Signal/noise separation and velocity estimation

William S. Harlan

ABSTRACT

To interpret noisy data one must recognize the lateral coherence of geologic events, their statistical predictability. We define "focusing" as increasing the statistical independence of samples. By the central limit theorem, focusing signal with some invertible, linear transformation L makes signal more non-gaussian; the same L must defocus noise and make it more gaussian. A measure F, defined from cross entropy, measures non-gaussianity from local histograms of an array and thereby measures focusing. Local histograms of the transformed data and of transformed, artificially incoherent data provide enough information to estimate the amplitudes distributions of transformed signal and noise--errors only increase the estimate of noise. With these distributions one can recognize and extract samples containing the highest percentage of signal. Iterative estimations of signal and noise improve the extractions of each.

If we remove bed reflections and noise, F will determine the best migration velocity for the remaining diffractions. Slant stacks map lines to points, greatly concentrating continuous reflections. We invert the strongest extracted events and subtract them from the data, leaving diffractions and noise. Next, we migrate with many velocities, extract focused events, and invert. We find the least-squares sum of these events best resembling the diffractions in the original data. Migration of these diffractions will maximize F at the best velocity. We successfully extract diffractions and estimate velocities for a window of data containing a growth fault. A spatially variable least-squares superposition allows variable velocities.

Localized slant stacks allow a laterally adaptable extraction of locally linear events. For a stacked section we successfully extract weak signal with highly variable coherency from behind strong gaussian noise. An more general local transformation expresses events as a sum of short second-order curves.

An algorithm similar to that for diffractions will allow wave-equation velocity analyses of a shot or midpoint gather. Unlike NMO, migration will image the skewed hyperbolas of dipping reflectors. This algorithm treats offset truncations as another removable form of noise and escapes artifacts usually plaguing the wave-equation method.

One may remove non-gaussian noise from shot gathers by first removing the strongest signal, then estimating amplitude distributions for noise and remaining signal. Without the strongest signal present, some samples may be recognized as containing a small percentage of signal. Extracting these events and then subtracting from the original data removes the strongest noise. This procedure successfully removes ground-roll and other noise from a common-shot gather.

INTRODUCTION

To interpret noisy data one must recognize the lateral coherence of geologic events, their statistical predictability. For example, reflections of continuous beds have roughly hyperbolic shapes in shot and midpoint gathers and show lateral continuity in offset or stacked sections; faults and other abrupt interruptions appear as diffraction hyperbolas. Statistical predictability overspecifies an event. Flat reflections could be specified by slopes and vertical intercepts rather than depths every 50m; hyperbolas could be specified by scattering locations and average rock velocities. We define noise then as components of the data whose description cannot be reduced, components showing no spatial coherence or predictability. We shall emphasize in this paper the extraction of events containing velocity information because they illustrate well two possible complications.

1. Before extracting the desired signal we must remove both coherent events containing no velocity information and noise.
2. We estimate unknown velocities from extracted signal; we extract signal by recognizing the coherence dependent on these velocities. The extraction of signal must not bias the estimate of the velocities.

FOCUSING AND EXTRACTION OF GEOLOGIC SIGNAL

We shall establish the following results before describing specific algorithms. Using these results, we may identify some data component as signal (with predetermined reliability) and identify the transform focusing this signal best. We define "focusing" as a linear transformation making the samples of an array statistically independent.

1. A transformation focusing signal must also defocus noise.
2. Focused signal becomes more non-gaussian; defocused noise becomes more gaussian.
3. A function exists to measure the local non-gaussianity, and thereby the focusing, of data.
4. One may estimate amplitude distributions of the transformed signal and noise.
5. With these distributions, one may recognize and extract those samples of the transformed data containing most of the focused signal.
6. To identify the best transform, one should extract signal with several transforms and construct a superposition best resembling the data. The focusing measure will indicate which transform focuses this superposition best.

A measure of focusing

Most often one does not know the best focusing transformation (that making the array samples most statistically independent) and prefers to add some unknown parameters. One then needs a measure to detect the quality of focusing.

Local histograms will provide us sufficient information for estimating the focusing of the data. Let $\bar{x}$ be an array of random variables, and let $p_{\bar{x}}$ be the corresponding joint probability density function (p.d.f.). $\int p_{\bar{x}} d\bar{x}$ gives the probability of a particular range of $\bar{x}$ values. One cannot estimate from a single array a joint p.d.f. allowing arbitrary dependencies between samples. A marginal density function (m.d.f.) is defined for each sample by integrating over the other dependent samples.

$$p_{x_i}(x_i) = (\prod_{j \neq i} \int dx_j) p_{\bar{x}}(\bar{x}) \ .$$

If statistics vary smoothly over array samples (local stationarity), then one may easily estimate the m.d.f.'s from local histograms of the data. A natural uncertainty principle exists, that accuracy in estimating the m.d.f.'s must be balanced against spatial resolution.

The central-limit theorem requires that a linear sum of independent random variables make the corresponding m.d.f.'s more gaussian. By the contra-positive of this theorem, focusing an array with a linear transformation must make the m.d.f.'s more non-gaussian. Thus, a measure of the local non-gaussianity of an array in turn measures the increase in focusing.

Cross entropy (defined first by Kullback as directed divergence) measures the unpredictability of a given $p_1(x)$ with respect to some reference $p_2(x)$.

$$I[p_1(x) : p_2(x)] \equiv \int p_1(x) \log[p_1(x)/p_2(x)]dx \tag{1}$$

If p_2 is uniform, then cross entropy becomes the negative of Shannon's statistical entropy. I approaches the minimum value of 0 when p_1 is most like p_2. We define a measure of non-gaussianity, F, as the minimum cross entropy of the data m.d.f.'s with respect to gaussian distributions of all variances σ^2. F increases with non-gaussianity and thereby with the focusing of the data. (We shall assume zero-mean processes for simplicity in notation.) For a single m.d.f. we define

$$F[p(x)] \equiv \min_{\sigma} I[p(x) : gaussian(\sigma,x)] \tag{2}$$

$$= \min_{\sigma} \int p(x) \log[p(x)/(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}})]dx$$

$$= \min_{\sigma} \left\{ \int p(x) \log p(x) dx + \frac{1}{2\sigma^2} \int x^2 p(x) dx + \log\sigma + \log\sqrt{2\pi} \right\}$$

Not surprisingly, F attains this minimum value when the gaussian distribution possesses the same standard deviation as $p(x)$:

$$\frac{d}{d\sigma} I[p(x) : gaussian(\sigma,x)] = -\frac{1}{\sigma^3} \int x^2 p(x) dx + \frac{1}{\sigma} = 0 \tag{3}$$

$$\rightarrow \sigma^2 = \int x^2 p(x) dx$$

$$\rightarrow F[p(x)] \equiv \int p(x) \log p(x) dx + \frac{1}{2} \log \int x^2 p(x) dx + C \tag{4}$$

$$where\ C = \log\sqrt{2\pi} + \frac{1}{2}$$

(4) provides a simpler, working definition of F. Notice that (4) is scale invariant: we may multiply the random variable x by a constant a but F remains constant.

$$F[\frac{1}{a} p(\frac{x}{a})] = F[p(x)] \tag{5}$$

Finally, we may demonstrate that a gaussian distribution minimizes (4). First we replace $p(x)$ in (4) by a perturbed $(1-\varepsilon)p(x) + \varepsilon\eta(x)$. Then, setting the ε derivative equal to zero yields

$$-\int p(x) \log p(x) dx + \int \eta(x) \log p(x) dx + \frac{1}{2\sigma^2} \int x^2 \eta(x) dx - \frac{1}{2} = 0 .$$

(3) again defines the variance. Let the arbitrary $\eta(x) = \delta(x - x_0)$.

$$\rightarrow -\int p(x)\log p(x)dx + \log p(x_0)dx + \frac{x_0^2}{2\sigma^2} - \frac{1}{2} = 0$$

$$\rightarrow \quad p(x_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x_0}{2\sigma 2}} \quad and\ F[p(x)] = 0 \tag{6}$$

(6) gives the equation of a gaussian. The constraint of unit area requires that the measure attain a minimum value of 0.

In practice one must evaluate (4) for discrete histograms functioning as m.d.f.'s. We may represent the sampled versions as $\{p_i\}$, defining each sample as an average over a short interval of $p(x)$. Assume that the $\{p_i\}$ are sampled N times per standard deviation.

$$p_i = \frac{1}{(\sigma/N)}\int \Pi\left(\frac{x - i\sigma/N}{\sigma/N}\right) p(x)dx \quad where\ \Pi(x) \equiv \begin{cases} 1 & -1/2 \le x \le +1/2 \\ 0 & else \end{cases} \tag{7}$$

The sampling reduces $F[p(x)]$ by some ε made arbitrarily small by large N.

$$F[p(x)] = \int p(x)\log p(x)dx + \log\sigma + C \tag{8}$$

$$= \sum_i \frac{\sigma}{N} p_i \log p_i + \log\sigma + C + \varepsilon$$

$$= \sum_i \left(\frac{\sigma}{N} p_i\right) \log\left(\frac{\sigma}{N} p_i\right) + \log N + C + \varepsilon$$

Sampling $p(x)$ a fixed N times per standard deviation yields a simplified form equal to the negative of Shannon's statistical entropy (plus constants).

Extracting geologic signal

A linear transformation focusing signal also defocuses noise. Noise, by our definition, has no coherence spatially, though possibly over time. Noise includes all events that the chosen transformation will not focus. Note that coherent noise such as multiples, side-swipe, ground-roll, and cable noise may require wave and geologic models similar to those for signal. Because focused signal becomes more non-gaussian, energy concentrates about narrow peaks, increasing the number of low values. A transformation L must add coherence to already incoherent noise. Because of the central limit theorem, the noise m.d.f.'s become more gaussian. Noise energy diffuses. Thus, an invertible transformation of noise will always decrease F. If one first removes noise, then the best L maximizes F.

We should extract the strongest transformed signal from samples with low percentages of noise. Let $\bar{d} = L^{-1}\bar{s} + \bar{n}$, where $\bar{d}$ contains data, $\bar{s}$ focused signal, and $\bar{n}$ laterally

incoherent events. L^{-1} is the right inverse of the transform L focusing signal. Write the linear transformation, $\bar{d}' = L\bar{d}$ as

$$d_i' = \sum_j a_{ij} d_j \ . \tag{9}$$

Primes will specify a component transformed out of its defined domain. Because we can do nothing with the coherence of diffused noise, and because we assume samples of transformed signal to be independent, we estimate the signal in each sample independently. If such independence were perfect, an ideal extraction would first perform an adaptive prediction-error filter to remove much noise. One cannot expect perfect focusing, however, so such a filter would also remove some signal. (A later modification will allow for remaining coherence in signal.)

With estimated m.d.f.'s of transformed data and noise, one may estimate the signal present in each sample. Consider now three components of the same transformed sample: $d' = s + n'$. Their distributions are related by

$$p_{d'}(x) = p_s(x) * p_{n'}(x) \tag{10}$$

The asterisk denotes convolution. With knowledge of two one could calculate the third distribution by a multiplication or division in the frequency domain. From these three distributions we estimate the expected value of s given that of d'.

$$E(s \mid d') = \int x p_{s \mid d'}(x \mid d') dx \ = \ \frac{\int x p_{s,d'}(x,d') dx}{p_{d'}(d')} \tag{11}$$

$$= \frac{\int x p_s(x) p_{d' \mid s}(d' \mid x) dx}{p_{d'}(d')} = \frac{\int x p_s(x) p_{n'}(d'-x) dx}{p_{d'}(d')}$$

We use Bayes' rule twice. Local histograms of the transformed data estimate $p_{d_i'}$. We shall now find a $\hat{p}_{n_i'}(x)$ with a variance guaranteed larger that of $p_{n_i'}(x)$. This estimated noise distribution, perfect in the absence of signal, results in a pessimistic estimation of signal. After extracting the strongest signal one may re-estimate this noise distribution with greater accuracy.

The m.d.f.'s of transformed, artificially incoherent data will overestimate the amount of noise present in the transformed data. After transformation (9), the m.d.f.'s for transformed noise become multiple convolutions of the originals--convolutions bringing the distributions nearer gaussianity.

$$p_{n_i'}(x) = \prod_j * \left[a_{ij} p_{n_j}\left(\frac{x}{a_{ij}}\right) \right] \tag{12}$$

Our notation specifies convolutions of all $p_{n_j}(x)$ with each other. If we could ignore or destroy coherence in the original data without harming marginal distributions, then the m.d.f.'s of transformed noise would not change. But signal would defocus, with resulting m.d.f.'s as double convolutions of the originals.

$$\tilde{p}_{s_i}(x) = \prod_{j,k} * \left[a_{ik}^{-1} a_{kj} p_{s_j}\left(\frac{x}{a_{ik}^{-1} a_{kj}} \right) \right] \tag{13}$$

The a_{ik}^{-1} correspond to the inverse transformation L^{-1}. Such adulterated signal approaches gaussianity very well. Transformation of artificially incoherent data produces the m.d.f.

$$\hat{p}_{n_i'}(x) \equiv \tilde{p}_{s_i}(x) * p_{n_i}(x) \tag{14}$$

The variances of the two right distributions add, guaranteed larger than the noise distribution. In the absence of signal, the signal m.d.f. becomes a delta function and $\hat{p}_{n_i'}(x) = p_{n_i'}(x)$. Randomly reversing the polarity of traces would artificially destroy coherence. We could then transform and calculate the estimates $\hat{p}_{n_i'}(x)$ from local histograms. Then sufficient information exists to evaluate a lower bound for $E(s \mid d')$ in (11).

The estimate (11) should be modified somewhat in practice. If intending later to extract noise or make the process iterative, then one should zero samples containing significant percentages of noise avoid the conversion of any noise to coherent events. Analytic envelopes should be used in extractions (though not when preparing histograms). Otherwise an extraction corrupts waveforms by deepening troughs between peaks. As another precaution one should smooth an array of the $E(s \mid d')/d'$ values both spatially and temporally before multiplying $\bar{d}'$, thus avoiding sharp edges on extracted events.

The highest amplitudes of the transformed data most commonly contain the highest concentration of signal. If transformation makes signal more non-gaussian than noise then $E(s \mid d')$ has a characteristic shape: for amplitudes of d above some abrupt cutoff, $s \approx d$. Below this cutoff, noise begins to contribute quickly and significantly (cf. Godfrey). We discourage, however, any *a priori* assumptions about the shape of this function. Since the data will provide the necessary statistics, we should use them.

Finding the best extraction

Let us next treat a more general transform L, perhaps of higher dimensions, like migration. We extract signal over a range of a transform's parameter, linearly combine inverted events, and find the parameter maximizing the focusing measure. In this way we both detect all signal and determine the best transform.

Let our transform L_v depend on v. We extract all signal focused at some v and then invert:

$$\bar{e}_v = L_v^{-1}\{Extract\{L_v\{\bar{d}\}\}\} \tag{15a}$$

where L_v^{-1}, the least-squares inverse of L_v, inverts all signal with the parameter v. Let $L_v^{-1} = (L^*L)^{-1}L^*$, where an asterisk denotes the adjoint. Having $\bar{e}_v$ for a physical range of v, we find the least-squares sum best resembling the data:

$$\hat{s} = \sum_v a_v \bar{e}_v \ ; \quad \min_{\bar{a}}(\hat{s} - \bar{d})^2 \ \rightarrow \ \sum_w <\bar{e}_v,\bar{e}_w> a_w = <\bar{e}_v,\bar{d}>. \tag{16a}$$

The brackets indicate a simple dot product. We solve (16a) for all coefficients a_v and for $\hat{s}$ by inverting a symmetric matrix with an order equal to the number of velocities used. $\hat{s}$ contains the strongest signal, without bias, for the chosen range of v. Transforming $\hat{s}$ over this same range will maximize F at the v focusing the signal best.

In this development we have assumed all signal in the data window corresponds to the same parameter v, when in fact this dependence may be spatially variable. We shall add such a dependence to equations (15a) and (16a) as the applications become clearer.

EXTRACTING DIFFRACTIONS FOR VELOCITY INFORMATION

Many velocity analyses extrapolate wavefields back in time, thereby concentrating signal and dispersing noise. For example, an NMO stack finds the image source of plane reflections. Inverse scattering extrapolates wavefields back to residual-velocity scatterers and divides out the effect of extrapolated sources. Genuine seismic events always begin simply and produce increasingly diffuse wavefronts--the increase of thermodynamic entropy requires it. Because the wave equation is symmetric in time, seemingly isolated arrivals, such as noise and missing data, require more diffuse wavefronts at time zero. That extrapolation concentrating the signal best, while dispersing noise, determines the best velocities.

Offset sections span greater distances on the surface and so detect deep velocity changes better than do shot and midpoint gathers. Diffraction events, such as reflections of bed truncations and point scatterers, contain all of the velocity information present in offset sections. Diffractions appear in the background of sections containing faults, such as Figure 1a, a section stacked over offset. Migration with correct velocities focuses these diffractions; however, reflections of continuous beds do not focus, and noise defocuses. We must first remove bed reflections (Figure 1b) from the data, exposing diffractions and noise. We must next extract the strongest diffractions from the noise as in Figure 1c. Note

that phase changes appear at peaks of the hyperbolas, as predicted by theory. When these events are migrated, the focusing measure F is maximized at the best velocity (Figure 2). The remaining component of the data (Figure 1d), noise, is neither focused by slant stacks or migration. Continuous beds, diffractions, and noise (Figures 1b,1c, and 1d) add together to produce the original data (Figure 1a). (Diffractions and noise are amplified for plotting.)

Separating bed reflections from diffractions and noise

A window of stacked, offshore Texas data (provided by Western Geophysical) contains a growth fault with weak diffractions off truncated beds (Figure 1a). One may remove reflections of continuous beds and retain diffractions by focusing with slant stacks. We define slant stacks by the inverse transform $d(x,t) = \int d'(p,\tau=t-px)$. (See the appendix on slant stacks.) Slant stacks map lines of constant dip into points, thereby focusing bed reflections laterally. Slant stacks may be said to remove the first-order predictability of gently curving events. Coherence remaining after a slant stack indicates curvature in the original events. Events with rapidly changing dips, such as noise and diffractions, do not focus. Points map approximately to lines, thereby defocusing noise and diffusing its energy.

We particularize the results of previous sections to the estimation and removal of bed reflections.

1. Slant stack artificially incoherent data (with randomly reversed traces) and estimate $\hat{p}_{n'}(x)$ locally from histograms.
2. Slant stack data and estimate $p_{d'}(x)$ locally from histograms.
3. Estimate $p_s(x)$ from (10).
4. Evaluate $E(s \mid d')$ and evaluate for each sample of the transformed data. Smoothly zero samples containing significant noise.
5. Invert the strongest signal and subtract from the original data.
6. Use this section to re-estimate $\hat{p}_{n'}(x)$ and repeat steps 3-5.

In practice, a full slant stack is not necessary for step 1. Summing should be performed at enough p values to indicate the local statistics of noise.

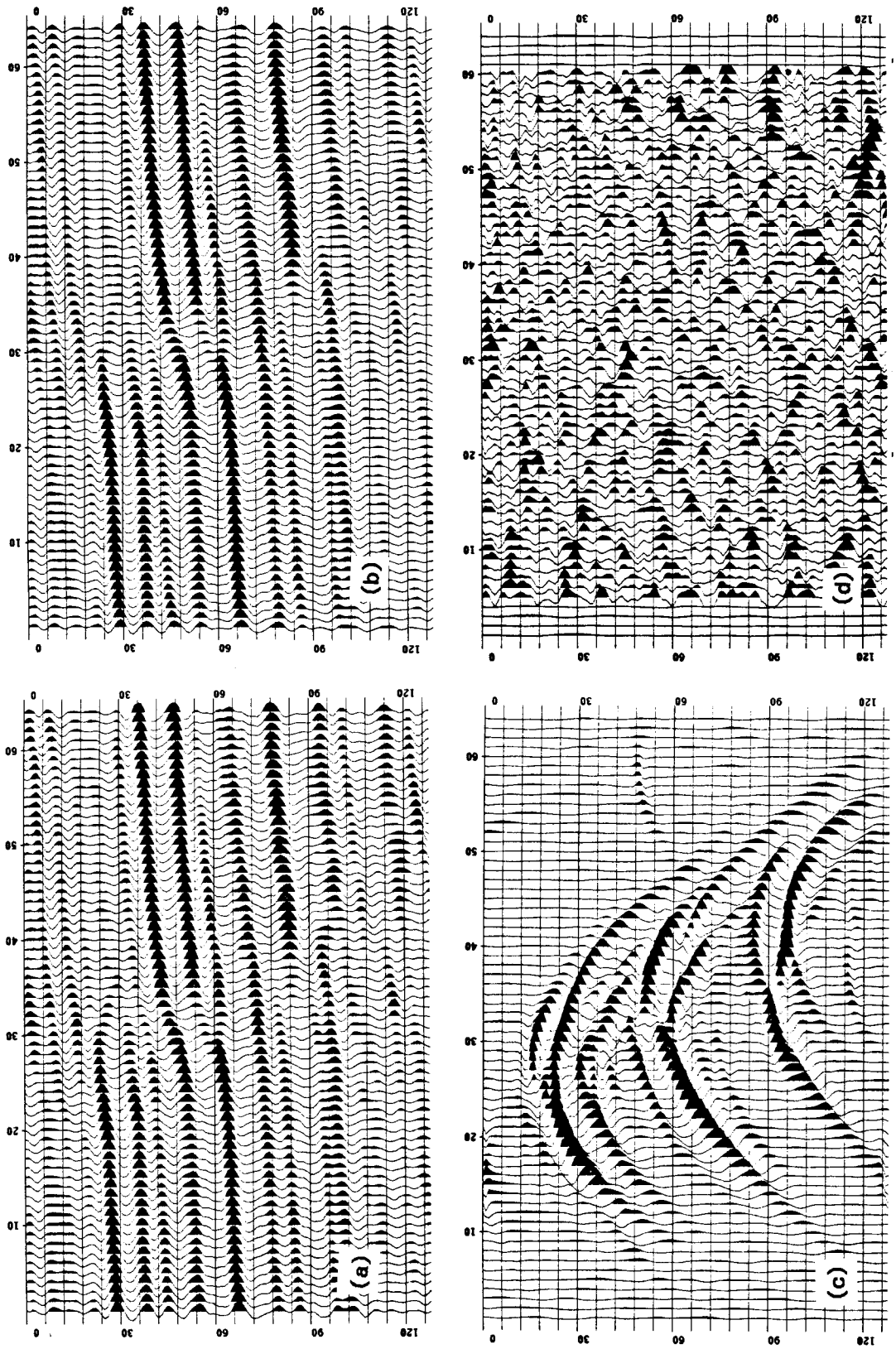


FIG. 1. (a) A window of stacked offshore Texas data contains a growth fault, with weak diffractions off truncated beds. Continuous reflections and noise obscure the velocity information in diffractions. (Data supplied by Western Geophysical: extending 1-1.5 s and 2 km, at 4 ms and 33 m sampling.) We decompose this window into the following three, which add up to the original data.
(b) Continuous reflections are easily described as a superposition of lines.
(c) Diffractions are easily described as a superposition of diffraction hyperbolas.
(d) Noise cannot be simplified by any invertible transformation.

Extracting diffractions from noise

To extract diffractions and estimate velocities, we use equations (15a) and (16a), with migration and velocity as the appropriate transform and parameter. For the present we assume that all diffractions within the window express the same unknown velocity v. For interval velocities one could estimate velocities in upper windows, then downward continue lower events. Velocities may be measured as locally as the density of diffractions permits. For particularly deep events the aperture of offsets is very small relative to the depth, so diffractions become the only source of velocity information.

We summarize the extraction of diffractions from noise. Previous methods are assumed for the estimation of the distributions in equation (11).

1. Migrate the data without bed reflections over a physical range of velocities.
2. For each migrated section, smoothly zero those samples containing significant noise.
3. Diffract (invert migration) each section at the extraction velocity.
4. Find the least-squares superposition of these diffracted sections best resembling the data without bed reflections.
5. Migrate this superposition over the previous range of velocities.
6. Determine the best migration velocity by evaluating the focusing measure.

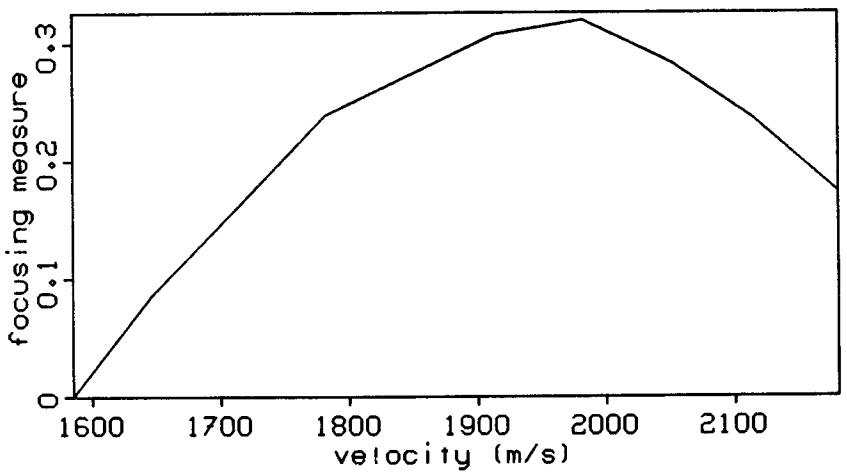


FIG. 2. Migration of the extracted diffractions in Figure 1c maximizes the focusing measure F at the best velocity.

Extracting diffractions with spatially variable velocities

We should like to extract diffractions with vertically and laterally varying velocities. One could use the results of the previous section directly by partitioning the migrated sections and extracting diffractions independently in each. This approach, however, requires a many-fold increase in expense and has particular problems at partition boundaries. Instead we shall assume smooth polynomial variations in rock velocities.

To allow polynomial variations of diffraction velocity with depth, we rewrite (15a) and (16a) to read

$$\bar{e}_v^n \equiv L_v^{-1}\{Gain\ by\ z^n\{Extract\{L_v\{\bar{d}\ \}\}\}\} \tag{15b}$$

$$\hat{s} = \sum_n \sum_v a_v^n\, \bar{e}_v^n\ ; \quad \min_{\mathbf{a}}(\hat{s} - \bar{d})^2 \ \rightarrow\ \sum_m \sum_w <\bar{e}_v^n, \bar{e}_w^m> a_w^m = <\bar{e}_v^n, \bar{d}>. \tag{16b}$$

We have added a gain to multiply by powers of the depth coordinate z. Lateral changes could be added with lateral gains. The least-squares solution may now smoothly change the weighting of extracted events spatially. Solving (16b) requires inverting a symmetric matrix of an order equal to the number of velocities times the order of gains used. Unfortunately, each additional order in the polynomial variation requires another inverse transform, L_v^{-1}, for each velocity.

If diffractions with substantially different velocities do not overlap in the time section, however, then we may expect the gain and the inverse transform to commute very well.

$$\bar{e}_v^n \equiv Gain\ by\ t^n\{L_v^{-1}\{Extract\{L_v\{\bar{d}\ \}\}\}\} \tag{15c}$$

Gains over the time coordinate t may be made implicitly in the scalar products. Thus, the additional inverse transforms may be avoided and variable velocities added with only the expense of additional scalar products.

To estimate the spatially variable velocities, one should migrate $\hat{s}$ over the proper range of velocities and evaluate the focusing measure in tapered windows.

FURTHER APPLICATIONS

We shall now exploit the generality of the previous sections. We employ localized slant stacks for a laterally adaptable extraction of locally linear events. By an algorithm similar to that for diffractions, wave-equation velocity analyses in common-shot and common-midpoint sections can treat offset truncations as a removable form of noise. Also, the interpretation of stacked sections and gathers need not be obscured by non-gaussian noise: the strongest non-gaussian noise may be subtracted from data without harming the weakest signal.

Finally, a treatment of static shifts demonstrates how hidden signal coherence can often be recovered.

Extraction of continuous events with localized slant-stacks

With localized slant stacks one may specify continuous events as a summation of short, tapered line segments of all dips. Global slant stacks require signal to be easily expressed as a sum of lines extending across the section--an assumption producing corresponding artifacts when the data do not agree. The localized slant stack maps a data set to a narrow cube with the transformation

$$d'(p,\tau,x_c) = \int\int W(\frac{x-x_c}{X_w})d[x,t=\tau+p(x-x_c)]dx \quad \textit{plus a "rho" filter.} \tag{17}$$

$W(x)$ represents a windowing function--in the simplest case, a rectangle function as in (7). x_c fixes the center of the window, and X_w the width. For an explanation of the rho filter, see the appendix on slant stacks. We may choose the inverse transform simply as

$$d(x,t) = \int d'(p,\tau=t,x_c=x)dp \tag{18}$$

(17) decomposes the array into planes containing narrow ranges of dip. Adding these planes together as in (18) reconstructs the original data.

The width X_w of the window specifies the number of traces over which an event should be approximately linear. A natural uncertainty principle results: lateral resolution must be traded for dip resolution. Wider windows see greater curvature, increasing the number of p values needed to describe an event. A larger X_w does allow greater concentration of linear signal.

One may implement the local slant stack very inexpensively. For a window of few traces, an $x-t$ implementation becomes less expensive than the Fourier. The loop over p should be the outermost, with sums over various windows and intercepts inside. If we take the weighting function W as a rectangle, then forward transformation without the rho filter requires linear moveout of the entire section, then horizontal sums over narrow windows of traces. The sum for one window gives that for the adjacent window after the addition of one trace and the subtraction of another [cf. Robinson]. This transformation becomes only twice as expensive as the equivalent global $x-t$ slant stack.

The great advantage of local slant stacks lies in its resistance to artifacts. Its lateral adaptability prevents these stacks from oversimplifying focused events, from straightening out or extending events. This transformation will easily extract events with higher

curvature than diffractions if X_w is small, on the order of 4-8 traces.

Local slant stacks successfully extract weak signal from behind strong gaussian noise in Figure 3. Strongly non-gaussian noise may be successfully removed without harming signal, as we find in a later section. Gaussian noise, however, never focuses or defocuses. Only extracting the strongest recognizable signal will improve the interpretability of this section. For the data window in Figure 3a, we extract those events showing significant coherence over at least four traces (Figure 3b). We took $X_w = four\ traces$ and $W(x) = \exp(-\pi x^2)$ in equation (17). Subtracting the strongest signal from the data leaves noise and weak signal (Figure 3c). The chosen range of p values excluded that of highly dipping diffraction tails, now appearing with the incoherent noise. In this way, extractions with limited ranges of slopes may separate overlapping coherent events.

An even more general transformation could express the data as a sum of short second-order curves. We define

$$d''(p,p',\tau',x_c') = \int\int W'\left(\frac{x-x_c'}{X_w'}\right) d'[p,\tau=\tau'+(p'+p)(x-x_c'),x_c=x]\,dx \qquad (19)$$

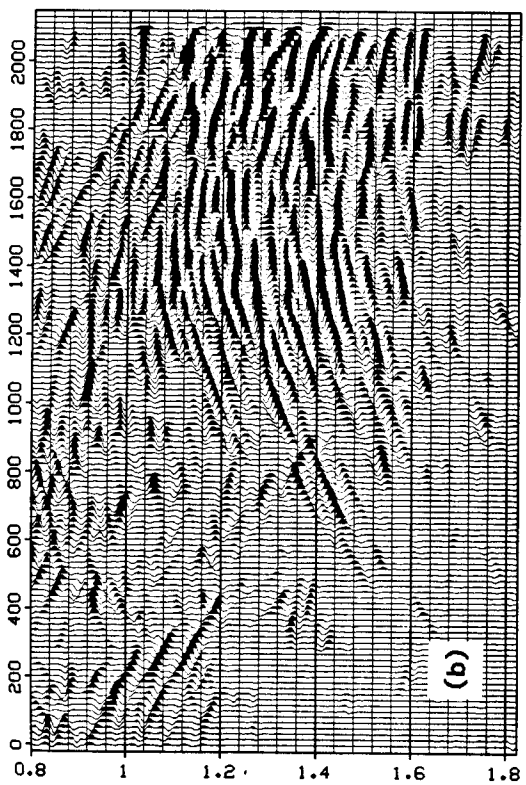
plus a rho filter.

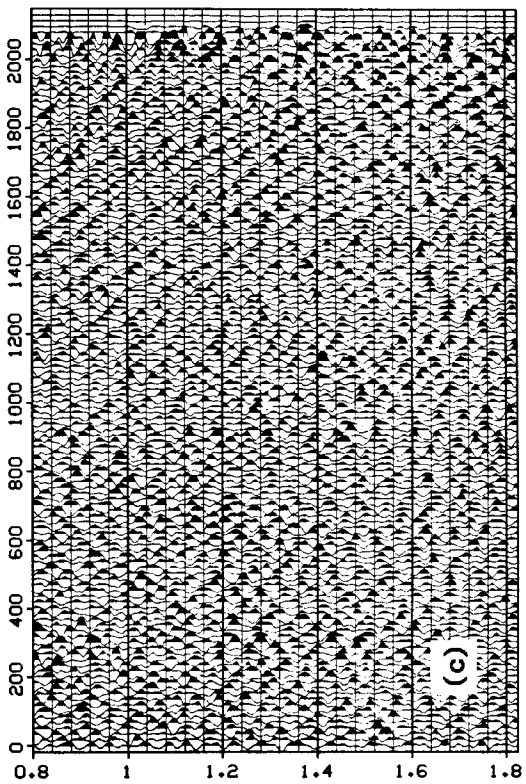
This transformation, defined in terms of (17), detects linear, spatial changes in the slope of events--that is, the curvature. This more general transformation will increase the focusing of curved events for a fixed window width. Alternatively, one may increase the width of windows and increase the signal-to-noise ratio in focused events.

Wave-equation velocity analyses for shot or midpoint gathers

Normal-moveout (NMO) velocity analyses have long been preferred to wave-equation methods because NMO ignores the truncation of events at the maximum offset. Wave-equation methods, on the other hand, can image arbitrarily dipping reflectors, whereas NMO implicitly assumes flat reflectors. A signal/noise decomposition makes the wave-equation method feasible by treating offset truncations as a removable form of noise.

Bed reflections in shot gathers may be extracted just as diffractions in offset sections, with important differences: 1) migration focuses reflections of planer beds at the corresponding image source; 2) two-way travel times must be used; 3) velocities must be allowed to vary spatially, at least with depth. The location of the migrated image indicates the slope of the reflecting boundary: level reflections should focus at zero offset. Truncations act as noise containing the negative of the events one expects at high offsets. Truncation noise disperses and avoids extraction; however, if too much signal is replaced by





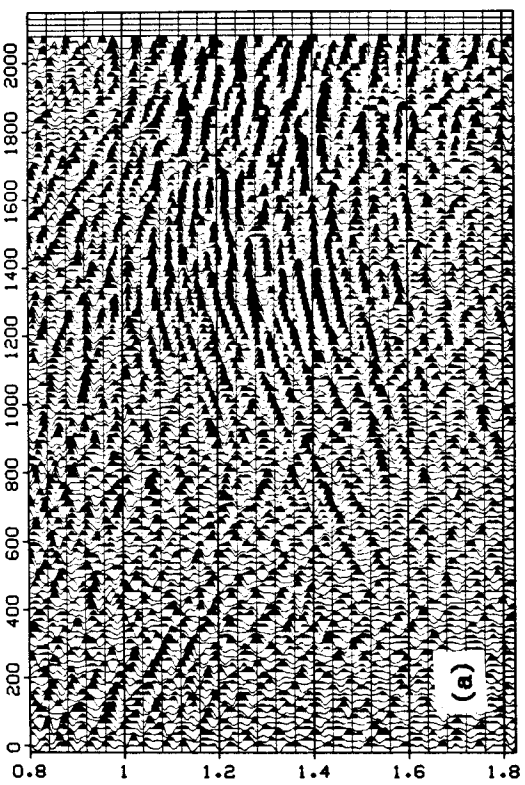


FIG. 3. (a) A window of stacked data contains weak, coherent reflections much obscured by strong gaussian noise. (Data courtesy of Southland Royalty; extending 0.8-1.8 s and 2.2 km.)
(b) We extract that component of the data easily expressed as a sum of short, tapered line segments--that signal focused best by localized slant stacks. All events also appear in 3a.
(c) Adding this section to 3b gives the original data 3a. Noise and weak signal remain. The chosen range of p values excluded highly dipping diffraction tails, now appearing with the incoherent noise.

truncation noise, focusing will be weak and hard to detect with the focusing measure, particularly for deep events. One can replace much of the truncated energy, but only with the limited information at near offsets. NMO, of course, also suffers from this loss of information.

We modify the procedure for diffractions with (15b) and (16b). The appropriate algorithm follows.

1. Migrate data over a physical range of velocities.
2. Extract the strongest signal in each migrated section.
3. Produce depth-gained versions of these sections to allow vertically varying velocities.
4. Diffract all sections at the extraction velocities.
5. Add the diffracted sections to make the least-squares superposition best resembling the data in the known range of offsets.
6. Migrate this superposition over the previous range of velocities
7. Evaluate the focusing measure locally for depth variable velocities.

The extracted events will possess longer tails than the original events, extending into the truncated offsets. By not constraining the tails in the least-squares superposition, only known offsets will affect the estimation of velocities.

Common-shot gathers are particularly adaptable to the streamlining indicated in (15c). With a very crude idea of the moveout of events one could gain the time section to very closely approximate the correct gain before diffraction.

Removing non-gaussian noise from shot gathers and offset sections

Often, for interpretation, one would like to remove the strongest noise from a section without harming the weakest signal. If one first estimates the strongest signal in a section and removes it, then an appropriate function $E(n \mid d)$ will indicate which samples contain the smallest percentage of signal (the strongest and most independent). Those samples containing a significant percentage of signal may be zeroed; what remains represents the strongest noise. This noise may then be subtracted from the original data. Gaussian noise, unfortunately, neither focuses or diffuses after an invertible transformation and cannot be distinguished from signal. The removable non-gaussian noise distracts the eye most.

For noise removal in the absence of full diffractions, localized slant stacks are more effective and less expensive than multiple migrations. Many common-offset or stacked sections contain smooth reflections whose interpretation is hindered by noise with comparable amplitudes. As a result, slant stacks focus these roughly linear events and diffuse noise

better than a single migration. Slant stacks, moreover, do not assume predetermined velocities, so resulting extractions may be used for velocity estimation. The following algorithm seems quite effective. We first remove the strongest signal as when separating bed reflections from diffractions.

1. Remove the strongest continuous events from the data, using a slant stack.
2. Artificially destroy coherence in the stack without strongest signal, invert, and make local histograms for $\hat{p}_s(x)$, overestimating remaining signal.
3. Recover the data without the strongest signal from 1. and zero samples containing significant signal; strongest noise must remain.
4. Subtract this strongest noise from the original data.
5. Repeat if necessary.

Once the strongest non-gaussian noise has been removed, the strongest signal may be estimated more accurately. For this reason, at least two iterations are recommended.

We apply this procedure to a shot gather with common noise problems. Figure 4a displays a common shot gather (provided by Western Geophysical) corrupted by very strong aliased ground-roll, sharp static shifts, and over-amplified traces. Figure 4b contains the extracted ground-roll and other noise; no reflections are visible. Figure 4c shows the data with ground-roll subtracted. Because ground-roll possesses substantial coherence of its own, a few residual events remain at dip aliases, but certainly the majority of the interfering waveforms are gone. Note that events not originally covered by ground-roll or other noise are unaffected. A dip filter, by contrast would have removed high dips from all events.

Correction of static shifts

Land data often contain traces shifted by uneven topography and variations in surface velocities. Almost all information remains; only the exact zero time is uncertain. Yet if not corrected, this trace will interfere as pure noise. A static correction should be that shift of the entire trace adding the most coherency to the data.

Let us first extract the strongest signal in the data by one of the methods of previous sections. High-frequency static shifts interfere as noise and do not contribute to the estimate. Let us cross-correlate original traces with the estimate of the strongest signal, shifting the original trace by varying amounts and taking the scalar product with the extracted trace. This correlation will appear largest at the static shift time placing events in the most coherent position. Traces may then be corrected accordingly, and the procedure repeated for second-order improvements.

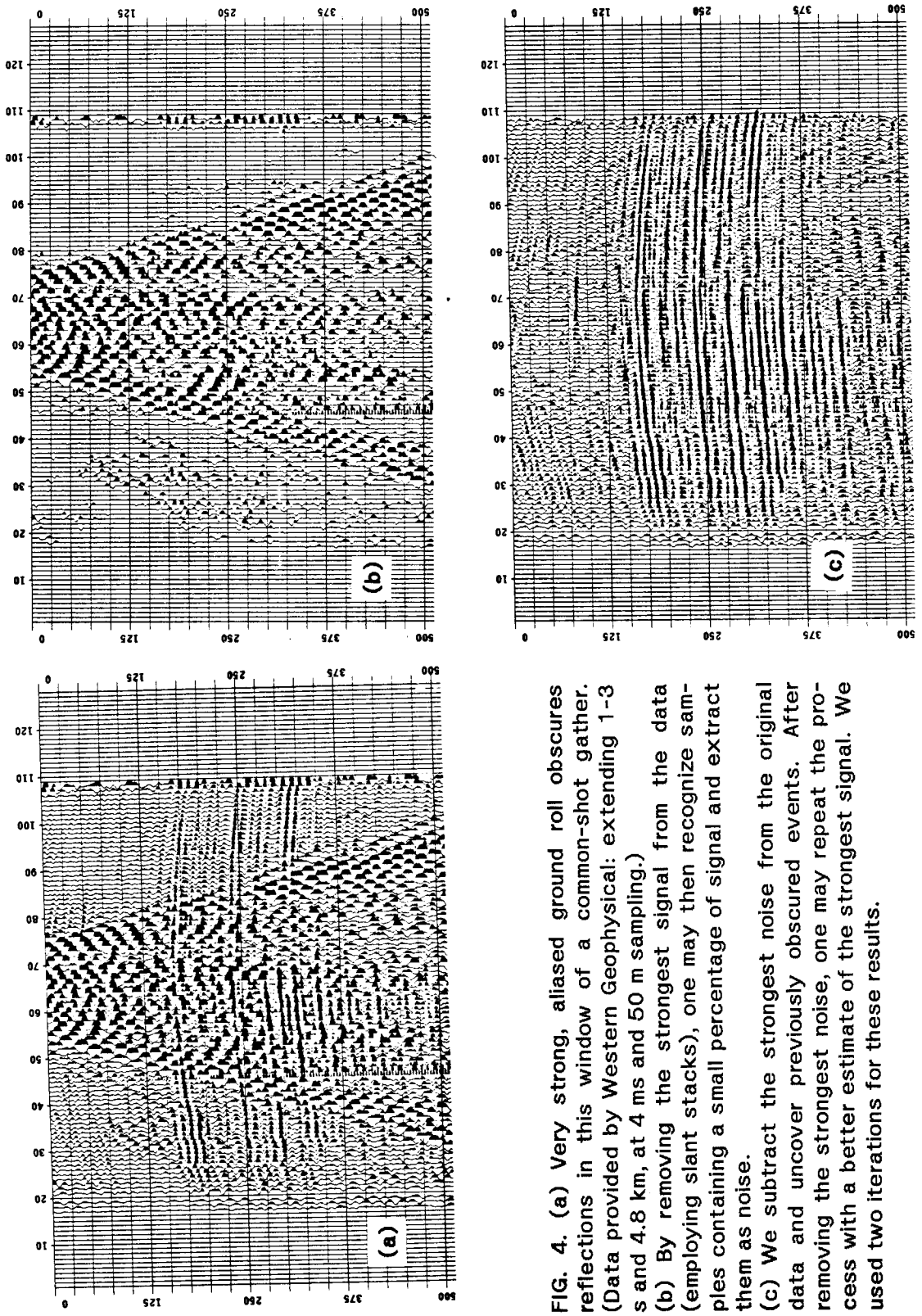


FIG. 4. (a) Very strong, aliased ground roll obscures reflections in this window of a common-shot gather. (Data provided by Western Geophysical: extending 1-3 s and 4.8 km, at 4 ms and 50 m sampling.)
(b) By removing the strongest signal from the data (employing slant stacks), one may then recognize samples containing a small percentage of signal and extract them as noise.
(c) We subtract the strongest noise from the original data and uncover previously obscured events. After removing the strongest noise, one may repeat the process with a better estimate of the strongest signal. We used two iterations for these results.

ACKNOWLEDGMENTS

I wish to thank Fabio Rocca for his most recent suggestion to generalize my local slant stacks to accommodate the curvature of events. Again, his and Jon Claerbout's help has added much coherence to these noisy thoughts.

REFERENCES

Berkhout, A. J., 1980, Seismic migration--Imaging of acoustic energy by wave field extrapolation: Amsterdam/New York, Elsevier/North Holland Publ. Co.
Claerbout, J. F., 1983, Imaging the earth's interior (in press).
Godfrey, R., 1979, A stochastic model for seismogram analysis: Ph.D. thesis, Stanford University, 88 p.
Harlan, W. S., 1982, Signal/noise separation with slant stacks and migration: SEP-32, p. 25-35.
Harlan, W. S., 1982, Wave equation event migration: SEP-32, p. 79-80.
Harlan, W. S., J. F. Claerbout, and F. Rocca, 1983, Extracting velocities from diffractions: SEP-35, p. 107-126.
Kullback, S., 1959, Information theory and statistics: New York, John Wiley.
Papoulis, A., 1965, Probability, Random Variables, and Stochastic Processes: New York, McGraw-Hill Book Co.
Robinson, J., and T. Robbins, 1978, Dip-domain migration of two-dimensional seismic profiles: Geophysics, v. 43, p. 77-93.
Shannon, C. E. and W. Weaver, 1963, The mathematical theory of communication: Chicago, University of Illinois Press.
Thorson, J., 1978, Reconstruction of a wavefield from slant stacks: SEP-14, p. 81-86.
Van Trees, H., 1968, Detection, estimation, and modulation theory, v.1: New York, John Wiley.

APPENDIX

We now derive formulations of global and local slant stacks. First we derive a global, frequency-domain implementation and find a sufficient sampling of dips. Then we discuss implicit operations performed by the frequency-domain interpolation.

Frequency-domain implementation of slant stacks

A frequency domain implementation becomes less expensive than the spatial-domain implementation for sections of many traces. Lateral wrap-around may be avoided with the proper interpolator. We define slant stacks by the inverse transformation.

$$d(x,t) = L\bar{d}' = \int d'(p,\tau=t-px)\,dp \qquad \text{(A-1)}$$

Thus, a single point in the slant stack domain $d'(p,\tau) = \delta(p-p_0)\,\delta(\tau-\tau_0)$ will map to a line in

the spatial domain. Two points map to two additively superimposed lines. Let the variables (x,t,p,τ) have the Fourier duals of (k,s,q,ν). By Fourier transforming to $d(x,s)$ and $d'(q,\nu)$, we find a simplified relation.

$$d(x,t) = \int\int\int e^{i2\pi\nu(t-px)} e^{i2\pi qp}\, d'(q,\nu)\,dp\,dq\,d\nu \tag{A-2}$$

$$= \int\int\int e^{i2\pi p(q-\nu x)} e^{i2\pi\nu t}\, d'(q,\nu)\,dp\,dq\,d\nu = \int e^{i2\pi\nu t}\, d'(q=\nu x,\nu)\,d\nu$$

$$\rightarrow\ d(x,s) = d'(q=sx,\nu=s) \tag{A-3}$$

This transformation may be implemented as a frequency domain stretch. Unfortunately, in the forward direction a rectangle is mapped to a triangle. Dips not in the range of calculated p's must be aliased. The adjoint transformation, however, implicitly zeros those dips omitted from $\bar{d}'$.

$$d'(p,\tau) = L^*\bar{d} = \int d(x,t=\tau+px)\,dx \tag{A-4}$$

$$\rightarrow d'(p,\nu) = d(k=-\nu p, s=\nu) \tag{A-5}$$

We may implement (A-3) by first performing (A-5), then following with an equivalent of the "rho" filter of Radon transforms. We recognize this filter as the inverse $(L^*L)^{-1}$.

$$\tilde{d}'(q,\nu) = L^*L\bar{d}' = \int\int e^{i2\pi(\nu p)x} e^{-i2\pi qp}\, d'(q=\nu x,\nu)\,dx\,dp \tag{A-6}$$

$$= \int \delta(\nu x - q)\, d'(q=\nu x,\nu)\,dx = \int \frac{1}{|\nu|}\delta(x-\frac{q}{\nu})\, d'(q=\nu x,\nu)\,dx = \frac{1}{|\nu|} d'(q,\nu)\,dx$$

$$\rightarrow\ d'(p,\nu) = L^{-1}\bar{d} = |\nu|\, d(k=-\nu p, s=\nu) \tag{A-7}$$

Forward transform with (A-7) to avoid aliasing dips and inverse transform with (A-3).

One should sample p values well enough to avoid aliasing frequencies in traces at high x. Equivalently the sampling in (A-7) should satisfy

$$\Delta p \cdot s_{\max} < \Delta k$$

The discrete equivalent becomes

$$\Delta p < \frac{\Delta t}{N_x \cdot \Delta x} \tag{A-8}$$

where N_x is the number of x samples.

Frequency domain interpolation

The discrete frequency-domain stretch over k requires an interpolation operator to implicitly zero replications responsible for "wrap-around." Interpolation operators are equivalent to convolutions. Convolving by a function over k is equivalent to multiplying by a function over x. We should like to multiply $d(x,t)$ by a rectangle function which drops to zero before the first trace and after the last. Let us derive a general interpolator and allow arbitrary implicit windowing functions in the other domain. Localized slant stacks should use a gaussian window. Stolt migration requires a similar stretch over temporal frequencies; the windowing rectangle shifts when the surface is to be placed at other than the first sample.

Define three parameters: x_z, the coordinate to be newly assigned the value of zero in the x domain; x_c, the center (in new coordinates) of the function to multiply the x domain; X_w, the width of the function to multiply the x domain. We assume the function $W(x)$ to be symmetric about zero. Windowing $d(x)$ with $W(x)$ gives the transform

$$W(\frac{x-x_c}{X_w})d(x) \quad transforms\ to \quad \int e^{-i2\pi x_c(k-k')} X_w\, W'[X_w(k-k')] e^{i2\pi x_z k'} d'(k')dk' \qquad \text{(A-9)}$$

(We suppress dimensions not being stretched.) $W'(s)$ is the Fourier transform of $W(x)$. Interpolation merely modifies (A-9) to the discrete case. Let n_z, n_c, and N_w be the samples of the important parameters.

$$d'_{n+\delta n} = \sum_{m=0}^{N_x-1} \frac{N_w}{N_x} W'\left[\frac{(n+\delta n-m)N_w}{N_x}\right] e^{-i2\pi n_c(n+\delta n-m)/N_x} e^{i2\pi n_z m/N_x} d_m' \qquad \text{(A-10)}$$

For a global slant stack and for Stolt migration with the surface at the first sample we may use $n_z = 0$, $n_c = N_x/2$, and $N_w = N_x$. We transform the rectangle function of equation (7) into the sinc function. The following simplification occurs.

$$d_{n+\delta n} = \frac{1}{\pi} e^{-i\pi\delta n} sin(\pi\delta n) \sum_{m=0}^{N_x-1} d_m/(n+\delta n-m) \qquad \text{(A-11)}$$

(A-11) may be tapered to a small number of terms.

When the windowing function $W(x)$ becomes a gaussian, we may minimize the number of terms in the interpolation operator. Let $W(x) = exp(-\pi x^2)$ so that X_w governs the distance between half-amplitude points ($W(0.5)\approx 0.5$). Then $W'(s) = exp(-\pi s^2)$. If we preserve all terms in (A-10) with coefficient magnitudes greater than 0.01, then

$$|n+\delta n-m| < 1.2\,\frac{N_x}{N_w}$$

The number of terms in the interpolator should be greater than $\sim 3 \cdot N_x/N_w$.

Note that smaller N_w allow larger Δp's. When n_c = 0 we may replace N_x by N_w in (A-8).