Chapter 2
Signal/noise separation and velocity estimation: modeling seismic coherence

2.1. SUMMARY

In Chapter 1 we examined the statistical tools of signal/noise separation for a simple mathematical model. I shall generalize these methods for non-linear physical models in Chapter 3 and in the appendices. But first, let us explore strategies that will help us choose mathematical models for a variety of coherent seismic events. These models should simplify the description of coherence created by geologic structure as well as by wave propagation. I shall emphasize the extraction of reflections containing velocity information because this problem presents a well-known dilemma in seismic processing: not all seismic reflections contain velocity information; however, velocities must be known before reflections with velocity information can be modeled and isolated. Let us take a quick look at the results of this chapter.

A signal/noise separation should recognize the lateral coherence of geologic reflections, their statistical predictability, before it extracts those components most useful for velocity analysis. Reflected events with recognizable coherence can be modeled as signal, the rest as noise. Let us define "focusing" as increasing the statistical independence of samples with some linear transformation. Using the central-limit theorem, we shall see that samples of signal must become more non-Gaussian after focusing—that is, less likely to have come from the Gaussian or "normal" probability distribution. The same linear transform must defocus noise and make samples of noise more Gaussian than they were before transformation. We can find a measure, defined from cross entropy, that measures non-Gaussianity from histograms of an array and thereby measures focusing. These histograms of linearly transformed data, with and without coherent signal, estimate the amplitude distributions of transformed signal and noise—the distributions can be guaranteed not to underestimate the noise. These

distributions allow us to recognize and extract transformed samples containing the highest percentage of signal. An iterative estimation of signal and noise improves the extractions of each. Such issues will be discussed in section 2.2.

The focusing measure determines the best migration velocity for diffractions that are extracted from behind undiffracted reflections and noise. Only diffracted reflections contain velocity information in zero-offset or "stacked" sections. Slant stacks allow us to model undiffracted reflections as sums of dipping line segments. Let us extract reflections that are linear over a distance greater than the Fresnel zone; unextracted events will include diffractions and noise. Seismic migration focuses diffractions best; we can use the inverse of migration (forward modeling) to model the signal. I extract events for a range of migration velocities and find the least-squares superposition that best resembles the diffractions in the original data. Migration of these diffractions maximizes the focusing measure at the best velocity. This algorithm will successfully extract diffractions and estimates velocities for a window of data containing a growth fault (section 2.3).

Local slant stacks allow locally linear reflections to be extracted in a laterally adaptable fashion. Weak signal with highly variable coherency will be extracted from behind strong Gaussian noise in a stacked section (section 2.4.1).

Unlike NMO stacks, wave-equation migration of a few adjacent common midpoint gathers can correctly image the skewed hyperbolas of dipping reflectors (section 2.4.2). Short local slant stacks will extract reflections with different dips over midpoint. A simple Stolt (f-k) algorithm migrates these dipping reflections with appropriate dispersion relations. This migration can then be used to extract reflections containing velocity information over offset. Previously, it has been necessary to examine migrated sections by eye to determine the best seismic velocities. Now, by reliably isolating the previously obscured diffractions, we can estimate seismic velocities numerically.

2.2. THE FOCUSING AND EXTRACTION OF SEISMIC REFLECTIONS

To use the signal/noise separation of Chapter 1 in other applications, an interpreter must prepare mathematical models that efficiently describe the coherent signal in the data. The interpreter must understand how the appearance of these events can be simplified by transformation, so that the fewest parameters will contain the most useful information. I shall now develop some physical and statistical principles that are useful in the derivation of good transforms.

2.2.1. The statistical tools

When interpreting noisy surface seismic data, one searches for the lateral coherence of geologic reflections, their statistical predictability. For example, the reflections of smooth, continuous beds appear hyperbolic in common-shot and common-midpoint (CMP) gathers and appear even more smoothly continuous in common-offset or stacked sections. Faults and other lateral irregularities in reflecting beds create coherent diffraction hyperbolas in stacked sections. Such consistent lateral predictability means that our reflected events are overspecified. For example, linear reflections could be specified by slopes and vertical intercepts rather than by their depths every 50 meters; diffraction hyperbolas could be specified by scattering locations and root-mean-square (rms) seismic velocities.

Noise then must include components of the data whose description cannot be simplified, components showing no spatial coherence or predictability. To be chosen as signal, a geologic component must have recognizable coherence. If our processing fails to recognize the coherence, some geologic reflections might then function as noise. A given component of the data, signal or noise, usually has a limited class of possible appearances; each possibility can be produced by a combination of more fundamental, statistically independent "events." It will be our business to describe and recognize these events as efficiently as possible.

We should remember that our goal is to extract rather than estimate signal. An estimate can be said to find the expected or most probable signal in the given data; all other events are assumed to be noise. An extraction should accept only signal with a pre-determined reliability and should eliminate details significantly contaminated with noise. Similarly, to extract noise, one should avoid the extraction of signal. Alternately extracting signal and noise will place information in its most reliable domain; unextracted events should not be distinguishable as either signal or noise.

Before describing specific algorithms, I shall establish a few statistical results. Using these results, we can satisfy the following goals: identify a component of the data as signal, with predetermined coherence and reliability; identify the transform that simplifies and describes this signal best. As a preview, the most important results will be listed.

1. Define "focusing" as a transformation making the samples of an array statistically independent. Inversions are most easily interpreted when the inverted parameters are focused.

2. An inversion transform that focuses signal must also defocus noise.
3. The samples of focused signal become more non-Gaussian; those of defocused noise become more Gaussian.
4. A function will estimate the non-Gaussianity of samples of an array and thereby measure their focusing.
5. Amplitude distributions of the transformed signal and noise can be estimated from histograms.
6. With these distributions, one can identify and extract those samples of the transformed data containing a high percentage of focused signal with a high reliability.
7. To identify the best focusing transform, extract signal with several transforms and find a superposition best resembling the data. The focusing measure will identify which of these transforms focuses the superposition best.

Mehta (1977) first proposed using a cross-entropy measure of non-Gaussianity to estimate migration velocities from unstacked data. De Vries and Berkhout (1982) substituted the varimax function (Wiggins, 1977) for the cross-entropy and applied the strategy to stacked data. Mehta made the powerful observation that extrapolation of a wavefield back to its source will make the wavefield "simpler" and more non-Gausssian. I shall elaborate on this idea in the following section, but first I shall argue the necessity of signal/noise separation. In brief, events containing velocity information should first be separated from other events, including noise. I shall also use, instead of the varimax function, a modification of Mehta's non-Gaussianity measure (in Appendix D).

2.2.2. Extrapolating waves to focus signal

A velocity analysis can implicitly extrapolate wavefields back in time, so that signal is concentrated and noise is dispersed. For example, an NMO stack finds the image source of reflections from planar interfaces. Wave-equation stacks and migration extrapolate to image sources (exploding reflectors). To simplify the appearance of reflections, each of these processes removes much of the coherence that resulted from wave propagation. As we shall see, these claims are equivalent to one first made by Mehta (1977): "The amplitude pdf [probability density function] for the wavefield ... will deviate more from a Gaussian pdf at the source point than at any other point in the migration process."

Extrapolations of wavefields help one recognize velocities because seismic wavefronts begin simply and produce increasingly diffuse wavefronts with time. The waves of seismic surveys begin from a confined source on the earth's surface. These waves expand, reflect, and create new waves, which in turn create others. The number and disorder of waves steadily increases with time.* An extrapolation of these waves backwards in time should recover increasingly simpler waves. The extrapolation that simplifies the waves best will determine the best seismic velocities.

Waves can be extrapolated backwards in time just as they are extrapolated forward in time, except for the noted changes in disorder. Linearly propagating waves obey the wave equation equally well if the time axis is reversed. (For instance, the same Green's functions will suffice. Huygen's principle of superimposed wavelets will predict new wavefields in either direction in time.) Changes in disorder are different, however. For example, backwards propagation creates extraordinary "coincidences": wavefronts steadily decrease in length and increase in curvature, and finally merge together at time zero into a single impulsive source. This propagation obeys the wave equation as if time were going forward. One would not believe, however, that forward-propagating waves were collapsing and becoming simpler with time.

If a recorded event appears late in time without the telltale coherence of a propagating wave, it must be noise. Because the wave equation is symmetric in time, recorded events that are isolated in space must be extrapolated to diffuse and coherent sources of seismic energy at time zero. Thus, the extrapolation that concentrates the signal best, while dispersing noise, determines the best velocities.

For instance, Figure 2.1a shows a synthetic zero-offset section, containing a coherent hyperbolic event and an incoherent spike. The hyperbola corresponds to an impulsive, spherically spreading wave at depth. The spike corresponds to noise. A wave-extrapolation procedure called migration will find the image sources at time zero that best explains these events (Figure 2.1b). The hyperbola maps to an impulsive source; the noise spike maps to a semi-circle (or to an ellipse because of vertical exaggeration). The semi-circle describes in the earth a source that fortuitously collapses at the correct time and recording position. We immediately recognize the improbability of such a source and conclude that the noise spike did not result from wave

*As the number of waves grows, the propagation of waves behaves similarly to a diffusion problem of statistical mechanics. When many particles diffuse in a volume, they lose any order they might have begun with: local concentrations of particles spread out and the density of particles begins to approach a constant. The "disorder" of the system (measured by its thermodynamic entropy) increases with time, according to the second law of thermodynamics.

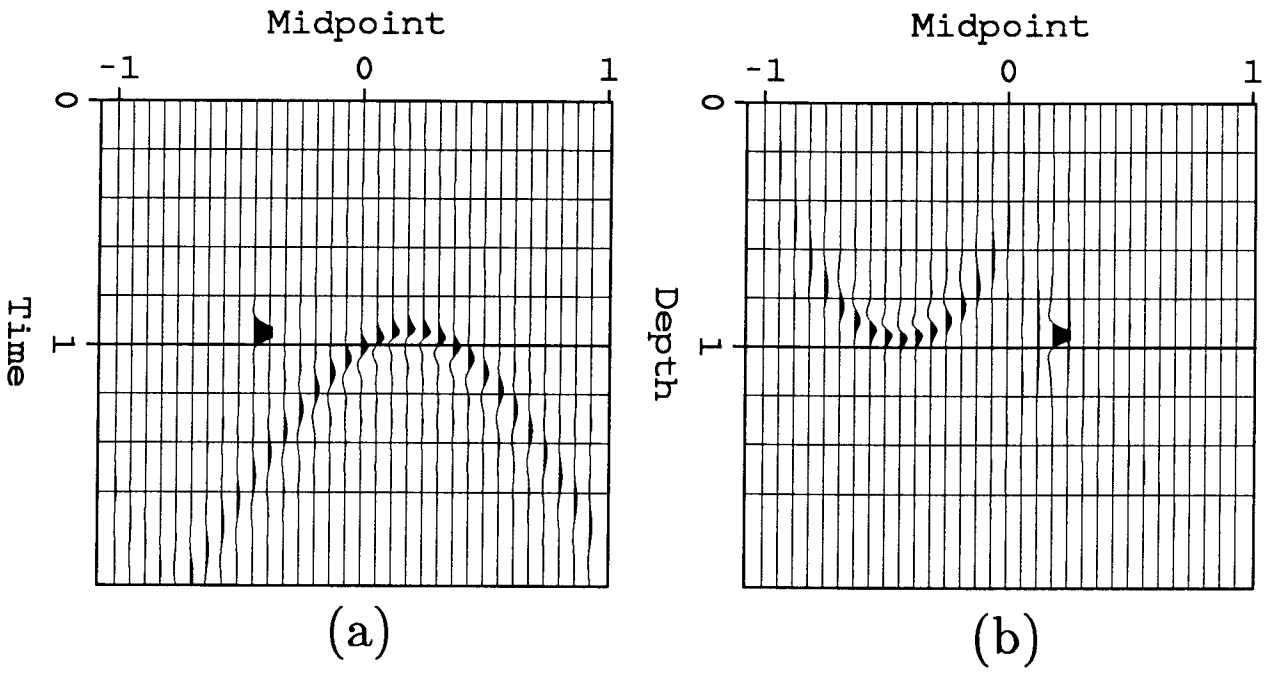

FIG. 2.1. (a) A zero-offset section contains a hyperbolic diffraction (signal) and a single point (noise). The hyperbola corresponds to an impulsive, spherically spreading wave at depth. (b) Migration at a velocity of 1 focuses the diffraction correctly into an impulsive image source, and defocuses the point into an ellipse. The samples containing the signal are of much higher amplitude than those containing the noise. In statistical terms, the noise has become more Gaussian and coherent, the signal less so.

propagation.

Our eyes recognize that wave extrapolation has simplified the synthetic signal and has dispersed and complicated the noise. Let us say that the extrapolation has "focused" the signal and has "defocused" the noise. In the following section I shall formulate these observations statistically and numerically.

2.2.3. A measure of focusing

To quantify the "interpretability" of an inverted model is no simple task. In this section I first explore statistical properties that seem necessary to the construction of simple models. I argue in particular that the most interpretable models have the most statistically independent and non-Gaussian parameters. I then describe a method for measuring these properties.

Not all inverted model parameters need describe useful information in the data. When we choose a mathematical model for our signal, we also choose default values for the model's physical parameters. For example, samples of the NMO stack default to

zero. These default values generally describe a featureless, uninteresting earth model. We prefer to assume this description of the earth unless the data tell us differently. Assume that we have a simple transformation of the data that will update parameters to describe the data better (a least-squares inversion, a gradient perturbation, etc.). Only altered model parameters give useful information about the data; parameters remaining at default values require no interpretation.

The most interpretable model would contain the largest percentage of unchanged parameters. The smaller the number of perturbed parameters, the higher the concentration of information in each. Let us return to Figure 2.1b. This model is defined as a migration of the data. If the data contained only hyperbolic signal, then the model would contain only impulsive spikes. Each spike occupies only a few samples; moreover, each spike can be interpreted independently. On the other hand, if the data contained only noise, then the model would contain a superposition of ellipses. Ellipses occupy a great many samples, and each ellipse must be interpreted in its entirety. Signal should be described by independent perturbations of model samples; noise, by dependent perturbations.

In other terms, signal can be described by statistically independent model parameters, the noise by dependent ones. Parameters that are inverted from recorded data can show all degrees of dependence and independence; we must discover how to measure these degrees. Let us assume for the moment that the data contain only signal, created by a linear transformation $\mathbf{F}^1$ of the statistically independent samples of a model array:

$$\text{data}_i = \sum_j \mathbf{F}^1_{ij} \, \text{model}_j \quad ; \qquad \mathbf{data} = \mathbf{F}^1 \cdot \mathbf{model} \quad .$$

For example, $\mathbf{F}^1$ could be the forward modeling transform that creates diffraction hyperbolas (the signal) from single points (the model). A single sample of the data is a weighted sum of many samples of the model.

In Chapter 1, I defined signal as that component of the data that is most "easily" described by the modeling transform—here, $\mathbf{F}^1$. For the easiest interpretation, the signal should be described by statistically independent model samples. If the samples were not independent, then it might be possible to reduce further the number of samples required to describe the signal. It is easiest to interpret individually each sample necessary to describe the signal and to make as little reference as possible to other samples.

Next assume that an inversion of the model is a linear transformation $\mathbf{F}^2$ of the data:

$$\hat{\mathbf{model}} = \mathbf{F}^2 \cdot \mathbf{data} = \mathbf{F}^2 \cdot \mathbf{F}^1 \cdot \mathbf{model}$$

For example, the migration of Figure 2.1 is the least-squares inverse of the forward transformation that creates diffraction hyperbolas from points. This inverse must also be a linear function of the original model. The inversion will, in general, be imperfect. The samples of the model are statistically independent. If $\mathbf{F}^2$ perfectly inverts $\mathbf{F}^1$, then the samples of the inverse will also be independent. If the inversion is less than perfect, the samples of the inverse will remain a weighted sum of samples of the model.

For example, a sample $\hat{m}_i$ of the inverted array could be the sum of one, two, or three independent samples of the original model array $\mathbf{m}$.

$$\hat{m}_i{}^1 = m_i$$

$$\hat{m}_i{}^2 = m_i + m_{i+1}/100$$

$$\hat{m}_i{}^3 = m_i + m_{i+1}$$

$$\hat{m}_i{}^4 = m_i + m_{i+1} + m_{i+2}$$

Each of these inversions represents an increasingly poor inversion. The first, $\hat{m}_i{}^1$, is perfect, exactly equaling a sample of the signal. The second, $\hat{m}_i{}^2$, includes a fractional contribution from another sample. The third, $\hat{m}_i{}^3$, includes equal contributions from a correct and an incorrect sample. The fourth, $\hat{m}_i{}^4$, is the sum of one correct and two incorrect samples.

Each of the four inverted arrays shows increasing dependence between samples. $\hat{m}_i{}^1$ and $\hat{m}_{i+1}^1$ are independent, but $\hat{m}_i{}^2$ and $\hat{m}_{i+1}^2$ are slightly dependent, and $\hat{m}_i{}^3$ and $\hat{m}_{i+1}^3$ are very dependent. $\hat{m}_i{}^3$ and $\hat{m}_{i+2}^3$ are independent, but $\hat{m}_i{}^4$ and $\hat{m}_{i+2}^4$ are dependent, and so on. Measuring the statistical dependence between samples is difficult because the data rarely provide enough redundancy. Fortunately, as the independence of inverted samples changes, so do their probability distributions, which can be estimated more easily.

Because of the central-limit theorem, a sum of independent random variables creates a more Gaussian, or normally distributed, random variable (see a text such as Papoulis, 1965). If the samples m_i are statistically independent and are taken from the same non-Gaussian distribution, then the perturbations $\hat{m}_i{}^1$, $\hat{m}_i{}^2$, $\hat{m}_i{}^3$, and $\hat{m}_i{}^4$ are increasingly Gaussian. We can see such results numerically.

Figure 2.2a contains an array of independent, non-Gaussian random numbers. Half the samples have zero amplitude; the other half are evenly distributed between the maximum and half the maximum amplitude. The variance of the amplitudes is 1; the expected value is zero. Figure 2.3a shows the amplitude histogram of this array and shows the Gaussian distribution with a variance of 1. The histogram strongly diverges from this Gaussian distribution.

Figure 2.2b displays a convolution of the first array with a three-sample triangle function ([1,2,1], scaled for a final variance of 1). Each new sample is the sum of three independent random variables. The central-limit theorem predicts that each new variable will be closer to Gaussianity than the old. A histogram of this array (Figure 2.3b) is visibly closer to the Gaussian distribution than was the first. A second convolution brings the array even closer to Gaussianity (Figures 2.2c and 2.3c).

By dispersing and overlaying independent information, each convolution increases the statistical dependence of samples. Because the convolutions are invertible, one can imagine a deconvolution that proceeds in the opposite order and restores the original data. Linear transformation of independent random variables created Gaussian variables; inversion would recover the non-Gaussianity.

Partial inversions can be ranked by the independence of their samples, but statistical independence is not directly quantifiable. Data do not easily contain the redundancy needed for us to estimate independence. We can, however, measure the non-Gaussianity of independent samples. To find enough redundancy for measuring the non-Gaussianity, we must observe certain assumptions.

Describing signal by the fewest random variables (parameters to be estimated) allows use of the simplest statistical tools. If a transformation has focused all signal in the data, then signal statistics can be completely described by independent probability density functions (pdf's), which measure the probability of a sample having particular amplitudes. For unfocused signal, one would be obliged to use "joint" pdf's that measure the probability of all possible combinations of sample amplitudes.

Focused data easily provide enough redundancy for the estimation of independent pdf's if we assume that the statistics of parameters do not change rapidly over spatial dimensions. A robust and unbiased statistical model should describe physical structures that might be found anywhere in a region. An inversion that was told where a physical structure was likely to appear could create a model that only reinforced the interpreter's prejudices. If a structure appears often in one location then one should assume that it can also appear nearby. Thus, one not only expects but desires that

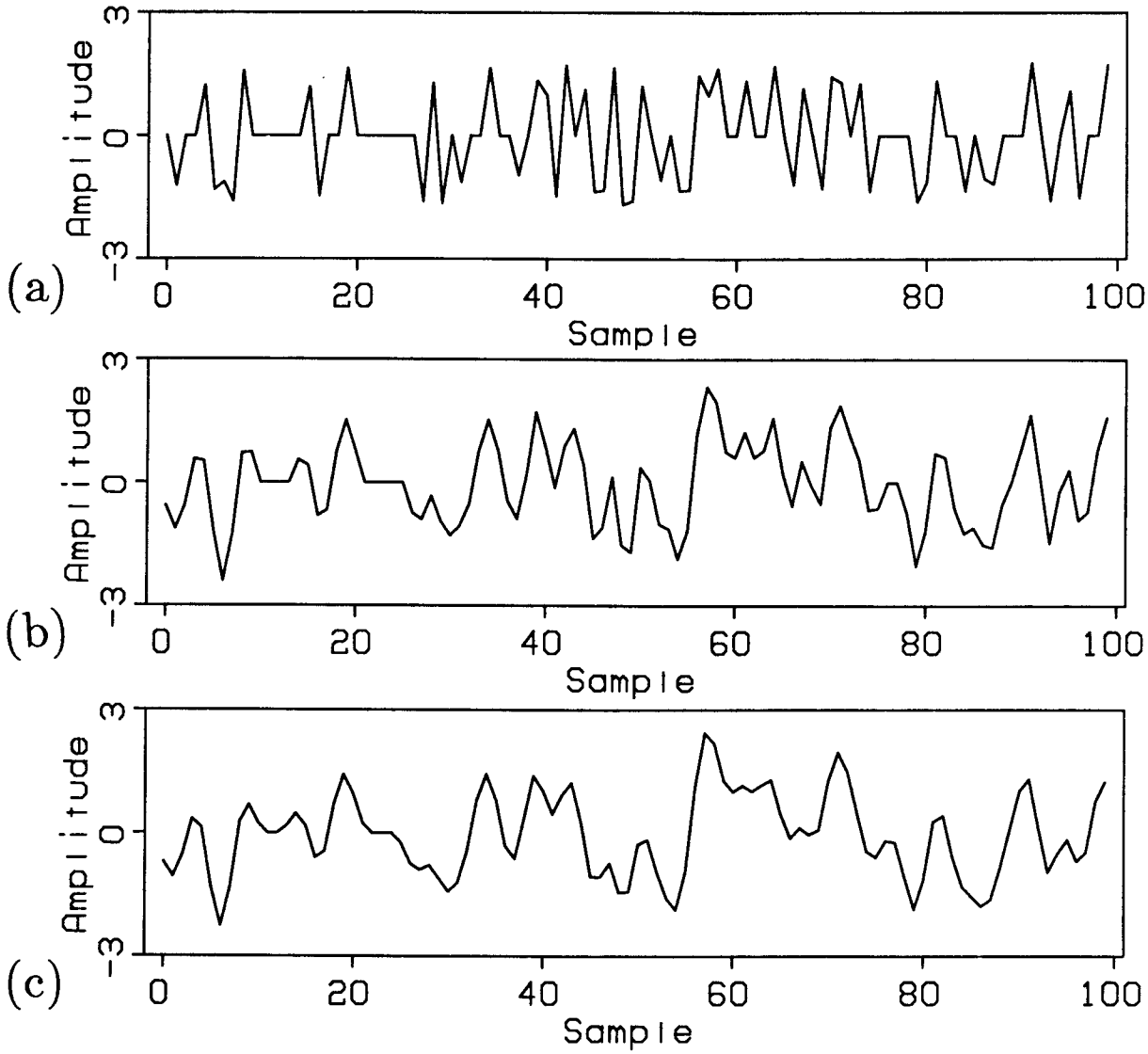

FIG. 2.2. (a) This array of independent, random numbers corresponds to the histogram of Figure 2.3a. Half the samples are zeros; the other half are evenly distributed between the maximum and half the maximum amplitude. (b) The array of 2.2a is convolved by a three-point triangle function, adding coherence. Each sample is now the sum of three independent random variables; the central-limit theorem predicts that these samples will be more Gaussian. (c) The array of 2.2b is convolved a second time by the triangle function. The array has again increased visibly in coherence; the non-Gaussianity of samples should increase further. Each convolution is invertible.

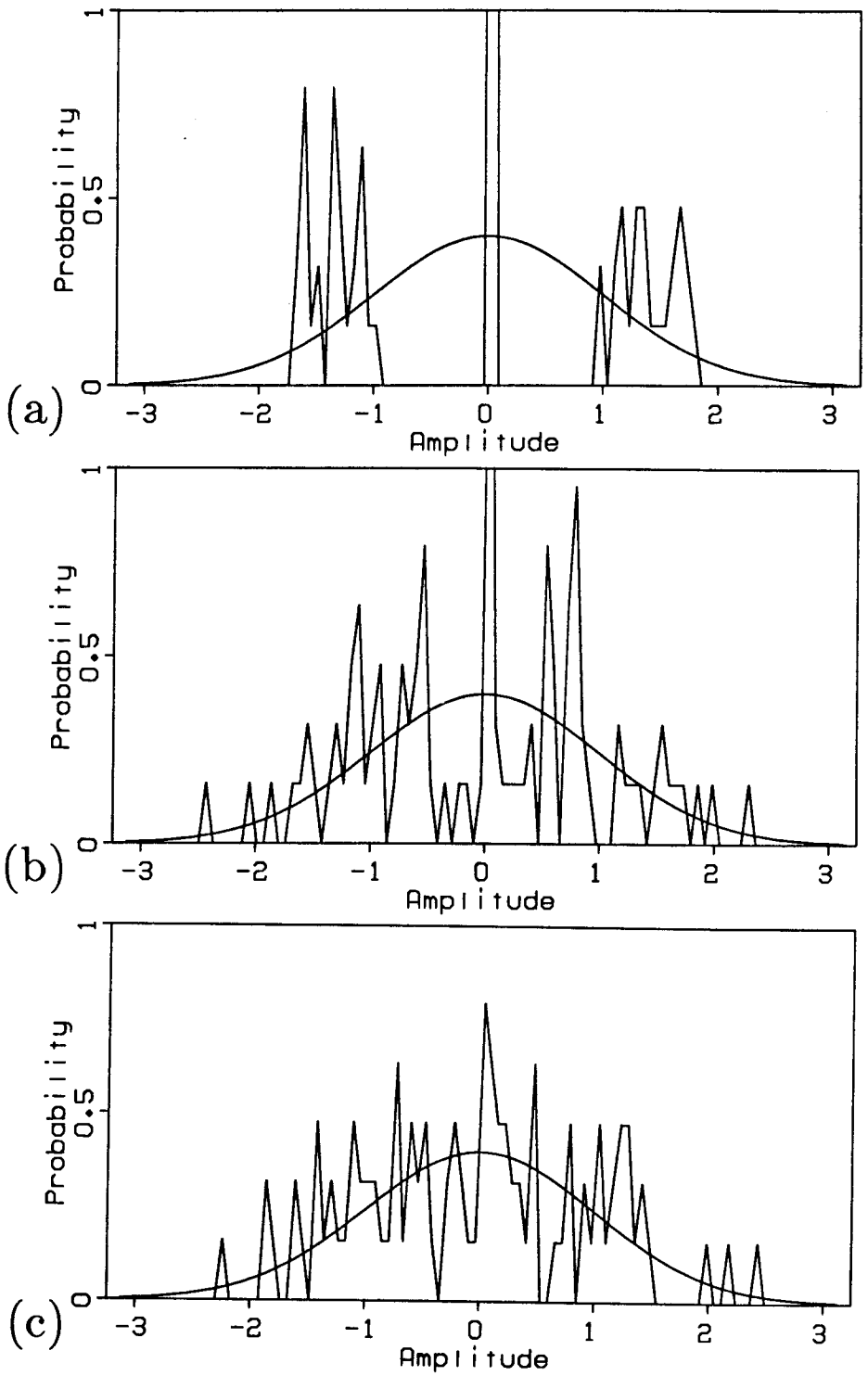

FIG. 2.3. (a) This non-Gaussian amplitude histogram is taken from the independent samples of Figure 2.2a. The best-fitting Gaussian distribution is plotted as well. (b) After one convolution the array of Figure 2.2b gives this histogram, visibly more Gaussian. (c) After a second convolution the array of Figure 2.2c shows a very Gaussian distribution. Linearly transforming independent variables creates increasingly Gaussian variables; inversion of these transformations recovers the non-Gaussianity.

estimated pdf's change slowly over spatial dimensions, and perhaps over other dimensions as well. Because of this stationarity, a histogram prepared from a great many samples with identical marginal pdf's will describe the possibilities open to them all.

In review, the central-limit theorem requires that a linear transformation of *independent* random variables make the corresponding marginal pdf's more Gaussian, that is, more like the Gaussian or normal distribution. By the contra-positive of this theorem, focusing an array with a linear transform must make the marginal pdf's more non-Gaussian. Thus, a measure of the non-Gaussianity of an array in turn measures the increase in focusing.

I derive a measure of focusing in Appendix D, equation (D.8b). This function uses Kullback's cross entropy (Kullback, 1959) to measure the divergence of an amplitude histogram from the best-fitting Gaussian. This function measures the likelihood that a histogram was taken from stationary, Gaussian random variables.

In general, noise must be removed before the focusing of signal can be measured numerically. A linear transform that focuses signal also defocuses noise, as we saw in Figure 2.1. Noise, by our working definition, has no *recognizable* spatial coherence, though it may have coherence over time. Noise includes all events that the chosen transform will not focus. Note that coherent noise such as multiples, side-swipe, ground-roll, and cable noise can require wave and geologic models similar to those used for signal. A linear transform must add coherence to previously incoherent noise and thereby increase the Gaussianity of the noise. Thus, applying a linear transform to noise will decrease the focusing measure. Only after the removal of noise can the transform that best focuses signal also maximize the focusing measure.

Sometimes one does not know how to perturb all model parameters so that the data is better described. Some parameters, such as seismic velocities, must remain undetermined. A measure of focusing can identify the values that best focus other parameters.

2.3. EXTRACTING DIFFRACTIONS FOR VELOCITY INFORMATION

Common-offset sections span greater distances on the surface and so detect deep velocity changes better than do shot and midpoint gathers. Diffractions contain all velocity information in common-offset sections; noise and undiffracted reflections must be removed before velocities are estimated.

The hyperbolic stacking algorithms of Chapter 1 make many common assumptions that greatly limit the resolution of seismic velocities. Midpoint stacks require

that reflections be from flat, featureless beds. Dipping beds, diffracting points, and beds with high curvature create reflections with misleading curvatures. When examining an event in a single midpoint gather, one cannot distinguish the curvature due to structure from the curvature due to wave velocities. Moreover, the range of recorded offsets is usually small compared to the depth of reflecting beds. With too small an aperture (the ratio of the offset to the depth of the reflector), the midpoint stack cannot recognize the curvature of hyperbolic reflections with enough accuracy to obtain useful seismic velocities.

A common-offset section displays seismic data as a function of midpoint and time for a fixed shot and geophone offset (see section 1.2.1). The range of recorded midpoints is much larger than the range of offsets, so common-offset sections can potentially resolve deep changes in velocity much better than common-shot or midpoint gathers. Structure-dependent coherence is easily distinguished from purely velocity-dependent coherence. Reflections from flat, featureless beds remain flat and featureless in offset sections if interval velocities are not too irregular. Such undiffracted reflections do not contain any velocity information over midpoint. Point scatterers, bed truncations, and beds with rough texture create diffracted reflections that have high curvature over midpoint. Such diffractions contain all of the velocity information present in constant-offset sections.

I shall use the zero-offset section (the simplest common-offset section) to illustrate the isolation of diffractions from other events. In these sections, diffracted reflections become simple hyperbolas over midpoint. In Chapter 1, by eliminating the offset dimension from field data, the NMO stack created a pseudo-zero-offset section. In this chapter, I shall treat a stacked section as a zero-offset section. When the aperture is small about zero-offset, the stacked section is almost identical to the zero-offset section; moreover the choice of stacking velocities affects the character of stacked events very little. If the laterally adaptable stack of section 1.4 is used, the aperture about zero offset can be kept small.

2.3.1. A strategy for isolating diffractions

As in Chapter 1, we should first propose a simple model for the useful reflections. A zero-offset section can be said to contain two types of coherent events that overlap very little: 1) reflections from laterally continuous and smooth beds; 2) diffractions from sharp lateral changes in reflection coefficients, including rough textures on beds, truncations of beds by faults and salt domes, etc. These two components rarely blur into one another: even if beds have a rough texture, they tend to have a modest overall

curvature. (Horizontal reflections on the earth's surface are an ambiguous case, but uninteresting.) We must first choose transforms that model both coherent reflections, then we can attempt a reliable separation of the two.

Let us assume that the seismic source is impulsive and linearize our reflected wavefield with respect to perturbed earth parameters: velocities, impedances, bulk moduli, rock densities, reflection coefficients, etc. The linearized reflected wavefields can be made to appear very similar for any parametrization of the wave equation. Different choices of earth parameters might add minor phase shifts to reflected waves but will not affect the number or geometry of the waves. Many common inversion schemes concern themselves entirely with the geometry of such perturbations and do not attempt to adjust the phase according to any particular physical parameter. Such perturbations can be referred to by a generic name, such as "reflectivity."

All linearly scattered wavefields can be expressed as a sum of waves reflected from single anomalous points in the earth (described analytically by Green's functions). Such points radiate spherical wavefronts as though they were buried sources and add appropriately delayed travel-times that account for the path between the true source and the point scatterer. When sources and geophones have the same location (zero-offset), the recorded reflections have double the travel time one would expect from the corresponding buried sources. The zero-offset model corresponds to the common description of "exploding reflectors" (Claerbout, 1985): one can imagine that the reflecting structures began producing seismic waves at time zero and that these waves traveled up to the surface at half the usual velocity.

Having accepted the assumptions of the exploding reflectors model, we can use the procedure called migration to invert reflectivity from zero-offset or stacked data. I shall use the frequency-domain algorithm defined by Stolt (1978) (see Appendix F). If a point scatterer stands alone, the scattered wavefield is spherical and is recorded at the surface as a hyperbolic function of midpoint and time. Figure 2.4a shows a zero-offset section that contains a synthetic diffraction hyperbola from a constant velocity medium. When this section is migrated with the correct velocity of 1, the corresponding earth image is a single point (Figure 2.4b).

When point scatterers fall on a continuous structure (a smooth, reflecting horizon), the scattered spherical wavefronts merge to form a continuous wavefield, which must show less curvature over midpoint and time than waves scattered from single points. Figure 2.5a shows the sum of diffraction hyperbolas that have been shifted horizontally over all negative midpoints. These hyperbolas cancel each other out except

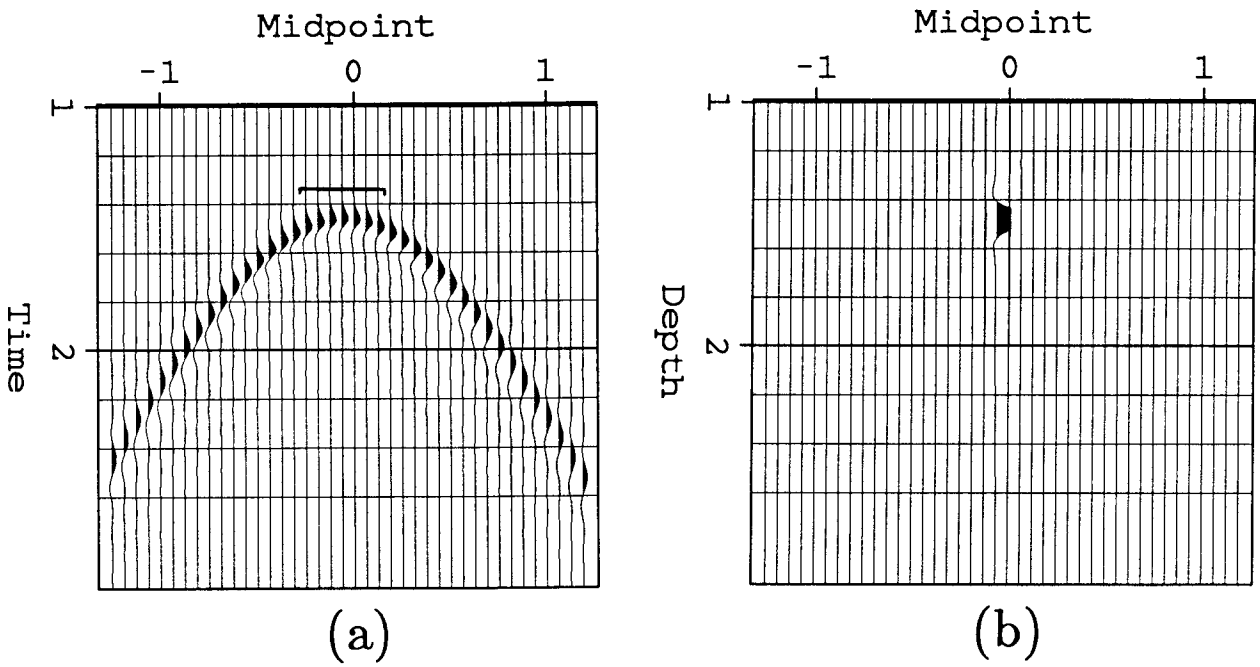

FIG. 2.4. (a) The Fresnel zone of a diffraction hyperbola is the interval over which adjacent hyperbolas might be expected to sum constructively. At midpoint 0 of this zero-offset section, the peak of the hyperbola appears to have nearly the same phase over approximately seven traces. (b) Migration of this section at the correct velocity of 1 focuses the diffraction completely; a spike is created at the location of a corresponding point scatterer.

along the line of their peaks and at the truncation of the sequence at midpoint 0, where a phase-shifted diffraction is visible. Migration of this time section reveals a line segment at constant depth (Figure 2.5b). Because the depth section was laterally coherent, hyperbola tails disappeared from the time section.

The curvature of reflected waves, or lack of it, distinguishes diffractions from undiffracted, smooth-bed reflections. Because the waves have a finite bandwidth, adjacent hyperbolas add constructively when they are separated by a limited range of midpoints (called the Fresnel zone). At midpoint 0, directly over the diffraction of Figure 2.4a, this Fresnel zone is about seven traces wide. Over this area, the hyperbola appears roughly linear. (Other narrow portions of the hyperbola add constructively when the reflecting bed is dipping. The tails add constructively only when the wave is traveling horizontally.) If a recorded reflection appears linear for a distance greater than the Fresnel zone, then it must correspond to the sum of more than one point scatterer—that is, to a coherent geologic structure, such as in Figure 2.5b.

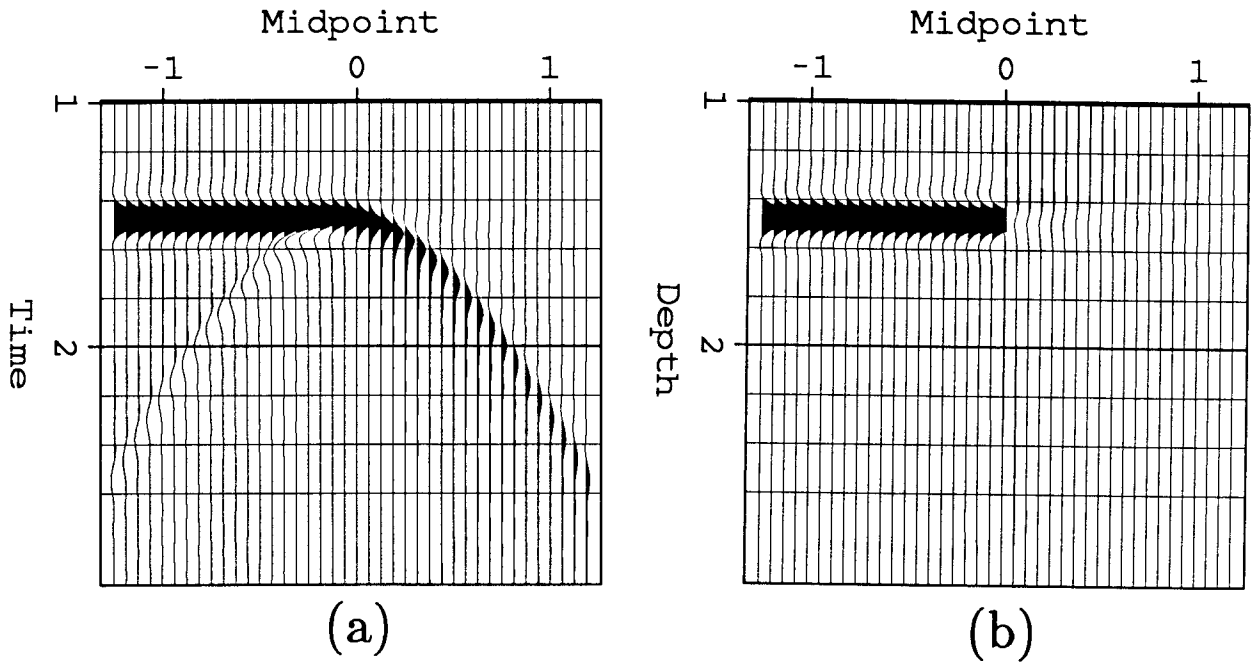

FIG. 2.5. (a) I sum diffraction hyperbolas with peaks at all negative midpoints. The hyperbolas sum constructively at their peaks, where their Fresnel zones overlap. The tails of the diffraction hyperbolas interfere destructively except at the truncation of the sum at midpoint 0. (b) Migration (at the correct velocity) maps the time section into a truncated line segment. We see that the undiffracted portions of reflections can nevertheless be described by a summation of diffraction hyperbolas.

Let us then define a model of the data as a sum of diffraction hyperbolas and of undiffracted, continuous reflections. Diffraction hyperbolas can be modeled from point scatterers with a Stolt, frequency-domain wave extrapolation. For a window of stacked data, let us describe undiffracted reflections by sums of tapered lines segments, with widths greater than the Fresnel zone of point scatterers. Thus, it will be impossible to express diffractions with the model for undiffracted reflections. Though all reflections can be expressed as a sum of diffraction hyperbolas, line segments will describe the bed reflections much more efficiently, without the unnecessary hyperbola tails. Let us estimate undiffracted bed reflections first since they tend to dominate reflection data. Ideally, we should like to invert for all the necessary parameters at once; but because the models overlap so little, we can avoid the expense and invert for the two types of coherent events individually.

2.3.2. Separating bed reflections from diffractions and noise

We have seen that undiffracted bed reflections can appear linear over a spatial distance greater than the Fresnel zone; diffractions and noise cannot. Let us describe undiffracted reflections as a sum of line segments that are too wide to describe diffractions or noise. A collection of line segments can be modeled by statistically independent parameters: their slopes and time intercepts. The statistical methods of Chapter 1 will allow us to extract those components of the data that are locally linear and that are unlikely to be a superposition of noise.

I shall begin with a simple modeling equation to describe dipping lines that extend across the data.

$$\text{data}(x\,,t\,) = \int \text{model}(p\,,\tau = t\, - px\,)dp \tag{2.1}$$

The data are assumed to be a function of time t and a spatial parameter x. (See Appendix E on slant stacks.) This transformation maps points in the model to lines with dip p and intercept times of τ at $x = 0$. The lines extend across the entire section. Because undiffracted reflections can have more curvature than this model will allow, this model should not be used for very large windows of seismic data; the window should only be large enough to exceed the Fresnel zone of point scatterers and to include most of a diffraction. I shall discuss later how, for larger windows of data, smooth beds can be described adaptably as a sum of short line segments (see section 2.4).

Unlike the NMO stack, the least-squares inverse of equation (2.1) can easily be calculated explicitly. The least-squares inverse equals the adjoint transform followed by a *rho* filter:

$$\text{model}_{LS}\,(p\,,\tau) = \int \text{data}(x\,,t\, = \tau + px\,)dx \;\; * \;\; \text{rho}(\tau) \quad . \tag{2.2}$$

rho(τ) is the inverse Fourier transform of the absolute value of the frequency. This particular inverse and *rho* filter were first derived by Thorson (1978, 1984). Claerbout (1985) first defined equation (2.2) as the slant stack but did not include the *rho* filter. I include the *rho* filter in the definition so that its corresponding model will map points into lines.

Slant stacks map lines of constant dip into points, and thereby focus bed reflections that are roughly linear over this interval. The mapping will not create perfect points because reflections are not linear over an infinite range of midpoints. Slant stacks can be said to remove the first-order predictability of gently changing reflections. If a reflection shows some curvature, then the corresponding model detail

will show coherence. Events with rapidly changing dips, such as noise and diffractions, do not focus. A single point in the data maps roughly to a line; the slant stack thereby defocuses incoherent noise, diffuses its energy, and increases its Gaussianity.

For example Figure 2.6a shows a hypothetical stacked section containing a single linear dipping event (signal) and an single spike (noise). After a slant stack is performed (Figure 2.6b), the line has mapped roughly to a point, and the point to a line. The point is not perfectly focused because of the limited number of midpoints. Nevertheless, the slant stack has effectively reduced the number of samples required to describe the signal and has increased the number for the noise. Moreover, samples containing signal are much stronger than those containing noise; much of the noise has mapped outside the range of dips used in the slant stack.

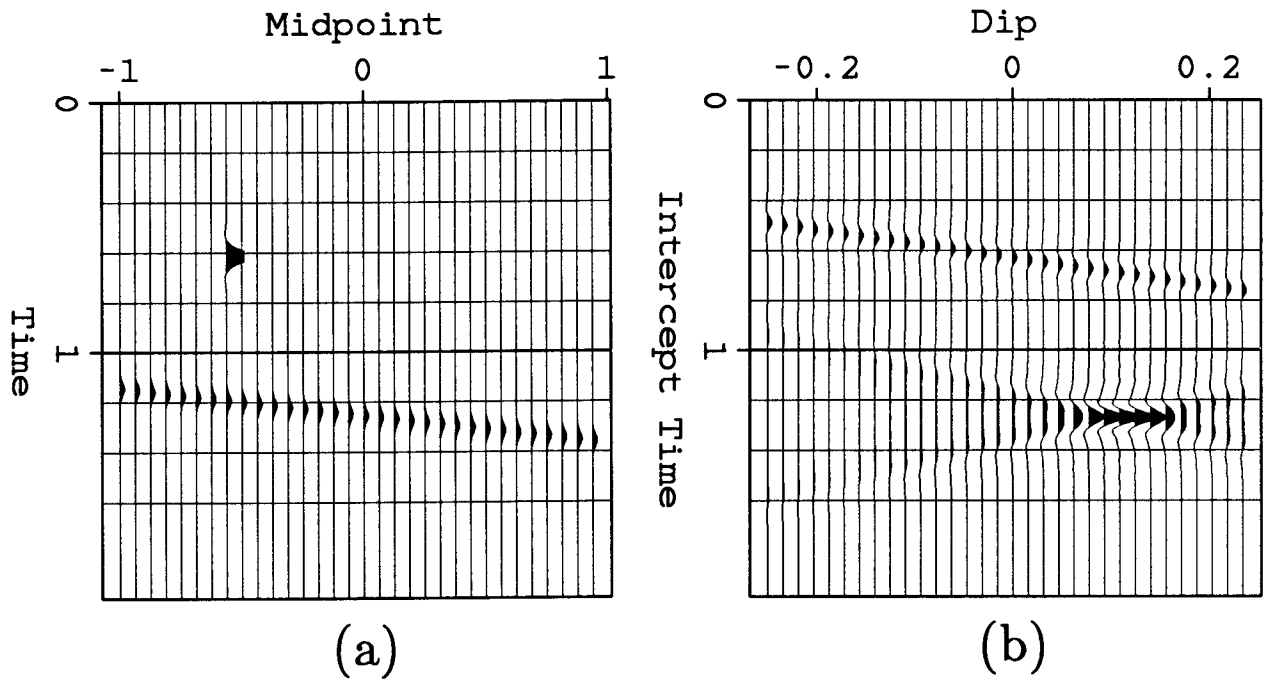

FIG. 2.6. (a) A zero-offset section contains a dipping line (signal) and a point (noise). (b) A slant stack focuses the line into to a slightly smeared point, and defocuses the point into a line. Again, the signal has become more non-Gaussian, and the noise less so.

Slant stacks also diffuse events with high curvature, such as hyperbolas. Figure 2.7a adds a diffraction hyperbola to the time section of Figure 2.6a. A slant stack (Figure 2.7b) diffuses the hyperbola into a nearly linear event, just like the slant stack of a point. Samples containing energy from the diffraction are even weaker than before. Most of the energy of the transformed hyperbola is outside the range of dips used.

If we are willing to assume that reflections of smooth beds are linear over a substantially greater range of midpoints than are noise or diffractions, then the slant stack can be used as a focusing transformation to isolate these smooth reflections. The results of Chapter 1 (derived in Appendices A, B and C) apply directly to the estimation and removal of undiffracted reflections. A new modeling transform changes only the definitions of signal and noise. The extraction algorithm follows:

1. Slant-stack the data and estimate the data pdf $p_{d'}(x)$ from histograms.
2. Slant-stack data whose spatial coherence has been destroyed and estimate the noise pdf $p_{n'}(x)$ from histograms. Deconvolve histograms for $p_{s'}(x)$.
3. Calculate the Bayesian estimate and reliability of signal in each sample of the transformed data. Set to zero those samples containing significant noise. (Preserve wavelets as in section 1.7.)
4. Reconstruct the data from the extracted signal and subtract the reconstruction from the original data.

In practice, a full slant stack is not necessary for step 2. Summing should be performed at enough dip values to provide the local statistics of noise. The noisy data set can be created by several methods—random time shifts of traces up and down, random reverses in trace polarity, random numbers taken from histograms, etc.

The above algorithm successfully extracted reflections from a small window of stacked, offshore-Texas data (Figure 2.8). This data is stacked over offset so that it appears like a zero-offset section. The window contains a linear series of beds, interrupted by a growth fault with weak diffractions off the truncations. The window is small enough that the reflections show little curvature, but the window is large enough to span the diffractions. The global slant stack of equation (2.2) describes these reflections adequately. Larger windows demand that laterally adaptable slant stacks be used; these will be discussed in section 2.4.1.

Reflections were extracted over a wide range of dips ($\pm$0.12ms/m); this range is much larger than necessary to describe the linear events. Figure 2.9 shows the most reliable linear reflections. When these reflections were subtracted from the data, the

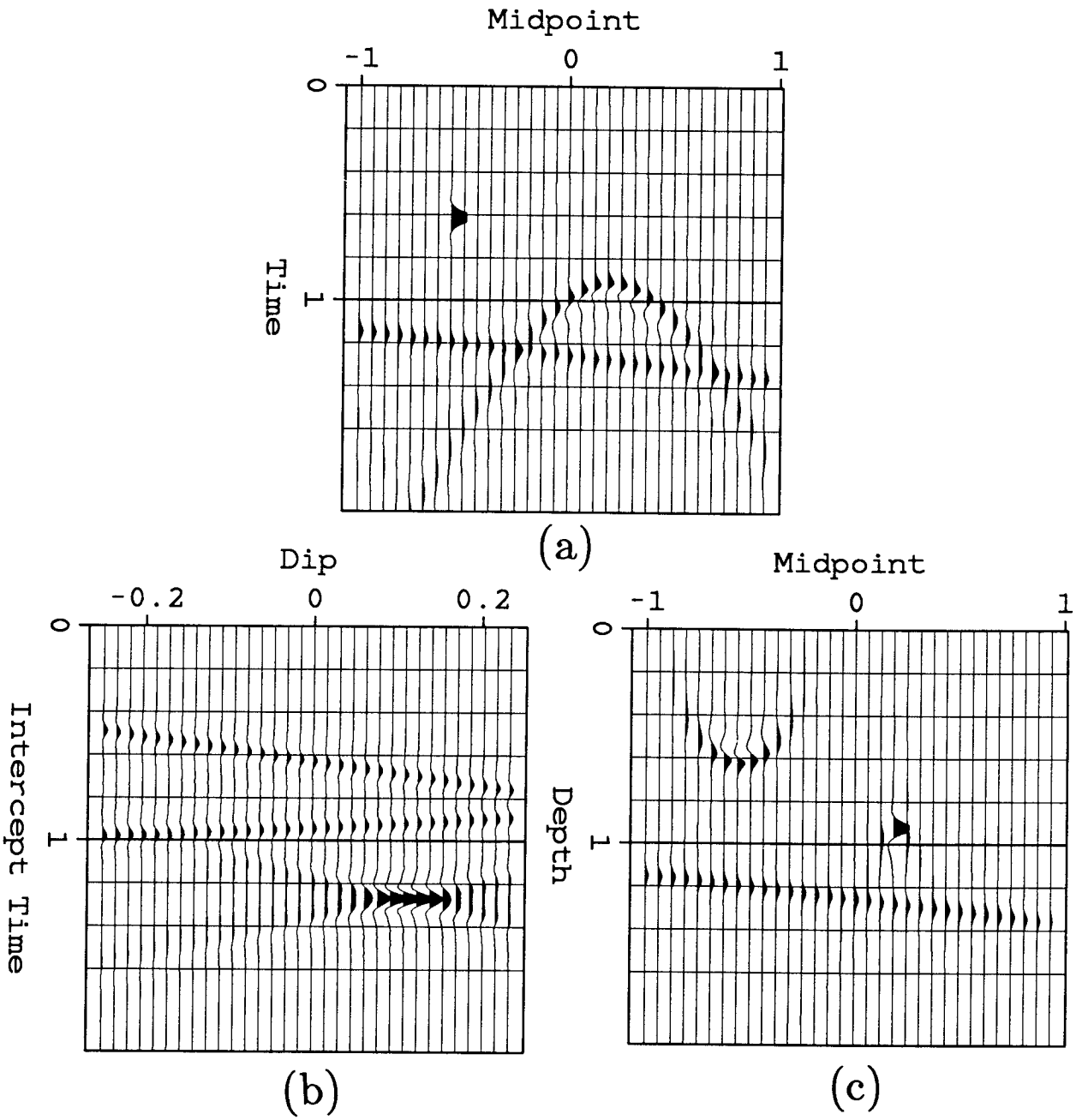

FIG. 2.7. (a) This zero-offset section combines the signal and noise of Figures 2.1a and 2.6a: a diffraction hyperbola, a dipping line, and a point. (b) A slant stack of 2.7a includes the features of Figure 2.6b plus a weak curve from the hyperbola. The slant stack treats only the dipping line as signal and defocuses the hyperbola and the point. (c) Migration of 2.7a maps the dipping line into another dipping line; the line is neither focused nor defocused. This experiment suggests that undiffracted events should be extracted before either diffractions or non-Gaussian noise are extracted.

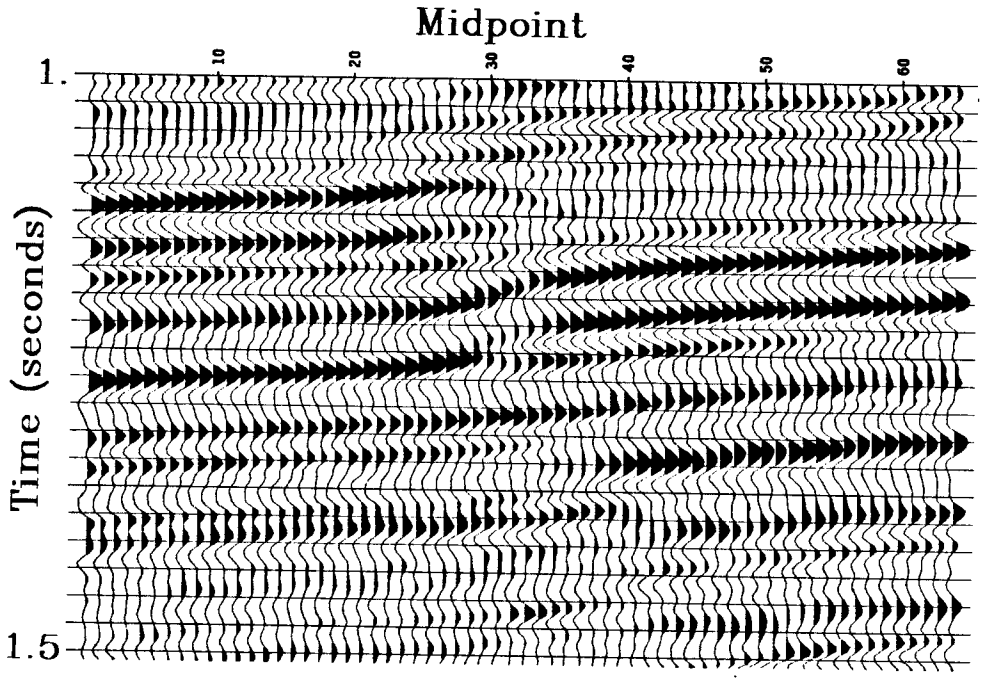

FIG. 2.8. A window of stacked, offshore-Texas data contains a growth fault, with weak diffractions off truncated beds. Smooth-bed reflections and noise obscure the velocity information in diffractions. I shall decompose this window into three components, Figures 2.9, 2.15, and 2.16, which add up to the original data. The extracted components are scaled differently for plotting. Data were supplied by Western Geophysical. This window extends 2 km at 33m sampling.

diffractions and noise were exposed (Figure 2.10). The slant-stack model included dips present in the diffractions; however, these dips changed too rapidly to be described as a sum of lines. The range of dips did avoid the linear asymptotes of the diffraction hyperbolas. Isolated asymptotic dips result only from waves reflected horizontally from scatterers on the surface—and so are trivial and uninteresting features. The unextracted residuals will be used in the following section as the data for estimation of hyperbolic diffractions.

2.3.3. Modeling isolated diffractions

Migration, as defined by Claerbout (1985), Stolt (1978), and others, successfully focuses diffraction hyperbolas in stacked or zero-offset sections, and diffuses noise. However, undiffracted reflections obscure the focusing of diffractions and must be removed first. Let us return to Figure 2.7a, a zero-offset section that contains a single

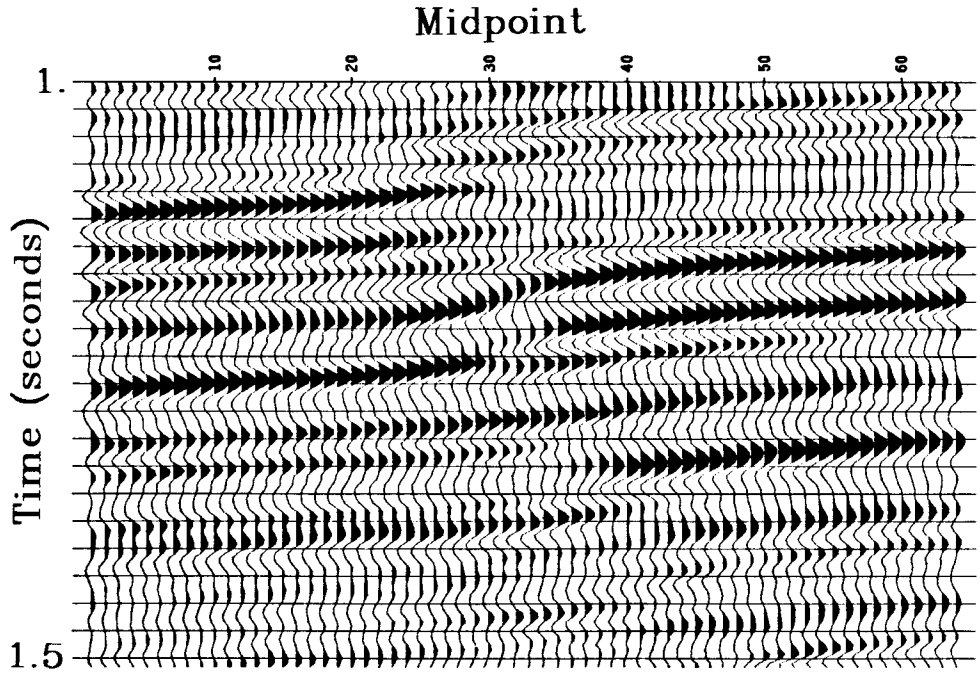

FIG. 2.9. The undiffracted reflections in Figure 2.8 are easily described as a superposition of lines. To create this section, I apply a slant stack, extract samples that are little contaminated by noise, and reconstruct the data.

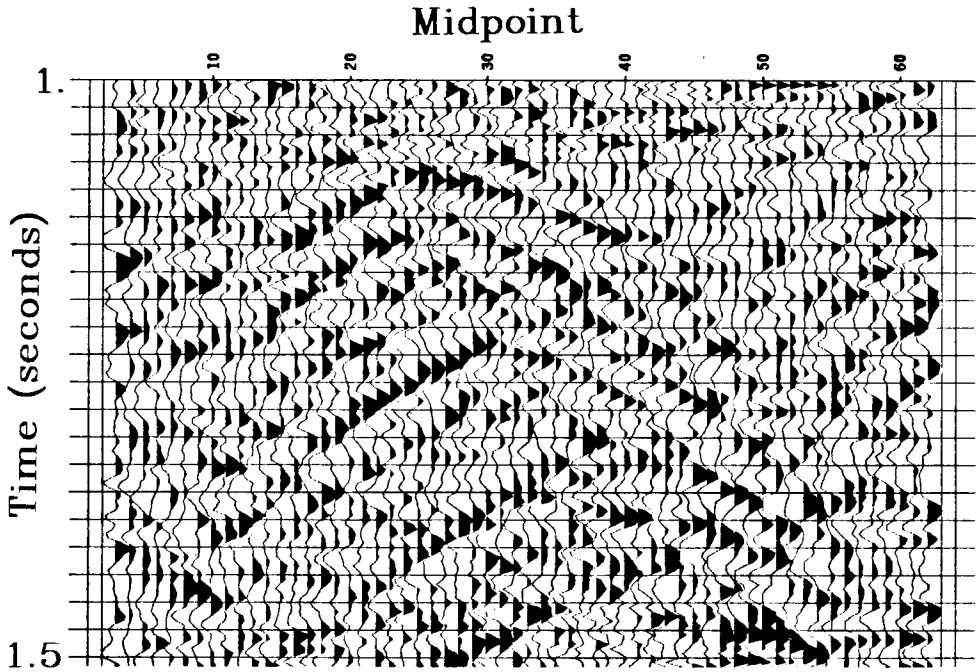

FIG. 2.10. A subtraction of undiffracted reflections from data (Figure 2.8 minus 2.9) leaves diffractions and noise; neither component is easily described as a superposition of lines.

point, a dipping line, and a diffraction hyperbola. Migration of this section (Figure 2.7c) maps the point into an ellipse, the hyperbola into a point, and the line into another line. The signal (the hyperbola) focuses, and the incoherent noise (the point) defocuses; but the coherent noise (the line) does not change its coherence or its Gaussianity. If the section contained many such undiffracted reflections, as do the data of Figure 2.8, then the statistics of individual samples would be dominated by this unchanging modulation, and the focusing measure would remain constant after migration.

Again, diffractions that result from the rough texture of reflecting beds can be isolated for useful velocity information, but only if the undiffracted portions of reflections are first suppressed. For example, the synthetic zero-offset section of Figure 2.5a contains a flat reflection from a flat, continuous bed and a diffraction from an abrupt truncation of the bed at midpoint 0. The migrated section of Figure 2.5b shows the corresponding depth section, a truncated linear bed. The time section represents the essential features of the data recorded over a growth fault in Figure 2.8. The recorded data, however, contains so many reflections that the diffraction hyperbolas are almost completely hidden beneath the linear events.

We wish to isolate the diffraction, but we cannot completely remove the undiffracted portion of the reflection without distorting the velocity information in the diffraction. For example let us subtract the truncated time event of Figure 2.11a from the data in Figure 2.5a. This subtraction isolates the diffraction hyperbola in Figure 2.12a. Note that the phase of the diffraction changes sign at the peak; no trace remains of the undiffracted portion of the reflection. However, after migration at the correct velocity of 1, this diffraction hyperbola diffuses into an ellipse. The explanation is that the truncated time event of Figure 2.11a, which we subtracted from the original time section, did not correspond to a simple depth section, but rather to the migrated section of Figure 2.11b. This migrated section contains a flat, continuous bed as before, but it also contains an ellipse. A subtraction of the truncated event from the original time section is equivalent to a subtraction of their corresponding migrations. Only the ellipse remains.

Alternatively, if undiffracted reflections are extracted without sharp changes in amplitude over midpoint, then the isolated diffractions will focus at the correct velocity after migration has been performed. If the undiffracted reflections are described as a sum of dipping lines as in the previous section (or of tapered line segments as in a later section), then the extracted reflections will appear smooth over midpoint. Figure 2.13a shows such an extraction: the reflection amplitude does not truncate abruptly at

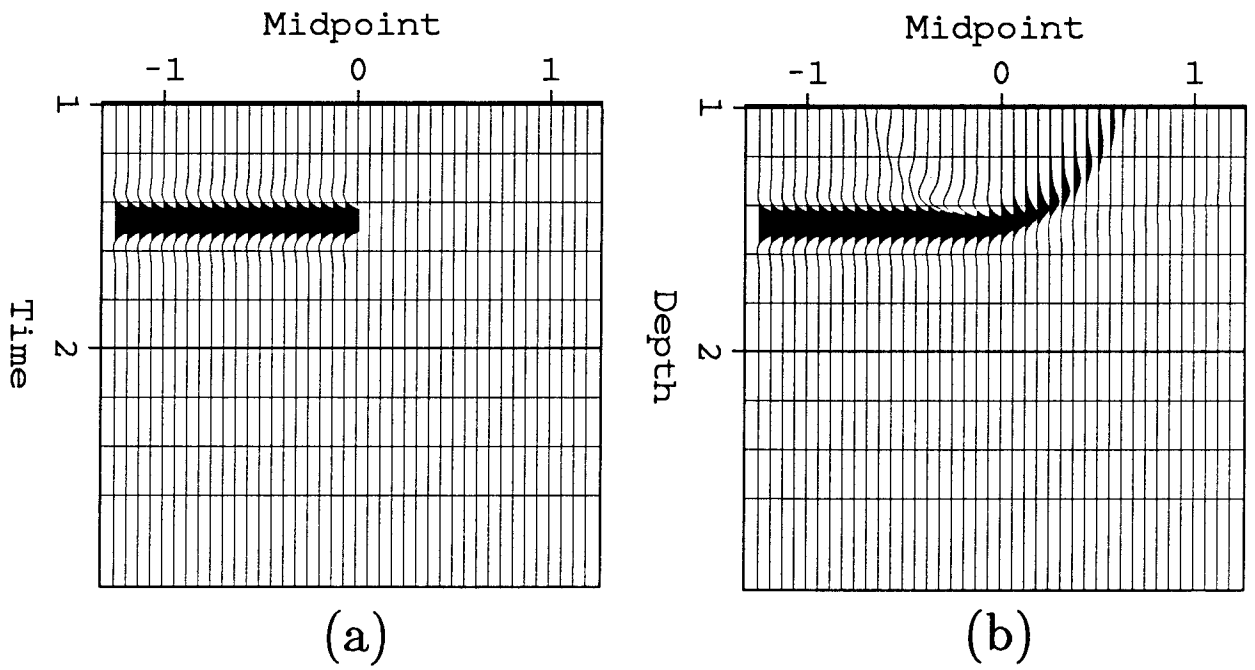

FIG. 2.11. (a) This time section contains a truncated time event corresponding to the undiffracted portion of the time section in Figure 2.5a. (b) This truncated time event migrates (at the correct velocity) into a non-physical depth section with an elliptical feature at the edge.

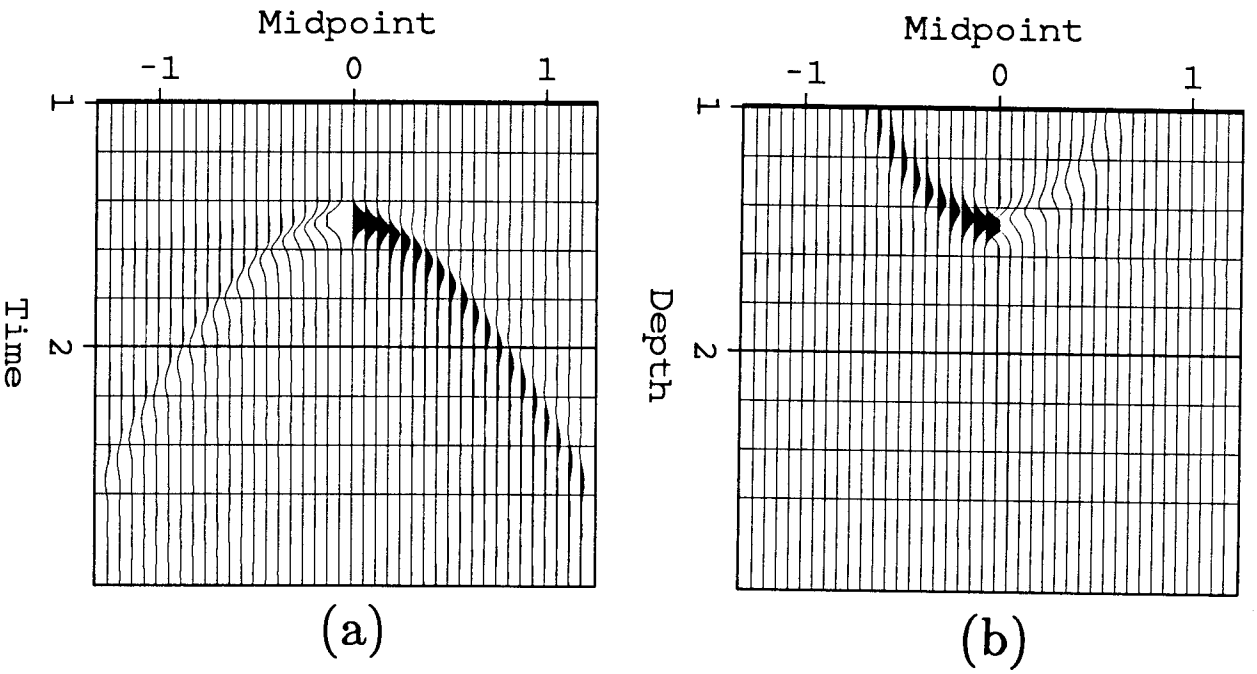

FIG. 2.12. (a) I subtract the truncated time event of Figure 2.11a from the original time section of Figure 2.5a. The isolated diffraction shows a phase reversal at midpoint 0. (b) Migration (also at the correct velocity) does not focus the diffraction, but gives an elliptical event, equal to the subtraction of Figure 2.11b from Figure 2.5b.

midpoint 0, but tapers smoothly to zero. Migration of this time section changes the coherence negligibly (Figure 2.13b); the edge does not disperse into an elliptical feature as did the truncated time event of Figure 2.11a. After subtraction of this smooth, extracted time event from the original time section, the isolated diffraction in Figure 2.14a appears little more complicated than after the subtraction in Figure 2.12a of the truncated event. Figure 2.14b shows the migration of the newly isolated diffraction; the diffraction has focused almost perfectly into two slightly blurred points. Migration at too high a velocity would map these points into small ellipses; and at too low a velocity, into hyperbolas. The isolated diffractions now unambiguously indicate the best migration velocity.

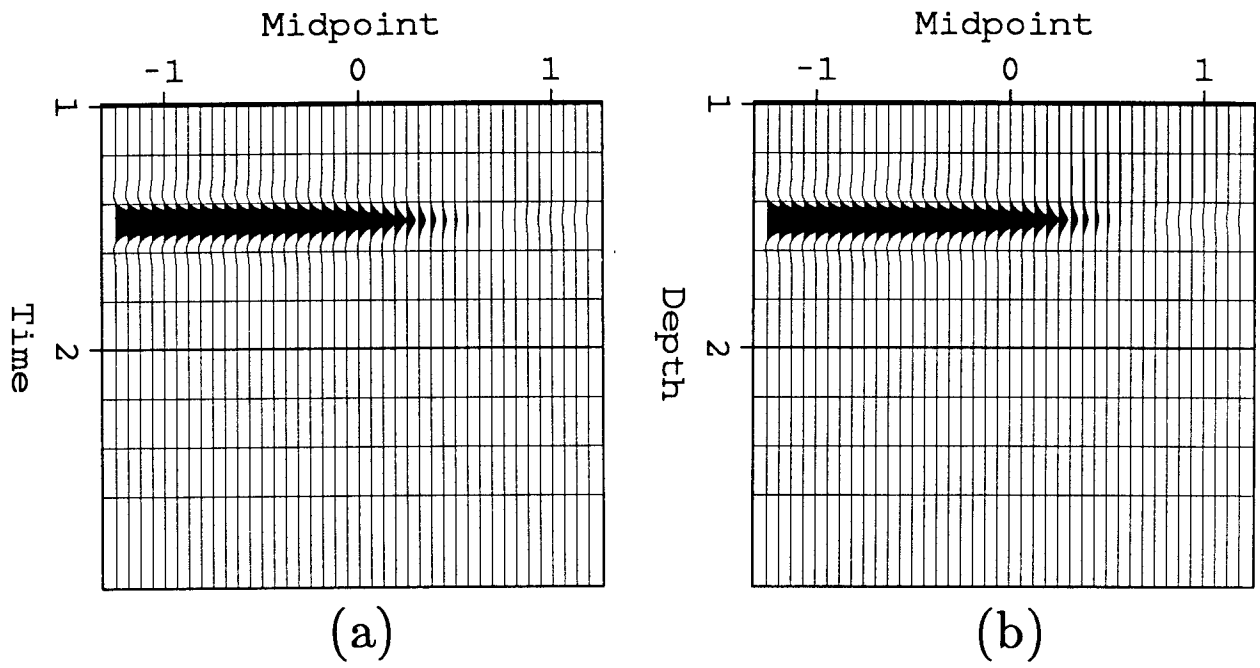

FIG. 2.13. (a) This time section displays the reflection that would result from a laterally smooth extraction of the undiffracted portion of the original time section of Figure 2.5a. (b) Migration creates a depth section without elliptical features; this section is little different, in fact, from the time section.

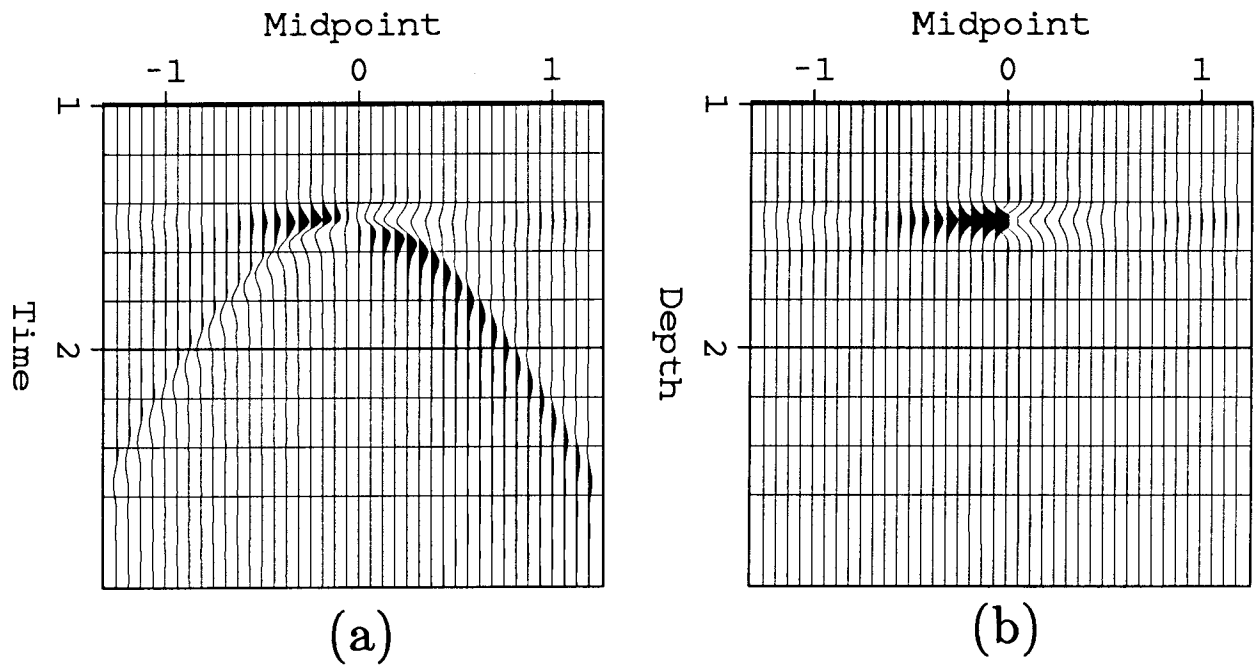

FIG. 2.14. (a) This section subtracts the smooth time event of Figure 2.13a from the original section of Figure 2.5a. The resulting isolated diffraction is little more complicated than the isolated diffraction of Figure 2.12. (b) Migration (at the correct velocity) focuses this diffraction into two slightly blurred points. A migration at an incorrect velocity would defocus the points into hyperbolas or ellipses.

2.3.4. Extracting diffractions from noise

Noise must also be eliminated from stacked or zero-offset sections before we can measure the focusing of diffractions by migration. For example, the synthetic zero-offset section in Figure 2.1a contains a diffraction hyperbola (signal) and a single point (incoherent noise). Migration at the correct velocity of 1 focuses the hyperbola into a point but defocuses the point into an ellipse. The focusing measure derived in Appendix D measures only the overall change in non-Gaussianity—an increase in non-Gaussianity from the migration of signal can be canceled by a decrease from the noise. If the time section contained only diffractions, then the best migration would maximize the focusing measure. Let us then extract the events that are most easily described as a sum of diffraction hyperbolas, that are unlikely to be the superposition of incoherent noise.

Unfortunately, migration depends on the undetermined velocity, and migration is the transformation that focuses diffractions best. As before, we should prefer to make the data a simple function of some model array. For example

$$\mathbf{data} = \sum_v \mathbf{F}_v \, \mathbf{model}_v \tag{2.3}$$

where v is the diffraction velocity. Let $\mathbf{F}_v$ be the modeling transform for which the least-squares inverse $\mathbf{F}_v^{-1}$ is the migration (the focusing transform). $\mathbf{F}_v$ then is the transform that models the events diffracted at the velocity v. (See Appendix F for definitions.) Because the inverted model has increased in size by the number of velocities, a full least-squares inverse of equation (2.3) is expensive to calculate.

Alternatively, we can obtain a satisfactory result with an approximate inverse. Let us extract signal over a range of migration velocities, linearly combine modeled events, and find the parameter value maximizing the focusing measure. This procedure both extracts the most reliable signal and determines the best transform.

To estimate the diffractions present at a particular velocity, extract all signal focused at some v and then model.

$$\mathbf{e}_v \equiv \mathbf{F}_v \cdot \text{Extract}\{\mathbf{F}_v^{-1} \cdot \mathbf{data}\} \tag{2.4}$$

Using the estimate $\mathbf{e}_v$ for a physical range of v, we can find the least-squares sum best resembling the data:

$$\hat{\mathbf{s}} = \sum_v a_v \, \mathbf{e}_v \;; \quad \min_{\mathbf{a}} \| \hat{\mathbf{s}} - \mathbf{data} \|^2 \rightarrow \sum_w a_w (\mathbf{e}_v \cdot \mathbf{e}_w) = (\mathbf{e}_v \cdot \mathbf{data}). \tag{2.5}$$

The dots designate simple scalar products. Set the scalar product equal to zero where the data are unrecorded, so that these values are unconstrained. To solve (2.5) for all coefficients a_v and for $\hat{\mathbf{s}}$, invert a small symmetric matrix, whose order equals the number of velocities used. $\hat{\mathbf{s}}$ contains the most reliable signal, without bias, for the chosen range of v. Transforming $\hat{\mathbf{s}}$ over this same range will maximize the focusing measure at the v focusing the signal best.

Let us summarize the extraction of diffractions from noise. Use previous methods to estimate the noise and signal pdf's.

1. Migrate the residual data (without undiffracted reflections, as in Figure 2.10) over a physical range of velocities.
2. Set to zero those samples containing significant noise in each migrated section.
3. Reconstruct data from each migrated section.
4. Find the least-squares superposition of the reconstructed data that best resembles the residual data (Figure 2.15).
5. Migrate this superposition over the previous range of velocities.

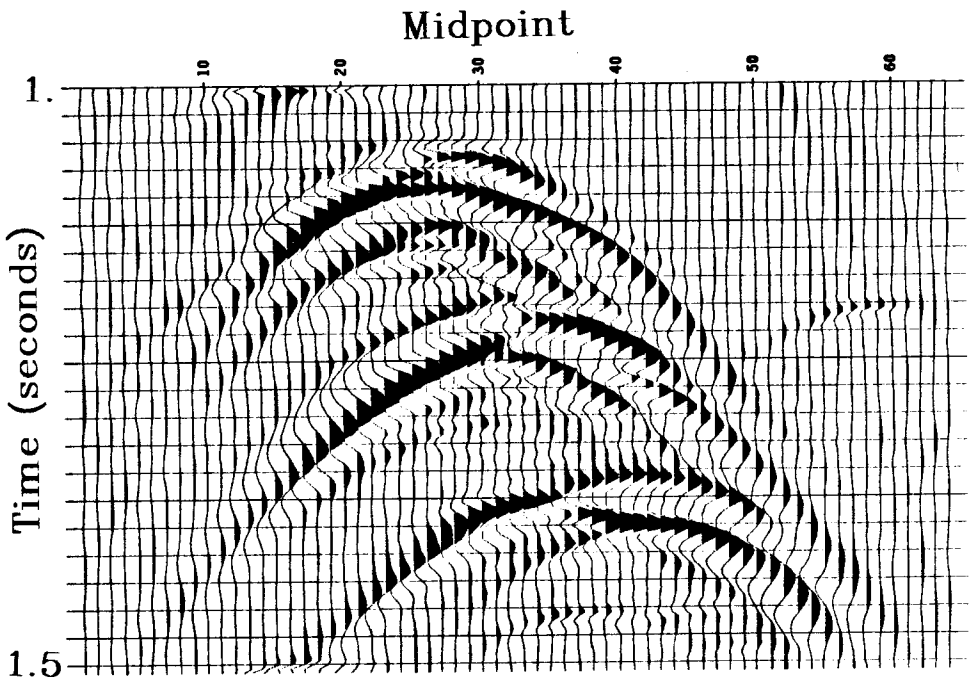

FIG. 2.15. Diffractions are easily described as a superposition of diffraction hyperbolas. I migrate the diffractions and noise (Figure 2.10) over a range of velocities, extract signal, and reconstruct diffraction hyperbolas for each velocity. These extracted diffractions are linearly combined for the best least-squares fit to the diffractions and noise of Figure 2.10.

6. Determine the best migration velocity by evaluating the focusing measure of each migration.

This algorithm requires that all diffractions within the window correspond the same unknown velocity v. To allow for vertical changes in velocities, one could begin with upper windows and then, to remove the effects of their interval velocities, extrapolate later windows (downward-continuation). Velocities can be estimated as locally as the density of diffractions permits. To allow the rms velocities of extracted diffractions to vary both vertically and laterally, use the modifications of Appendix G.

The diffractions of Figure 2.15 were extracted from the hyperbolas and noise of Figure 2.10. (The diffractions and noise were amplified for plotting.) Notice that phase changes appear at peaks of the hyperbolas, as expected from truncated beds. When these diffractions were subtracted from the original data, the incoherent, Gaussian noise in Figure 2.16 was revealed. Little coherence is visible in the central region. Diffraction tails from a nearby growth fault are visible on the left side.

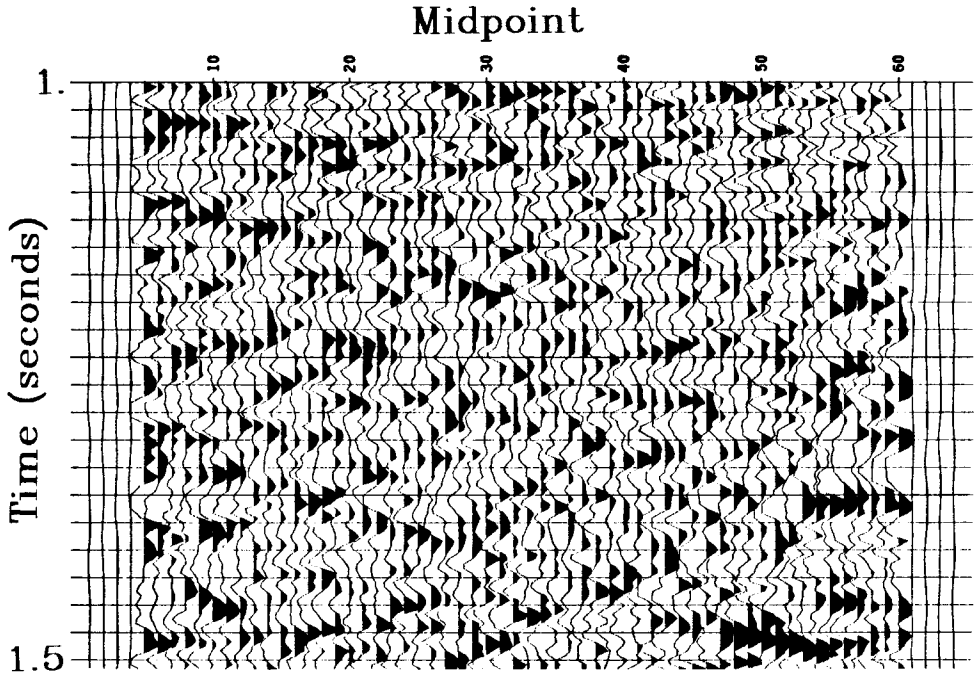

FIG. 2.16. Noise (equal to Figure 2.10 minus 2.15) cannot be simplified by slant stacks or migration. Some diffraction tails are visible from a growth fault to the left of this window.

When these diffractions are migrated, the focusing measure is maximized at the best velocity (Figure 2.17). This focusing measure did not depend on the range of velocities used in the extraction. Migration of the diffractions and noise of Figure 2.10 had no effect on the focusing measure, nor on the varimax function used by De Vries and Berkhout (1982).

In the following section I shall explore ways in which the methods of this section can be generalized to larger quantities of data. A laterally adaptable slant stack will allow us to model undiffracted reflections with high curvature. With this new model and an appropriately defined migration, seismic velocities can be estimated from unstacked data, for all recorded offsets.

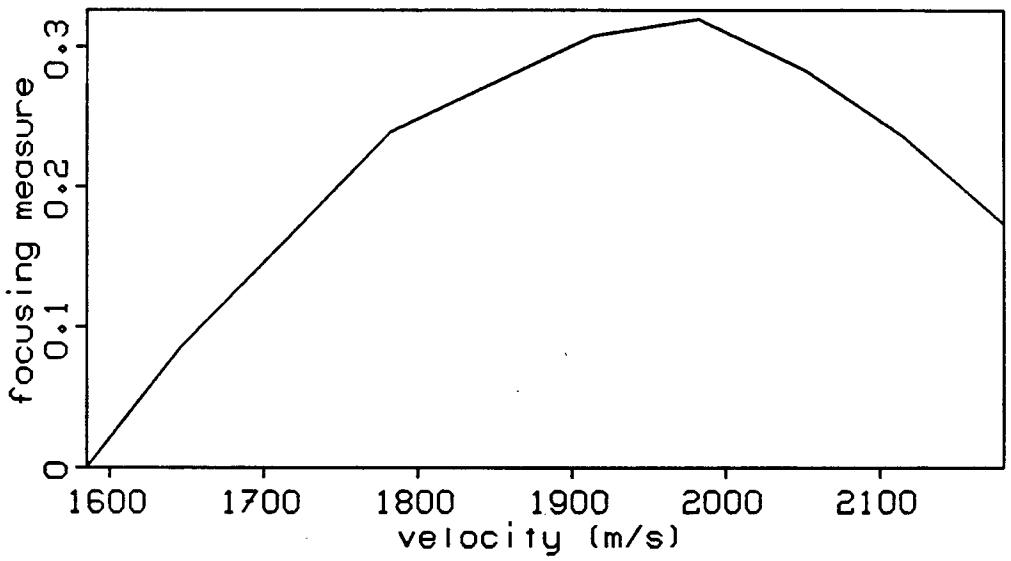

FIG. 2.17. Migration of the extracted diffractions in Figure 2.10 maximizes the focusing measure at the best velocity.

2.4. GENERALIZATIONS OF REFLECTION MODELS

I shall now relax some of the physical constraints of the previous section. A local slant stack will allow locally linear reflections to be extracted in a laterally adaptable fashion. The extraction algorithm of the previous section can then extract undiffracted reflections over larger windows of data. The methods of the previous section also generalize to unstacked data. By incorporating both slant stacks and migration, velocity analyses of common-midpoint sections can avoid noise and truncation problems and can image dipping beds. Finally, a treatment of static shifts will show how some hidden signal coherence can be recovered.

2.4.1. Laterally adaptable reflections and local slant stacks

The global slant stack of the previous section will distinguish laterally continuous reflections from noise only if the window of data is small or if the reflections are very flat. In some data, background noise can be very strong and Gaussian, as in Figure 2.18. The non-Gaussian noise in Figure 1.21 of Chapter 1 could be largely extracted without harming the surrounding signal, but Gaussian noise appears equally Gaussian after linear transformation and cannot be recognizably focused or isolated. In the data of Figure 2.18, the visible reflections change rapidly in coherence over only a few traces.

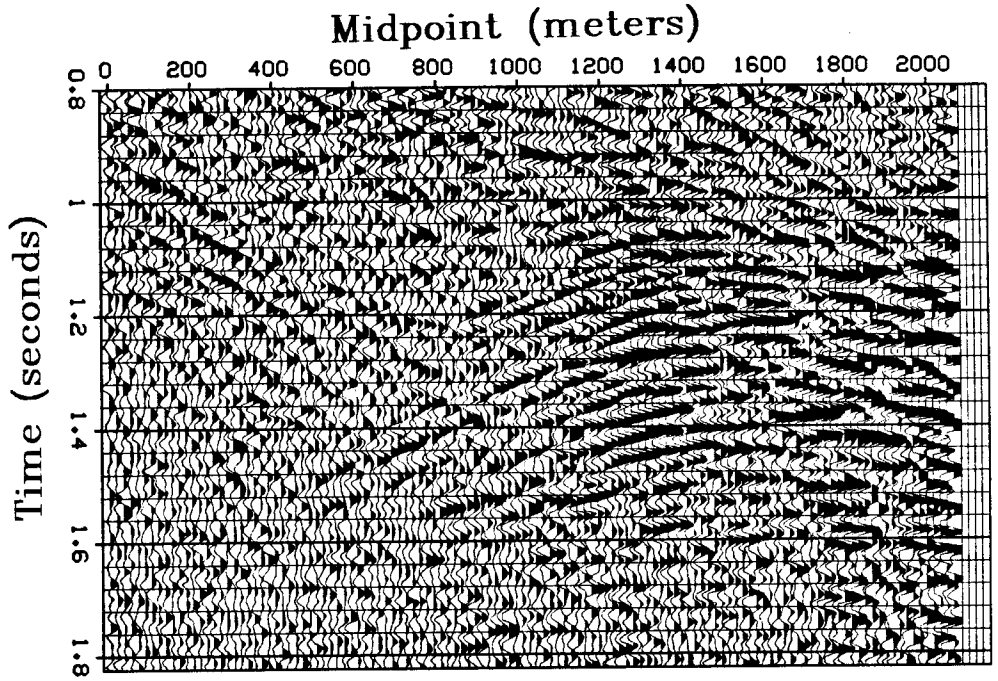

FIG. 2.18. A window of stacked data contains weak, coherent reflections, much obscured by strong Gaussian noise. (Data courtesy of Trans-Pacific Geothermal.)

One might ask how much of this coherence could be trusted and how much could be accidental. Because the noise would be difficult to suppress, let us aim only to extract the most reliable signal. First, we must define a more flexible model of the coherent reflections.

Let us make the slant stack laterally adaptable as for the hyperbolic stacks in section 1.4. Global slant stacks require signal to be expressed as a sum of lines extending across the section—an assumption that produces artifacts or prevents extraction when the reflections are shorter. Let a local slant stack specify reflections as a sum of short, tapered line segments of all dips. Define the data as the transformation of a narrow cube: $\text{model}_1(p\ ,x_c\ ,\tau)$.

$$\text{data}(x\ ,t\ ) \equiv \int W\left(\frac{x-x_c}{X_w}\right) \int \text{model}_1[p\ ,x_c\ ,\tau=t-p\ (x-x_c\ )]\,dp\,dx_c \tag{2.6}$$

$W(x\ )$ represents a windowing function—in the simplest case, a rectangle function as in equation (D.7). x_c fixes the center of the window, and X_w the width. p is the slope of the short linear event, and τ the vertical time intercept at $x=x_c$. Each point maps into a line that is tapered in a window about x_c.

Inverting this transformation explicitly, without iteration, is difficult. Instead let us perturb the model with the adjoint of (2.6), plus a *rho* filter.

$$\text{model}_2(p\ ,x_c\ ,\tau) \equiv \{\int W\,(\frac{x-x_c}{X_w})\ \text{data}[x\ ,t=\tau+p\,(x-x_c\,)]\,dx\ \}\ *\ rho\,(\tau) \qquad (2.7)$$

For an explanation of the convolution by a *rho* filter, see Appendix E on slant stacks. The adjoint plus the *rho* filter solves most of the least-squares inversion of equation (2.6). For a more accurate inverse, the residual uninverted data can be transformed by equation (2.7) and added to the previous model.

Each frame of $\mathbf{model}_2$ with constant x_c represents the ordinary inverse slant stack of a tapered window of the full data. Shifts are applied to place the zero offset in the center of the window. The forward transformation needs only invert the above for the center trace of each window. Thus,

$$\text{data}(x\ ,t\,) = \int \text{model}_2(p\ ,\ x_c=x\ ,\ \tau=t\,)\ dp \quad . \qquad (2.8)$$

Equation (2.7) decomposes the array into planes containing narrow ranges of dip. Adding these planes together as in (2.8) reconstructs the original data. Rieber (1936) first began applying similar local stacks, called sonograms, to unstacked seismic data in the 1930's. I merely present a simple invertible expression for these stacks.

The width X_w of the window specifies the number of traces over which an event should be linear. A natural uncertainty principle results: lateral resolution must be balanced with dip resolution. A larger X_w will improve the dip resolution of genuinely linear events. But wider windows can see greater curvature, which will increase the number of p values needed to describe an event. At some critical X_w, dip resolution will no longer improve for events with some curvature. For diffractions this critical X_w corresponds to the Fresnel zone. Undiffracted reflections can be separated from diffractions by setting X_w larger than this Fresnel zone.

A few simplifications can reduce the cost of the local slant stack. For a window of few traces, an $x-t$ implementation becomes less expensive than the Fourier implementation. The loop over p should be outermost, with sums over various windows and intercepts inside. The *rho* filter can be performed first. If the weighting function W is a rectangle, then forward transformation requires only linear moveout of the entire section and horizontal sums over narrow windows of traces. The sum for a new window can be calculated from an adjacent window by the addition of one trace and the subtraction of another (cf. Robinson and Robbins, 1978). The local transformation becomes only twice as expensive as the equivalent global $x-t$ slant stack.

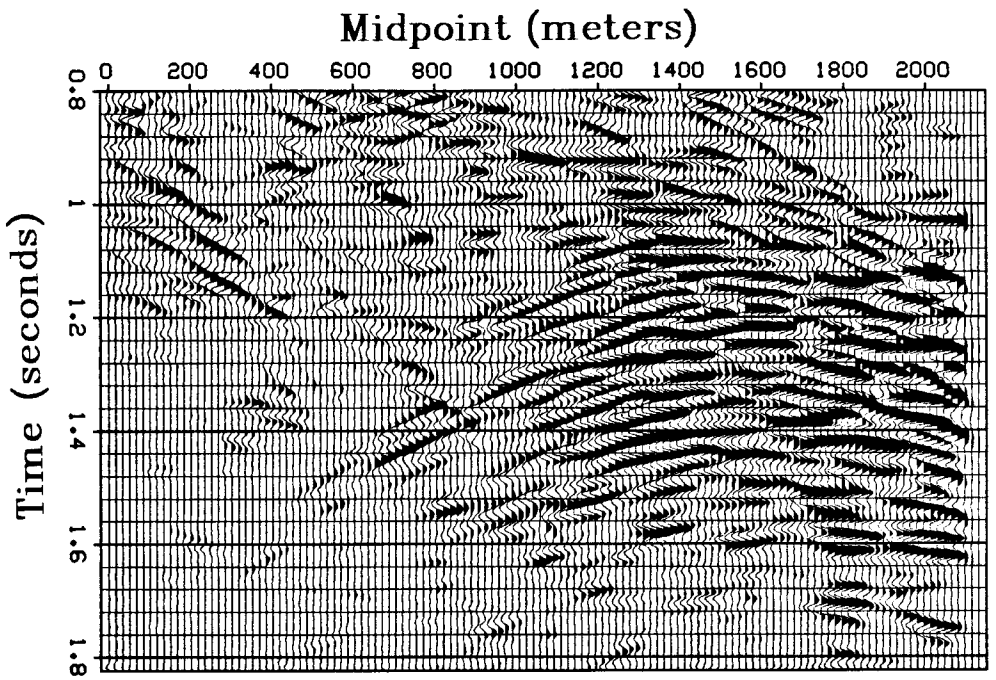

FIG. 2.19. This section contains the data component of Figure 2.18 that is most easily expressed as a sum of short, dipping line segments—that signal focused best by a local slant stack.

The great advantage of a local slant stack lies in its resistance to artifacts. Lateral adaptability prevents the oversimplification of focused reflections and prevents the straightening out or extension of their inverses after extraction. This transform will easily extract events with high curvature, such as shallow diffractions, if X_w is small enough, say 4-8 traces for 50 m sampling. If one wishes to avoid diffractions, as in the previous section, then X_w should be larger than the Fresnel zone of the diffraction hyperbolas.

From the data window in Figure 2.18, I extract events showing significant coherence over at least four traces (Figure 2.19). I took X_w equal to four traces and $W(x) = \exp(-\pi x^2)$ in equation (2.7). The extracted events must contain some noise, but one can assert that the percentage and likelihood are low. I subtract the most reliable signal from the data (Figure 2.20). The chosen range of p values excludes that of highly dipping diffraction tails that appear in the upper left corner of Figure 2.18 and throughout Figure 2.20. Any remaining coherence at lower dips appears very unreliable to the eye. In this way, extractions with different ranges of slopes can also separate overlapping coherent reflections and remove unwanted coherent events.

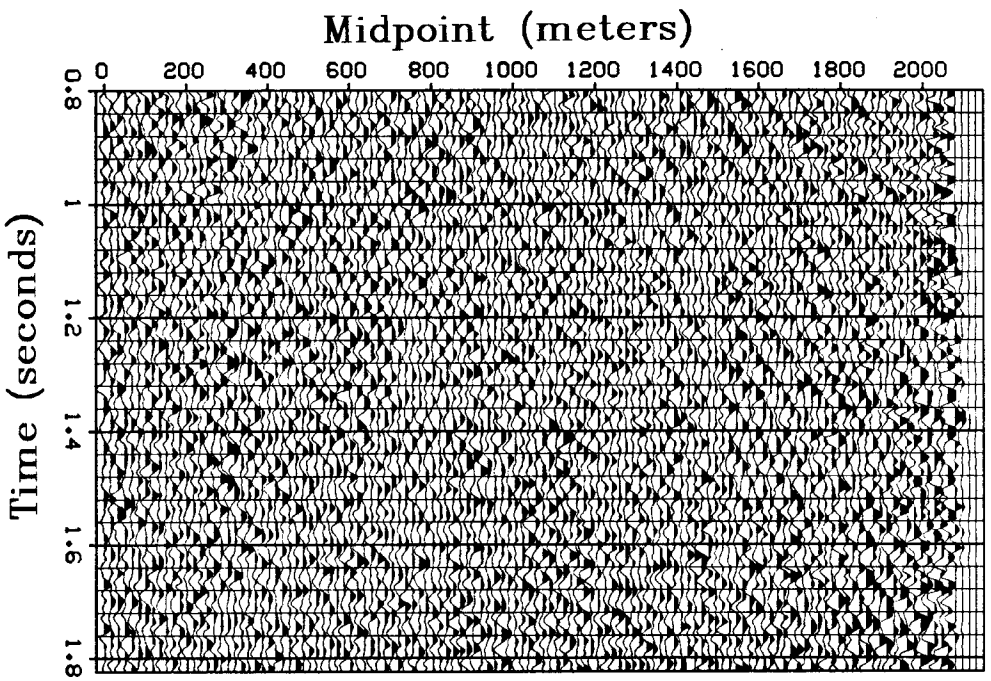

FIG. 2.20. I subtract Figure 2.19 from 2.18. The chosen range of p values excluded highly dipping diffraction tails, which now appear with the incoherent noise. (All panels are scaled the same.)

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

$$\text{model}_3(p\,,p'\,,x_c'\,,\tau') \equiv$$

$$\{\int\int W'\left(\frac{x-x_c'}{X_w'}\right)\,\text{model}_2[p\,,x_c=x\,,\tau=\tau'+(p+p')(x-x_c')]\,dx\,\}\;*\;\text{rho}(\tau')\;. \quad (2.9)$$

This transformation, defined in terms of (2.7), detects linear, spatial changes in the slope of reflections—that is, the curvature. This more general transformation increases the focusing of curved events for a fixed window width. Alternatively, the same amount of focusing can be achieved from wider windows. Summing planes over p and p' inverts the transformation.

Local extractions of coherent events considerably improve the appearance of stacked midpoint gathers. Stacking is generally expected to improve the signal-to-noise ratio; however, analysts of noisy data often notice that reducing the number of traces in the stack improves the result. They follow a simple guide: *do not stack beyond the coherent width of a reflection.* Otherwise, samples containing only noise will add to the

sum. As signal reduces to the widths of a few traces, of course, it becomes indistinguishable from random alignments of noise. A local extraction should not make signal wider or narrower than is statistically reliable. A stack of locally extracted signal will then be superfluous for the reduction of noise. For example, rather than stack a cube of data over offset (the other two dimensions being midpoint and time, or shotpoint and time), one could slice the extracted signal at constant offsets and search for changes in amplitude and waveform. Such changes can give much stratigraphic information.

A closely related process, interpolation, attempts to fill in missing data intervals in a way consistent with known data. One naturally prefers the simplest possible description of the missing events. Interpolating a wide region of the data requires wider modeled events than for narrow regions. Events that are extrapolated far from the known data must be modeled with increasingly simple models. The window parameter X_w should increase as the extrapolated distances increase.

Laterally adaptable modeling transformations are most useful when extracted signal must be interpreted directly. This situation arises when the noise is strong and Gaussian and cannot be reliably extracted, or when the signal must be interpolated. The laterally adaptable stack avoids systematic over-simplifications of the coherence of signal. By choosing a width for the stacking window correctly, an interpreter can maximize either the flexibility of the modeled coherence or the redundancy that will discriminate against noise.

2.4.2. Wave-equation velocity analyses

Wave-equation velocity analyses of common-midpoint gathers are easily plagued by numerical artifacts. The narrow recorded range of offsets truncates reflections abruptly. Much non-Gaussian noise, including glitches, dead or overamplified traces, and static shifts, appears in gathers. A wave-equation migration can image and focus signal, but truncations and non-Gaussian noise defocus into meaningless coherent events that overwhelm the eye. The truncation of reflections at the maximum offset acts like severe non-Gaussian noise containing the negative of the reasonable reflections one expects. Large amounts of data would seem necessary so that the truncation artifacts could be avoided. Finally, without information orthogonal to the midpoint gather, a migration must assume that reflections have no dip, as do NMO velocity analyses; estimated velocities suffer enormously.

Let us tackle each problem with tools already discussed. A signal/noise separation can avoid noise artifacts and allow the focusing measure to estimate velocities. A local slant stack can detect the dip of reflections along midpoint; 4 to 8 adjacent midpoint gathers would suffice. A dip-dependent migration can recognize velocity-dependent coherence over offset. An algorithm incorporating these tools follows.

1. Using the local slant stack (2.7), decompose data that is a function of (y, t, h) into reflections dipping differently over midpoint: $\text{model}_2(p_y, y_c, t, h)$, where p_y is the dip over midpoint, y_c the midpoint, h the offset, and t the time at which the reflection intercepts the gather (equivalent to τ).
2. Extract samples containing a high percentage of signal with a high probability.
3. For each p_y, use the dispersion relation (F.4) in Appendix F to migrate the corresponding gather.
4. For various velocities and dips extract those reflections focused by migration.
5 Find the least-squares superposition of these reflections best resembling the gather before extraction. Use the depth-variable versions (G.1) and (G.2) in Appendix G.
6. Use the focusing measure to estimate migration velocities locally.

If the least-squares superposition does not constrain the missing tails of reflections, then only known offsets will affect the estimation of velocities. If too much signal is replaced by truncation noise, however, focusing will be weak, particularly for deep reflections with little curvature. A rough compensation for the moveout of reflections would makes the simplification of (G.3) equally effective.

Rieber's Controlled Directional Receptivity (1936) first used many of the principles exploited here. Russian seismologists, particularly Dr. Rjabinkin and Boris Zavalishin (pers. comm.), have made his work the basis of their standard processing. Sonograms (local slant stacks without *rho* filters) assign an angle to the energy arriving at a given time in a narrow window of the data cube. Ray tracing finds the corresponding images. Refinements certainly exist, but imaged reflections usually do not make use of waveforms. They successfully image dipping structures in data too noisy for Western methods.

2.4.3. Correction of static shifts

Land data often contain traces shifted by uneven topography and variations in surface velocities. Almost all information remains in these traces; only the exact zero time is uncertain. Yet if not corrected, these traces 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 most identifiable signal in the data by a method from previous sections. High-frequency static shifts must interfere as pure noise and cannot contribute to the estimate of signal. Cross-correlate original traces with the extracted 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 time shift that places events in the most coherent position. Traces can then be corrected, and the procedure repeated for second-order improvements.

2.5. CONCLUSIONS

In this chapter we found that surface seismic reflections show two distinguishable varieties of spatial coherence: one results from the spreading of propagating waves; the second results from the continuity of reflectors. The two types of coherence can be described by different mathematical models, with constraints that allow each to be estimated independently. By distinguishing the two types of coherence, one can isolate those reflections that are most useful for velocity analysis.

The techniques of Chapter 1 apply to different varieties of coherence if the modeling transforms obey certain statistical assumptions. The statistics of individual samples can be estimated from a stationary data array, but statistical dependencies between samples are usually underdetermined. Thus, the statistics of signal are most easily estimated when the data samples are made statistically independent ("focused") by an invertible transformation. A transformation that focuses signal necessarily defocuses noise. Because of the central-limit theorem, focusing makes signal more non-Gaussian and noise more Gaussian. Using a cross-entropy function, one can define a measure of non-Gaussianity and thereby of focusing.

As was done in Chapter 1, we can use Bayesian methods to estimate focused signal. Histograms of focused data estimate the necessary probability density functions for signal and noise. To extract reliable signal accept those focused samples having a low percentage of noise with a high probability.

To estimate velocities from diffractions in stacked CDP sections, first remove undiffracted reflections containing no velocity information. Slant-stack models can describe undiffracted events as a sum of dipping lines whose widths are greater than the Fresnel zone of diffractions. Slant stacks focus reflections from smooth, continuous beds but not reflections with large curvature, such as diffractions or noise. By making the slant stacks laterally adaptable, one can apply this model to large windows of data.

After removing undiffracted reflections from the data, one can separate diffractions and noise. Migration focuses diffractions best for unknown seismic velocities. Extract diffractions over a range of velocities and find the least-squares superposition best resembling the original data. The focusing measure shows which velocity focuses this superposition best. A spatially variable least-squares superposition allows velocities to be estimated locally.

A local slant stack allows the extraction of that signal that is most easily expressed as a sum of short dipping lines. I successfully extract weak reflections with laterally variable coherency from behind strong Gaussian noise. The signal must be extracted for the data to become more interpretable; Gaussian noise cannot be extracted. A more general transformation expresses reflections as a sum of short second-order curves.

The procedure used to extract diffractions can also be used for wave-equation stacks of field (shot) gathers and midpoint (CDP) gathers. Using local slant stacks in the orthogonal direction allows one to recognize various reflector dips and image them correctly in the stack. (Normal move-out does not.)

The expense of our specific applications depends only on the chosen linear transformations (such as Stolt migration and the slant stack), each repeated several times, but easily vectorizable. Statistical calculations manipulate one-dimensional histograms and contribute negligibly to the cost.

The statistical tools of Chapter 1 have now been successfully applied to a variety of data sets and applications. The assortment of modeling transformations introduced in this chapter provide us with a set of tools that can model and separate many more combinations of coherent and incoherent events. In the following chapter I shall review the definition and separation of signal and noise for a non-linear modeling equation, the acoustic wave equation.