Chapter 1
Robust velocity stacks: an illustration of signal/noise separation

1.1. OVERVIEW AND INTRODUCTION

To introduce the techniques of this thesis, I shall examine a familiar and simple geophysical process, the NMO or velocity stack, and reconstruct the procedure as an inversion and signal/noise separation. The NMO stack suppresses much noise and considerably reduces the number of samples in surface seismic data, hence its popularity. The NMO stack recognizes that many reflections appear hyperbolic in the raw data and can be described uniquely by three parameters: an amplitude, a seismic velocity, and the minimum reflecting time. Even if one assumes that the data are composed entirely of such reflections, the NMO stack fails in two important ways to describe the data accurately:

- Hyperbolic reflections cannot be accurately reconstructed from the stack.
- Once the data have been stacked, one cannot distinguish hyperbolic reflections (signal) from non-hyperbolic events (noise).

Both of these objections make interpretation of the stack more difficult. The first objection is addressed in section 1.2, the second in 1.3.

To solve the problem of reconstructing data, section 1.2 first defines equations that map single points to hyperbolas. The stack is then replaced by a linear, least-squares inversion of these equations, after Thorson (1984, 1985). This section reviews least-squares techniques in a form that will be useful elsewhere for more complicated models of seismic coherence. The least-squares stack reconstructs hyperbolic events very well. Because it is linear, however, the least-squares stack attempts to describe noise by a superposition of hyperbolic signal. The second objection must be answered in the following section.

Section 1.3 prepares statistical tools to distinguish signal and noise in the stacked model. To "extract" signal, the algorithm eliminates details of the stack that can be too easily created by noise. The errors in the estimates of signal and noise can be quantified with known reliability. Later sections and chapters will use this algorithm to separate different varieties of coherent signal and noise.

Section 1.4 develops modeling equations to describe the signal more flexibly than does the hyperbolic model. By modeling reflections as a superposition of hyperbola segments, these equations allow the signal to adapt spatially. Such improvements have wide applicability. The essential character of a reflection can remain the same but still show slow modulations over spatial dimensions. Using laterally adaptable stacks, one can extract much more coherent signal than previously, with little change in the interference from noise. Such stacks are well suited to the interpolation of irregularly sampled data, as discussed in section 1.6.

The algorithm used to extract hyperbolic signal is easily modified to extract non-hyperbolic noise. In section 1.5 I extract ground roll and other noise that is not easily described by the laterally adaptable stacks. This procedure applies to other modeling equations; for instance, tube-wave noise will be extracted from a vertical seismic profile in Chapter 3.

It is particularly difficult to model the coherence of seismic data over time. Seismic wavelets can appear very non-stationary, with little redundancy for signal/noise separation. Section 1.7 explains how the procedure of signal extraction can be modified so that the shapes of extracted wavelets are preserved.

Let us first turn to a simple example and see pictorially how a stack fails. Figure 1.1 contains synthetic data (a midpoint gather) from a hypothetical seismic survey. A thin spherical wave expands from a point source at the earth's surface, propagates through a constant velocity medium, reflects from a horizontal, flat layer, and returns to be recorded at the surface. The horizontal axis (offset) is the distance from the point source to the point at which the wave is recorded; the vertical axis is the time elapsed from the activation of the source. When the earth model contains only one flat bed, the data contain a single obvious feature: a spike of unit amplitude appears as a hyperbolic function of the offset.

The hyperbolic or velocity stack offers an alternative description of the hyperbolic reflections in the data. Each hyperbola can be uniquely described by its intercept time at zero offset and by the constant propagation velocity that determines the curvature of the hyperbola. This constant velocity must average the effect of other unknown

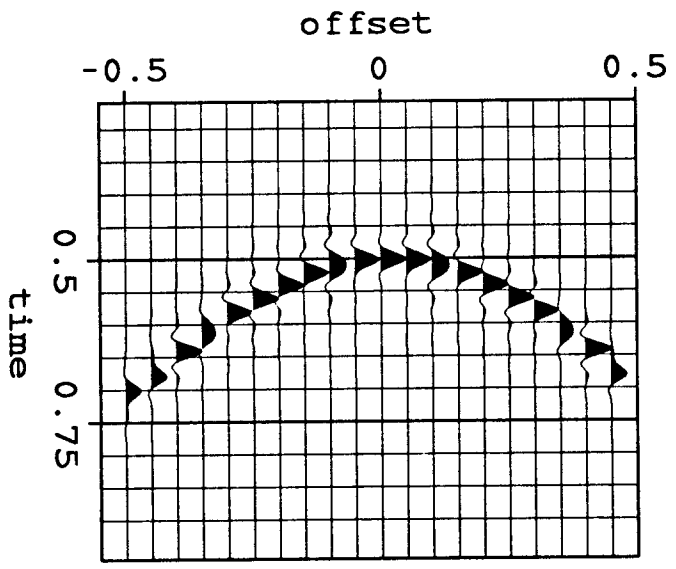


FIG. 1.1. A synthetic midpoint gather contains a single spike moved out along a hyperbolic path. A few parameters will describe this event: an amplitude of 1, a moveout velocity of 1, and a zero-offset intercept time of 0.5 . Sinc $(\sin[x\,]/x\,)$ interpolations were used; coarse sampling over time distorts the plotted waveforms but not the frequency content.

quantities: the dip and curvature of the reflecting horizon, the average and variability of the true propagation velocities. Reflections with different curvatures can arrive simultaneously, so the model must allow multiple velocities at all intercept times.

Each point of the hyperbolic stack ideally should describe a single hyperbola in the data. Figure 1.2 contains one such model of the data in Figure 1.1. A single spike of unit amplitude appears at an intercept time of 0.5 seconds and at a moveout velocity of 1. All other samples have zero amplitude, including those sampled at incorrect velocities, 0.75 and 1.25. This model appears simplest to an interpreter because only one sample appears at a non-zero value; the data are accurately described by a single hyperbola. Unfortunately, such a model cannot be expressed as a simple function of the data.

The NMO stack is the usual procedure to estimate a hyperbolic model. To prepare each point of the model, the NMO stack sums, or averages, over the corresponding hyperbolas. Figure 1.3a shows the result of stacking the synthetic data at three velocities. The largest amplitude appears at a coordinate giving the correct velocity and zero-offset travel time. The amplitudes are far from zero elsewhere,

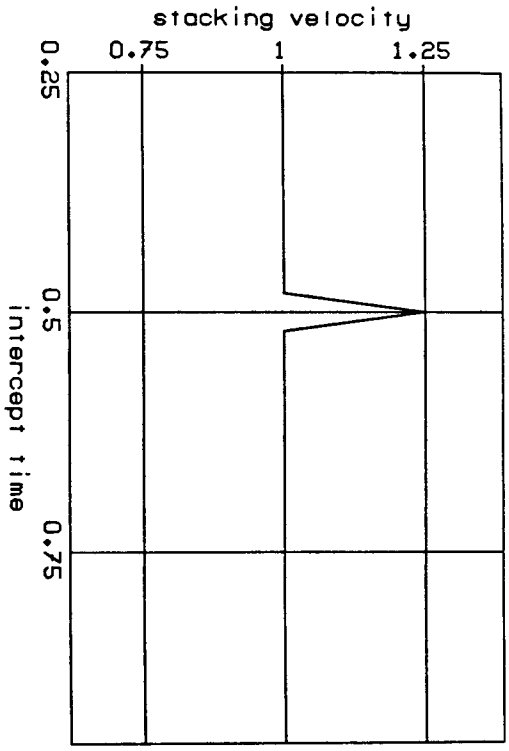


FIG. 1.2. Each sample of this model describes a unique hyperbola that could appear in the data of Figure 1.1. The single spike of unit amplitude describes a hyperbola with a zero-offset intercept time of 0.5 and a moveout velocity (stacking velocity) of 1. The moveout velocity is the constant seismic velocity that best explains the measured hyperbolic curvature. All other samples have zero amplitude, including those at incorrect velocities. This model is simplest to an interpreter because the fewest hyperbolic events have been used to describe the data. Unfortunately, such a model is not a simple function of the data.

however. This model would appear to claim, misleadingly, that the data contain many hyperbolic events with a variety of amplitudes. It is difficult to visualize the corresponding data. If more hyperbolic reflections were present in the data, these extra non-zero amplitudes, or "artifacts," could considerably obscure an interpretation of the stack. Each trace in the stack is independent of the other velocities used, so the resolution cannot be improved by increasing the number of velocities. A perfect model of this synthetic data would stack only at the single correct velocity. But since this velocity is "unknown," each extra velocity adds another trace with artifacts to the model. A single hyperbola will contribute to the model at all velocities, even at a non-physical negative velocity.

We are dissatisfied with the poor resolution of the NMO stack because we can easily visualize the false hyperbolas corresponding to the artifacts. The NMO stack should assume, as we do, that each point in the model corresponds to a distinct hyperbola. Let us test this assumption and compute the data corresponding to the model. To create the data of Figure 1.4a, map each point of the model of Figure 1.3a into a hyperbola and add them together. Artifacts now appear as two incorrect hyperbolas flanking the correct one. The NMO stack evidently violates the assumptions by which we interpret it.

The NMO stack has failed in two crucial ways, even for the simplest of synthetic data. First, we are unable to reconstruct hyperbolic events correctly from the model, so we do not know how to interpret the model. Second, the model describes hyperbolic events that do not exist in the data, so we do not know which details of the model to trust.

1.2. STACKING AS AN INCOMPLETE INVERSION

The usual definition of an NMO stack reconciles two competing goals. The first is to enhance signal: the stack should adequately describe the important reflections in the data; information should be preserved. The second is to suppress artifacts and noise: the stack should distinguish overlapping reflections and should be robust for noisy and irregularly sampled data. To avoid conflict between the two goals, I shall raise these issues separately.

If the stack is to be physically meaningful, the interpreter should be able to visualize the data that correspond to the stack. Rather than modify the NMO stack immediately, I shall propose modeling equations to create the data from the stack—for example, a single point in the stack could represent a full reflection hyperbola in the data. Modeling equations define the signal we wish to capture; no assumptions need be made yet about the noise or sampling of the data.

Define an improved stack then as an inversion of the equation that models hyperbolic reflections. The choice of inversion algorithms is crucial to the issue of robustness. Can the stack model the data accurately? Is the stack no more complicated than necessary? Thorson (1984, 1985) first formulated velocity stacks as an inversion. He defined a transformation that mapped points to hyperbolas, then inverted it by linear and non-linear algorithms that encouraged simplicity in the final stack.

This section illustrates a least-squares inversion which recreates the data very well. I shall first present Thorson's least-squares solution in a form that will generalize

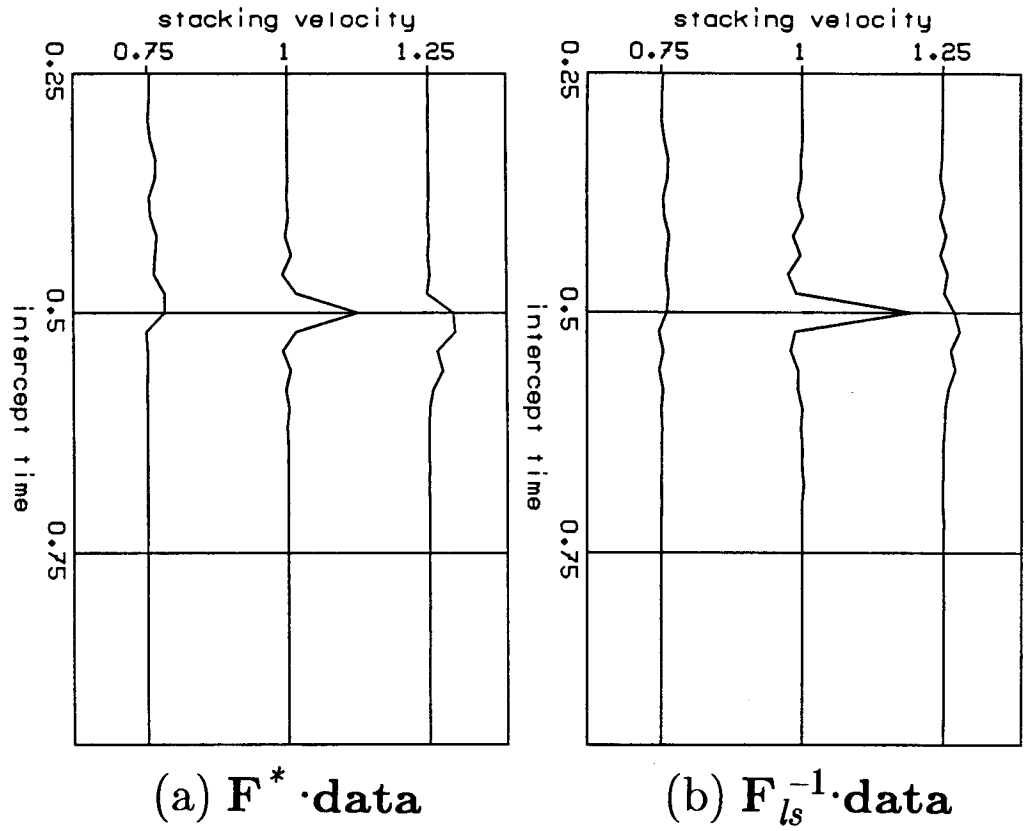


FIG. 1.3. Two methods approximately "invert" the data of Figure 1.1 for parameters to describe hyperbolic events. Samples of these models should give the amplitudes of hyperbolas with zero-offset times between 0.25 and 1.0, and with one of three possible moveout velocities, 0.75, 1, and 1.25. Two methods are used: (a) an NMO stack that averages over all possible hyperbolas (equation [1.1]); and (b) the least-squares inverse of an equation that maps points into hyperbolas (equations [1.2] and [1.3]). An ideal inversion would contain a single point as in Figure 1.2. Neither method succeeds, but the least-squares inversion gives better resolution than NMO stacking. The unwanted amplitudes at incorrect velocities and zero-offset times describe false hyperbolas, artifacts, not present in the data.

later to more difficult transformations. I shall not explicitly encourage further simplicity in the inverted model, as do Thorson, Wiggins (1978), Gray (1979), and others. Rather, in the following section I shall introduce statistical tools to suppress stacked noise and mis-interpreted signal. Because such noise is responsible for losses in resolution, the suppression of noise will improve resolution indirectly.

Figure 1.3b shows the least-squares stack of Figure 1.1. The least-squares stack appears better resolved and easier to interpret than is the NMO stack of Figure 1.3a. Note that the amplitudes at incorrect velocities are smaller than those produced by the NMO stack in Figure 1.2. Most importantly, the modeling equation recreates the data

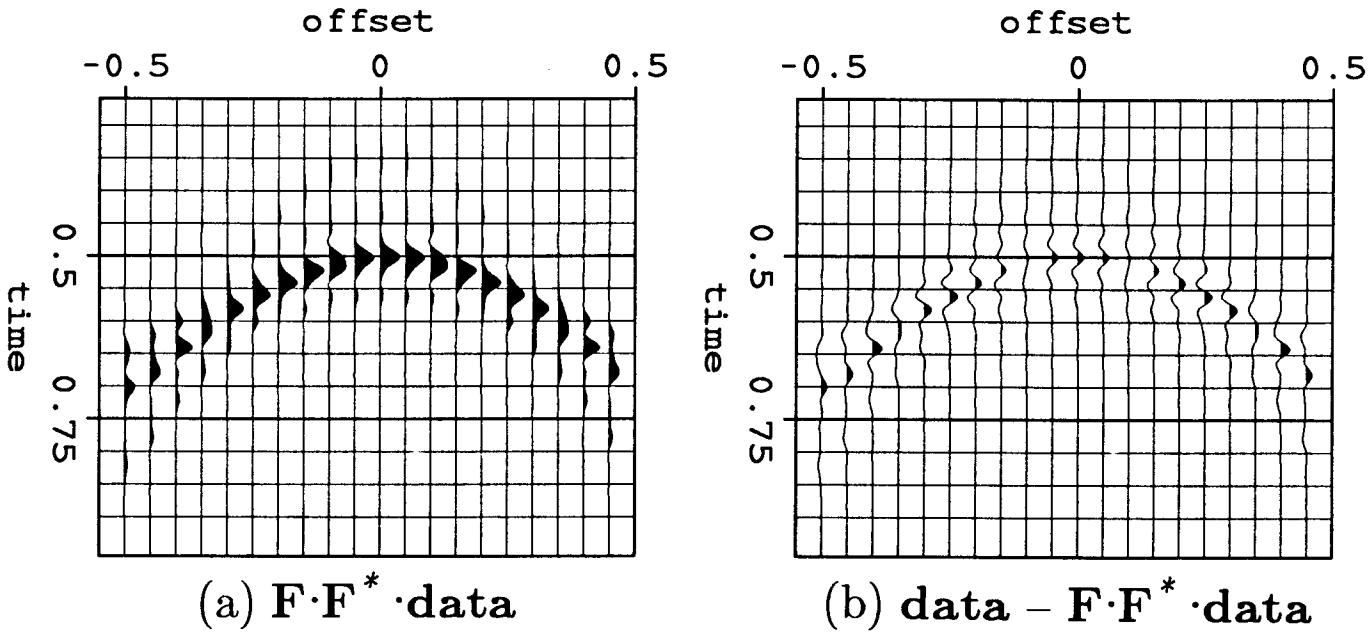


FIG. 1.4. Data are modeled from the NMO stack of Figure 1.3a. The artifacts appear as two hyperbolas flanking the original hyperbola, with visibly incorrect velocities. If more velocities had been used in the model, more artifacts would have been visible.

(Figure 1.5) without the false hyperbolas of the NMO stack. The least-squares stack is a linear function of the data; non-linear methods will be necessary before details of the stack can be interpreted reliably.

1.2.1. The NMO stack

Conventional surface reflection seismic surveys activate a seismic source, such as dynamite, released compressed air, or a truck equipped with a vibrating plate, and then record the reflected waves at regularly spaced distances from the source. The vertical displacement of the earth's surface is recorded with geophones placed along the line of the survey; the spacing of sources equals or doubles that of geophones along the line of the survey. The procedure is then repeated for many source positions. The displacement of the earth's surface can be parametrized as a function of the source position x_s, the geophone position x_g, and the time t elapsed after the source first becomes active. Positions are in absolute distances from the first point along the line of the survey.

For a typical NMO stack, the recorded data are first re-sorted into "midpoint-offset" coordinates defined by

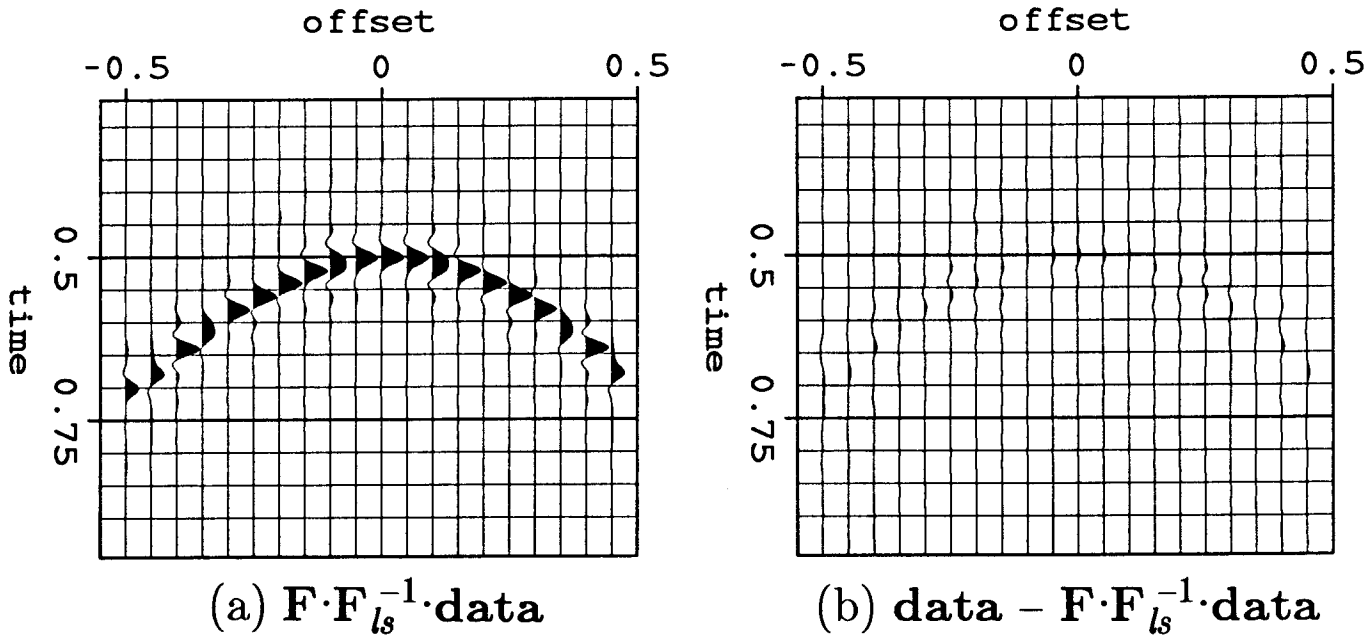


FIG. 1.5. (a) The least-squares model of Figure 1.3b creates data that are very similar to the original data of Figure 1.1a. (b) The residuals show negligible artifacts, unlike the adjoint inversion of Figure 1.4b. Mapping points into hyperbolas, equation (1.1), destroys some information, so we do not know how many details of the model in Figure 1.3b were necessary to create this data.

$$y = (x_s + x_g)/2 \;\; ; \;\; h = (x_g - x_s)/2 \; .$$

y is the midpoint coordinate, and h the half-offset. Individual traces, the one-dimensional functions of time, are simply reindexed without any rescaling. If the seismic waves are reflected from a series of horizontal, planar interfaces between rocks of comparable seismic velocities, then the arrival times of recorded wavefronts will approximate hyperbolic functions of offset.

By treating the predictable arrival times of reflections as a redundancy, the NMO stack removes the offset dimension of recorded data. Define the stack as an average of amplitudes along the hyperbolas:

$$\hat{\text{model}}(\tau, v) \equiv \frac{1}{N_h} \sum_h \text{data}(t = \sqrt{\tau^2 + 4h^2/v^2},\, h) \; . \tag{1.1}$$

τ is the time at which a reflected wave arrives at zero offset. N_h is the number of offsets. The caret over $\hat{\text{model}}$ indicates that this is an estimate of the model, not a definition.

The "stacking velocity" or "moveout velocity" v is comparable in value to the velocity at which the pressure or P wave travels through the earth. The moveout velocity must become a function of τ when the seismic velocity varies as a function of depth. For the hyperbolic approximation and a laterally invariant earth, the stacking velocity is equal to the root-mean-square (rms) velocity down to the depth of the reflector (the square root of the square of the velocity integrated from the surface to the reflector).

Let us assume that changes in amplitude with time, caused by the spherical spreading of wavefronts, have been compensated for approximately by a multiplication with the travel time t. When seismic velocities change with depth, this correction must be changed as a function of the curvature of the wavefront. Energy is usually diffused and dissipated by other mechanisms, such as the inelastic absorption of high frequencies, dispersion of high frequencies by thin sedimentary beds, incoherent scattering of waves by irregularities in reflectors, etc. A scaling of amplitude by the square of the travel time often corrects best for such losses of energy. See Claerbout (1985, p. 233) for a discussion of this correction.

1.2.2. Insufficiency of the NMO stack

Summing common-midpoint (CMP) data along hyperbolic paths does not discriminate between overlapping hyperbolic events optimally. For simplicity each sample of the model should describe an independent hyperbolic reflection in the data. As a result of this independence, an ideally inverted model could possess the fewest "details"—the fewest non-zero samples would be necessary to describe the data. Since details could be interpreted independently, a physical interpretation of the model should be easiest.

The samples of the stacked model in equation (1.1) cannot be statistically independent, even when reflections are perfectly hyperbolic. We must first surmount two difficulties:

- All hyperbolic reflections appear flat at zero offset and overlap at low offsets because of their finite bandwidth. Thus, a single hyperbolic event contributes to the stack at many incorrect velocities.
- Each trace of the data contributes equally to the stack, even though data are unevenly sampled, with truncations and missing traces.

The result is a description of the data with artifacts—hyperbolas not found in the data are described by non-zero samples, or unreliable details, in the model.

We see these defects clearly in the synthetic model. The gather was created by placing a single spike of unit amplitude along a hyperbolic path (Figure 1.1). The correct moveout velocity is 1, and the zero-offset travel time is 0.5. Sinc ($\sin[x]/x$) interpolations were used to preserve all frequencies (Bracewell, 1978); coarse sampling over time distorts the plotted waveforms but not the frequency content. Figure 1.3a contains a stack at three velocities: 0.75, 1, and 1.25. As noted, the stack has far from zero amplitudes at the incorrect velocities. Some resolution is also lost over time: the central peak has gained sidelobes.

More importantly, the stack makes no attempt to account for the effects of incoherent, non-hyperbolic noise on the model. At best, one can hope that enough redundancy exists so that stacking produces a high signal-to-noise ratio. Incoherent noise is often strong and non-Gaussian, however, and can create stacked models that imply strong, false hyperbolas on the gather. Moreover, most artifacts result from misinterpreted signal that behaves as incoherent noise. Suppressing the effect of noise on the stack would also improve resolution. I shall consider the issue of noise in section 1.3.

Many objections to the oversimplifications of the physical model can be raised. In particular the stack assumes that reflections appear globally hyperbolic, with amplitudes determined strictly by the spreading of spherical wavefronts. Instead of describing the reflections as a sum of full hyperbolas, we can describe them as a sum of hyperbola segments, limited to a narrow range of offsets. I shall add this lateral adaptability before treating any recorded field data.

Lastly, the model takes no advantage of the coherence of seismic data over time. Seismic reflections, unlike our simple synthetic illustration, also contain recognizable wavelets. Wavelets are predictable and add additional redundancy to the data. In section 1.7, I shall modify the sample-by-sample techniques to recognize entire wavelets.

1.2.3. A model for hyperbolic reflections

Interpretation of the velocity stack requires a modeling equation that can recreate the data from the stack. To simplify interpretation, each sample of the stack should represent independent information, a single elementary feature of the data. The following equation satisfies these requirements.

$$\text{data}(t,h) = \sum_{v} \text{model}(\tau = \sqrt{t^2 - 4h^2/v^2}\,, v) \quad . \tag{1.2}$$

This equation describes the data as a sum of hyperbolas of all curvatures. Single

points of the model array map to unique hyperbolas. Assume that amplitude changes due to the spreading of wavefronts have been removed, as stated before. If the data of equation (1.2) are divided in amplitude by the arrival time, then this equation corresponds to the spherical Green's function of a three-dimensional, constant velocity earth.

Admittedly, the choice of transformations is somewhat arbitrary; other wave-propagation effects could be incorporated. An interpreter should find it easy to visualize how every point of this model contributes to the data. A more complex equation must increase the difficulty of interpreting the model. We shall see that the least-squares methods of this section can be just as easily applied to a more complicated model (as in section 1.4).

We can write this equation for discretely sampled data as a matrix multiplication: $\mathbf{data} = \mathbf{F} \cdot \mathbf{model}$. The stacking equation (1.1) does not invert for the model, but rather multiplies the data by the transpose of this matrix, scaled by a constant: $\hat{\mathbf{model}} = a\,\mathbf{F}^* \cdot \mathbf{data}$ where $a = N_h^{-1}$, the reciprocal of the number of offsets. The asterisk indicates the transposed matrix. In general, $\hat{\mathbf{model}} \neq \mathbf{model}$. The continuous equivalent of a transposed matrix multiplication is called the *adjoint* equation. Equation (1.1) then is a scaled adjoint of equation (1.2).

The scale factor in the stacking equation was chosen to make the adjoint as close as possible to an inverse for purely hyperbolic events, the signal. In many applications, particularly tomography, this scaled adjoint is described as a "back projection." The back projection recognizes that each point in the data can be created by summing along a distinctive curve in the model. The back projection approximately inverts this transformation by redistributing the amplitude of a data point evenly along its summation curve in the model. Thus, the scaled adjoint equation (1.1) becomes an average rather than a simple sum. Application of the back projection after the forward transformation creates artifacts because the amplitude at every model point is redistributed over all curves that pass through it.

The synthetic data set of Figure 1.1 could have been created by the model of Figure 1.2, with a single spike at $v = 1$ and $\tau = 0.5$, written with delta functions as $\text{model}(\tau, v) = \delta(\tau - 0.5)\delta(v - 1)$. The artifacts in the NMO stack of Figure 1.3a are back-projection artifacts. Figure 1.4 shows the recreation of the data from the NMO stack with equation (1.2). Artifacts are now visible as two false hyperbolas flanking the correct one. Subtracting this model from the original data leaves strong residuals.

The NMO adjoint stack does reconstruct the data well, considering that the stack was not originally intended for this purpose. However, the remaining disagreement with the data offers an opportunity to improve upon the interpretability of the stack. In the following section I shall show that the adjoint stack is the first step in a least-squares inversion, which reconstructs the data very well. In turn, the least-squares stack will be the first step of a robust inversion, which suppresses noise.

1.2.4. An alternative least-squares inverse

Least-squares inverses are more difficult to compute than adjoints, but they do offer some valuable properties. The solution can be defined by an objective function that is independent of the algorithm used to solve it. An objective function measures the relative quality of an inverse: the best inversion minimizes the function. Programs to minimize such functions are being continually improved. See Luenberger (1984) for a good discussion of optimization methods.

The objective function can balance two competing goals: the inversion should minimize the differences between the modeled data and the original data; the model should not have amplitudes larger than are necessary to fit the data well. Other objective functions could achieve these goals, but only by sacrificing some useful properties.

The form of the least-squares objective function is particularly simple. The function is a quadratic function of the model and the data. Because of this form, the best model is a linear function of the data. This linearity will considerably simplify some statistical calculations shown in section 1.3. A quadratic objective function is also easily minimized. If desired, eigenvector methods can be used to evaluate resolution and stability. Stability is easily guaranteed.

The least-squares objective function also has a simple statistical interpretation. Let us assume that samples of the model and the residual noise are Gaussian random variables. Then minimization of the function equivalently maximizes the probability of the model (a "maximum-likelihood" estimate). When there is no other information, Gaussian statistics can be assumed more safely than others. See Appendix A for more discussion.

When minimized, the following least-squares objective function will optimize the data fit and the model amplitudes (e.g.):

$$\min_{\text{model}(\tau,v)} \left(\sum_{t,h} \{\text{data}(t,h) - \sum_{v} \text{model}[\tau(t,h),v]\}^2 + c^2 \sum_{\tau,v} [\text{model}(\tau,v)]^2 \right) \tag{1.3}$$

$$\text{or} \quad \min_{\mathbf{model}} [\, || \mathbf{data} - \mathbf{F} \cdot \mathbf{model} ||^2 + c^2 || \mathbf{model} ||^2]$$

c is a small dimensionless constant. Double bars represent Cartesian magnitudes, defined by the sums that precede them. The first term minimizes the sum of the squares of the uninverted data; the second term minimizes the sum of the squares of the model. Samples of signal and noise are assumed independent, so no covariance matrices are included.

The unique minimum, could it be explicitly calculated, defines the damped least-squares inverse:

$$\mathbf{model} = \mathbf{F}_{ls}^{-1} \cdot \mathbf{data} = (\mathbf{F}^* \cdot \mathbf{F} + c^2 \cdot \mathbf{I})^{-1} \cdot \mathbf{F}^* \cdot \mathbf{data} \tag{1.4}$$

$\mathbf{I}$ is the identity matrix. Notice that this explicit solution is a simple linear function of the data. However, we cannot easily invert the large matrix between parentheses.

The second term of equation (1.3) is a "penalty function" that adds stability to the minimization of the objective function. This term adds what is sometimes called "prewhitening" to the data. In equation (1.4), prewhitening adds a small constant to the diagonal of the matrix $\mathbf{F}^* \cdot \mathbf{F}$, increases the smallest eigenvalues, and guarantees that an inverse exists. The solution must attempt to minimize the magnitude of the model as well as the magnitude of the data errors. Otherwise, a large perturbation of the model could be allowed for a negligible improvement in the modeled data. An infinitesimal c gives analytic stability, and a small c suffices for numerical stability.

I shall not attempt to find a closed-form solution equivalent to equation (1.4). Thorson (1984) gives an explicit solution for the least-squares inverse of a hyperbolic model that is evenly sampled over the inverse of velocity. This inverse is equivalent to the adjoint followed by a two-dimensional convolution, called a *rho* filter in tomography applications. As soon as irregularly sampled data are introduced, however, such an algebraic calculation of these closed-form solutions become difficult.

Let us avoid solutions that will not generalize to other modeling equations. Explicit least-squares solutions will become difficult to find for more complicated hyperbolic models, as in section 1.4, and especially for such complicated transformations as the acoustic wave equation in Chapter 3.

Rather, let us minimize (1.3) to an acceptable convergence with an iterative algorithm. A steepest-descent algorithm iteratively perturbs a current model in the

direction of the gradient (multiplied by a scale factor a).

$$\mathbf{model}_{new} = \mathbf{model} + a\,\delta\,\mathbf{model}; \tag{1.5}$$

$$\text{where } \delta\,\mathbf{model} = \mathbf{F}^* \cdot (\mathbf{data} - \mathbf{F} \cdot \mathbf{model}) + c^2 \mathbf{model}$$

If the second term is dropped ($c = 0$), or if the starting model is zero, then the gradient is just the adjoint operation $\mathbf{F}^*$.

The scale factor a of the gradient can be calculated if the expression for the new model (1.5) is substituted into the objective function (1.3). The result should be differentiated with respect to the scale factor a and set equal to zero. The best value of a then appears as the function of a few simple summed products (dot products).

Two iterations of the steepest-descent algorithm are equivalent to two successive back projections (if we set $c = 0$).

$$\mathbf{model} = a_1 \mathbf{F}^* \cdot \mathbf{data} + a_2 \mathbf{F}^* \cdot (\mathbf{data} - a_1 \mathbf{F} \cdot \mathbf{F}^* \cdot \mathbf{data})$$

The first term is the back-projection of the data, the original stacking equation. The second term adds the back projection of the residual data (between parentheses) that were not fit by the first iteration. Though this effect is not at all obvious, further back-projections can grow increasingly unstable. The prewhitening penalty function is necessary to prevent a divergence of amplitudes after many iterations.

The preceding infinite sum of back-projections (scaled adjoints) is just a particular and suboptimum method of minimizing function (1.3)—the method of steepest-descent. Other gradient methods, particularly the method of conjugate gradients, will minimize equation (1.3) much more efficiently (see Luenberger, 1984). The first step of most gradient methods (including conjugate gradients) is equivalent to the scaled adjoint of equation (1.1).

1.2.5. Advantages and limits of the least-squares stack

The inversion of a simple forward model by the minimization of an objective function makes two improvements over NMO stacking:

- The model can be constrained for stability, so that components of the model that have negligible effect on the data are suppressed.
- The inversion can fit irregularly sampled data, without assumptions being made elsewhere, by limiting sums in the objective function.

The synthetic data of Figure 1.1 illustrates how the least-squares stack diminishes artifacts. Figure 1.3a shows the result of applying conventional stacking with equation

(1.1). Figure 1.3b shows the least-squares solution that minimizes equation (1.3) (from a conjugate-gradient algorithm). Notice that the least-squares stack has much smaller amplitudes at incorrect velocities than does the conventional stack. The resolution of the central peak has considerably improved.

Figure 1.5 shows data recreated from the least-squares model with equation (1.2). Residual artifacts are considerably weaker than those obtained in Figure 1.4 from the conventional stack. Neither the model nor the data changes significantly with further iteration. The modeling equation (1.2) blurs some information, so the inversion is not obliged to invert for a perfect spike. As the velocities are sampled more finely, one can see this lost resolution more clearly.

The single spike that we prefer cannot be obtained with this linear inversion: too much information is destroyed when the original model is mapped into hyperbolas. Model components that have no effect on the data belong to what is called the "null space" (see Strang, 1980). More importantly, some components of the model have a *negligible* effect that falls below the accuracy of the data measurements (beneath the level of background noise). Whether or not such poorly-determined components should appear in the inverted model depends on their ability to simplify the overall appearance of the model. The least-squares norm penalizes only the total energy of model amplitudes and so uniformly suppresses all components of the null and near-null space. Non-quadratic measures of simplicity might add some components of the null space to the model, though perhaps in an arbitrary way.

On the other hand, some linear components of the model might help fit the data better, and yet complicate the appearance of the model too greatly. Such components, though well-determined by the data, must be called unreliable because we cannot believe the corresponding structure. For example, it might be possible to create a single noise spike in the data by combining a large number of hyperbolic events. The least-squares inversion of these hyperbolic events might be unique and well-determined, but the combination is too unlikely to accept.

The least-squares objective function does define a unique inverted model, as do non-linear objective functions by Wiggins (1978), Gray (11979), and Thorson (1984). Unfortunately, a very different model might *almost* minimize the chosen objective function and reconstruct the data *almost* as well. Before accepting a new perturbation of the model, we should decide whether the perturbation is describing signal or noise in the data.

1.3. INVERTING FOR RELIABLE SIGNAL

The least-squares inversion is capable of fitting the data very well, as we saw with the simple synthetic example. The chosen parameters are not well constrained, however; a simpler and more interpretable stack than the least-squares stack could model the data equally well. Wiggins (1978), Gray (1979), and Thorson (1984) each proposed alternative objective functions that encourage the inversion to maximize some measure of simplicity in the inverted model. The amount of simplicity can be adjusted to the taste of the interpreter. Rather than follow their strategy, I shall ask questions that bear directly on the usefulness of the model: how reliable are the inverted parameters? Do they model signal or noise? If we can answer such questions, then we can perhaps decide how much simplicity to trust.

Let us consider the data to be the sum of genuinely hyperbolic events (signal) and non-hyperbolic, incoherent events (noise). Such events cannot be entirely distinguished from each other; we can only attach different probabilities to our choices. A sum of incoherent noise events could explain a single hyperbolic event—but with a low probability. A single hyperbolic event is much more likely to be the sum of hyperbolic signal. Alternatively, a sum of hyperbolic signal could destructively interfere and create an incoherent event—but with a low probability. Many different combinations of signal and noise can model the same data. By choosing statistical distributions for the signal and noise, however, we can quantify our expectations for the simplicity of the model and for the fit with the data.

Whatever our assumed statistical distributions, some events in the data might be only slightly more probable as signal than as noise. A minimized objective function must determine whether an event will be treated as signal or as noise, though the preference might be marginal. An objective function gives little opportunity to introduce statistics or decision-making criteria that will improve the choice. An objective function determines model perturbations from data that include signal and noise; Thus, an inverted model is also a function of signal and noise. Any measured statistics about the signal and noise in the data will also tell us about the statistics and reliability of the inverse.

Let us say that an inverted parameter is "reliable" if a negligible part of its amplitude could have resulted from noise in the data, rather than from signal. Noise can include misinterpreted signal that behaves as noise. Let us identify and eliminate unreliable perturbations of the model to prevent them from obscuring reliable perturbations.

Reliably inverted model parameters are simpler and better resolved than unreliable parameters because reliable parameters, by definition, model mostly signal and little noise. We chose a modeling equation (1.2) whose parameters should describe independent events (signal) in the data. The perturbation of a single sample of the model will create a single hyperbolic event in the data. An incoherent noise event, on the other hand, requires perturbations of many model parameters, obscuring any well-resolved signal.

If one assumes that the model parameters are Gaussian, the least-squares inversion finds the most probable signal and noise parameters to model the data (see Appendix A for a demonstration). Often, independent physical parameters are not Gaussian, and the assumption of Gaussianity distorts them. Appendix A shows how other global objective functions can be derived as *maximum-likelihood* estimates when more is known about the statistics of noise and model parameters. Without this additional information, we can assume Gaussian statistics and keep the convenient least-squares form. (One can also argue that Gaussian statistics make the fewest prior assumptions.) Again, the objective function provides only part of the answer: in addition to finding the most probable parameters, we must determine which parameters are sufficiently reliable. Even if we used the non-quadratic objective functions of Wiggins, Gray, or Thorson, we could ask if the simple model parameters were reliable. Instead of replacing the least-squares objective function, let us restrict the models used to minimize this objective function.

1.3.1. Observing the effects of noise on the inverse

A simple experiment will estimate the possible effects of noise on our least-squares model. The synthetic data of Figure 1.1 contain no visible noise: all non-zero amplitudes fall along a single hyperbolic path. Yet, it is possible (though unlikely) that this hyperbolic event is merely a random alignment of incoherent noise. To measure our confidence, let us find the probability that independent noise spikes might fall on a hyperbolic trajectory. I shall use the least-squares stack to measure this probability: stacked noise should show very different statistics from the stacked data.

Let us first examine the least-squares stack (Figure 1.3b) for clues. Non-zero amplitudes appear as artifacts at incorrect velocities. Artifacts result from misinterpreted signal that behaves as noise. Because recorded samples are sparse, portions of the single hyperbolic reflection appear to belong to hyperbolas with other velocities. Artifacts can be regarded as signal that has accidently aligned in a misleading way. Noise then includes artifacts as well as any component that does not have the

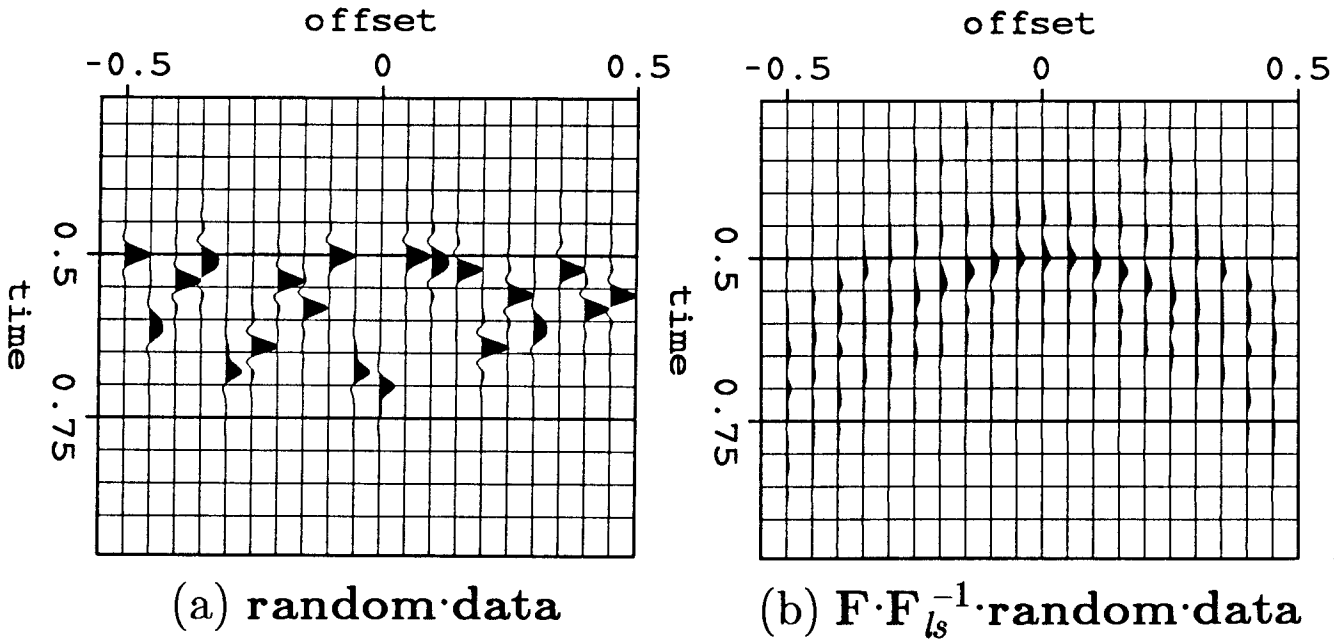


FIG. 1.6. (a) These noisy data were created by randomly reordering the traces of the original data in Figure 1.1. The distribution of amplitudes is the same as that of the data, but without the hyperbolic coherence. (b) The least-squares stack fitted many false hyperbolas to random alignments of noise. If the original data contained only noise, then the reconstructed data of Figure 1.5b would show a comparable distribution of amplitudes.

coherence of the signal being inverted.

To over-estimate the effect of noise on the least-squares model, I first assume that all data are noise and then measure how well the least-squares inversion can fit such noise. To create the noisy data in Figure 1.6a, I randomly reorder the original traces of Figure 1.1. Figure 1.6b shows hyperbolic reflections that were fit by the least-squares inversion of equation (1.3). The inversion fits many false hyperbolas to random alignments of noise. These false hyperbolas help diminish the objective function (1.3), but the result is clearly meaningless.

If the original data were entirely noise, then the distribution of amplitudes in the reconstructed data of Figures 1.5b and 1.6b would be comparable. The artificial noise has the same number of zero and finite amplitude samples as the original data—only the coherence has changed. If the original data were incoherent, then scrambling traces would be unlikely to change this incoherence. A small chance does exist that two arrays of scrambled data would have considerably different amounts of accidental hyperbolic alignment. For the comparison to be conclusive, the artificial noise should

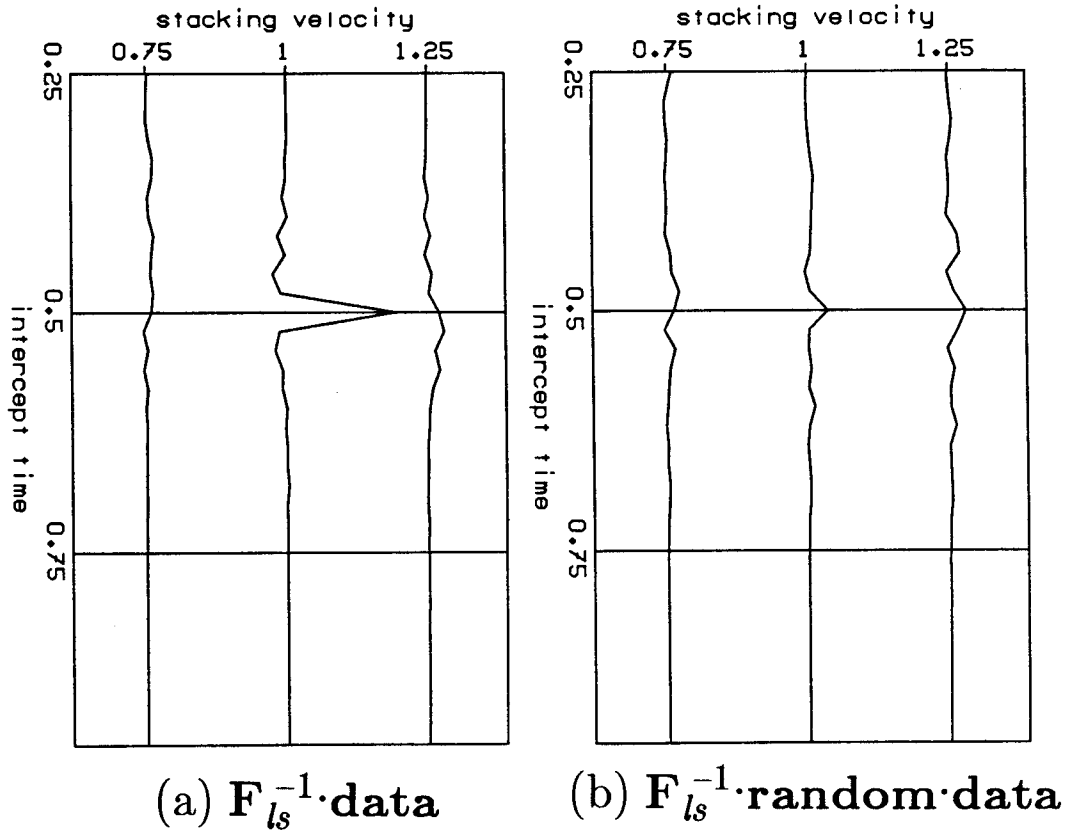


FIG. 1.7. (a) The least-squares stack of the original data (same as Figure 1.3b) contains one strong spike and sidelobes of uniformly low amplitude. (b) A least-squares stack of the noise in Figure 1.6a contains many non-zero samples, equal in strength to the sidelobes of the stacked data. The large peak in the stacked data (a) corresponds to the genuine hyperbola in the original data of Figure 1.1. Incoherent noise of equal strength could not create a comparable peak in (b). Examining only (a) and (b), we can conclude that the original data contained only one event that could be reliably attributed to hyperbolic signal and not to random alignments of noise.

show a *typical* amount of accidental coherence. The number of data points (the redundancy) should be high, or the scrambling and inversion should be repeated many times.

A comparison of the stacked data and artificial noise distinguishes the signal from noise (and artifacts). Figure 1.7b displays the least-squares stack of the artificial noise alongside the least-squares stack of the original data (Figure 1.7a). The stacked noise contains many non-zero samples, equal in strength to the sidelobes (artifacts) in the stacked data. The stacked data contain a large central peak corresponding to the genuine hyperbola in the original data; no comparable amplitude exists in the stacked noise. All other details of the two models are similar. By comparing the two stacks,

we can conclude that the original data contained only one event that could be reliably attributed to hyperbolic signal and not to random alignments of noise. In the following sections I shall develop a method of quantifying such observations statistically.

1.3.2. Estimating the statistics of stacked signal and noise

The demonstration of the previous section took advantage of a implicit assumption in our definitions of signal and noise. We defined signal as a sum of *independent* hyperbolas, and noise as the component of the data whose samples were *independent* along hyperbolic paths.

Independent parameters can be described by a simple statistical function, the marginal probability density function (pdf). Define the pdf $p_y(\cdot)$ of a random variable y by the following:

$$\text{probability } [y_1 < y < y_2] = \int_{y_1}^{y_2} p_y(x)dx \quad .$$

An integration of the pdf over a range of amplitudes gives the probability that the random variable will have an amplitude within that particular range.

To estimate statistics for the transformed signal and noise, note two properties:

- If signal and noise are additive in the data, they remain additive in the least-squares inversion, a linear transformation.
- Adding independent random variables convolves their pdf's.

See an elementary statistics text such as Papoulis (1965) for a demonstration of these properties. A sample of linearly transformed data (let us use d') can be written as the sum of independent signal and noise random variables (s' and n'). I use primes to emphasize that these components have been transformed. The pdf's convolve:

$$d' = s' + n' \; ; \; p_{d'}(x) = p_{s'}(x) * p_{n'}(x) \quad . \tag{1.6}$$

The asterisk indicates convolution.

Randomly scrambling the original data did not change the number of samples at any given amplitude. Because the data no longer showed coherence, however, samples added destructively rather than constructively in the least-squares inversion. If the original data had been entirely noise, then scrambling the data would not have affected the statistics of the least-squares model. The differences between the distribution of amplitudes in the stacked data and that of the stacked noise can be explained only by the presence of signal in the original data.

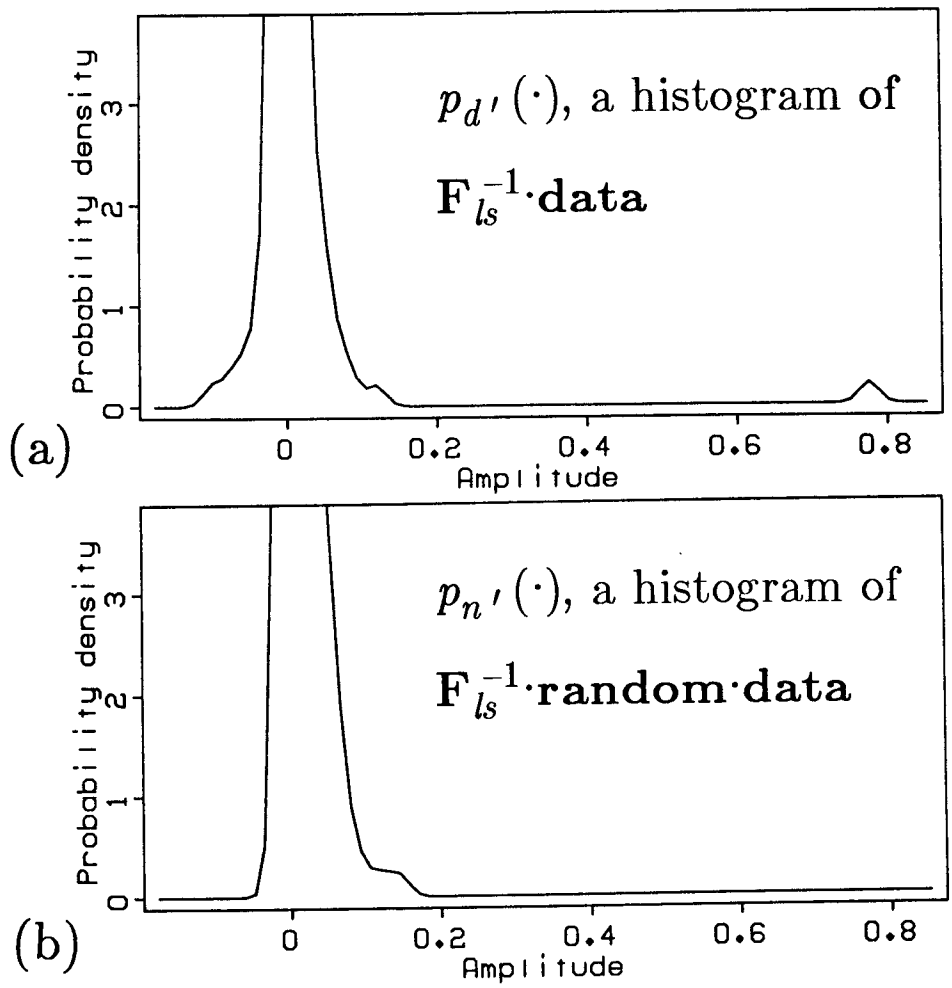


FIG. 1.8. These histograms show the distribution of amplitudes in the stacked data and noise of Figures 1.7a and 1.7b. When the number of samples is large, amplitude histograms approximate the probability density functions (pdf's) of the samples. (a) is the histogram of the stacked data of Figure 1.7a, and (b) that of the stacked noise of Figure 1.7b. The transformed data contains a sample at high amplitude (near 0.8), which corresponds to the one true hyperbola, but the noise does not. The noise histogram (b) shows the amplitudes that could be achieved if the data were assumed pessimistically to be entirely of noise. Comparison of the two histograms shows the amplitude 0.8 to be improbable as transformed noise and therefore probable as signal.

Figure 1.8 contains amplitude histograms of the transformed data and noise in Figure 1.7. These histograms display the measured frequency of amplitudes. As the number of samples increases, these histograms will approach the shapes of their corresponding pdf's, $p_{d'}(\cdot)$ and $p_{n'}(\cdot)$. The inner peak of the noise distribution is somewhat narrower than that of the data. Most significantly, the data distribution

also contains a second peak at the high amplitude of 0.8; this peak corresponds to the one true hyperbola in the data.

The differences between the data and noise histograms must be explained by additive signal that affects the convolution of equation 1.4. Let us assume that the histogram for the noise accurately estimates the noise pdf. This pdf implicitly assumes that all the data were noise, so I shall call it a *pessimistic* estimate. If all signal had zero amplitudes, then its pdf would be a delta function ($p_{s'}[x] = \delta[x]$), and the noise and data pdf's would be identical. But the pessimistic pdf for noise does not easily explain all features of the data histogram. Let us now find the signal pdf that, when combined with the noise pdf, will best explain the data histogram.

Let us find the *maximum-likelihood* estimate of the signal pdf. If the noise pdf is assumed to be exact, the best signal pdf should maximize the probability (likelihood) of the data histogram. Equivalently, the maximum-likelihood estimate of $p_{s'}(\cdot)$ minimizes the divergence of the data histogram from the convolved pdf $p_{s'}(\cdot) * p_{n'}(\cdot)$. Appendix C shows how this divergence can be measured and minimized with a cross-entropy function. Figure 1.9a contains such a "deconvolved" estimate of $p_{s'}(\cdot)$. Most samples appear likely to contain zero signal; a small fraction might have amplitude 0.8, to account for this amplitude in the data distribution. The large number of amplitudes with zero probability reflect the conservatism of the estimate. Other amplitudes might be possible; however, they are unnecessary to explain the data. Recorded seismic events will tend to show a continuous range of possible amplitudes because wavelets are coherent and smoothly oscillating over time. In section 1.7 I shall discuss how sample-by-sample histograms can be used to describe wavelets.

Figure 1.9b displays the convolution of the estimated signal and noise pdf's. This convolution fits the data histogram of Figure 1.8a well, but a perfect fit is impossible. The width of the noise distribution prevents the narrow peak at amplitude 0.8 in the data histogram from being fit perfectly by the convolution. Only an optimization can deconvolve such histograms; an exact solution will usually be impossible. The uncertainty is, however, due to insufficient data, not to the method of estimating the pdf. Explicit estimates of pdf's discriminate more differences between signal and noise than do statistics with fewer degrees of freedom, such as means, variances, and other moments.

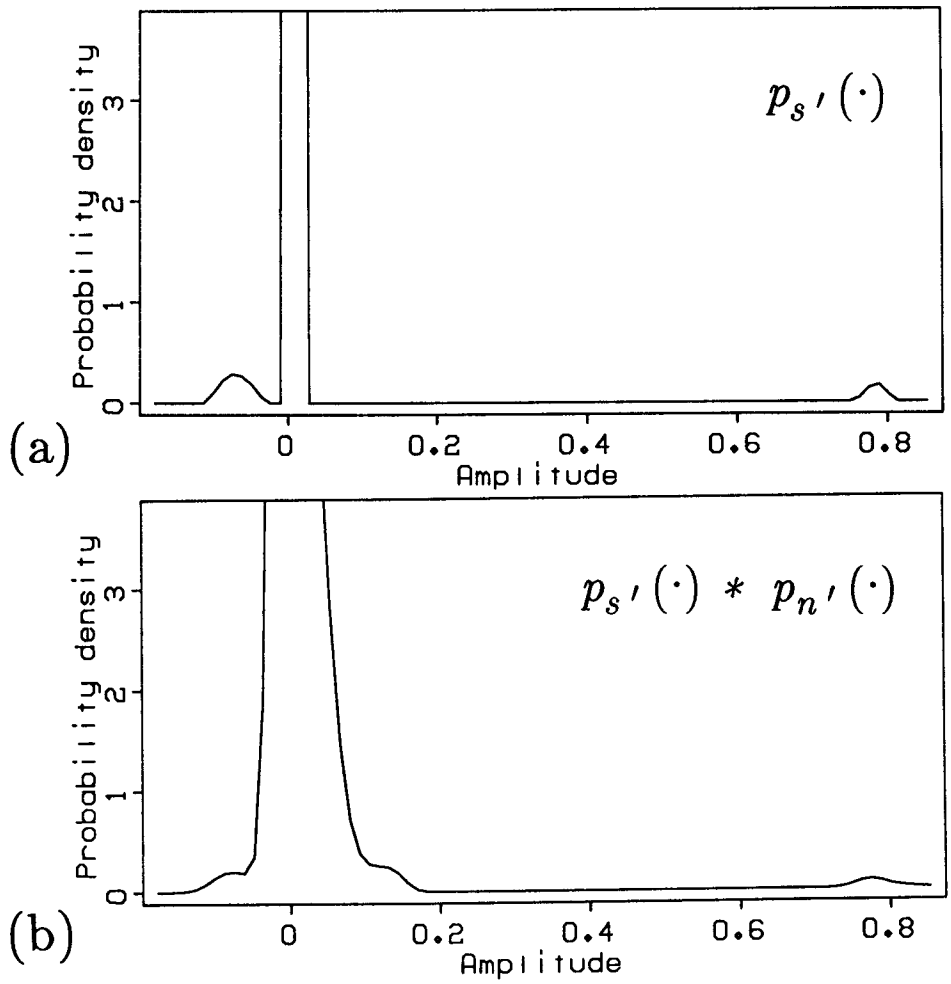


FIG. 1.9. Let us assume that the noise pdf of Figure 1.8b accurately (though pessimistically) estimates the additive noise to the stacked data. If signal and noise are additive in the original data, then they will remain additive after the linear least-squares stack. The pdf's of stacked signal and noise will convolve to determine the pdf of the stacked data: $p_{d'}(\cdot) = p_{s'}(\cdot) * p_{n'}(\cdot)$. Let us find a $p_{s'}(\cdot)$ for which the corresponding $p_{d'}(\cdot)$ will maximize the probability of the actual data distribution. (a) shows this "maximum-likelihood" estimate $p_{s'}(\cdot)$: most amplitudes appear likely to be near zero, excepting a small fraction near 0.8. The implied $p_{d'}(\cdot)$ in (b) fits the data histogram in Figure 1.8a well; however, the peak near 0.8 is necessarily wider in (b) because of the width of the noise distribution.

1.3.3. Separating stacked signal from noise

We have pessimistic estimates of three distributions, $p_{s'}(\cdot)$, $p_{n'}(\cdot)$, and $p_{d'}(\cdot)$, and we can now estimate, with known reliability, the amount of stacked signal in the stacked data. Let us first find the "expected" amplitude of signal in a sample of

stacked data—that is, the "average" amount of signal in a particular sample of data.

$$\hat{s}' = E(s' \mid d') = \int x \; p_{s'|d'}(x \mid d')dx$$

$$= \frac{\int x \; p_{s'}(x)\, p_{n'}(d'-x)\, dx}{p_{d'}(d')} \quad . \tag{1.7}$$

This estimate is a Bayesian estimate; it uses Bayes' rule twice to calculate the conditional probability. See Papoulis (1965) for a discussion of Bayesian methods. Figure 1.10 displays the Bayesian estimate of the expected signal for various amplitudes of the transformed data. For most amplitudes of data, the most probable signal is zero. The striking exception appears at the high amplitude peak (near 0.8) corresponding to the true hyperbola. This peak is determined to be entirely of transformed signal. Application of the Bayesian estimate to the model (Figure 1.11) suppresses much of the amplitude of sidelobes, but some spurious details remain.

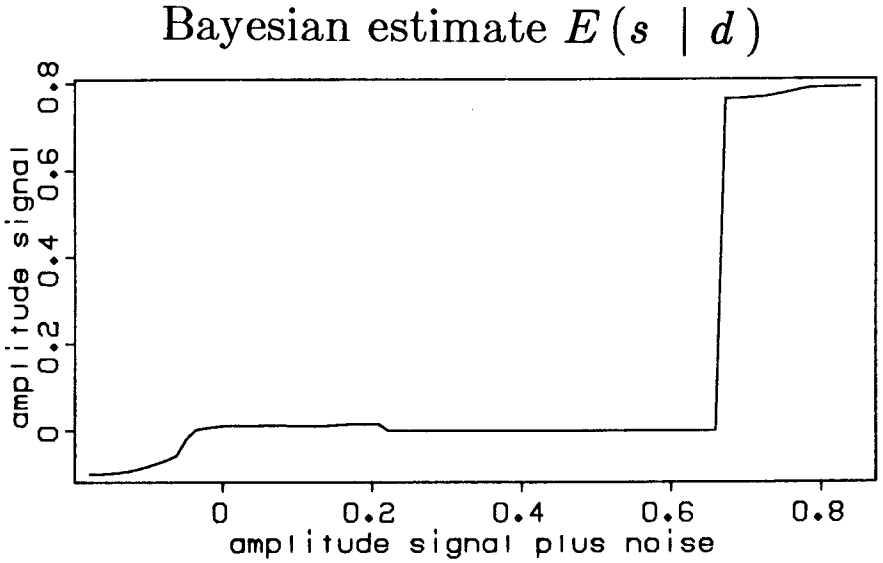


FIG. 1.10. The estimated pdf's of Figures 1.8 and 1.9 can be used to estimate the signal present in the least-squares stack of Figure 1.7a. The Bayesian estimate gives the expected value of signal when the signal plus noise is known. Most transformed data amplitudes are expected to contain nearly zero signal, excepting a fraction with amplitude 0.8.

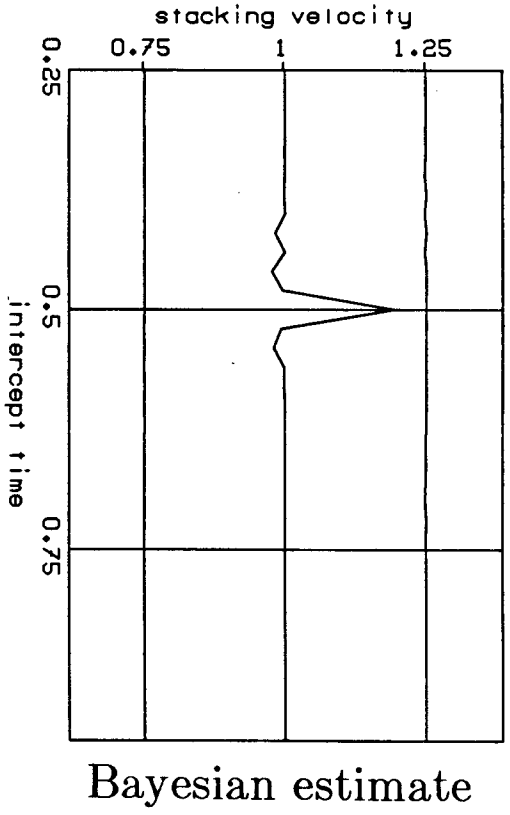


FIG. 1.11. This model shows the Bayesian estimate of signal in the least-squares stack of Figure 1.7a. Most sidelobes have diminished, but some spurious events remain.

We have found the most expected values of signal, but we do not yet know *how* probable these values are. Because we have pdf's for the signal and noise, we could explicitly (and expensively) calculate the probability of every possible value of signal for every possible value of the transformed data. Instead I shall define a simple measure of the accuracy of the Bayesian estimate. Let us define the "reliability" of the Bayesian estimate as the probability that the percentage error is less than 5%.

$$\begin{aligned} \text{reliability} &= \text{probability}[0.95\hat{s}' \leq s' \leq 1.05\hat{s}' \mid d'] \\ &= \frac{\int_{0.95\hat{s}'}^{1.05\hat{s}'} p_{s'}(x)p_{n'}(d'-x)dx}{\int_{-\infty}^{\infty} p_{s'}(x)p_{n'}(d'-x)dx} \end{aligned} \tag{1.8}$$

The denominator of the above is a value of the implied data distribution $p_{d'}(d')$. Figure 1.12 displays this probability, computed for all amplitudes of the transformed data.

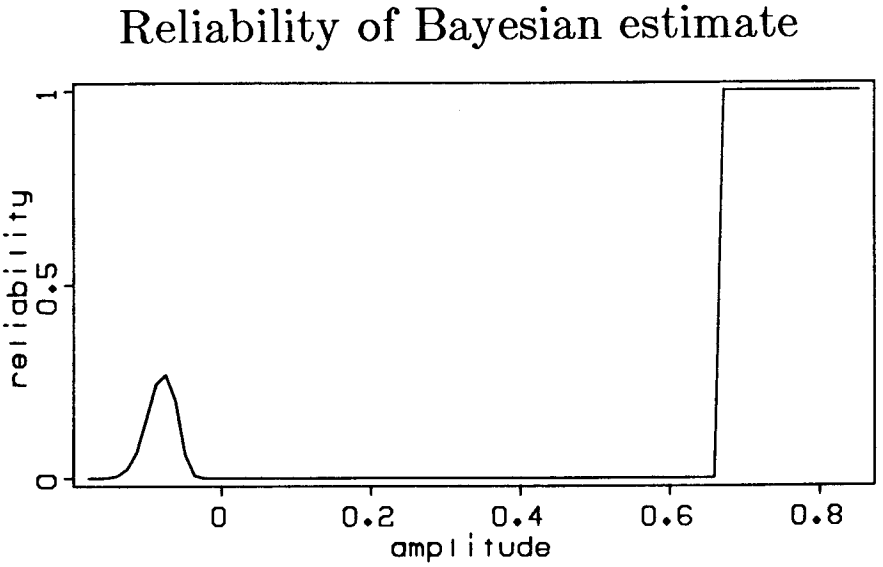


FIG. 1.12. Define the reliability of the Bayesian estimate as the probability that the percentage error is less than 5%. A reliability of 1 is perfect; 0 is completely unreliable. The peak with an amplitude near 0.8 has a reliability of almost 1; other events are much less reliable.

The reliability of the peak with amplitude 0.8 is seen to be almost unity; other amplitudes are much less reliable. Of course the 5% error can be replaced with other acceptably small percentages.

I extract those samples of the model whose estimated signal has greater than 95% reliability (Figure 1.13). Only the central correct peak survives, without sidelobes. I rescale the entire model to fit the data best according to objective function (1.3); the central peak changes to an amplitude of 1.0. This rescaling improves the fit without altering the simplicity of the model. Data (Figure 1.14) prepared from this model fit the original data even better than did the original least-squares inversion of Figure 1.5.

If our data events show coherence over time, then we must modify this procedure in order not to distort the interior of wavelets. In section 1.7 I discuss how the analytic envelope estimates the amplitude of locally monochromatic data and more correctly indicates the signal/noise content of samples. If we smoothly extract samples as a function of their analytic envelope, then a wavelet will be found to be equally reliable over its entire length.

We cannot not increase resolution indefinitely by adding more velocities to the model. Model traces that correspond to almost identical velocities will model almost identical hyperbolas; the inverted traces would be affected by the same signal and

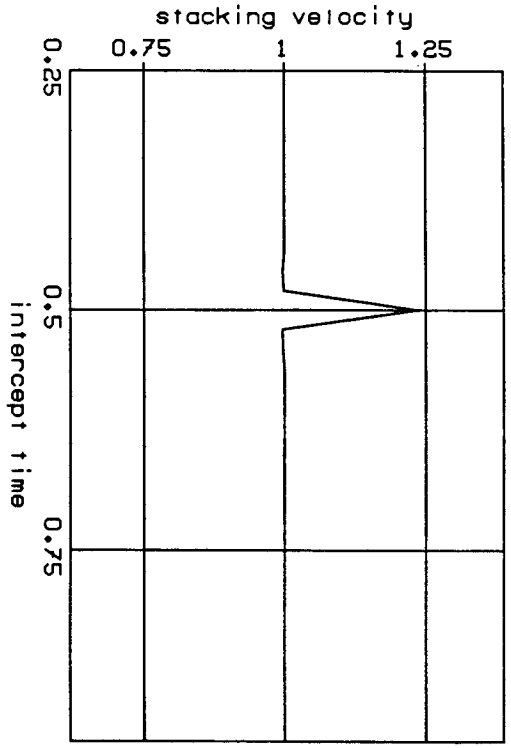
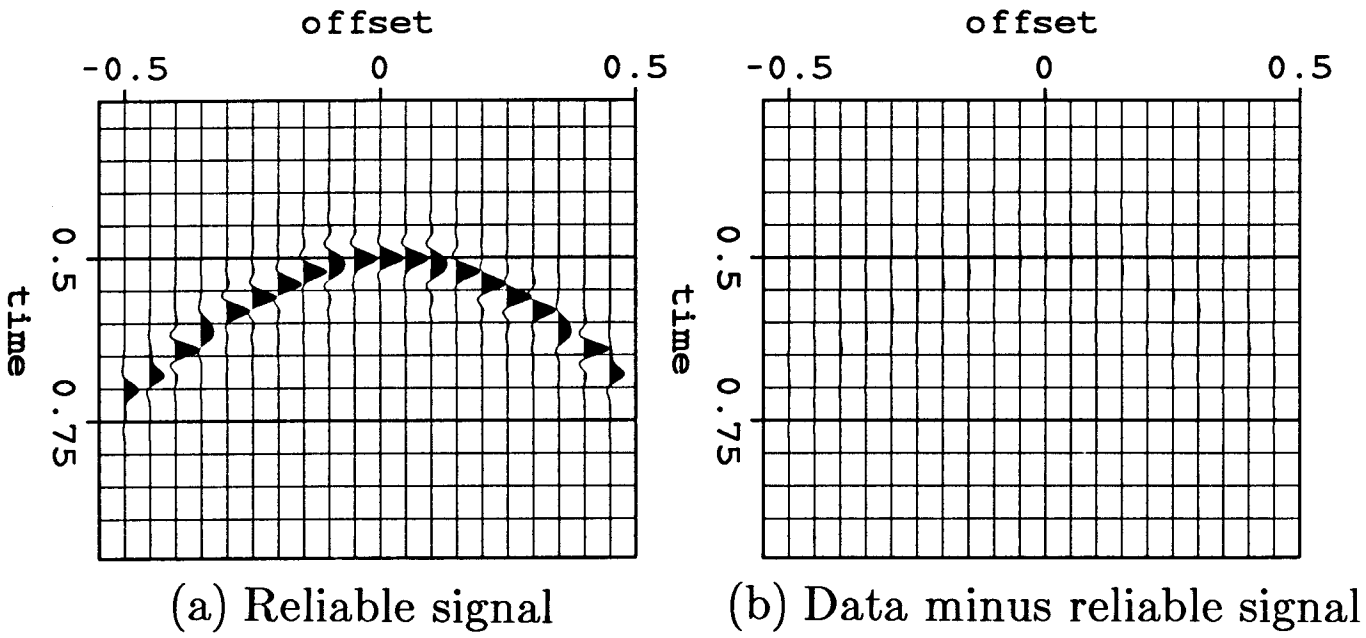


FIG. 1.13. A extraction from Figure 1.11 of those details with greater than 95% reliability eliminates all but the central correct peak, without sidelobes. When events are coherent over time, we must apply this estimate to analytic envelopes to avoid distorting the near-zero portions of wavelets.

noise. Inverted parameters with the same statistics are equally reliable. The suppression of noise increases resolution only to the degree that noise decreases resolution. One cannot recover information about signal that the data have not preserved. An inversion method that explicitly encouraged simple models might simply create misleading and non-existent details.

The reliable model of Figure 1.13 is "simpler" to the eye than the least-squares model of Figure 1.3; however the two solutions are little different according to objective function (1.3). In fact, the least-squares model minimized the objective function a little better than did the reliable model. Because the least-squares objective function incorrectly assumed a Gaussian model, the resulting model was much more Gaussian than necessary: energy was dispersed, and one large amplitude was replaced by many smaller amplitudes. Again, I do not recommend modifying the global objective function to assume other prior statistics; an incorrect guess of non-Gaussian statistics will



(a) Reliable signal (b) Data minus reliable signal

FIG. 1.14. (a) A recreation of the data from the reliable model of Figure 1.13 fits the data almost perfectly. (b) The residual events are negligible. This fit is certainly no worse than that of the least-squares stack in Figure 1.5, but the reliable stack is simpler and less likely to be dismissed as noise.

only bias the data more unpredictably. A maximum-likelihood objective function can maximize only the probability of a result, not its reliability. The details of a maximum-likelihood estimate should be judged by their sensitivity to noise in the data and should be kept or discarded accordingly. This strategy is much like usual interpretation methods, and so is intuitive to an interpreter.

1.4. LATERALLY ADAPTABLE VELOCITY STACKS

In this section I define two alternative NMO stacking operators, each of which elaborates on the algorithm of the previous section. For a variety of reasons, reflections are rarely predictable over the entire range of recorded offsets. For example, the arrival times and amplitudes of reflections in midpoint gathers are perturbed by lateral velocity anomalies, strong vertical velocity anomalies, angle-dependent reflection coefficients, and anisotropic rock velocities. Therefore, much information cannot be preserved by the simple hyperbolic model of the previous section.

Field (common-shot) gathers also contain roughly hyperbolic reflections, but stack less well than midpoint gathers. A dipping bed will move the peak of a hyperbolic

reflection from zero-offset in a shot gather, though not in a midpoint gather. On the other hand, a single shot gather exhibits the same source waveform on each trace. A single source simplifies many processes, including deconvolution, migration, and the removal of near-surface static shifts.

Some interpreters might hesitate to increase the number of parameters to describe reflections because the reduction in data size is a chief justification for use of the stack. On the other hand, purely hyperbolic stacks will prevent us from seeing much of the complexity of the information we record. For this reason I illustrate two possible strategies for a generalization of the least-squares stack.

The first generalization, a radial stack, attempts foremost to represent amplitude changes with offset clearly in the inverted model parameters. The new dimension of the model parameters corresponds roughly to the angle of reflection. The number of additional parameters is kept to a minimum, and the new signal parameters have useful physical meaning. Unfortunately, rapid changes in signal coherence are not captured so flexibly by this generization as we should like.

The second generalization, an offset-local stack, will stress fitting the lateral non-stationarity of reflections, so that as little signal as possible will be treated as noise. The additional signal parameters, however, must become too numerous to be examined directly. This model should be reserved for the application of noise extraction, when the description of the signal must be flexible. The radial stack can still be applied after noise extraction, so that non-stationary information is placed in the most manageable form.

1.4.1. A radial stack

Our first alternative is designed to improve the quality and interpretability of hyperbolic stacks. A good reconstruction of the data from the model will be of secondary importance.

If changes in amplitude with offset are of use to interpreters, then stacked sections should preserve this information. Of the causes for non-hyperbolic reflections, the most useful information comes from changes in reflection coefficients with offset. These changes, as well as those from ray bending, are most evenly distributed in arrival angles. For example, in a constant-velocity, stratified medium, constant arrival angles are equivalent to constant radial parameters, defined by $r = h/t$, where h is arrival offset and t arrival time (Claerbout, 1985, p.311). If an rms velocity is to vary with τ, then one could use a new r made dimensionless by division with the velocity (raypaths

must be straight).

Let us begin with stacked models that represent known ranges of r. As before, these stacked models should invert an easily interpreted modeling transform. Let the modeling equation window and sum hyperbolas over several different ranges of r:

$$\text{data}_{h,t} = \sum_{r} \text{weight}_{r,h,t} \sum_{v,\tau} \text{moveout}_{v,\tau,h,t}\, \text{model}_{v,\tau,r} \tag{1.9}$$

where

$$\text{moveout}_{v,\tau,h,t} \equiv \begin{cases} 1 & \text{if } t = \sqrt{\tau^2 + 4h^2/v^2} \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

$$\text{weight}_{r,h,t} \equiv W\left(\frac{r - h/t}{\Delta r}\right)$$

Let $W(\cdot)$ be a symmetric windowing function with unit area, such as the triangle function:

$$\Lambda(x) \equiv \begin{cases} 1 - |x| & |x| \le 1 \\ 0 & |x| > 1 \end{cases} . \tag{1.11}$$

This simple modeling equation allows us to visualize a hyperbola segment in the data for every point in the model. After multiplication with the moveout matrix, a single point in the model maps to a full hyperbola. The weighting matrix then tapers the amplitude of the hyperbola to zero away from the point's value of r. The final summation adds hyperbola segments from all overlapping ranges of r.

Figure 1.15 windows a common-shot gather supplied by Western Geophysical. The reflecting beds are not all horizontal, so this gather, unlike a midpoint gather, will not have perfectly symmetric hyperbolic reflections. Near-surface velocity anomalies distort the hyperbolas locally and make the reflection amplitudes vary with offset. The ground roll is present as strong, non-Gaussian additive noise.

Figure 1.16 displays a decomposition of the data into three ranges of r, without the summation over r shown in equation (1.9):

$$\text{decomposition}_{r,h,t} = \text{weight}_{r,h,t} \sum_{v,\tau} \text{moveout}_{v,\tau,h,t}\, \text{model}_{v,\tau,r} \tag{1.12}$$

Δr is set to 1. Three velocities, 2.5, 3.0, and 3.5 km/s, are used to create three stacked traces. I shall not examine the quality of the stacked traces here: many stacked gathers must be compared with the geology they describe before these stacks can be properly interpreted. As was done with the NMO stack, I do not ask whether the stack describes earth properties better than the gather does—it cannot. Rather,

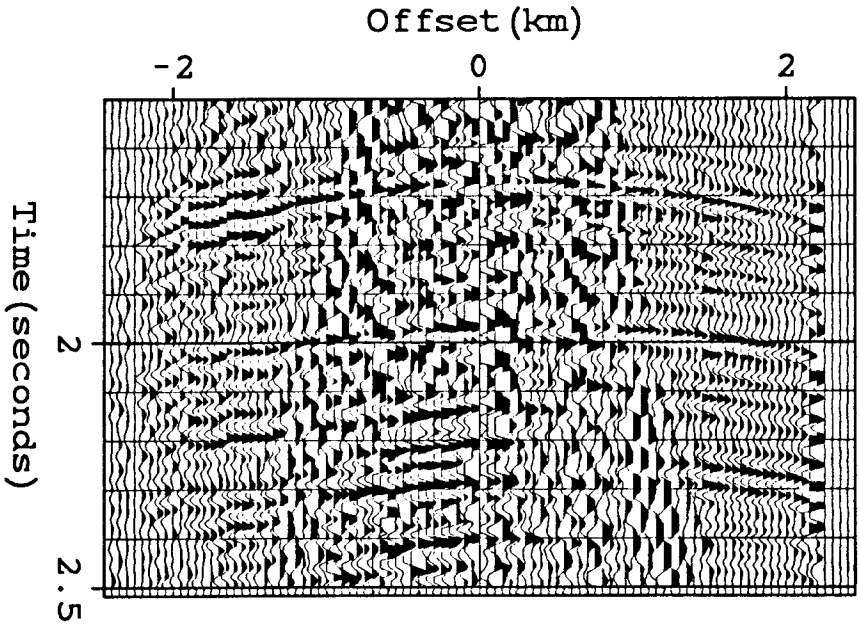


FIG. 1.15. A common-shot gather supplied by Western Geophysical. Reflecting beds are not perfectly horizontal. Near-surface anomalies distort hyperbolas locally and make amplitudes vary with offset. Ground roll is strong additive noise.

the stack should reduce the number of data samples that must be interpreted simultaneously; the stack should model only the interesting reflections. These conditions are satisfied if the radial model fits the reflections well, does not create false events or obscure important events.

When the panels of Figure 1.16 are summed together as shown in equation (1.9), the coherence of the hyperbolic reflections changes visibly with offset (Figure 1.17). The recreated data fit the original data very well: subtracting the sum from the original data leaves only ground-roll and weak hyperbolic events behind (Figure 1.18).

A trade-off exists between the complexity of the model and the fidelity to the data. The three ranges of r do not merge as well as they might; the changes in coherence seem abrupt. However, the NMO stack assumes a single infinite range of r (r =0; $\Delta r = \infty$) and could not fit the data so flexibly. As the number of r increases, the reconstructed data become freer to adapt to lateral changes in coherence. The number of r can be increased until the coherence changes arbitrarily fast, though the stack will first become too large for examination. If signal coherence changes too fast,

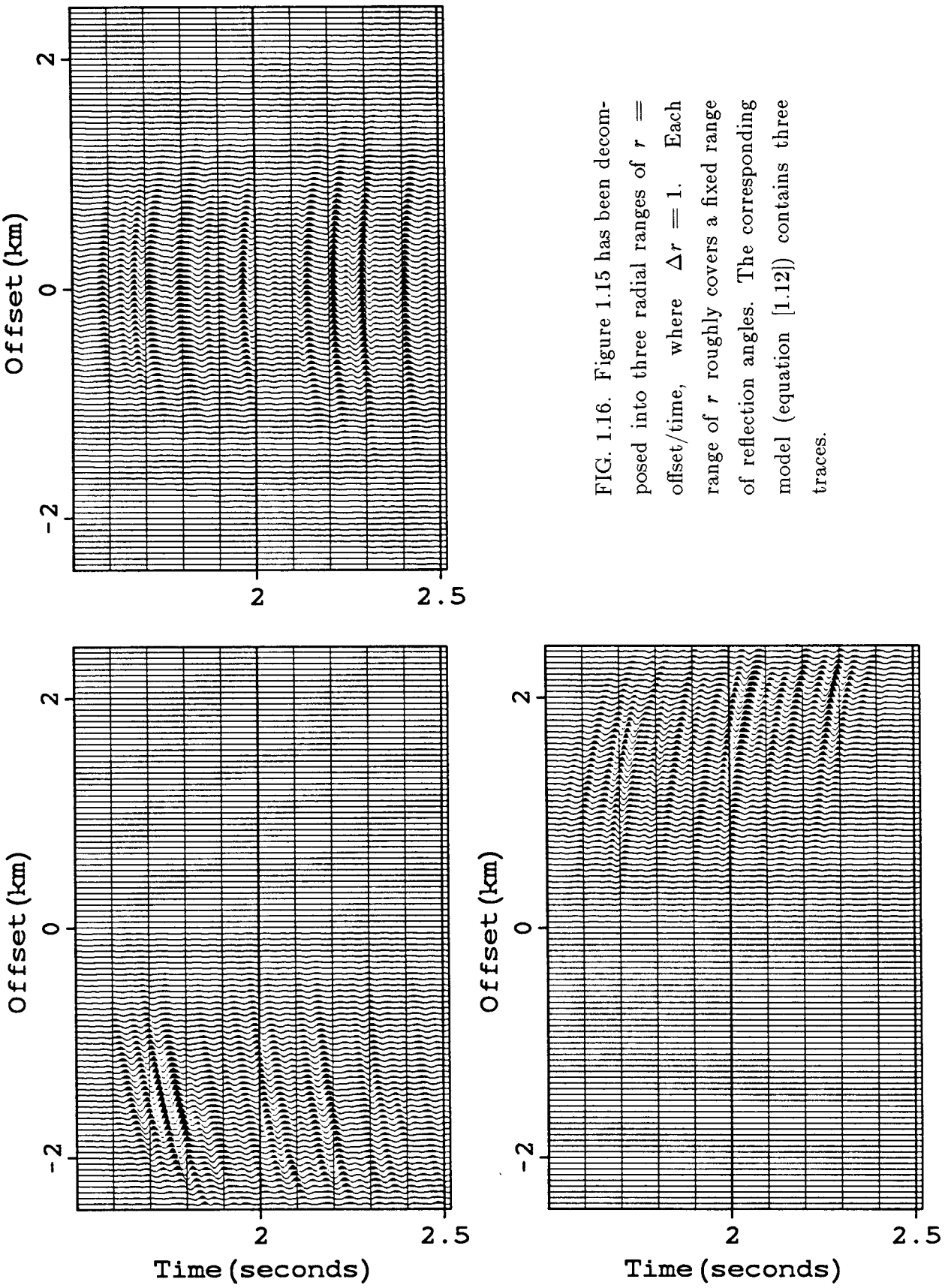


FIG. 1.16. Figure 1.15 has been decomposed into three radial ranges of r = offset/time, where $\Delta r = 1$. Each range of r roughly covers a fixed range of reflection angles. The corresponding model (equation [1.12]) contains three traces.

however, it will appear incoherent.

If we wish to extract noise, such as the ground roll, then a preceding extraction of hyperbolic signal should be very flexible, though not so flexible that the modeled signal resembles noise. Unextracted signal will interfere with a subsequent extraction of noise. (If the extraction of noise is reliable, little residual signal should actually be lost. However, we might be prevented from extracting any noise.) To account for most of the non-stationarity of the hyperbolic reflections, we must require considerably more samples in the model. I recommend that, rather than increase the number of r arbitrarily, we use the offset-local stack of the following section.

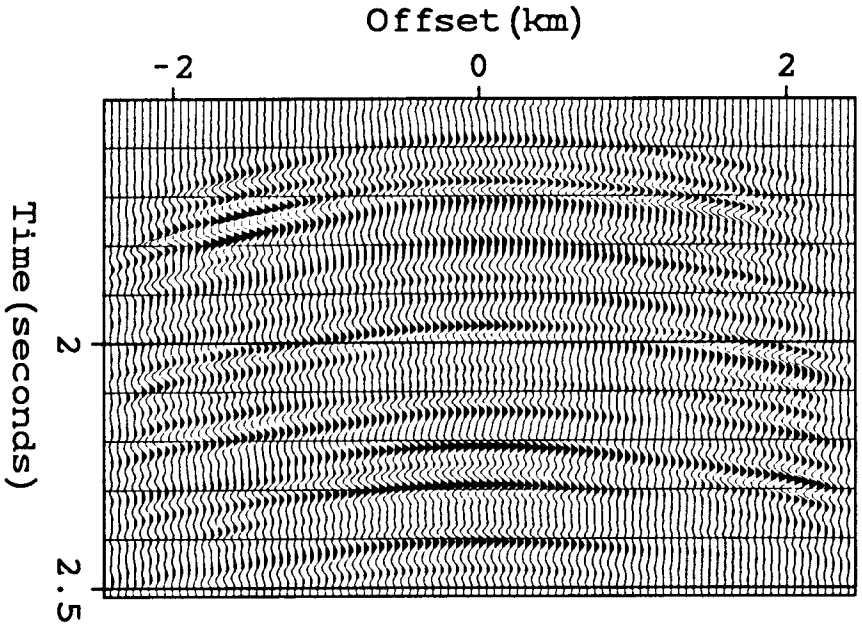


FIG. 1.17. The three ranges of r in Figure 1.16 sum together as shown in equation (1.9) to create a flexible model of the data. The three ranges of r do not merge as well as they might; the changes in coherence seem abrupt. As the number of r increases, however, the reconstructed data become freer to adapt to lateral changes in coherence.

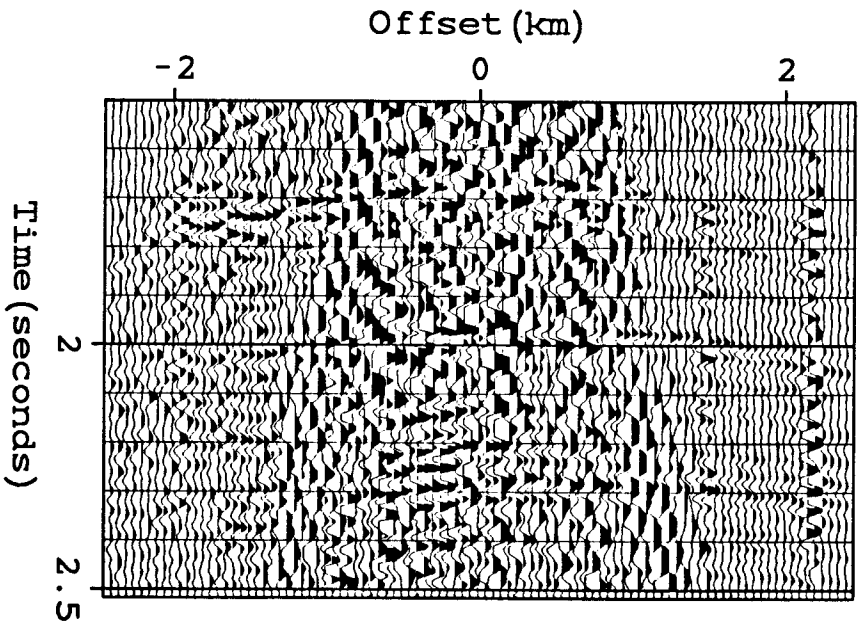


FIG. 1.18. A subtraction of the modeled data (Figure 1.17) from the original data (Figure 1.15) leaves only ground-roll and weak hyperbolic events behind. A finer sampling of r would have left fewer hyperbolic events.

1.4.2. An offset-local stack

If the object of our stack is not an interpretable model, but a flexible description of the signal for signal/noise separation, then the number of modeling parameters is unimportant. Although physical variations in amplitude and non-hyperbolicity are evenly distributed in the radial coordinate r, the redundancy (the data sampling) necessary for the discrimination of signal and noise is evenly distributed in offset. The limited sampling over offset limits the lateral resolution of anomalous amplitudes whether or not the wave propagation has preserved them.

Let us define an alternative equation for the modeling of the hyperbolic reflections.

$$\text{data}_{h,t} = \sum_{v,\tau} \text{moveout}_{v,\tau,h,t} \sum_{h'} \text{smooth}_{h,h'} \ \text{model}_{h',v,\tau} \tag{1.13}$$

"smooth" is defined as a low pass convolution such as

$$\text{smooth}_{h,h'} = W\left(\frac{h - h'}{\Delta h}\right) \tag{1.14}$$

Each point in the model is mapped first to a tapered line segment whose width is

constant for all arrival times. The moveout summation then maps these line segments to segments of hyperbolas. These hyperbola segments overlap so continuously that seams cannot appear. This model will contain considerably more samples of h' than equation (1.9) contained of r. Stacks of constant h' will also have less physical basis for interpretation.

Figure 1.19 shows the result obtained when the data are modeled with this equation. I used a Butterworth lowpass filter (Oppenheim and Schafer [1975]) with Δh approximately equal to ten traces. A damped least-squares algorithm was used, without an extraction of reliable events. Figure 1.19 thus shows those events the model is capable of fitting. Events are indeed more continuous and more adaptable to the changing coherence of the hyperbolic signal than were those obtained with the radial stack. No ground roll has been extracted as locally hyperbolic. Figure 1.20 displays the data minus Figure 1.19. Fewer hyperbolic events remain than in the residuals of Figure 1.18 after the radial stack. Thus, this second alternative will leave less residual noise to interfere with a reliable estimate of the noise.

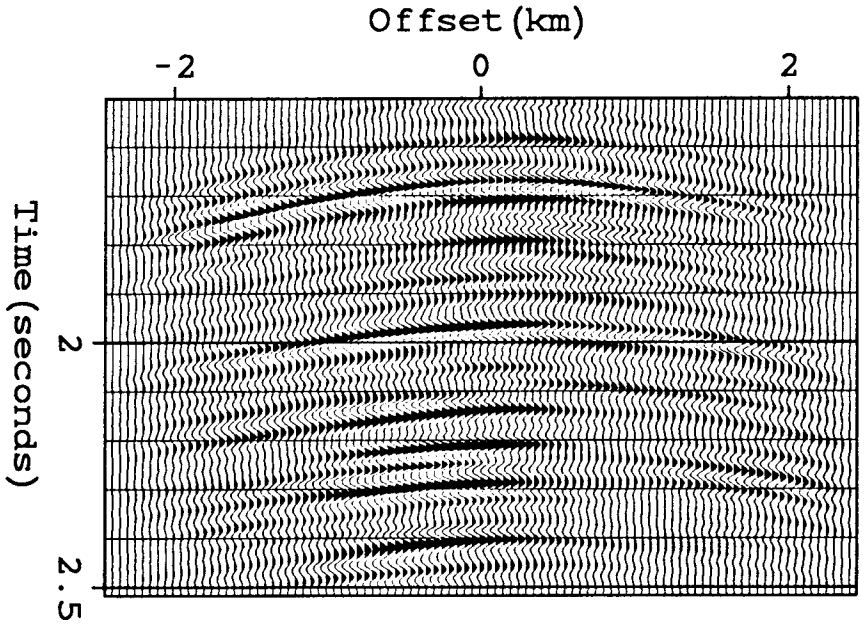


FIG. 1.19. An offset-local stack (fitting the data with equation [1.13]) fits the data more smoothly than did the radial stack.

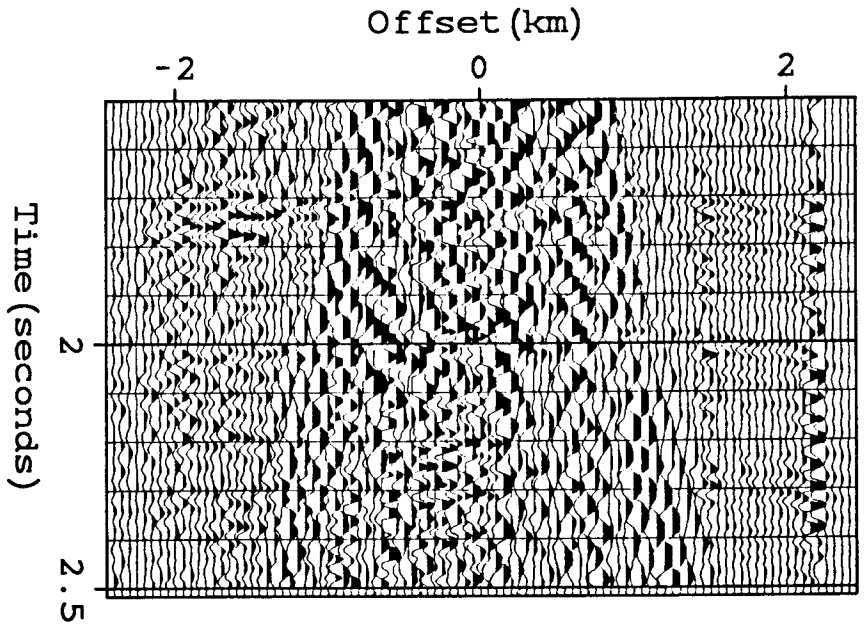


FIG. 1.20. A subtraction of Figure 1.19 from the data of Figure 1.15 leaves fewer hyperbolic events than are shown in Figure 1.18. The offset-local stack describes lateral variations in reflections more flexibly than do radial stacks.

1.5. EXTRACTING GROUND ROLL AND OTHER NOISE

After locally hyperbolic reflections have been extracted, ground roll and other incoherent noise can be extracted by a similar procedure. The ground roll does have visible coherence, but it is undersampled and strongly aliased with offset. Moreover, ground roll shows a strong phase dispersion with time. This apparent incoherence prevented the ground roll from interfering with the modeling of hyperbolic reflections. Unfortunately these complications make the coherence of the ground roll difficult to model for extraction.

To our advantage, however, the ground roll is relatively isolated in time and offset and appears very non-Gaussian. (See Chapter 2 for a discussion of the importance of non-Gaussianity.) By modeling such noise with the identity transformation, we can extract them as if they had no lateral coherence at all. Each sample of the noise can be assumed independent of the others. Above all, we should avoid extracting the locally hyperbolic events with the ground roll. Because the offset-local stack fits the hyperbolic signal most flexibly, let us extract from the residual data only those samples that contain a small percentage of events that can be modeled by the offset-local stack.

Define the residual signal (residual hyperbolas) to be all events that can be modeled by the offset-local stack (1.13) in the residual data (Figure 1.20). If the previous signal extraction eliminated some events as unreliable, then the amount of residual signal will be large. We accepted all least-squares inverted events as reliable; nevertheless, residual signal remained for two reasons. First, convergence was not complete for the minimization of the least-squares objective function. Second, the signal damping term (e.g. the second term in equation [1.3]) tried to minimize energy in the estimated signal.

Because we can overestimate the residual signal, the extraction of reliable noise is parallel to the extraction of reliable signal. The estimation of signal and noise pdf's, however, can avoid the cross-entropy deconvolution used to extract signal in section 1.3.2. First, estimate the signal pdf $p_{s'}(\cdot)$ from a histogram of all events that equation (1.13) fits to these residuals. Then estimate the noise pdf $p_{n'}(\cdot)$ from a histogram of the remaining residuals. Extract as reliable noise (ground roll) all samples of the data that cannot be easily described as a sum of hyperbola segments.

As I explain in section 1.7, the coherent wavelets can be extracted if the analytic envelopes are treated as the true amplitudes. I accepted those residual samples whose analytic envelopes contained less than 5% hyperbolic amplitude with greater than 95% probability. The extraction was also smoothed somewhat over time. Figure 1.21 displays the extracted events, which contain, in diminishing strength, ground roll, overamplified traces, and static shifts of strong reflections. Subtracting the unwanted events from the original data of Figure 1.15 exposes much of the previously hidden hyperbolic reflections (Figure 1.22). Yet most samples of the data have remained untouched. Any bias introduced into the data by the assumptions of equation (1.13) appears only in the interpolated values left after the unwanted noise has been extracted.

1.6. INTERPOLATION AS SIGNAL EXTRACTION

A seismic process that minimizes an objective function can usually be defined for irregularly sampled data. For instance the least-squares objective function of equation (1.3) can be constrained so that summations occur only over recorded samples of the data. On the other hand, regular sampling often allows more efficient programming than does irregular sampling, as for wavefield extrapolation or migration. Such programs might create artifacts if unrecorded samples were replaced by zeros. To solve this problem, an interpolation can fill a regular grid with values that are consistent

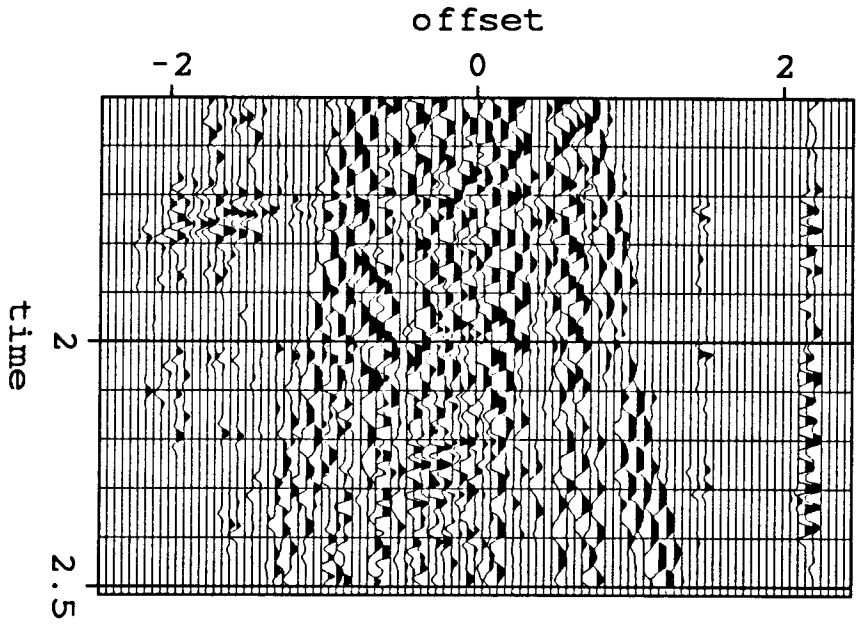


FIG. 1.21. I extract those samples containing a sufficiently low percentage of signal with a sufficiently high probability. Ground roll dominates, but over-amplified traces appear in the right-most traces, and strong events with static shifts appear in the upper-left corner. (Analytic traces were used to calculate the true amplitudes of the oscillations.)

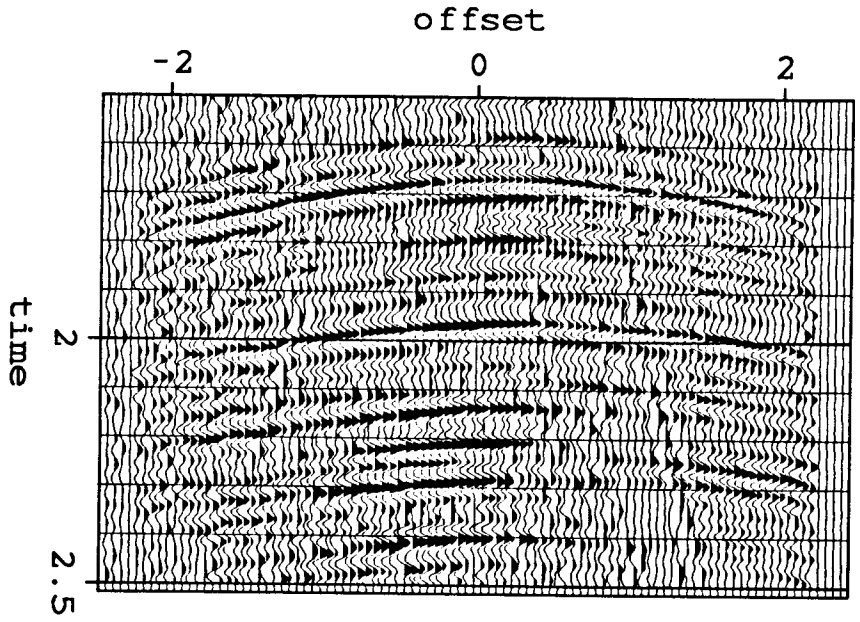


FIG. 1.22. Subtracting the unwanted events of Figure 1.21 from the original data of Figure 1.15 uncovers much of the previously hidden hyperbolic events. Most samples remain untouched. The assumptions of the offset-local stack affect only the values interpolated under the extracted noise.

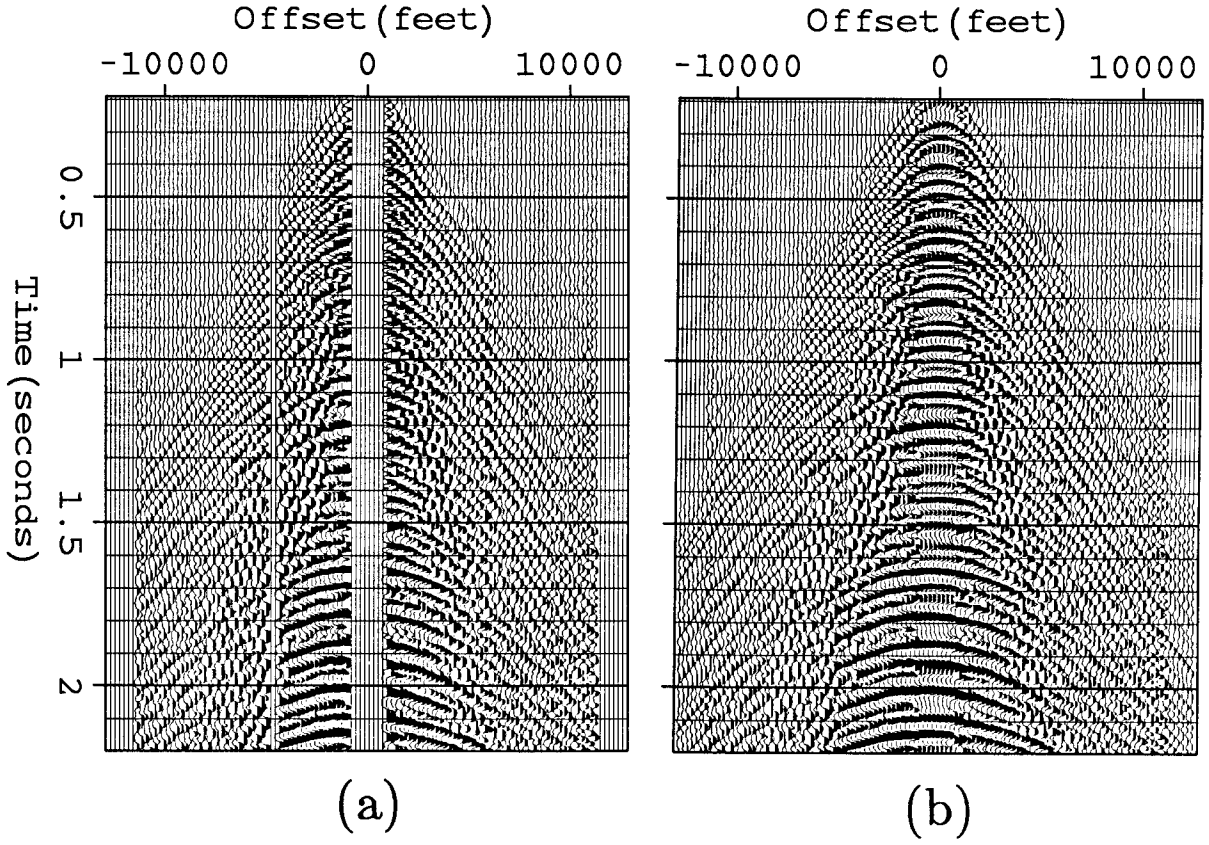


FIG. 1.23. (a) A shot gather (supplied by Western Geophysical Corp.) has been interpolated with zeros at inner and outer offsets. Two intermediate traces have been replaced with zeros as well. Many wavefield processing techniques would create artifacts by treating such zeros as genuine. (b) I interpolate instead with data reconstructed from a least-squares inversion of the offset-local model. The model used one constant velocity; the inversion constrained only the recorded traces and did not constrain the model at all.

with recorded values.

As we saw in section 1.5, if only a few traces are missing, they can be interpreted as additive noise that exactly cancels the reasonable events one expects. Such noise was extracted along with ground roll in the data of Figures 1.15 and 1.21. If not too many consecutive traces are missing, the preliminary estimation of signal will not suffer.

For example, Figure 1.23a contains a shot gather (field profile). The inner offsets were not recorded. Zeroed traces have been added at the abrupt truncation at the outer offsets. Two consecutive traces have been zeroed at an intermediate offset. Most artifacts of wavefield extrapolation could be avoided if these zeroed traces were

replaced by values that were roughly consistent with the locally hyperbolic reflections.

Let us find a model to describe the signal in the missing traces and then invert for this signal as a constrained optimization problem. Describe the signal by the offset-local model of equation (1.13). Find the least-squares inverse model that best fits the recorded data samples and reconstruct the data from this model. Use this reconstructed data to fill in the unrecorded traces.

Figure 1.23b displays such an interpolation. For the offset-local model, I used a single, constant velocity and a smoothing operator approximately 20 traces wide. The model was left entirely unconstrained, and the least-squares objective function was minimized by only a few gradient iterations. Even with such crude approximations, the unrecorded traces have been adequately interpolated so that abrupt truncations in events are avoided. The interpolated traces are not numerous enough to bias a later velocity estimation. The tails of the interpolated hyperbolas are not perfectly consistent, but they will avoid artifacts more effectively than do zeros.

The adjoint (or back-projection) of the offset local stack is equivalent to a normal moveout correction followed by lateral smoothing. Further back-projections do little to change the resulting model. Each back-projection must be scaled by a constant, calculated from simple summed products, to fit the recorded traces best.

The least-squares solution is linear, and for most applications, has no need of the robust procedures of section 1.3. The number of adjacent traces that can be interpolated is limited by the width of the events described by the model—in this case 20 traces. If events are aliased enough, linear least-squares methods can also create false events. Robust methods should be used when many traces must be interpolated and whenever the model appears under-determined.

1.7. EXTRACTING COHERENT WAVELETS

A non-linear extraction of signal must consider whether model samples should contain coherence over time, so as not to distort the interior of wavelets. For example, the single trace of Figure 1.24a, like much of seismic data, appears locally monochromatic. The true amplitude of the central wavelet might be said to be near 1 over its entire length, including the near-zero portion between the large peaks. As we have noticed previously, the most reliable estimates of signal tend to be found in samples with the highest amplitudes. Assume that all amplitudes over 0.7 are found to be reliable. A sample-by-sample extraction would find the near-zero portion of a wavelet to be as unreliable as a low-amplitude interval filled with noise. As in Figure 1.24b, peaks

are extracted, but not the low-amplitude intervals between peaks. The wavelet's shape and frequency content are severely distorted.

The simplest solution would simply smooth extractions over time. Reliable and unreliable samples of the model could be marked by ones and zeros in a matrix with the dimensions of the model. The matrix could then be smoothed over time before a multiplication with the model. However, the smoothing might be insufficient, as the one shown in Figure 1.24c. Broader wavelets would be penalized. In short, such smoothing requires that many prior assumptions be made about the expected coherence of the wavelet.

Alternatively, the time coherence could be included in the original modeling equations. One could attempt to solve the deconvolution problem as part of the full inversion. For example, a linear prediction-error filter can first make samples of the model white (independent for Gaussian random variables); the estimated wavelet can be reconvolved with extracted samples. However, this solution is much too elaborate and trouble-prone for our purposes. A prediction-error filter must make many restrictive assumptions about the wavelet (stationary, short, etc.). Instead let us examine envelopes of the wavelets.

Seismic traces measuring ground motion tend to appear locally monochromatic: a segment has a single mean frequency with a low-frequency modulation as its envelope. The square of the envelope gives the true energy of displacement. For instance, if the data contained just a sine wave, we would consider the true amplitude to be constant. A wavelet appears locally to be just a modulated sine wave. A zero at the interior of a wavelet corresponds to nearly the same amplitude displacement as those of nearby peaks.

For example, a phase shift will preserve the local information in a seismic trace while changing its sample-by-sample amplitudes. A linear 90° phase shift (Fourier sines to cosines and cosines to sines) causes the zeros of wavelets to map to peaks and peaks to zeros, as shown in Figure 1.25a. A sample-by-sample extraction of a phase-shifted wavelet would fail where previously it had succeeded, and vice-versa (Figure 1.25b). When we undo the phase shift of the extracted samples (Figure 1.25c), we find that the peaks are poorly extracted and that the highly dipping portions are much better represented. Intermediate phase shifts would maximize the visible amplitude of the intermediate portions of wavelets and so would improve the extraction elsewhere.

Let us imagine a much larger set of model parameters—the previous set of traces plus these traces shifted over all phases between plus and minus 90°. Every sample in

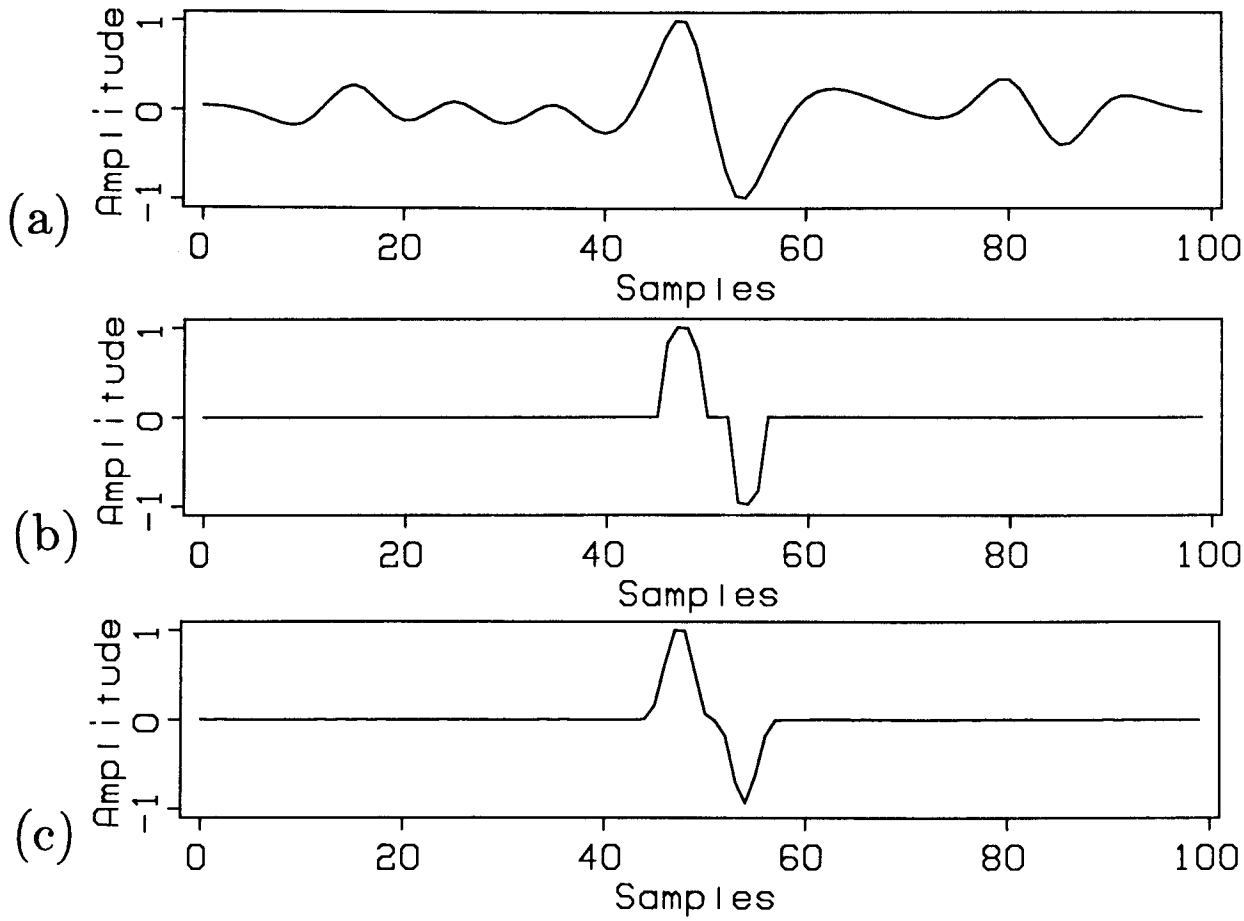


FIG. 1.24. (a) This short trace, like much of seismic data, appears locally monochromatic. (b) Extract all samples with amplitudes greater than 0.7. The central wavelet is greatly distorted in shape and frequency: peaks are extracted, but not the low-amplitude interval between. (c) Merely smoothing the extraction over time might be insufficient for the wavelet's original character to be retained.

this enlarged model space is a linear function of the original model, so signal and noise remain additive. We can visualize these shifts geometrically. Figure 1.26 shows an "analytic" version of the trace in Figure 1.24a: it is a three-dimensional spiral with a right-hand twist. One projection of this spiral corresponds to the original trace. A perpendicular projection corresponds to the trace with a 90° phase shift. All other phase shifts can be obtained from planar projections of this spiral at other angles. Define this "analytic" trace as a complex function of the original model.

$$\text{analytic}(t) = \text{model}(t) * [\delta(t) + \frac{i}{\pi t}] \quad . \tag{1.15}$$

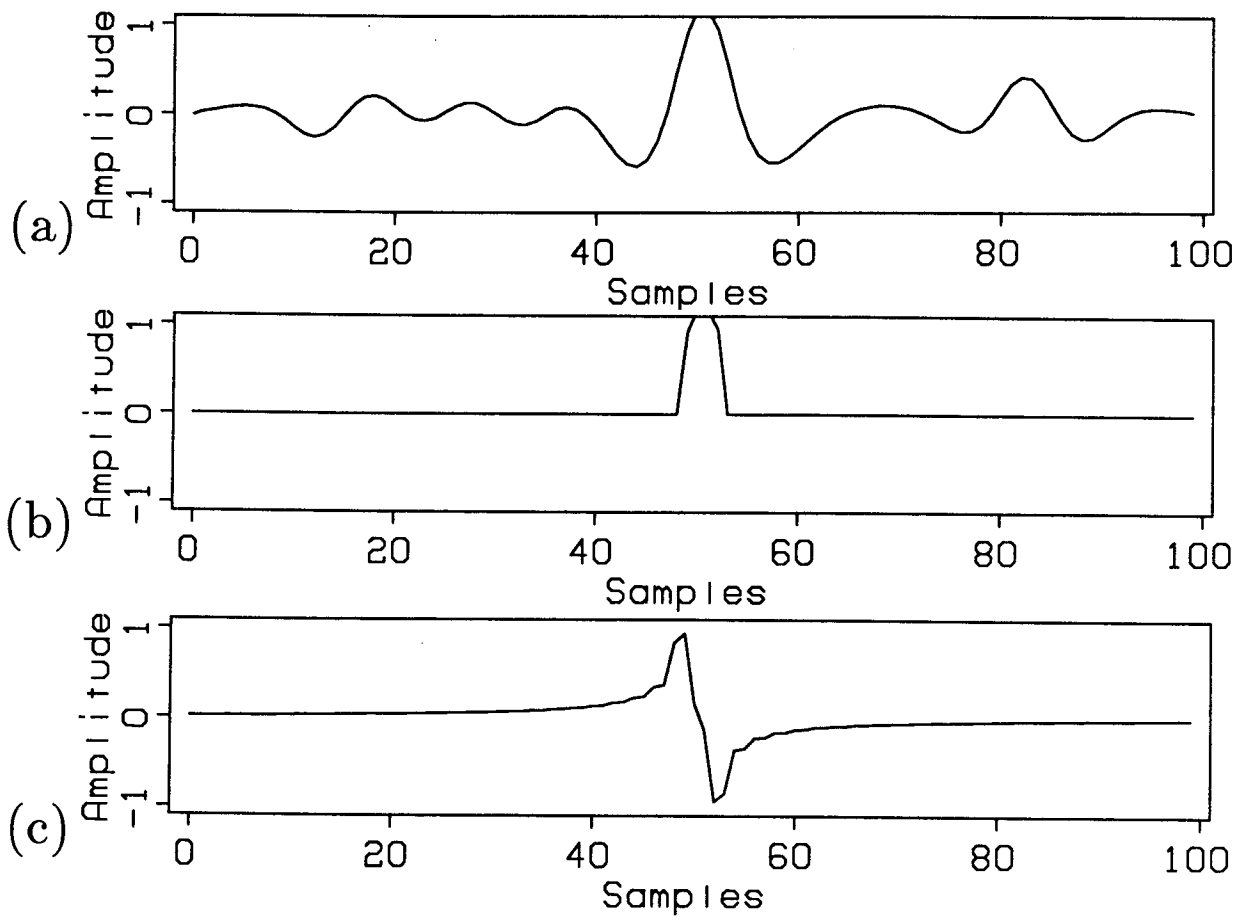


FIG. 1.25.. (a) A 90° phase shift of Figure 1.24a creates this trace. Peaks map to zeros, and zeros map to peaks. (b) A sample-by sample extraction succeeds where previously it failed, at the interior of the central wavelet. (c) Undoing the phase shift of (b) gives poor reconstruction of the wavelet. Though the interior of the wavelet is much better than in Figure 1.25b, the peaks are now distorted. Intermediate phase shifts would succeed and fail elsewhere in the wavelet, but no one phase shift is enough.

The analytic trace is the original trace plus a purely imaginary trace with a 90° phase shift (the Hilbert or quadrature transformation). The real part of the analytic trace is precisely equal to the original trace. Equivalently in the Fourier frequency domain

$$\text{analytic}(f) = \begin{cases} 2 \cdot \text{model}(f) & f > 0 \\ \text{model}(f) & f = 0 \\ 0 & f < 0 \end{cases} . \tag{1.16}$$

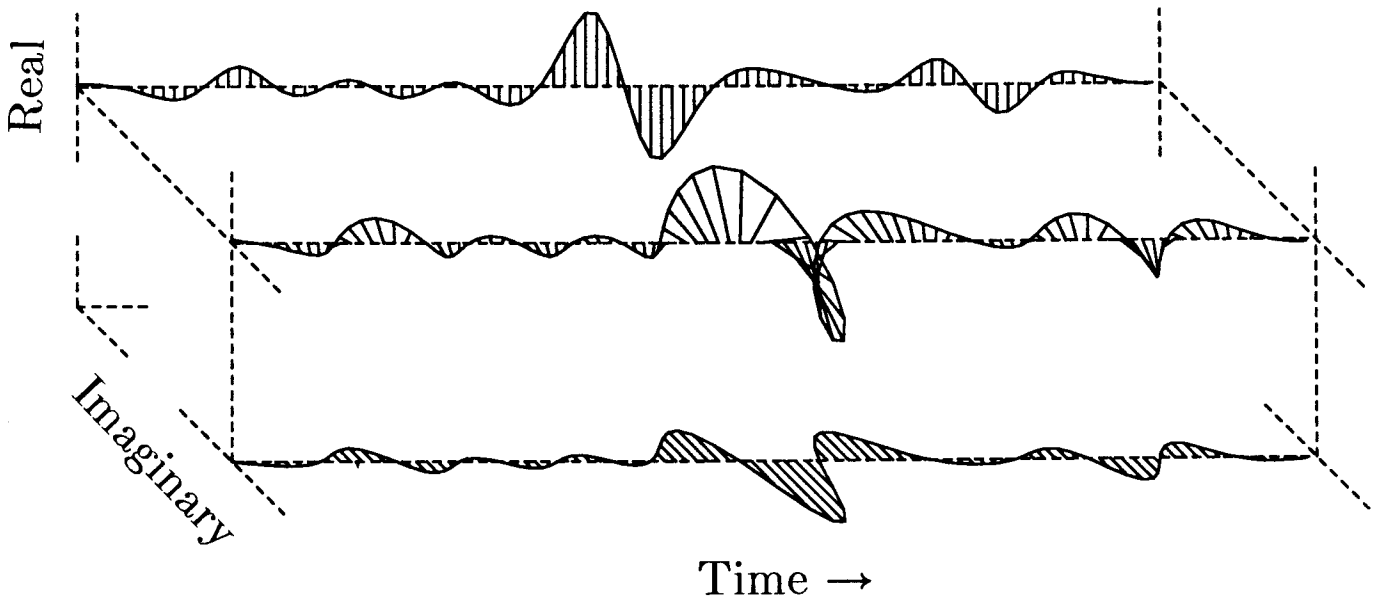


FIG. 1.26. The trace of Figure 1.24a corresponds to a unique analytic trace, a complex spiral with a right-hand twist. The real part of this trace projects into the original trace of Figure 1.24a; the imaginary part is a 90° phase shift. The zeros of the real part correspond to the peaks of the imaginary, and vice-versa. All other constant phase shifts can be obtained from planar projections of this spiral. The analytic envelope, the true magnitude of the analytic trace, is the distance of the spiral from its axis.

To calculate the analytic trace we need only apply the Fourier transform, zero the negative frequencies, double the positive, and invert the Fourier transform. See Bracewell (1978).

Let us not estimate the reliability everywhere but rather only for the phase shift at which a time sample has the largest amplitude. This largest amplitude appears as the distance of the analytic spiral from its axis in Figure 1.26. If this time sample is reliable enough for one phase shift, we can treat it as reliable for all phase shifts. The phase shift does not add or destroy information; it merely reveals the true amplitude of the monochromatic wave at that point. The maximum amplitude obtained by the phase shifts is called the "analytic envelope," the complex magnitude of the analytic trace.

$$\text{envelope}(t) = \pm[\text{analytic}(t)\cdot\text{analytic}^*(t)]^{1/2} \tag{1.17}$$

The asterisk indicates the complex conjugate. The sign of the original sample should

be retained for a minimum phase shift.

We can calculate the reliability of a model sample from its analytic envelope. To estimate the statistics of all linearly phase-shifted traces, one could take histograms from the original traces, or from several phase-shifted traces. Assume that we have chosen the linear phase shift for which a sample's amplitude equals its envelope. We can then estimate the probability that an envelope contains a small fraction of noise, and save the actual model amplitudes accordingly.

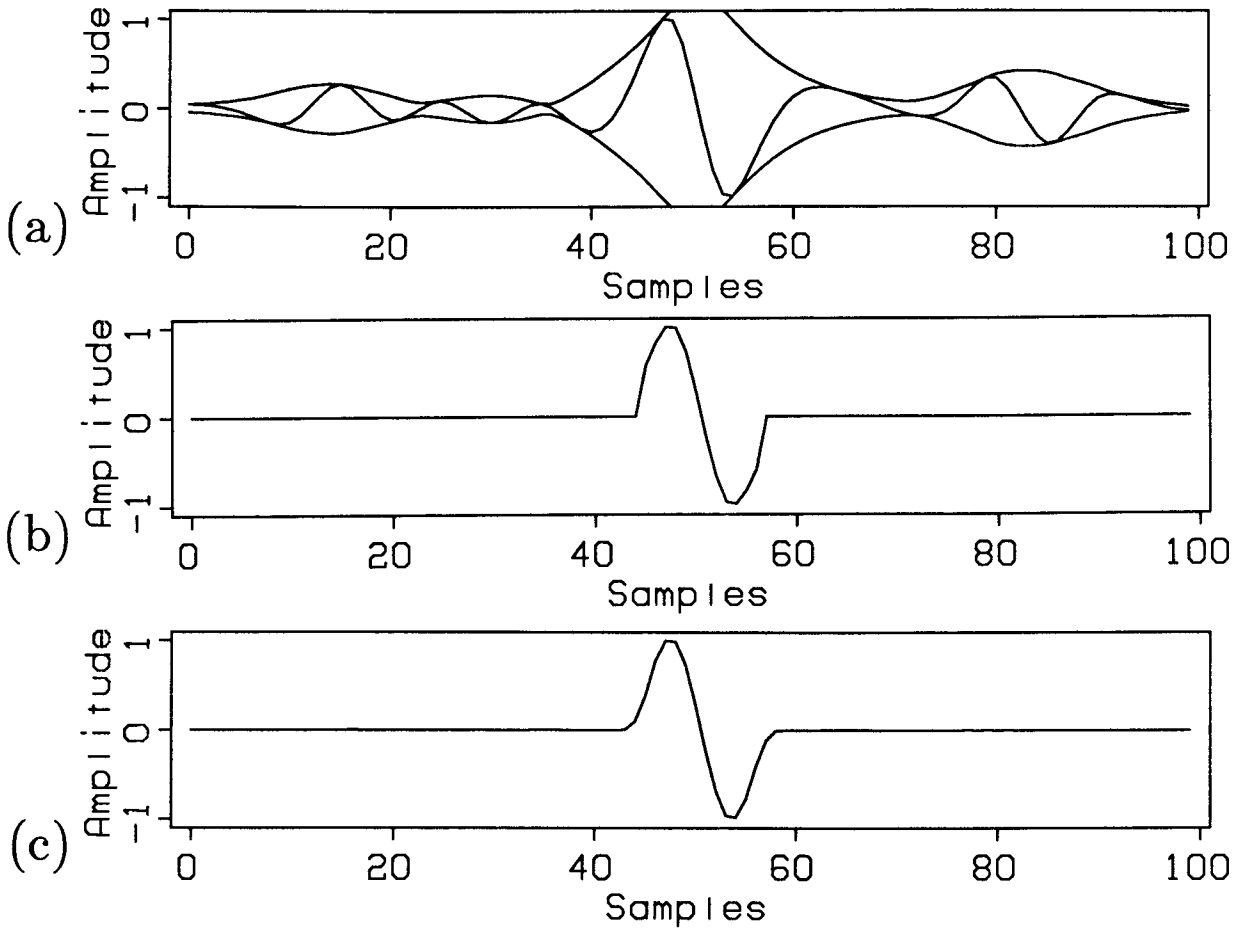


FIG. 1.27. (a) The "analytic envelope" (plotted with the original trace) accurately represents the true amplitude of locally monochromatic traces. (See Figure 1.26.) (b) If we extract samples as a function of the analytic envelope, the interior of the central wavelet is preserved with little distortion. (c) Some smoothing of the extraction reduces the sharpness of the wavelet's outer edge and does not affect the interior.

For example, Figure 1.27a shows the analytic envelope and the original trace of Figure 1.24a. Assume again that all amplitudes over 0.7 are found to be reliable. If we extract samples as a function of the analytic envelope, we find that the central wavelet is preserved with little distortion (Figure 1.27b). A little smoothing of the extraction reduces the sharpness of the wavelet's outer edge but does not affect the interior (Figure 1.27c). The largest, most reliable wavelet has been isolated from its weaker neighbors, as we desired.

1.8. DEFINING AN INVERSION

We shall be examining many varieties of signal and noise in the following chapters. Let us first summarize the methods of this chapter without referring to the particular modeling transformation.

An inversion must satisfy several concurrent goals. Inversion should place seismic data in a form that is most easily compared to earth properties. The inversion should efficiently describe the useful recorded events, the signal, with the fewest independent parameters at unexpected values. The signal should be inverted reliably, so that it does not accidentally describe useless events, the noise. Define inversion then as a choice of model and parameters that efficiently and reliably describe the recorded data.

1.8.1. Limits to inversion

The processing of seismic data always takes us farther from the physical properties of the earth: processing tends to destroy, not add information. We must remember this disagreeable fact before we claim to have "inverted" intrinsic earth properties from indirect observations. In fact, processed data observe the earth less directly than do the original data. A seismic recording is by itself a property of the earth, strongly related to other physical properties, but never a mere function of these: properties of the earth are defined by the methods of their measurement.

Processing can, however, place recorded information in a form that is more easily compared to other earth properties. An interpretation must predict unknown earth properties from others. Processing can reduce the number of observations and comparisons that must be made, but it cannot tell us what we did not measure. Rather than avoid the word inversion, let us define it in an unmisleading way that will include already useful methods.

First, an inversion must choose a mathematical model for the data set, with a set of as yet undetermined parameters. Such a model describes an interpreter's schematic

view of the earth. Second, an inversion must choose parameter values that describe the important events in the data. Many sets of parameter values might model the data equally well; they can be distinguished only by their statistical properties. For this reason the model parameters should be defined as random variables. Probabilities measure such uncertainties and ignorance explicitly.

When we speak of the probability and reliability of a parameter, we are quantifying the acceptable variations of the parameter and the extent to which the data should affect the parameter. This statistical language leads us to speak of the data set as a particular outcome of the random variables. Estimation theory can optimize the statistical trade-offs and find the most acceptable values of the random variables. Let us now describe the desirable properties of inversion in statistical terms.

1.8.2. The goals of inversion

To describe data efficiently, an inversion should describe only the most important features of the data and give unexpected values to the fewest physical parameters. These goals are possible if the parameters describe independent events in the data, and if the parameters describe only the most reliable events (those which are not likely to be noise).

Let us design our model so that an interpreter can assume that the parameters are statistically independent. In other words, one parameter's value should not be prejudiced by values chosen for other parameters. The modeling equations should be modified when the interpreter is unwilling to see the parameters treated independently of one another. For instance, rocks at various depths show dependent seismic velocities because rocks tend to appear in homogeneous layers. Knowing the velocity at 1000m improves the chances of accurately predicting the value at 1005m. Reflection coefficients, however, are very independent with depth: reflection coefficients detect the edges of layers, and layers rarely appear in a predictable order. The reflection coefficient at 1000m tells little or nothing about the coefficient at 1005m. The assumption of independence should be treated as a *prior* assumption, which describes properties of the modeling equations and their random variables, not of the measured data.

After choosing the modeling equations, we must choose default parameter values that describe the "simplest" physical structure. If other values do not describe the data sufficiently better, then we prefer to assume this simplest model. Perturbations of this default model are immediately recognizable as information extracted from the data.

Their independence allows the important data events to be reduced to the smallest number of parameters at other than default values. If parameters were predictably dependent, they could contain some redundancy and could be further reduced in number by transformation. Most importantly, independence allows an interpreter to examine the inverted parameter details locally, without being obliged to look for global patterns.

Let us now make a few crucial terms more precise. Define a model "detail" as a parameter that has been changed from its default value. Let an "event" be those differences that appear in the modeled data when a parameter detail is added. For example the spike in Figure 1.13 is a model detail; the hyperbola in Figure 1.14 is the corresponding event. Details must be independent. If the modeling transformation is linear, then events will also be independent, though not necessarily distinguishable from one another. The recorded data can be said to contain an event if a particular modeled event easily fits the recorded data (a fit measured statistically). Let the "signal" be those details and events that easily describe the data for a particular physical model. Let "noise" be that component of the data that can be easily described by a much simpler model, particularly one without any physical coherence at all.

Define inversion as a choice of reliable model details, both signal and noise, that efficiently describe the recorded data. A "reliable" signal detail should not create an event that can be easily described by a chance combination of noise events. Similarly, a reliable noise detail should model signal events poorly. An iterative inversion should converge on the most reliable model details first.

1.9. CONCLUSIONS

In this chapter I covered the essential elements of a signal/noise separation algorithm. The following two chapters will focus in turn on the extraction of non-stationary reflections and on the reliable inversion of rock parameters. In Chapter 2, I shall explore principles and examples that will help us to choose models for the most effective extraction of wave events, particularly those models dependent on geologic structure. In Chapter 3, I shall generalize these methods for use with non-linear models. However, we have already seen most of the necessary tools while inverting the simple hyperbolic model.

We saw first that a stacking procedure could be improved if it were replaced with a simple linear inversion. We might first search for a transformation, such as the stack, that simplifies the data by summing over predictable events. Instead we should

find a simple modeling transformation that maps independent parameters into these events. Then we can pose the "stack" as an optimized inversion of this transformation. The inverse should find the simplest combination of the predictable events to describe the data. Such an optimization can improve the resolution of the modeling parameters: it reduces the number of false events (artifacts) fitted to the data, even when the data are irregularly sampled.

Let us choose the global objective function as a maximum-likelihood estimate. If we have no information on the statistics of signal and noise in our data, we should assume Gaussianity. This assumption of Gaussianity gives a quadratic or least-squares functional; the inverted model will be a linear function of the data. If the signal and noise are additive in the data, they will remain additive in the linearly transformed data—this additivity simplifies considerably our calculations of probabilities. Let us define our noise as those events that will be transformed as though the samples were independent, so that the samples will sum with destructive interference. This assumption allows us to estimate the probability distribution functions (pdf's) of this transformed noise and of the transformed component that adds constructively, to be called our signal. With these distributions we can calculate the probability that noise makes up less than a small fraction of a given transformed sample's amplitude. Samples with a high probability can be preserved as "reliable" inversions (or perturbations) of model parameters.

This summary includes few details that are specific to the NMO hyperbolic model. We can expect many of the method's strengths and limitations to be preserved in other applications. For example, the extraction of reliable signal cannot discover detail that has been destroyed in the data. Coherent reflections have a finite bandwidth in time and a finite spatial aperture. The exact coherence of reflections can never be perfectly resolved, and the extraction of reliable signal cannot increase this resolution. If samples at two different velocities in the model seem equally able to describe a particular reflection, then they will both be extracted as signal.

On the other hand, much resolution can be gained if one avoids blurring existing information. Merely summing along coherent events, as do the simplest stacks, causes much energy from irrelevant events to be included. The least-squares inversion can fit signal more accurately in the data, but its linearity can allow large amounts of noise to enter the inverted model. Linear methods treat a sample of noise in the data as they would a sample of signal, and so obscure details due only to signal. The non-linear estimate of reliability can guarantee that a sufficiently small percentage of noise will obscure the inverted signal.

Other applications can imitate our various extensions of the hyperbolic model. The coherence of recorded events can change unpredictably over a large spatial range because of oversimplifications in the modeling equations. Slow spatial changes in amplitudes and other parameters can describe these events well, without great complications of the model. Laterally adaptive transforms can respect the essential character of coherent events and still allow the coherence to change over large distances.

Many processes require that irregularly sampled data be interpolated into a regular grid. Once the spatial coherence of the signal is recognized, the missing samples can replaced with an extrapolation of signal that was extracted from nearby samples. If the assumptions of the modeling equations are acceptable, then we should extrapolate the measured information no farther than the recognizable coherence of signal. Reliably interpolated events will not bias further processing.

Though I have preferred to define each point in a physical model as an independent physical parameter, it is often impractical for the coherence of an event to be modeled over time. Time coherence can be non-stationary and non-convolutional because of frequency absorption, wavelet dispersion, etc. Rather than alter the modeling equations, we should alter the extraction procedure so that it recognizes the true amplitude of seismic wavelets, which are locally monochromatic.

The discussion of this chapter uses the familiar velocity stack as an illustration of the principles of signal/noise separation. Appendices A, B, and C review the statistical tools when applied to arbitrary modeling transformations. It remains to be shown that usable modeling transformations can be chosen for a variety of coherent events and for a variety of data. I explore some useful strategies for these choices in Chapter 2. Finally, such methods should be applied to the inversion of intrinsicly interesting physical parameters and to non-linear physical models. In Chapter 3, I shall examine the inversion of acoustic impedance from the vertical seismic profile.