Geophysics, 43, 49-76

Ziemer, R.E. and Tranter, W.H., 1995, Principles of Communications, Wiley