GEOPHYSICS, VOL. 53, NO. 7 (JULY 1988); P. 932-946, 11 FIGS.

Separation of signal and noise applied to vertical seismic profiles

William S. Harlan*

ABSTRACT

Inversion of the band-limited one-dimensional VSP response is nonunique because impedance functions with very different statistics produce equivalent responses. Least-squares methods of inversion linearly transform noise and tend to produce impedance functions with a Gaussian distribution of amplitudes. I modify a least-squares inversion procedure to exclude nonzero impedance derivatives that are significantly influenced by noise. The resulting earth model shows homogeneous intervals unless the data have reliable information to the contrary.

The data are modeled with a one-dimensional wave equation and three invertible functions: acoustic impedance, a source wavelet, and the traces' amplification. First, a linearized least-squares inverse perturbs the source function to model the downgoing wave. A relinearized inverse finds perturbations of all three modeling functions to account for first-order reflections. Further iterations explain higher order reflections.

To estimate the reliability of impedance perturbations, each linearized inversion is repeated for pure noise that equals or exceeds the noise in the data. Amplitude histograms are used to estimate probability density functions for the amplitudes of the signal and of the noise in the perturbations. Nonzero impedance derivatives are accepted as reliable if, according to the probability functions, the perturbations contain, with a high probability, only a small amount of noise.

For a set of VSP data provided by L'Institut Francais du Petrole, four iterations allowed only a few nonzero impedance derivatives and modeled a recorded VSP as well as did a least-squares inversion that accepted all proposed perturbations. Estimated probability densities for the remaining signal and noise were used to extract a tube wave that contained little signal.

INTRODUCTION

Amplitudes of reflected seismic waves contain much information on high-frequency changes in acoustic properties, particularly impedance; but inversions of surface surveys lack information on the phase of the source waveform and local background wave velocities. Vertical seismic profiles (VSPs) contain sufficient redundancy to invert for source waveform and background velocities.

VSP interval velocities have proven to be easily determined by inversion of the arrival times of the first arriving waves (Stewart, 1984). This velocity set differs significantly from that measured by a high-frequency sonic log (Stewart et al., 1984). Separation of upgoing and downgoing waves allows one to compute an effective source waveform at every recorded depth. With these source waveforms, one can invert for the impulse response at every depth and thereby begin to estimate reflection coefficients and high-frequency impedance changes (Grivelet, 1985). Questions common to other inversions then arise: will the inverted model misinterpret irrelevant information or present more detailed information than the data contain?

Previous work has been aimed at answering these questions. For example, Lanczos' (1961) singular-value decomposition and damped least-squares inversion (Menke, 1984) suppress eigenvectors that are poorly preserved by a linear modeling equation. Stewart (1984) found damped least-squares inversion to be sufficient for the well-determined inversion of VSP interval velocities. As a drawback, eigenvectors rarely correspond to simple physical structures.

Claerbout and Muir (1973), Wiggins (1978), Gray (1979), and Thorson and Claerbout (1985) assumed that their data resulted from independent, non-Gaussian parameters. They optimized functions that encouraged details of the inverted parameters to be isolated, sparse, spikey, and parsimonious. Similarly, Macé and Lailly (1984) used an ℓ_1 constraint on

Manuscript received by the Editor December 23, 1985; revised manuscript received October 29, 1987.
*Formerly Department of Geophysics, Stanford University; presently Consultant, 3201 Cumberland, San Angelo, TX 76904.

impedance derivatives to guarantee a unique inversion of the VSP. Grivelet (1985) used detection to invert a VSP for an impedance function with sparse nonzero reflection coefficients.

In this paper, I suggest a statistical inversion step whose purpose can be summarized with the following statement: do not complicate the model with details that may explain only noise in the data. The definition of a simple model is crucial. As an example, this paper suppresses unnecessary nonzero derivatives in an initially homogeneous impedance function. The following section shows the magnitude of the problem posed by nonuniqueness.

NUMERICAL NONUNIQUENESS IN ACOUSTIC INVERSION

Band-limited seismic data inversion can be nonunique even with redundant VSP data. Similar vertical-incidence VSP responses can be generated from one-dimensional acoustic impedance functions that differ considerably from one another.

Figures 1a and 1b show two alternative impedance functions. Some details correlate, but on the whole they are strikingly different. Figures 2a and 2b show the corresponding modeled VSPs. Each VSP is modeled with a one-dimensional (1-D) wave equation and three other input functions: a lengthy source waveform, irregular geophone depths, and unequal amplification of traces. The two modeled VSPs have negligible differences, as shown when the second is subtracted from the first (Figure 2c).

The first impedance function (Figure 1a) has a blocky quality, with wide intervals of constant impedance. The second function (Figure 1b) appears more periodic and monochromatic. The two can be compared by mentally smoothing the first function; mentally sharpening the second is more difficult. The depth derivative of impedance, roughly a measure of reflectivity, displays the qualitative difference better. Derivatives of the first impedance function (Figure 3a) have fewer nonzero values than those of the second function (Figure 3b).

If one assumes a homogeneous impedance function as the simplest default model, then the nonzero derivatives contain all unpredictable details. To model the VSP, the first impedance function requires fewer nonzero derivatives than does the second function. The zero derivatives can express either a lack of information or a genuinely homogeneous interval.

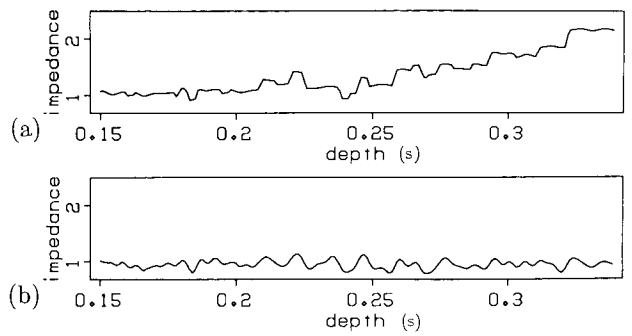


FIG. 1. Acoustic-impedance functions obtained by inversion of a recorded VSP: (a) impedance resulting from a method that suppresses impedance derivatives when noise influences the value too greatly; (b) impedance resulting from a least-squares inversion. [(a) is blockier, with wide intervals of constant impedance; (b) appears more monochromatic, with few striking details.]

The impedance function in Figure 1b is a least-squares inversion of a VSP and so shows ℓ_2 or Gaussian statistics. An amplitude histogram of the derivatives in Figure 3b is fit very well by a Gaussian distribution, but a histogram of Figure 3a is not. Incoherent noise in the data is transformed linearly by least-squares inversion and so tends to have a Gaussian distribution. The impedance function in Figure 1a is from a modified least-squares algorithm that accepts perturbations of impedance derivatives only if these perturbations are unlikely to have resulted from incoherent noise. The following section describes the statistical tools used for this inversion.

DEFINING AN INVERSION

Interpretation of a nonunique inversion should begin with the most reliably determined aspects of the model. The following definitions and assumptions will be used to define reliability:

(1) Model data as the sum of two random processes: noise and (nonlinearly) transformed signal. Let the signal determine the parameters of the physical modeling equations.

(2) Choose the modeling transformation so that the signal parameters can be treated as statistically independent random variables (e.g., impedance derivatives).

(3) Define events as the changes that appear in modeled data when a signal parameter is changed from its simplest default value (e.g., reflections from an impedance derivative).

(4) Assume that samples of noise show none of the spatial statistical dependence (coherence) of signal events. (For example, tube waves and P waves have unequal velocities and coherences; i.e., if P waves are signal, tube waves are noise.)

(5) A reliable perturbation of a signal parameter should not model an event that can be easily described by a chance combination of noise. (A weak modeled reflection may attempt to fit a random alignment of strong noise.)

Not all data need be inverted as signal or noise. An iterative inversion should converge on the most reliable parameters first because these limit the reliability of later perturbations. The following outline describes the algorithm to be used.

(1) Find perturbations of model parameters from a linearized least-squares inverse of the recorded data (cf., Macé and Lailly, 1984).

(2) For comparison, invert pure noise that is equal to or greater than the noise in the data (cf., Harlan et al., 1984).

(3) Take amplitude histograms of the results of steps (1) and (2) and use these to estimate probability density functions for the signal and noise in the model perturbations.

(4) Using these probability functions, accept perturbations that have a low percentage of noise with a high probability. [Compare with the detection step of Grivelet (1985).]

(5) Relinearize the modeling equations and repeat.

The least-squares inversion involves by far the most programming and computation time of the algorithm. The statistical test of reliability modifies the least-squares algorithm at only one point, just before the addition of the perturbations to the reference model. The following section will construct the least-squares inverse.

A GLOBAL OBJECTIVE FUNCTION

This section will discuss the modeling equations and the least-squares objective function used to prepare linear perturbations of the physical parameters. This procedure makes two important modifications to the model and objective functions of Bamberger et al. (1982) and Macé and Lailly (1984): (1) the amplification of traces is included as an invertible parameter and (2) the objective function is placed in quadratic form to facilitate an iterative least-squares linearization.

The differential system

As in Macé and Lailly (1984), the 1-D acoustic wave equation is used to model zero-offset VSPs. All interbed multiple reflections are modeled, to the limit of the time sampling rate. The effect of spherical spreading will be approximated by prescaling data amplitudes with arrival time. The following system of equations models the wave field:

$$\sigma \frac{\partial^2 y}{\partial t^2} - \frac{\partial}{\partial x}\left(\sigma \frac{\partial y}{\partial x}\right) = 0; \tag{1}$$

$$\sigma \frac{\partial y}{\partial x} = -g \qquad \text{for} \quad x = 0; \tag{2}$$

$$y = \frac{\partial y}{\partial t} = 0 \quad \text{for} \quad t = 0; \tag{3}$$

and

$$\frac{\partial y}{\partial x} = 0 \qquad \text{for} \quad x = X. \tag{4}$$

$y(x, t)$ is the time derivative of the vertical displacement measured by geophones at time t and at a pseudodepth x. x is measured by the direct traveltime from the depth of the wave source:

$$x = \int_0^z \frac{dz}{v_p(z)}.$$

z is the true depth; and $v_p(z)$ is the P-wave velocity as a function of depth. $\sigma(x)$ is the acoustic impedance, equal to the product of velocity and density. $g(t)$ is the seismic source, pro-

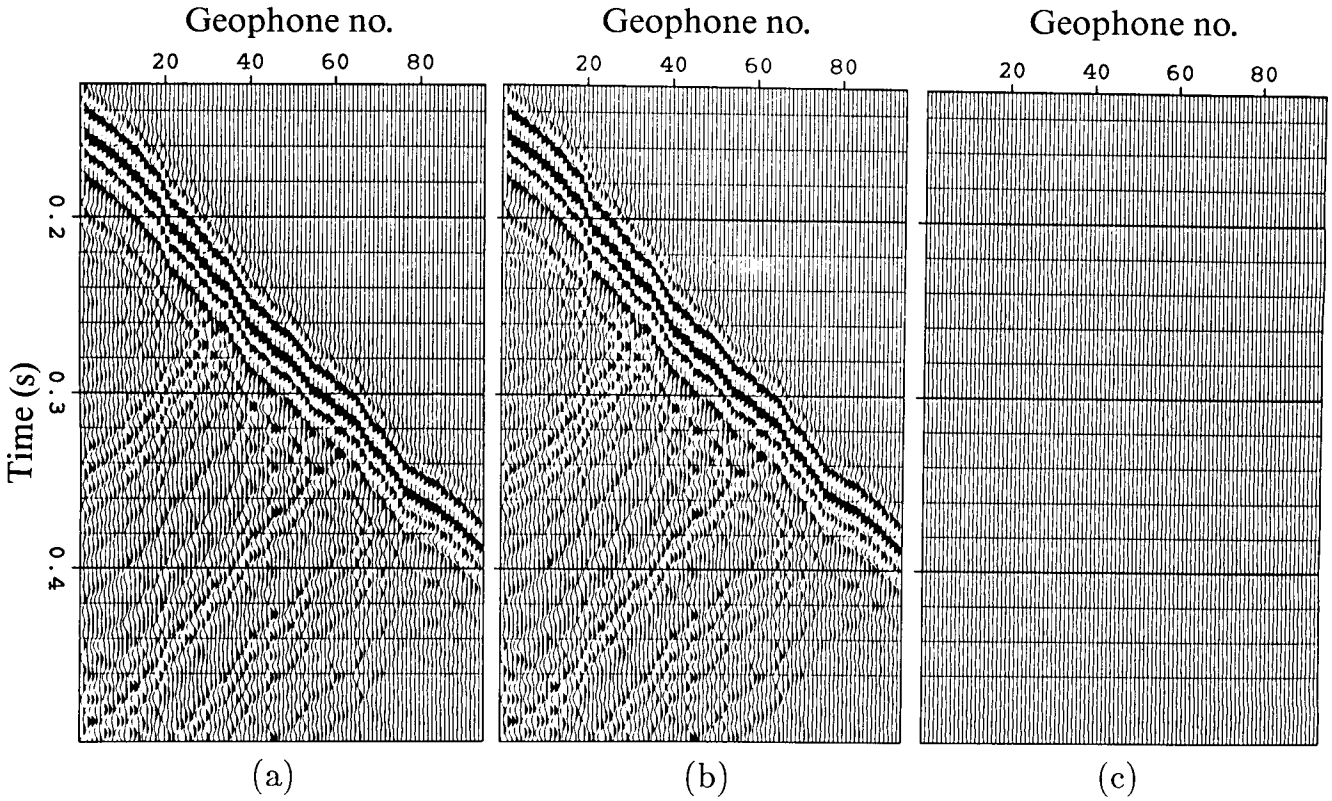


FIG. 2. VSPs modeled from the two impedance functions of Figure 1 using the acoustic wave equation and a long source function. (a) and (b) correspond to Figures 1a and 1b, respectively. Subtracting these two VSPs shows their negligible differences in (c).

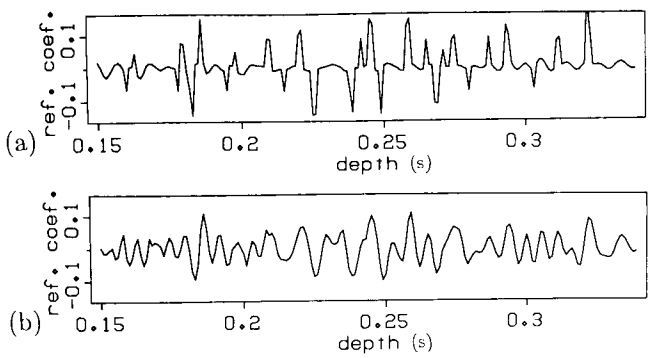


FIG. 3. The first derivative of impedance with depth, roughly a measure of reflectivity, highlights differences between the two functions. (a) and (b) correspond to Figures 1a and 1b, respectively. (a) is sparser, with a higher number of near-zero values.

portional to the time derivative of the traction at the surface. Assume that the source and its derivative are zero at time zero. Equation (4) is necessary to complete the Neumann boundary condition of equation (2), appropriate to a free surface. X is a depth sufficiently large so that no reflections can be received in the recorded time.

The data are assumed to be measured at a set of depth points $\{x_i\}$, which can be irregular and sparse, for $0 \le t \le T$. The source is assumed to be slightly above the first geophone. Such a source describes all near-surface reverberations and also compensates for any reflections from an incorrectly placed free surface.

Notice that the impedance and source functions can be scaled by any two constants as long as the product of the constants is 1. For convenience, the impedance will be given an expected value of 1.

Trace amplification and nonuniqueness

Irregular coupling of geophones and variable shot strengths can vary the amplification of traces considerably. Such changes introduce some nonuniqueness into the estimates of acoustic impedance. The VSP used in this paper showed amplitudes that varied as much as 30 percent for equivalent events on neighboring traces. Changes in impedance cannot model such rapid amplitude changes without creating spurious reflections.

For this reason a third function, the trace amplification $r(x)$, is introduced for inversion. After the data $y(x, t)$ are modeled from the impedance and source functions with equations (1) through (4), each trace is scaled by the amplification:

$$y'(x, t) = r(x)\, y(x, t). \tag{5}$$

Unfortunately, low-frequency changes in the amplification can also be explained by low-frequency changes in the impedance function; such changes affect only the recorded amplitude of displacement and do not create reflections. Low frequencies should be subtracted from inverted impedance logs if deceptive structure from uneven trace amplification is to be avoided.

A least-squares objective function

To find least-squares estimates of the impedance, source waveform, and trace amplification, I minimize

$$J_1 = C_n^{-2} \sum_i \int_0^T \left[d(x_i, t) - y'(x_i, t) \right]^2 dt + C_g^{-2} \int_0^T g^2\, dt + C_\sigma^{-2} \int_0^X \left(\frac{d\sigma}{dx} \right)^2 dx + C_r^{-2} \sum_i \left[r(x_i) - 1 \right]^2. \tag{6}$$

The Cs are the assumed standard deviations of the random variables. $y'(x, t)$ is a function of the three arrays of model parameters. $0 \le x \le X$ is the interval over which the impedance is to be inverted. Numerically, the first term of equation (6) minimizes the error between the measured data and the modeled data; the other terms encourage simplicity (less energy) in the inverted parameters.

Bamberger et al. (1982) and Macé and Lailly (1984) used an inequality constraint on the ℓ_1 norm of the differentiated impedance rather than the least-squares penalty term used in equation (6). An ℓ_1 constraint (or penalty function) has consistently required a great many optimization iterations before the impedance function is noticeably simplified. Most importantly, an impedance penalty function or constraint ensures stability in the optimization algorithm; for this purpose, the least-squares form serves equally as well as an ℓ_1 constraint. The quadratic form is more convenient for gradient optimization.

The depths of the geophones must be treated as a fourth 1-D array requiring inversion. This function is estimated in a preliminary process that windows and crosscorrelates the wave that arrives first, just as is done by automatic statics programs. Since the depths of the geophones are quantified in terms of the direct arrival times, the measured depths of the geophones (in meters) cannot be used to calculate these values.

Assumptions about impedance

The inversion defined by the objective function (6) does not make use of all prior information available on the impedance function. The most important omission is that interval velocities can be calculated from the earliest arrival times and the measured depths of geophones; interval velocities are strongly related to impedances. In addition, high-frequency P-wave acoustic impedance logs are routinely measured in wells, particularly when VSPs are recorded.

This information has been omitted in order to examine the information available from the relative strengths of reflected waves. The impedance function cannot be inverted from amplitude changes in the direct arrival alone because such changes can also be explained by changes in trace amplification. This situation better resembles the less redundant data recorded at the surface, where direct arrivals and logs are unavailable.

Let us make the following statistical assumptions: Assume that differentiating impedance with depth creates a function whose samples have negligible statistical dependence. In other words, assume that knowledge of the derivative at one depth tells little about the magnitude of derivatives at other depths. Assume also that a particular magnitude is equally likely at all depths (stationarity). The samples of a differentiated impedance function will thus be treated as an independent, identically distributed (IID) random process.

The least-squares penalty function in objective function (6) treats $d\sigma(x)/dx$ as IID but additionally assumes that the samples are Gaussian [see Kendall and Stuart (1979) for a derivation of such objective functions as maximum-likelihood solutions]. Inversion with this objective function takes advantage of the nonuniqueness to encourage a very Gaussian impedance function. Similarly, I found that the ℓ_1 norm produces impedance derivatives with an exponential distribution, which may be closer to the true distribution. The objective function will not be modified for a third arbitrary distribution (although well logs should allow a good guess). Instead, perturbations of the impedance function will be refused if their effect on the data is poorly distinguishable from that of noise.

ESTIMATION OF SIGNAL AND NOISE

The modeling equations and objective function of the previous section define an inverse for the modeling parameters, but complete optimization can be impractical. Very different solutions (such as in Figure 1) can appear almost equally optimum in a flat-bottomed objective function. This section describes how a partial optimization can be modified to minimize the effect of noise on the estimated signal.

Iteratively linearized least-squares inverses

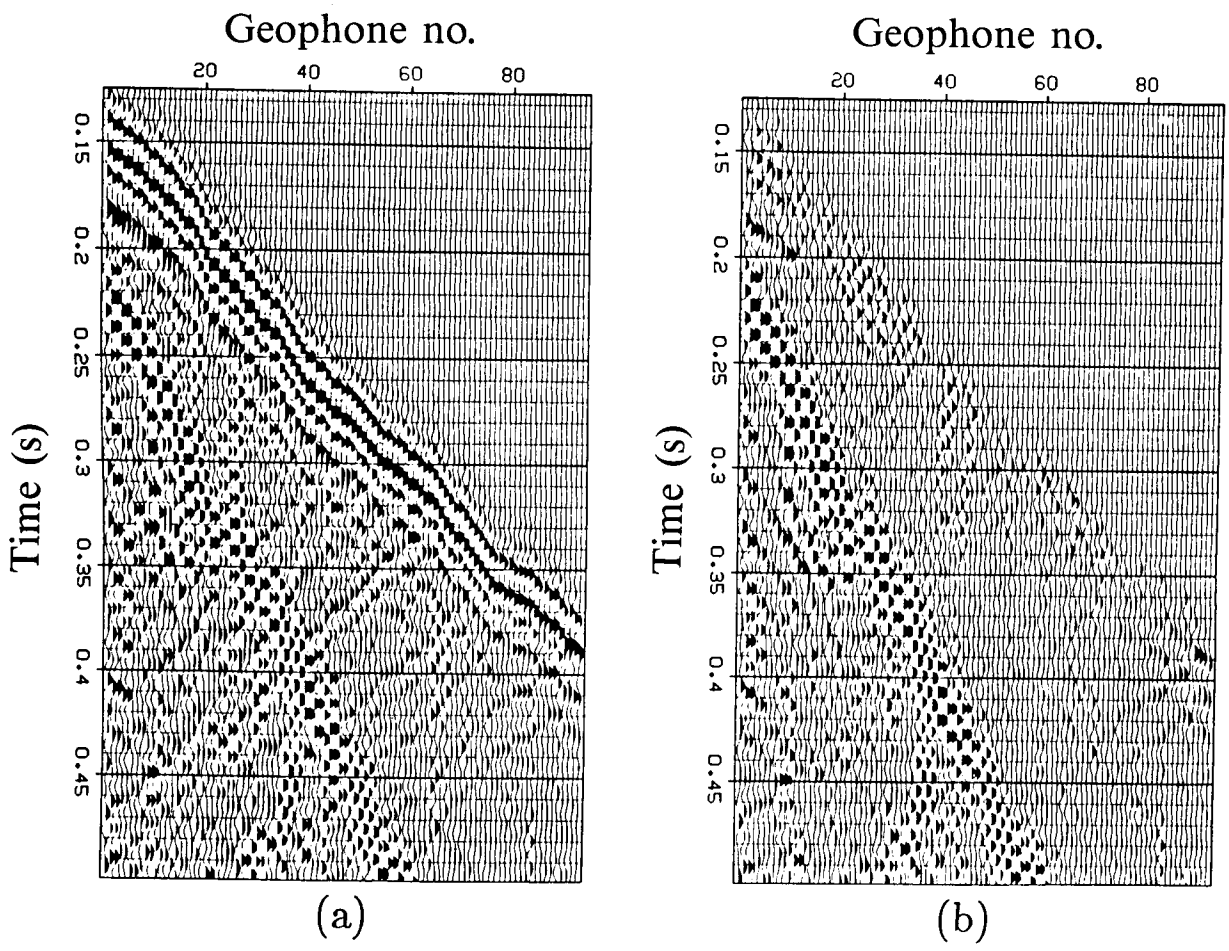
Let the data, the recorded VSP, be a sum of noise and transformed signal:

$$d_i = f_i(\mathbf{s}) + n_i$$

or

$$\mathbf{d} = \mathbf{f}(\mathbf{s}) + \mathbf{n}. \tag{7}$$

The signal array $\mathbf{s}$ contains three sets of parameters from the impedance, source, and trace amplification. The nonlinear transform f is implicitly defined by a stable finite-difference solution of equations (1) through (5). The noise includes all data components not modeled by this system of equations, such as Gaussian background noise (incoherently scattered



FIGS. 4a, 4b. (a) A recorded VSP. First arrival times parameterize the irregular depths of traces. (b) Gaussian noise, a tube wave, and weak uninverted signal obtained by subtracting modeled waves from (a).

waves), bad traces, tube waves, dispersed and absorbed wavelet frequencies, etc.

If we write a linearized transformation as

$$f_i(\mathbf{s}^0 + \Delta\mathbf{s}) \approx f_i(\mathbf{s}^0) + \sum_j F^0_{ij}\Delta s_j, \tag{8}$$

then the least-squares estimate of the perturbed signal minimizes a quadratic objective function:

$$\min_{\Delta\mathbf{s}} \left\{ \sum_i C_{si}^{-2}(s_i^0 + \Delta s_i)^2 + \sum_i C_{ni}^{-2}\left[d_i - f_i(\mathbf{s}^0) - \sum_j F^0_{ij}\Delta s_j \right]^2 \right\}. \tag{9}$$

This quadratic objective function (9) can be minimized with the conjugate-gradient algorithm (see Luenberger, 1984) and with the adjoint and linearized modeling equations (Appendix A). If a minimum of objective function (6) is to be found (though not guaranteed), then this linearization must be repeated.

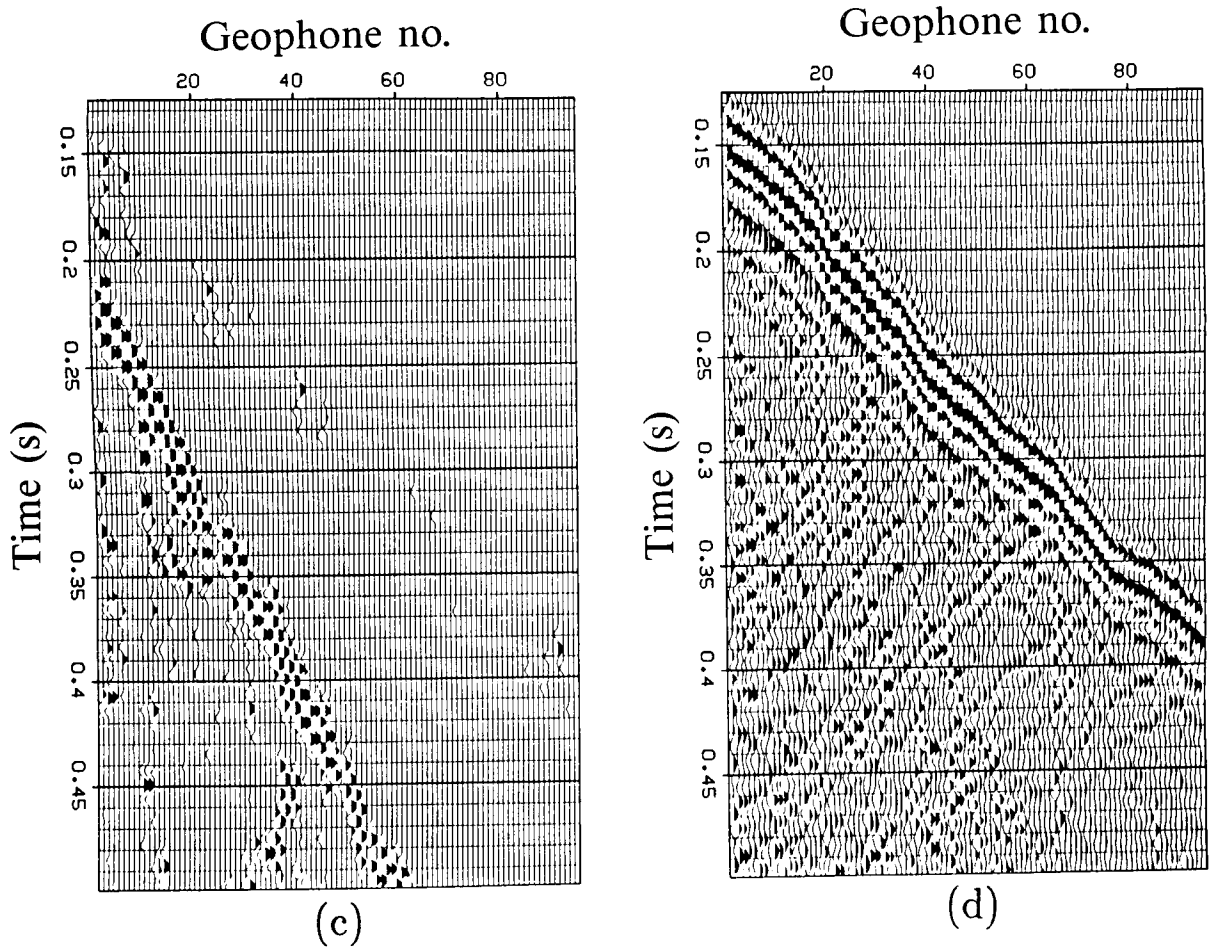
An example of noise extraction

Estimates of reliable signal will improve if reliable noise is iteratively estimated and removed from the data. Noise estimates likewise improve after removing estimated signal. I shall now illustrate this outermost loop of the inversion.

Figure 4a contains a portion of a VSP provided by L'Institut Francais du Petrole. This section contains considerable Gaussian noise. A strong tube wave violates the physical assumptions of the modeling equations and acts as strong additive non-Gaussian noise.

To estimate the signal present in these data, objective function (6) is minimized with four linearized perturbations, as defined by objective function (9). Because noise is the object, the appearance of the impedance function is not considered. Subtracting the modeled signal (similar to Figures 2a and 2b) from the original data reveals the Gaussian noise, tube wave, and weak uninverted signal (Figure 4b).

After the modeled signal is removed, the most reliable noise is estimated. Using the statistical techniques elaborated in the



FIGS. 4c, 4d. (c) Samples of (b) that contain, with a high probability, a high percentage of noise. (d) The difference between the original data (a) and the noise (c) used as another estimate of signal.

following sections, I zero those samples of the residual data that do not contain, with a sufficient probability, a high percentage of noise. Only the most non-Gaussian noise, the tube wave, has been extracted (Figure 4c). The Gaussian noise is too easily explained as the sum of coherent Gaussian signal. The most reliable noise is subtracted from the original data (Figure 4d), removing a source of error in the first estimate of signal. This cleaner VSP is now used in place of the recorded data in subsequent iterations.

Insufficiency of the linear estimate

For the first linearized inversion of the VSP in Figure 4d, no knowledge of the signal parameters is assumed. Impedance is given a constant value of 1, and its derivatives, a small variance of 0.01. The source is set to zero, and the trace amplification function $r(x)$ is set to one—both with large variances. The variance of the noise is pessimistically set equal to that of the recorded data.

The first minimization of the quadratic objective function (9) perturbs only the source function and accounts only for downgoing waves without scattering (Figure 5a). The source function changes little in later iterations, though it is free to change. The first and higher-order scattered waves appear clearly in the residuals of Figure 5b (equal to Figure 4d minus Figure 5a).

Each linearization allows the inversion to consider an additional order of scattering. Because the downgoing wave field is now nonzero, the second least-squares linearization can perturb impedance and account for first-order scattering. Figure 6a displays the new first-order reflections modeled by the linearized acoustic equations (A-1) through (A-5). All three signal functions receive perturbations. The perturbations of the impedance derivatives appear in Figure 6b. The uneven amplification of traces is evident in the modeled reflections. The amplification, like the source function, changes little in later iterations.

A simple test shows the effect of noise on the impedance perturbations. To make the coherent signal become incoherent and behave as noise in the inversion, I randomly reorder the traces of Figure 5b in Figure 7a. This reordering does not affect the statistics of the incoherent noise. I repeat the linearized least-squares inversion on the noise and find the perturbed impedance derivatives of Figure 7b. Note that the

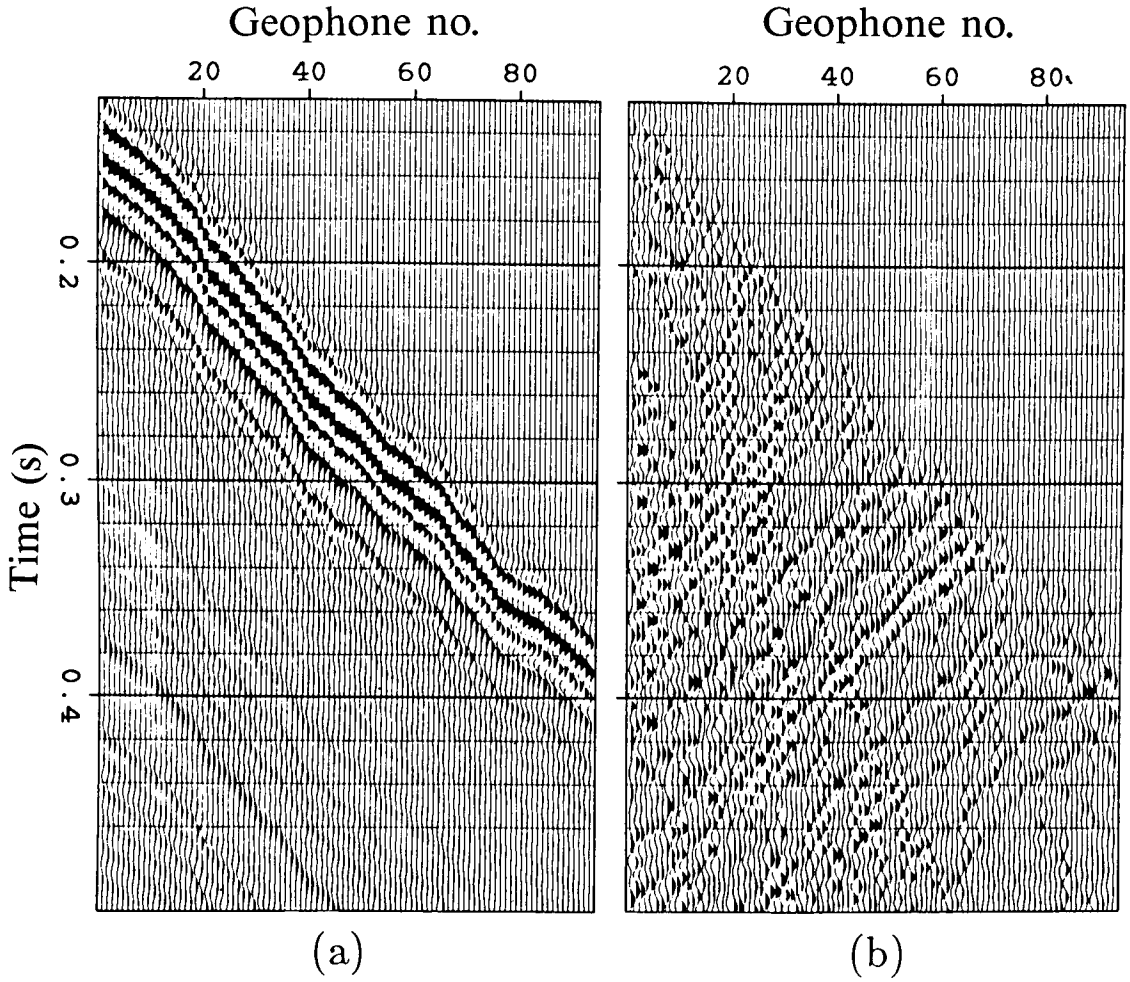


FIG. 5. (a) The downgoing wave field obtained by perturbing only the source waveform (first iteration). (b) The residuals obtained by subtracting (a) from the data of Figure 4d. (All scattered waves are left.)

perturbations are far from zero. These false details result from random alignments of noise. This much amplitude could be attributed to noise in the perturbation of Figure 6b. More complicated schemes of generating noisy data sets (with the same data amplitude distributions) gave similar results.

Distinguishing transformed signal and noise

The advantage of the least-squares linearization is that the optimum perturbation from objective function (9) remains a linear function of the uninverted events in the data. The linear perturbation can be written as the following transformation:

$$d'_i = \Delta s_i = \sum_j F^{-1}_{i,j}\left[d_j - f_j(\mathbf{s}^0)\right]. \tag{10}$$

This notation indicates a least-squares rather than a perfect inverse.

Probability density functions will be used to quantify the distribution of signal and noise amplitudes in the perturbation. The probability that a random variable falls within a given range of amplitudes is found by integrating the probability function over that range. Because of the assumption that signal and noise are stationary, many parameters should result from the same probability functions. Subscripts will be dropped from individual samples.

The transformed signal and noise remain additive in a transformed data sample: $d' = s' + n'$. Define $\mathbf{s}' = \mathbf{F}^{-1}[\mathbf{f}(\mathbf{s}) - \mathbf{f}(\mathbf{s}^0)]$ and $\mathbf{n}' = \mathbf{F}^{-1}\mathbf{n}$. When independent random variables add, their probability density functions convolve:

$$p_{d'}(x) = p_{s'}(x) * p_{n'}(x). \tag{11}$$

A probability function is estimated for the transformed data (Figure 6b) with an amplitude histogram (Figure 8a). The transformation of the artificially incoherent data (Figure 7b) effectively overestimates the amplitudes of the incoherent noise; a histogram gives a pessimistic estimate of the noise probability function (Figure 8b).

A deconvolution of equation (11) estimates a probability function for the transformed signal. Let us assume that the pessimistic estimate of the noise probability function is correct

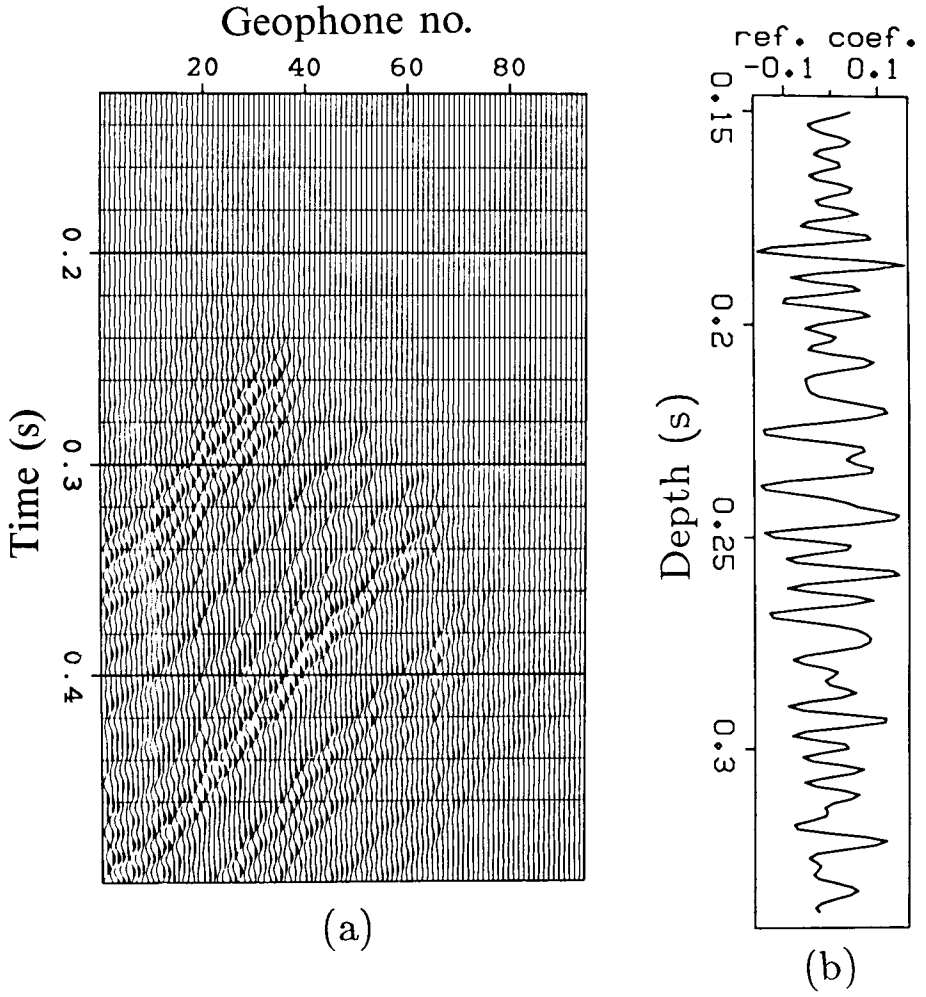


FIG. 6. A relinearized inversion of the residuals of Figure 5b models the first-order scattered waves in (a) with linearized modeling equations. The impedance-derivative perturbations (b) account for these new events.

and find the signal probability function that maximizes the probability of the data histogram. This estimate equivalently minimizes the following cross-entropy function, with appropriate constraints of positivity and unit area:

$$\min_{p_{s'}(\cdot)} \int p_{d'}(x) \ln \left[p_{d'}(x)/p_{s'}(x) * p_{n'}(x) \right] dx. \tag{12}$$

Figure 8c displays the resulting signal probability density function. The convolution of the estimated signal and noise functions (Figure 8d) cannot equal the data histogram because the inner peak of the noise function is wider than that of the data. The tails of the convolved function match the tails of the data histogram very well. See Appendix B for more details on the deconvolution method.

Extracting reliable signal

The expected amount of signal in samples of the transformed data can now be calculated from the probability functions for the signal and the noise. A Bayesian estimate (see Papoulis, 1965) of s', when d' is known, is

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

$$= \frac{\int x p_{s'}(x) p_{n'}(d' - x)\, dx}{p_{d'}(d')}. \tag{13}$$

For the probability functions of Figure 8, the expected value of s' is very close to the value of d', essentially an equality. The expected value of noise, for all data amplitudes, is almost zero. For other applications or other data, this estimate could alter amplitudes considerably [see Godfrey (1979)].

Now let us calculate the reliability of the Bayesian estimates. Accept an estimate as reliable if the percentage error (equivalent to the percentage noise) is less than say 5 percent with greater than 95 percent probability. Define the reliability

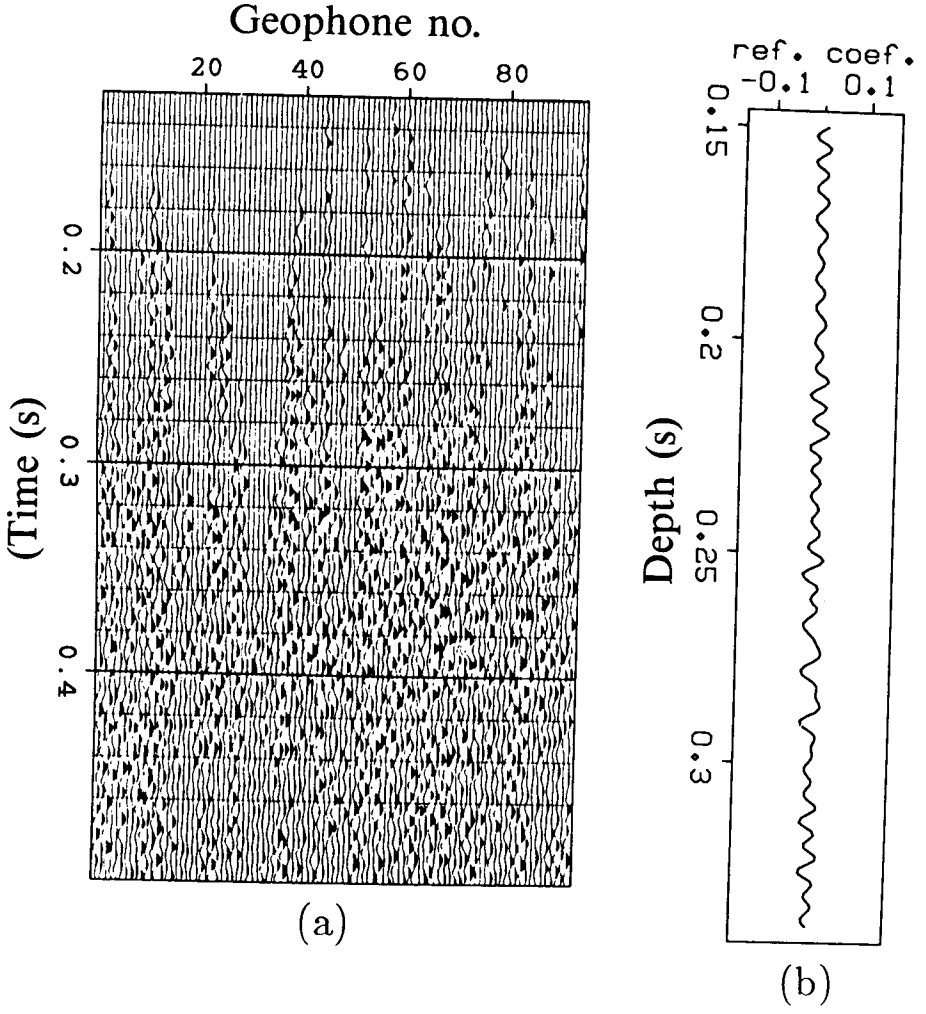


FIG. 7. The influence of noise on the impedance perturbation of Figure 6b. (a) The residual data of Figure 5b rearranged to destroy all coherence. (b) Impedance-derivative perturbations calculated from the noise as before.

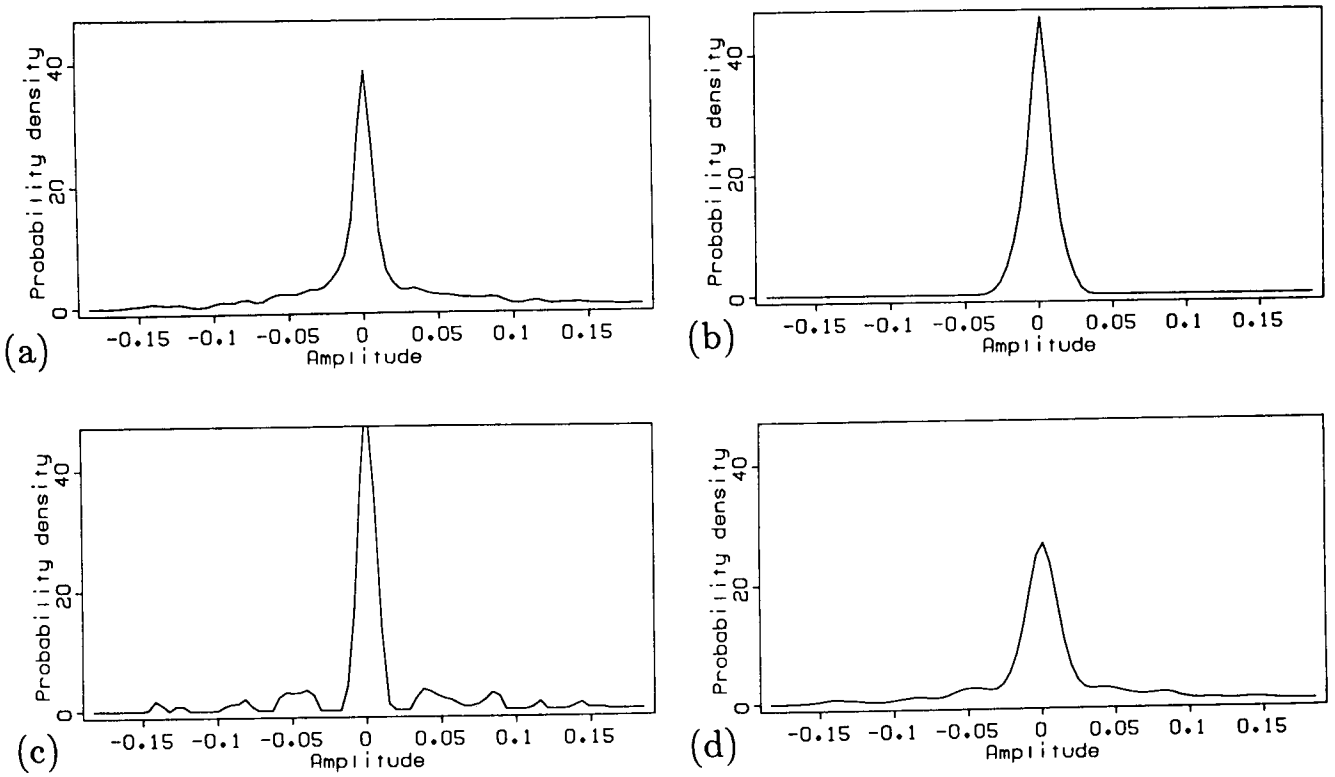


FIG. 8. (a) Histogram of the perturbation of Figure 6b estimating the probability density function of the transformed signal plus noise. (b) Histogram of Figure 7b overestimating the probability function of the transformed noise. (c) The signal function estimated by deconvolving (a) with (b). (d) The convolution of (b) with (c). [(a) is imperfectly reproduced because (b) has a wide central peak.]

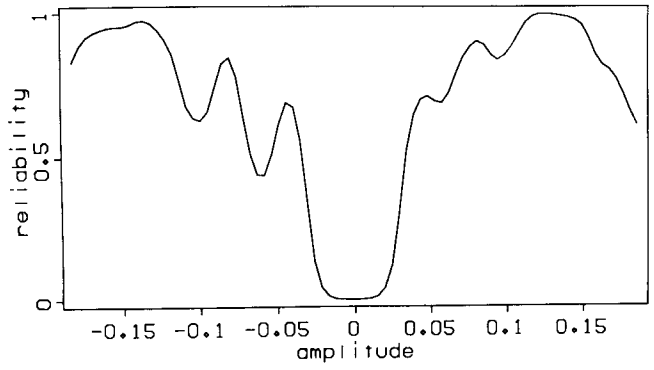


FIG. 9. A measure of reliability. (The largest amplitudes are the most reliable.) The probability functions of Figure 8 are used to calculate the probability that the impedance derivatives in Figure 6b contain less than 5 percent noise.

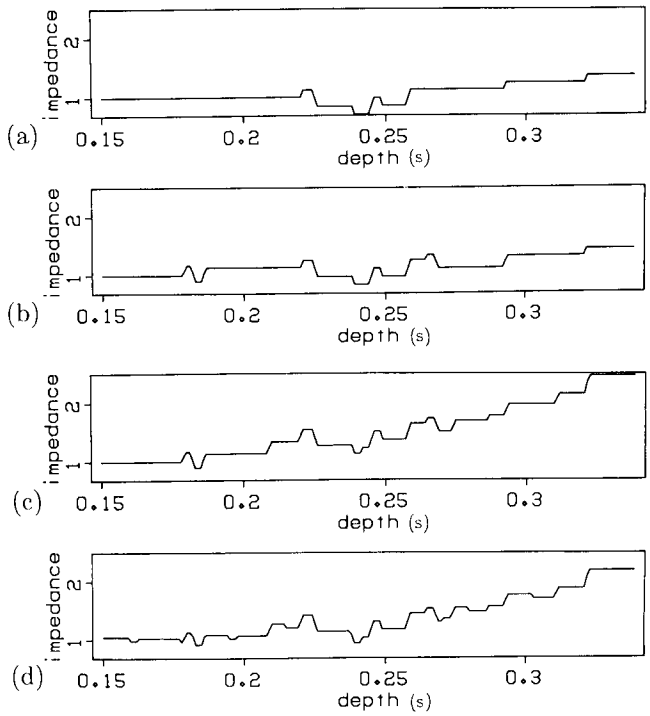


FIG. 10. (a) The impedance function perturbed with the eight most reliable derivatives of Figure 6b. (b), (c), and (d) show the results of further iterations. Further perturbations have less than 95 percent reliability of having less than 5 percent noise.

by

$$\text{reliability} \equiv P\left[-c\hat{s}' < s' - \hat{s}' < c\hat{s}' \mid d'\right] = \frac{\int_{\hat{s}'-c\hat{s}'}^{\hat{s}'+c\hat{s}'} p_{s'}(x)p_{n'}(d'-x)\,dx}{\int_{-\infty}^{\infty} p_{s'}(x)p_{n'}(d'-x)\,dx}. \quad (14)$$

$c = 0.05$ or some other small fraction of allowable noise.

Figure 9 shows the reliability of the Bayesian estimate as a function of the amplitude d'. The high amplitudes are the most reliable, and the low amplitudes the least. Quite a few samples of the perturbations in Figure 6b show greater than 95 percent reliability. To present a slow convergence, only the eight most reliable perturbations were accepted, and others were set to zero. Figure 10a shows the corresponding impedance function. Figures 10b, 10c, and 10d show the impedance functions of later iterations. Figure 11 shows the VSPs that are modeled for each of these four iterations. Each iteration adds details of decreasing reliability to the impedance. Further perturbations are weak and have less than 95 percent reliability (for less than 5 percent noise).

To obtain the impedance function of Figure 1a, I perform one last iteration that accepts all perturbations, whether reliable or unreliable. The resulting model (Figure 2a) is little better, and the structure of the impedance function is only slightly blurred. The small gradient shows that, in effect, objective function (6) has reached a minimum. The impedance function of Figure 1b was obtained by accepting the perturbations of all iterations as reliable, giving a least-squares solution. The two solutions of Figure 1 minimize objective function (6) equally well, showing that the function has a relatively flat bottom. The constraint of reliability on the signal perturbations effectively limits the class of impedance functions to be considered.

Extracting reliable noise

The early extraction of the tube wave from the VSP used the same statistical tools as the extraction of the signal. The algorithm follows.

(1) Subtract previously modeled signal from the data and take a histogram of the residual data amplitudes.

(2) Estimate the uninverted signal by a relinearized

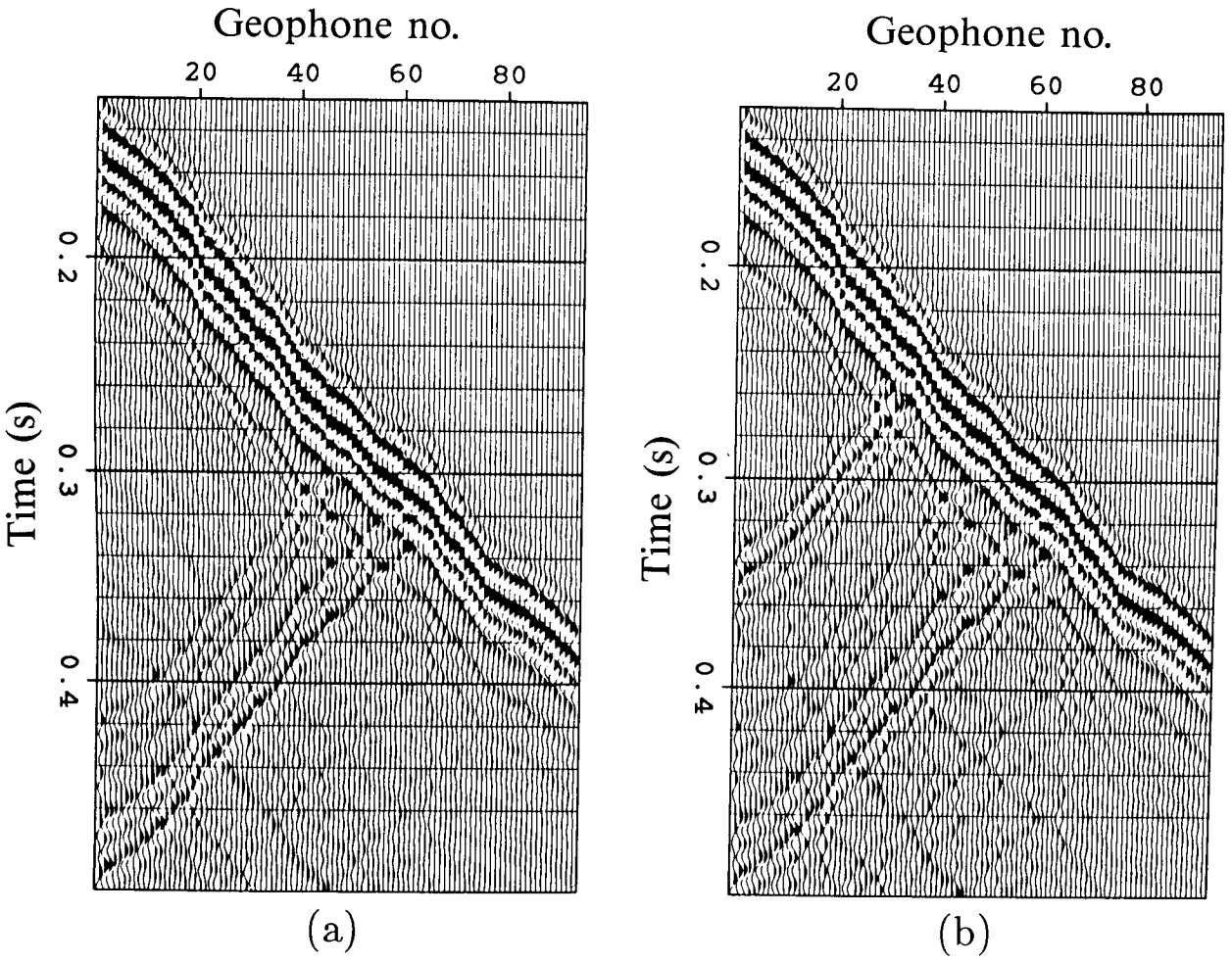


FIG. 11. Modeled VSPs corresponding to the impedance logs of Figure 10. The unlinearized acoustic wave equation is used. Further iterations change the modeled data negligibly.

inverse of the residual data [minimize objective function (9)].

(3) Estimate the probability density function of the uninverted signal amplitudes with a histogram of the linearly modeled residual data [from equations (A-1) through (A-5)].

(4) Estimate the probability function for the noise by deconvolving the probability functions from steps (1) and (3).

(5) Extract samples of the residual data that contain a large percentage of noise with a large probability.

The residual data appear in Figure 4b, and the extracted noise in Figure 4c.

To avoid distorting the waveforms of the noise, the extraction of reliable samples must be smoothed over time. The peaks of wavelets have the largest amplitudes and are more easily recognized as noise than are the high-derivative zero crossings between peaks. A linear phase shift of 90 degrees maps the wavelet zeros to peaks and the peaks to zeros. Different constant linear phase shifts will map other points of the wavelet to peaks. Let us consider any particular point to be reliable if the highest amplitude resulting from a linear phase shift is reliable. This maximum amplitude is equal to the analytic envelope calculated with the Hilbert transform (see Bracewell, 1978). The reliability of a sample of noise was calculated from its analytic envelope rather than from its actual amplitude. The histograms were not altered because they sufficiently represented the amplitudes of phase-shifted data.

CONCLUSIONS

The acoustic-wave equation ignores many physical parameters that must affect VSP data, yet the equation models the transmitted and reflected *P* waves very well. Moreover, very different 1-D acoustic-impedance functions can model the two-dimensional data equally well. To emphasize the information available from reflected waves, well logs and measurements of interval velocities were not used to constrain the impedance function. To choose among equally satisfactory models of the data, the inversion avoided unnecessary nonzero derivatives in the impedance. Unless the data contained reliable information to the contrary, the inversion assumed that impedance was homogeneous.

Statistical assumptions about the signal and noise should be kept in mind when interpreting the results or when applying

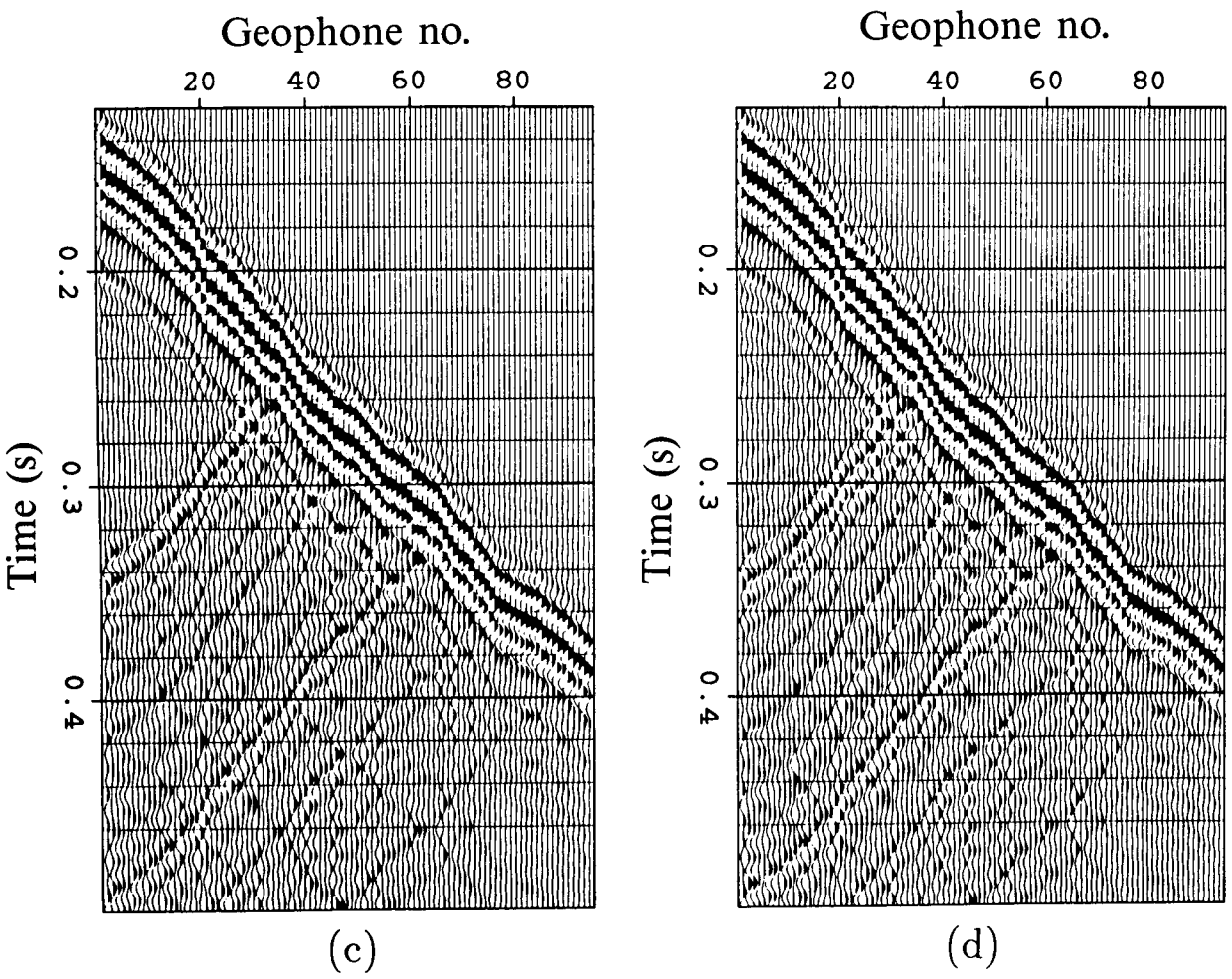


FIG. 11. Modeled VSPs corresponding to the impedance logs of Figure 10. The unlinearized acoustic wave equation is used. Further iterations change the modeled data negligibly.

the method to other data (see Harlan et al., 1984). The estimates of reliability are effective only if the events of physical interest can be modeled by statistically independent parameters of the signal. Furthermore, enough signal parameters should have similar statistics for histograms to be taken. Only non-Gaussian signal parameters will be distinguishable from noise. Gaussian signal remains Gaussian after a linear transformation has been performed, as does any unextracted noise.

Noise was assumed to have none of the spatial coherence of the signal. The recorded tube wave has a much slower velocity than do other waves and so acts as strong additive non-Gaussian noise. Ground roll behaves similarly in surface field gathers. Many varieties of coherent noise, such as sea-floor diffractions in ocean data, can be extracted if they are defined with their own modeling equations.

Not all signal and noise parameters should or need be encouraged toward sparseness and non-Gaussianity. A least-squares inversion gives the maximum-likelihood estimate of genuinely Gaussian signal and noise. The least-squares objective function sufficed for the VSP source wavelet and the trace amplification functions, both without useful geologic information. Also, signal and noise can be considered to contain both Gaussian and non-Gaussian components. Once the non-Gaussian component has been successfully extracted, least-squares methods will invert the Gaussian residuals.

In summary, this inversion does not perturb a signal parameter if the corresponding event in the data can be easily described by noise. Signal adds constructively in the calculation of the least-squares perturbations; noise adds destructively. To estimate the signal present in the perturbation, histograms of the perturbations are compared to histograms of inverted artificial noise. The histograms help estimate probability functions for the signal and noise, which in turn help estimate the probability that a perturbation contains a small percentage of noise.

ACKNOWLEDGMENTS

Above all I wish to thank Patrick Lailly and Danielle Macé for patient explanations of their work on the inversion of VSPs. I thank L'Institut Francais du Petrole for allowing me to visit for six months and for providing the data. I also thank the sponsors of the Stanford Exploration Project for their support of this research. I particularly thank Jon Claerbout and Fabio Rocca for the guidance that originally made this work possible.

REFERENCES

Bamberger, A., Chavent, G., Hemon, C., and Lailly, P., 1982, Inversion of normal-incidence seismograms: Geophysics, **47**, 757–770.
Bracewell, R. N., 1978, The Fourier transform and its applications: McGraw-Hill Book Co.
Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: Geophysics, **38**, 826–844.
Godfrey, R., 1979, A stochastic model for seismogram analysis: Ph.D. thesis, Stanford Univ.
Gray, W. C., 1979, Variable norm deconvolution: Ph.D. thesis, Stanford Univ.
Grivelet, P. A., 1985, Inversion of vertical seismic profiles by iterative modeling: Geophysics, **50**, 924–930.
Harlan, W. S., Claerbout J. F., and Rocca F., 1984, Signal-noise separation and velocity estimation: Geophysics, **49**, 1869–1880.
Kendall, M., and Stuart, A., 1979, The advanced theory of statistics, **2**, Inference and relationship, 4th ed.: Charles Griffin and Co. Limited.
Kullback, S., 1959, Information theory and statistics: John Wiley and Co.
Lanczos, C., 1961, Linear differential operators: Van Nostrand.
Lions, J. L., 1968, Contrôle optimal de systèmes gouvernés par des equations aux dérivées partielles: Dunod-Gauthier Villars (English trans., 1972, Springer-Verlag).
Luenberger, D. G., 1984, Linear and nonlinear programming: Addison-Wesley Publ. Co., Inc.
Macé, D., and Lailly, P., 1984, A solution of an inverse problem with the 1D wave equation applied to the inversion of vertical seismic profiles, *in* 6th internat. conf. on analysis and optimization of systems: Springer-Verlag.
Menke, W., 1984, Geophysical data analysis: discrete inverse theory: Academic Press, Inc.
Papoulis, A., 1965, Probability, random variables, and stochastic processes: McGraw-Hill Book Co.
Stewart, R. R., 1984, VSP interval velocities from traveltime inversion: Geophys. Prosp., **32**, 608–628.
Stewart, R. R., Huddleston, P. D., and Kan, T. K., 1984, Seismic versus sonic velocities: A vertical seismic profiling study: Geophysics, **49**, 1153–1168.
Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 2727–2741.
Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexplor., **16**, 21–35.

APPENDIX A

MINIMIZING THE VSP OBJECTIVE FUNCTIONS

The inversion of a VSP introduced a global objective function (6) and a quadratic objective function (9). This appendix presents the equations that are necessary for gradient methods to minimize these functions.

Let us invert for a source function whose length equals the recorded time of the data: $(0 \le t \le T)$. The impedance function can be inverted down to the maximum depth X, from below which no reflections can be received in the recorded time. Varying the impedance at the deeper points will not affect the modeled data. This maximum depth equals $X = x_{max} + (T - x_{max})/2$, where x_{max} is the depth of the deepest geophone.

I use the strategy of optimal control (Lions, 1968) for the minimization of objective function (6). The application to the 1-D inversion of seismic waves derives from Bamberger et al. (1982) and Macé and Lailly (1984). These previous applications use the ℓ_1 norm to constrain impedance and do not iteratively linearize modeling equations during inversion.

For each of the iterations, the modeled data $y'(x, t)$ can be linearized with respect to perturbed model parameters:

$$\sigma \frac{\partial^2}{\partial t^2}\,\delta y - \frac{\partial}{\partial x}\left(\sigma \frac{\partial}{\partial x}\,\delta y\right) = -\delta\sigma \frac{\partial^2 y}{\partial t^2} + \frac{\partial}{\partial x}\left(\delta\sigma \frac{\partial y}{\partial x}\right); \quad \text{(A-1)}$$

$$\sigma \frac{\partial}{\partial x}\,\delta y = -\delta g - \delta\sigma \frac{\partial y}{\partial x} \quad \text{for} \quad x = 0; \quad \text{(A-2)}$$

$$\delta y = \frac{\partial}{\partial t}\,\delta y = 0 \quad \text{for} \quad t = 0; \quad \text{(A-3)}$$

$$\frac{\partial}{\partial x}\,\delta y = 0 \quad \text{for} \quad x = X; \quad \text{(A-4)}$$

and

$$\delta y' = r\,\delta y + y\,\delta r. \tag{A-5}$$

All unperturbed functions remain at reference values.

The gradient of objective function (6) can be calculated independently:

$$\nabla_\sigma J_1 = \int_0^T \left(\frac{\partial^2 y}{\partial t^2} q + \frac{\partial y}{\partial x}\frac{\partial q}{\partial x}\right) dt - C_\sigma^{-2}\frac{d^2\sigma}{dx^2}; \tag{A-6}$$

$$\nabla_g J_1 = q\Big|_{x=0} + C_g^{-2} g; \tag{A-7}$$

and

$$\nabla_r J_1 = \frac{-1}{C_n^2}\int_0^T y(d - y')\,dt + \frac{r-1}{C_r^2}. \tag{A-8}$$

$q(x, t)$ is determined by the following system of equations:

$$\sigma\frac{\partial^2 q}{\partial t^2} - \frac{\partial}{\partial x}\left(\sigma\frac{\partial q}{\partial x}\right) = \frac{1}{C_n^2}\sum_i r(d - y')\delta(x - x_i); \tag{A-9}$$

$$q = \frac{\partial q}{\partial t} = 0 \qquad \text{for} \quad t = T; \tag{A-10}$$

and

$$\frac{\partial q}{\partial x} = 0 \qquad \text{for} \quad x = 0 \quad \text{and} \quad x = X. \tag{A-11}$$

The system of equations (A-6) through (A-11) is the adjoint of the linearized system. [Vector norms are implicitly defined by function (6).]

Using conventional gradient methods (steepest descent, Fletcher-Reeves, etc.) and gradients (A-6), (A-7), and (A-8), we can iteratively locate minima of objective functions (6) and (9) (see Luenberger, 1984). At any given iteration we have estimates of the impedance, the source, the trace amplification, and their corresponding wave field. Subtraction of this wave field from the data gives an error which the adjoint system uses to calculate new gradients. For the nonquadratic objective function (6), perturbations must be scaled by a line search before they are added to the reference parameters.

Minimization of the quadratic objective function (9) additionally requires the linearized differential system of equations (A-1) through (A-5). A steepest-descent or conjugate-gradient minimization requires only a few scalar products for calculation of the perturbations' scale factor.

APPENDIX B

ESTIMATING PROBABILITY DENSITY FUNCTIONS WITH CROSS ENTROPY

To estimate the probability density function for transformed signal, one must use the estimated functions for data and noise. The three are related by the convolution of equation (11). The necessary deconvolution can be posed as an optimization: which of a limited class of probability functions maximizes the probability of a histogram taken from the same random variable?

Let $\{p_i\}$ be a histogram, and let $\{\hat{p}_i\}$ be the discrete probability density function approximating it. The subscript indexes a narrow range of amplitudes (a bin). p_i is the frequency of the bin in the data, and $\hat{p}_i$ is its assumed probability. If the parameters sampled by a histogram are statistically independent, then the probability of the ensemble is equal to the product of the probabilities of the individual samples. A histogram of N samples will have the probability within brackets:

$$\max_{\{\hat{p}_i\}}\left[\prod_i C_i(\hat{p}_i)^{p_i N}\right], \tag{B-1}$$

where

$$C_i = \frac{(p_i N)!}{i!(p_i N - i)!}.$$

Exclamation points indicate factorials. The optimum probability function will maximize function (B-1). Taking the natural logarithm, adding a constant, and reversing a sign allows us equivalently to minimize

$$\min_{\{\hat{p}_i\}}\sum_i p_i \log(p_i/\hat{p}_i). \tag{B-2}$$

We discover, in the continuous limit, Kullback's (1959) directed divergence or cross-entropy:

$$\min_{\hat{p}(x)}\int p(x)\log\left[p(x)/\hat{p}(x)\right]dx. \tag{B-3}$$

x is now the index of amplitude.

Assuming that the estimated probability function for noise is correct, I define a maximum a posteriori estimate of the signal probability function as one that maximizes the probability of the data histogram. Equivalently, one can minimize the following functional of the signal probability function:

$$J_2\left[p_s(x)\right] = \int p_d(x)\ln\left[p_d(x)\Big/\int p_s(x - y)p_n(y)\,dy\right]dx. \tag{B-4}$$

In addition, the probability function must be positive and have unit area. Many nonlinear methods will minimize this functional, but the method of steepest descent is easiest to explain. The histograms contain fewer than a hundred samples, so the iterations are fast.

To calculate the gradient of functional (B-4) with respect to each point of the signal probability function, perturb the pre-

vious estimate with an infinitesimal delta function and differentiate with respect to the delta function's amplitude:

$$G(x_0) = \frac{\partial}{\partial \varepsilon} J_4\left[p_s(x) + \varepsilon\delta(x - x^0)\right]\Big|_{\varepsilon=0}$$

$$= -\int \frac{p_d(x)}{\int p_s(y)p_n(x - y)\,dy} p_n(x - x^0)\,dx. \quad \text{(B-5)}$$

This gradient divides the data histogram by the convolved function and crosscorrelates with a shifted noise function. The crosscorrelation indicates changes in the signal function that would compensate for a nonuniform divergence between the two data functions. This gradient is negative everywhere. A steepest-descent perturbation of the signal function can be written as

$$(1 - a)p_s(x) + aG(x) \Big/ \int G(x')\,dx'. \quad \text{(B-6)}$$

The optimum value of the scalar a can be found by searching the interval [0, 1] with a golden-section line search (Luenberger, 1984). The gradient can then be recalculated for the updated signal function.