GEOPHYSICS, VOL. 60, NO. 4 (JULY-AUGUST 1995); P. 1169-1177, 4 FIGS.

High-resolution velocity gathers and offset space reconstruction

Mauricio D. Sacchi* and Tadeusz J. Ulrych*

ABSTRACT

We present a high-resolution procedure to reconstruct common-midpoint (CMP) gathers. First, we describe the forward and inverse transformations between offset and velocity space. Then, we formulate an underdetermined linear inverse problem in which the target is the artifacts-free, aperture-compensated velocity gather. We show that a sparse inversion leads to a solution that resembles the infinite-aperture velocity gather. The latter is the velocity gather that should have been estimated with a simple conjugate operator designed from an infinite-aperture seismic array. This high-resolution velocity gather is then used to reconstruct the offset space.

The algorithm is formally derived using two basic principles. First, we use the principle of maximum entropy to translate prior information about the unknown parameters into a probabilistic framework, in other words, to assign a probability density function to our model. Second, we apply Bayes's rule to relate the a priori probability density function (pdf) with the pdf corresponding to the experimental uncertainties (likelihood function) to construct the a posteriori distribution of the unknown parameters. Finally the model is evaluated by maximizing the a posteriori distribution. When the problem is correctly regularized, the algorithm converges to a solution characterized by different degrees of sparseness depending on the required resolution. The solutions exhibit minimum entropy when the entropy is measured in terms of Burg's definition. We emphasize two crucial differences in our approach with the familiar Burg method of maximum entropy spectral analysis. First, Burg's entropy is minimized rather than maximized, which is equivalent to inferring as much as possible about the model from the data. Second, our approach uses the data as constraints in contrast with the classic maximum entropy spectral analysis approach where the autocorrelation function is the constraint. This implies that we recover not only amplitude information but also phase information, which serves to extrapolate the data outside the original aperture of the array. The tradeoff is controlled by a single parameter that under asymptotic conditions reduces the method to a damped least-squares solution. Finally, the high-resolution or aperture-compensated velocity gather is used to extrapolate near- and far-offset traces.

INTRODUCTION

Conventional velocity analysis is performed by measuring energy along hyperbolic paths for a set of tentative velocities. The analysis of the results in the $\tau - v$ (two-way zero offset time and velocity) plane serves to estimate the stacking velocity that is later used to construct the zero-offset section. The semblance (Neidell and Taner, 1971) is one of the most popular measures of coherent energy along hyperbolic trajectories in common-midpoint (CMP) gathers. The semblance measures the ratio of the signal energy within a window to the total energy in the window. Noise with nonzero mean and closely-spaced events in the same window sometimes cause a poor velocity resolution when using this measure. The poor resolution of the semblance has lead to more sophisticated techniques based on the eigenstructure of the data covariance matrix (Biondi and Kostov, 1989; Key and Smithson, 1990). In these techniques, the data covariance matrix is decomposed into signal and noise space contributions. Different metrics based on the eigenvector of the signal space are then used to measure coherent energy along hyperbolic paths.

The semblance, or any other velocity measure, can be displayed as a contour map where each maximum corresponds to the pairs $\tau - v$. Mapping the original data back

Manuscript received by the Editor February 17, 1994; revised manuscri t received November 17, 1994.
Dept. of Geophysics and Astronomy, The University of British Columbia, 129-2219 Main Mall, Vancouver, B.C., V6T 1Z4, Canada.

from this space is not possible since these energy measures do not contain phase information. This short-coming may be overcome, however, by using constant-velocity stack gathers that consist of constant-velocity CMP-stacked traces. For an infinite-aperture array, the summation along hyperbolas should map into a point in the velocity space (for a spike-like wavelet). Limited aperture, however, spreads the information and not only makes the identification of each seismic event difficult, but also degrades any reconstruction of the offset space beyond the original aperture. Hence, there is a practical importance for estimating high-resolution or aperture-compensated velocity gathers.

To reduce amplitude smearing in the velocity space, Thorson and Claerbout (1985) performed the inversion of a set of constant-velocity stacks in t - h (time-offset) space. The problem involves the inversion of large matrices that is a difficult task. Thorson and Claerbout developed a stochastic inversion scheme that converges to a solution with minimum entropy. The latter is achieved by defining a Gaussian prior pdf with variable variance. To avoid the inversion of such huge matrices, Yilmaz (1989) posed the problem in an elegant manner in f - h (frequency-offset) space. First, the data are stretched along the time axis and then are Fourier transformed to the f - *h* domain. Finally, a system of complex equations is solved at each frequency with a singular-value decomposition (SVD) routine. The reconstructed CMP gather from the SVD velocity gather exhibits an important signal-to-noise-ratio enhancement and a considerable reduction of amplitude smearing. However, SVD or damped least-squares do not provide enough resolution in model space to correctly extrapolate near- and far-offset traces. Damped least-squares can be derived using the Bayes rule assuming a Gaussian prior. Clearly, a Gaussian prior will impose smoothness on our model.

We seek a velocity gather that resembles the infinite-aperture velocity gather. Therefore, mapping the data from velocity space to offset space not only enhances hyperbolic events, but also extrapolates far- and near-offset traces. In our approach, we follow Yilmaz (1989) but instead of using SVD we derive a maximum a posteriori (MAP) estimator of the velocity gather at each single frequency. The MAP solution leads to velocity gathers consisting mainly of isolated pairs in the $\tau - v$ domain, a procedure which, when mapped back to the CMP space, extrapolates hyperbolic energy outside the original aperture.

VELOCITY GATHERS

Forward and conjugate mapping

Velocity estimation from a CMP gather can be regarded as a linear inverse problem. We are looking for feasible velocity models capable not only of enhancing hyperbolic energy but also of extrapolating near- and far-offset traces. Before dealing with the formalism of the inverse problem, we must understand how aperture in the data space and discretization in the model space affect forward and inverse mapping.

First we describe the transformations that map the offset space into the velocity space and vice-versa. For this purpose let $m(v,\tau)$ be the velocity gather (the model that we are seeking) and *d(h, t)* the data (the CMP gather). Let the variables *h*, v, t, and τ designate the offset, velocity, time, and intercept time, respectively. The forward transformation from offset to velocity involves summation along offset

$$m(v, \tau) = \sum_{h=h_{\min}}^{h_{\max}} d\left(h,\ t = \sqrt{\tau^2 + h^2/v^2}\right). \tag{1a}$$

Similarly, the conjugate transformation (mapping from velocity space to offset space) involves summation along the velocity axis

$$\hat{d}(h,\ t) = \sum_{v=v_{\min}}^{v_{\max}} m\left(v,\ \tau = \sqrt{t^2 - h^2/v^2}\right). \tag{lb}$$

In equation (lb), $\hat{d}(h,\ t)$ is the reconstructed CMP gather. Letting $\alpha = 1/v^2$ we obtain

$$m(\alpha, \tau) = \sum_{h=h_{\min}}^{h_{\max}} d\left(h,\ t = \sqrt{\tau^2 + h^2\alpha}\right), \tag{2a}$$

$$\hat{d}(h, t) = \sum_{\alpha=\alpha_{\min}}^{\alpha_{\max}} m\left(\alpha, \tau = \sqrt{t^2 - h^2\alpha}\right). \tag{2b}$$

In the f - *h* domain, the transformation from offset to velocity may be computed at each frequency leading to an easy to handle system of complex equations. To achieve this goal we have to eliminate the square root in equations (2a) and (2b). A stretching transformation on the time axis serves our purpose

$$t' = t^2 \qquad \tau' = \tau^2.$$

With this transformation, hyperbolic traveltimes are transformed into parabolic traveltimes

$$t^2 = \tau^2 + h^2\alpha \rightarrow t' = \tau' + h^2\alpha.$$

We can now write down the forward and transpose transformations in stretched coordinates

$$m(\alpha, \tau') = \sum_{h=h_{\min}}^{h_{\max}} d(h,\ t' = \tau' + h^2\alpha), \tag{3a}$$

$$\hat{d}(h, t') = \sum_{\alpha=\alpha_{\min}}^{\alpha_{\max}} m(v, \tau' = t' - h^2\alpha). \tag{3b}$$

Equations (3a) and (3b) describe a special form of the generalized Radon transform (Beylkin, 1987), which in our particular problem represents a parabolic transform (Hampson, 1986). Finally, taking Fourier transforms of expressions (3a) and (3b) we end up with our system of complex linear equations in the f - h plane

$$m(\alpha,\ \omega') = \sum_{h=h_{\min}}^{h_{\max}} d(h,\ \omega')e^{i\omega' h^2\alpha}, \tag{4a}$$

$$\hat{d}(h, \omega') = \sum^{\alpha_{\max}} m(\alpha,\ \omega')e^{-i\omega' h^2\alpha}. \tag{4b}$$

$$\hat{\mathbf{d}} = \underset{\sim}{\mathbf{L}}\mathbf{m}, \tag{5b}$$

where d and m are vectors of lengths ND and NP, respectively. The elements of the finite aperture forward operator are given by

$$L_{i,j} = e^{-i\omega' h_i^2 \alpha_j}. \tag{6}$$

Equation (5a) and (5b) define the conjugate and forward mapping of the problem. The operator $\underset{\sim}{\mathbf{L}}^H$ is the conjugate or transpose operator, also called the back-projection operator (Claerbout, 1992). Many problems in geophysics are solved with conjugate operators. Claerbout gives an interesting list of forward operators and their conjugate counterparts. A good example is migration, where we attempt to undo phase shift without modifying the power spectrum of the data and thus without danger of noise amplification. Although this is an obvious advantage, the tradeoff degrades the resolution. Substituting equation (5a) into (5b) we obtain

$$\hat{\mathbf{d}} = \underset{\sim}{\mathbf{L}}\underset{\sim}{\mathbf{L}}^H\mathbf{d}. \tag{7}$$

Let us now examine under what circumstances the transformation L is unitary, or in other words, $\underset{\sim}{\mathbf{L}}\underset{\sim}{\mathbf{L}}^H = \underset{\sim}{\mathbf{I}}$. The elements of $\underset{\sim}{\mathbf{G}} = \underset{\sim}{\mathbf{L}}\underset{\sim}{\mathbf{L}}^H$ are given by

$$g_{i,j} = \frac{1}{k_{\max} - k_{\min} + 1} \sum_{k=k_{\min}}^{k=k_{\max}} e^{i\omega'\delta\alpha k(h_j^2 - h_i^2)}, \tag{8}$$

where $(\sqrt{(k_{\max} - k_{\min} + 1)}$ is a normalization factor. Expression (8) corresponds to a geometric series with the following sum

$$g_{i,j} = \frac{1}{k_{\max} - k_{\min} + 1} \times \frac{\sin\left(\dfrac{\omega'\delta\alpha(h_j^2 - h_i^2)(k_{\max} - k_{\min} + 1)}{2}\right)}{\sin\left(\dfrac{\omega'\delta\alpha(h_j^2 - h_i^2)}{2}\right)} \times e^{i\omega'\delta\alpha(h_j^2 - h_i^2)(k_{\max} + k_{\min})/2}. \tag{9}$$

The kernel given in equation (9) arises from effects of discretization in the velocity space. We note that G depends only on the velocity range scanned and on $\delta\alpha$. Sampling the velocity axis with a dense grid yields the identity matrix, $\underset{\sim}{\mathbf{G}} = \underset{\sim}{\mathbf{I}}$. It must be pointed out that, for a fixed velocity range, decreasing the discretization interval also increases the number of unknowns. It follows that the problem is naturally underdetermined. Finally, the original CMP gather and the predicted one (by means of the conjugate operator) are related by a special type of convolution

$$\hat{d}(h_i, \omega') = \sum_j d(h_j, \omega') g(h_j^2 - h_i^2). \tag{10}$$

This expression explains the smearing of the reflections within the original aperture and the blurred continuation of the data outside the original aperture when this model is used for trace extrapolation. Clearly, these are end effects and not extrapolated traces.

In the previous analysis we have studied the effects of discretization. We proceed by considering the finite-aperture problem itself. To simplify the analysis we will assume that G = I or what is equivalent, that the discretization does not introduce any artifacts in the reconstructed CMP. Let $\mathbf{d}_\infty$ be the infinite-aperture CMP gather, in other words, the gather we should have recorded for an infinite cable length. Let $\underset{\sim}{\mathbf{L}}_\infty$ be the corresponding forward mapping for this unrealistic array

$$\underset{\sim}{\mathbf{L}}_\infty \mathbf{m} = \mathbf{d}_\infty. \tag{11}$$

Now let A be the aperture matrix that transforms an infinite aperture CMP into a limited aperture CMP. This matrix is equivalent to a box-car window and its form is

$$\underset{\sim}{\mathbf{A}} = \begin{pmatrix} \cdots & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \end{pmatrix}. \tag{12}$$

The matrix $\underset{\sim}{\mathbf{A}}$ operates like a window retrieving only a limited number of traces from the infinite-aperture CMP gather. Multiplying both sides of equation (11) by A yields the system of equations for the limited aperture gather

$$\underset{\sim}{\mathbf{A}}\underset{\sim}{\mathbf{L}}_\infty \mathbf{m} = \underset{\sim}{\mathbf{A}}\mathbf{d}_\infty. \tag{13}$$

We recall that AL, = L and Ad, = d. Letting the velocity gather computedby means of the conjugate operator be $\mathbf{m}_c$, we obtain

$$\mathbf{m}_c = \underset{\sim}{\mathbf{L}}_\infty^H \underset{\sim}{\mathbf{A}}^H \underset{\sim}{\mathbf{A}}\mathbf{d}_\infty. \tag{14}$$

The reconstructed and extrapolated CMP gather will be $\mathbf{d}_c = \underset{\sim}{\mathbf{L}}_\infty \mathbf{m}_c$. Then by equation (14)

$$\mathbf{d}_c = \underset{\sim}{\mathbf{L}}_\infty \underset{\sim}{\mathbf{L}}_\infty^H \underset{\sim}{\mathbf{A}}^H \underset{\sim}{\mathbf{A}}\mathbf{d}_\infty. \tag{15}$$

Finally, because we assume a problem free from discretization artifacts $\underset{\sim}{\mathbf{L}}_\infty \underset{\sim}{\mathbf{L}}_\infty^H = \underset{\sim}{\mathbf{I}}$, we can write

$$\begin{aligned} \mathbf{d}_c &= \underset{\sim}{\mathbf{A}}^H \underset{\sim}{\mathbf{A}}\mathbf{d}_\infty \\ &= \underset{\sim}{\mathbf{A}}^H \mathbf{d}. \end{aligned} \tag{16}$$

Equation (16) shows that the model computed with the conjugate operator recovers the original aperture plus a zero value extrapolation

$$\mathbf{d}_c = \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ \mathrm{d} \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{17}$$

When the discretization is not negligible, $\underset{\sim}{\mathbf{G}} \neq \mathbf{I}$ and the extrapolated CMP gather will be

$$\mathbf{d}_c = \underset{\sim}{\mathbf{G}} \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ \mathbf{d} \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{18}$$

We now illustrate our discussion by means of a synthetic example. Three primary events corresponding to a velocity of 3300 m/s, with zero-offset times of 0.4,0.8, and 1.2 s, were used to simulate a CMP gather. Superimposed on these events is a primary reflection corresponding to a velocity of 3000 m/s at 0.2 s with its corresponding multiples at 0.4,0.6, and 0.8 s, ... The CMP gather is plotted in Figure la. Near- and far-offset traces are located at 0 and 3500 m, respectively. The spatial sampling in the offset space is 50 m. Only traces with offsets ranging from 1000 to 2500 m were used to estimate the velocity gather. This spatial window was contaminated with random noise with a standard deviation of 0.1 (Figure lb). In Figure 2a, we show the velocity model computed with the conjugate operator. The spatial window was used to reconstruct the offset space. The results are shown in Figure 2b. We recall equation (16) where we showed that this approach leads to zero trace extension.

The frequency distortion of shallow events is caused by the t2 transformation that compresses the data before 1 s and stretches the data after 1 s. To compute the velocity gather in (τ, α), we must apply a $\tau^{1/2}$ transformation with the consequence that data before 1 s are stretched and data after 1 s are compressed. Because of alias introduced by the t^2 compression before 1 s, the reconstruction is not exact and leads to the aforementioned distortion.

BAYESIAN INVERSION

Inverse problems are naturally ill-posed since they do not satisfy the conditions of existence, uniqueness, and stability of the solution (see for example, Tikhonov and Goncharsky, 1987). The technique that permits the construction of a unique and stable solution by introducing some type of prior information is called regularization. The prior information can be given in a deterministic form or in a stochastic form. Positivity is a common deterministic constraint that is useful for solving a variety of inverse problems (e.g., magnetic susceptibility inversion, density inversion, etc.). A stochastic prior assumes some relevant information about the model in terms of moments of corresponding distributions.

In many problems, the regularization is carried out regardless of the nature of the model we are seeking. That is the

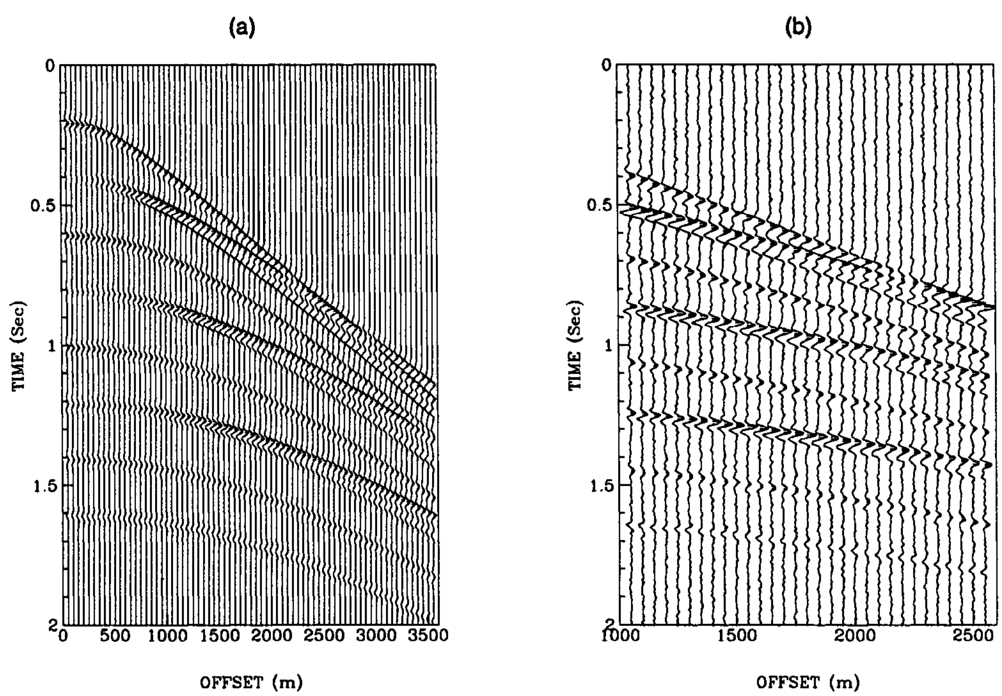

FIG. 1. (a) Synthetic CMP gather. (b) Spatial window extracted from panel (a). Panel (b) was contaminated with Gaussian noise (σ^2 = 0.1).

case of the widely used quadratic regularization. This method has the advantage of imposing smoothness to the model that is the common way to avoid the amplification of random errors associated with each observation. The well known prewhitening technique used in spiking and predictive deconvolution is an example of quadratic regularization. However, in many situations, we may wish to use other types of regularization that permit us to incorporate some relevant information about the model. If the additional information is in the form of a pdf, it can be combined with the data using Bayes's rule. In the next section, we will show how the prior probability induces a particular regularization on the inverse problem, i.e., a Gaussian prior leads to quadratic regularization.

To clarify some concepts, let us start with Bayes's rule. If p(A/I) denotes the probability of proposition A given proposition I, then Bayes's rules establishes the way of updating p(A/I) when additional information B is incorporated,

$$p(A|BI) = p(A|I)\frac{p(B|AI)}{p(B|I)}. \tag{19}$$

In equation (19) the proposition I is the prior information, and p(A/I) is the prior probability of *A* conditional only on the information given by I. The term on the left-hand-side of equation (19) is called the posteriori probability. For Bayesians, probability is a measure of the degree of plausibility of a proposition. Basically, Bayes's rule serves to update the plausibility of a proposition when our state of information changes because new data are acquired.

Bayesian approach to inversion

The relationship between m`odel space and data space when noise is considered is given by

$$\underset{\sim}{\mathbf{L}}\mathbf{m} + \mathbf{n} = \mathbf{d}, \tag{20}$$

where n stands for the noise term. Letting m = *A* be the proposition we want to assess and d = *B* represent our data, Bayes's rule can be written as

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \tag{21}$$

where for simplicity we have omitted I in the notation. To clarify notation, p(d/m) is the probability or likelihood of obtaining the data d assuming that the model m is true and given the prior I, p(m) is the prior probability of the model, p(d) is the data likelihood and enters into the problem as a normalization factor, and p(m/d) is the posterior probability of the model.

To use the Bayesian formalism we must consider two remaining problems. First, given p(m/d), how is m determined, and second, how is p(m) determined. The first problem can be solved by an appropriate choice of a decision rule. For example, we could use the MAP solution, m_{MAP}, which maximizes p(m/d). The second problem, which is probably one of the most cumbersome problems in inverse theory, is how to translate our prior appraisal about the model into a probability density function.

Various attempts have been made to find prior probabilities that represent a state of total ignorance about the model (Jeffreys, 1961; Jaynes, 1968). These priors are derived from mathematical arguments of symmetry and invariance. In our approach, we first derive a global constraint in terms of moments of the underling model pdf, and then we use the principle of maximum entropy to compute the prior probability.

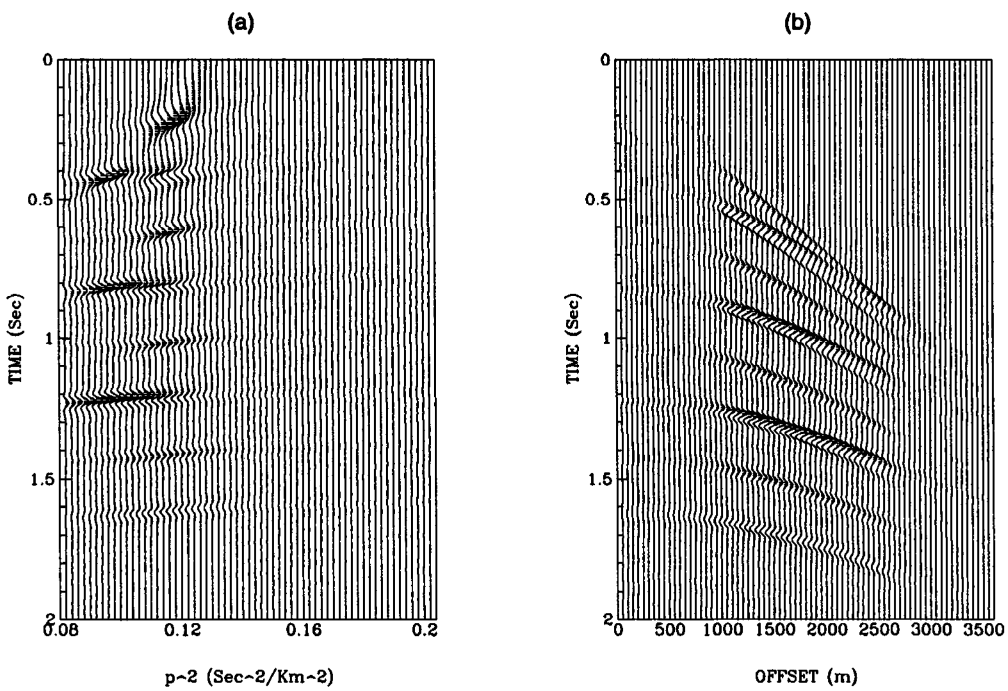

FIG. 2. (a) Velocity processing with the conjugate operator. Panel (b) shows the reconstructed CMP gather.

Prior probabilities and maximum entropy

Consider a continuous variable m with pdf p(m). The entropy h given by

$$h = -\int p(\mathbf{m}) \log [p(\mathbf{m})]\, d\mathbf{m}, \qquad (22)$$

expresses the uncertainty associated with the distribution P(m). Therefore, the distribution that most honestly describes our data, given only what we know without assuming anything else, is the one with maximum entropy (Jaynes, 1968). It is natural, therefore, that when we attempt to make inferences based on incomplete information, that we should draw them from that probability distribution that has the maximum entropy allowed by the available information. Suppose we have information about m in the form of some global constraint S(m). Then, the corresponding maximum entropy probability distribution will have the following form

$$p(\mathbf{m}) = e^{[-\lambda_0 - \lambda_1 S(\mathbf{m})]}, \qquad (23)$$

where the constants λ_0, λ_1 are Lagrange multipliers associated with this constrained maximization problem. In fact, λ_0 is a normalization factor that yields $\int p(x)dx = 1$. The constraint to the problem, S(m), contains our prior information about the model. [As pointed out by Jaynes (1985), the constraints need not be in terms of mathematical expectations.] Among all the possible solutions to the underdetermined inverse problem described in the previous sections, we seek solutions in which a reasonable feature of the model is reached (feasible solutions). For an infinite-aperture array, the velocity gather will contain isolated wavelets at each pair (τ, v). Each wavelet will occupy a single point on the velocity axis with support along τ depending on the frequency content. A solution with minimum structure (a sparse solution) will fulfill our objectives. A constraint of the form

$$S(\mathbf{m}) = \sum_{i=1}^{NP} \ln\, [m_i m_i^* + b], \qquad (24)$$

can be used to quantify the amount of sparseness of a vector. In equation (24), b is a small additive perturbation that represents the default power in absence of hyperbolic events. It also enforces continuity to the gradient of S(m) that will be used in the derivation of our algorithm. When the signals are linear events, the formalism leads to a 2-D spectral estimation problem in which the velocity axis (in fact the ray parameter axis) can be replaced by the wavenumber, and the argument $m_i m_i^*$ represents a spectral power. Consequently, when $b = 0$ the metric defined by equation (24) is analogous to Burg's entropy (Burg, 1975).

The prior probability according to equation (23) and (24) is given by

$$p(\mathbf{m}) = e^{[-\lambda_0 - \lambda_1 S(\mathbf{m})]} = \frac{e^{-\lambda_0}}{\prod_i (|m_i|^2 + b)^{\lambda_1}}. \qquad (25)$$

For a single element of the vector of unknowns, the 1-D prior is given by

$$p(m_i) \propto e^{-\lambda_1 \ln(|m_i|^2 + b)} \propto \frac{1}{(|m_i|^2 + b)^{\lambda_1}}, \qquad (26)$$

where parameter λ_1 controls the amount of sparseness in the model. Large values of λ_1 lead to sharp distributions. On the other hand, when λ_1 is very small, the last expression yields a uniform priori.

Having defined our constraint, we recall equations (23) and (25) to compute the posterior distribution

$$p(\mathbf{m}|\mathbf{d}) = K e^{[-\lambda_0 - \lambda_1 S(\mathbf{m})]} e^{-1/2(\underset{\sim}{\mathbf{L}}\mathbf{m}-\mathbf{d})^H \underset{\sim}{\mathbf{C}}^{-1}(\underset{\sim}{\mathbf{L}}\mathbf{m}-\mathbf{d})}. \qquad (27)$$

To derive the last equation, we assumed that the noise vector is normally distributed. In other words, the conditional probability of the data, given the model, is Gaussian. K is a normalization constant and $\underset{\sim}{\mathbf{C}}$ is the covariance matrix of the noise. It turns out that the MAP solution is achieved by minimizing the following objective function

$$\Phi = \lambda S(\mathbf{m}) + (\underset{\sim}{\mathbf{L}}\mathbf{m} - \mathbf{d})^H \underset{\sim}{\mathbf{C}}^{-1} (\underset{\sim}{\mathbf{L}}\mathbf{m} - \mathbf{d}), \qquad (28)$$

where $\lambda = 2\lambda_1$.

Minimization of the objective function

For simplicity we separate the objective function as follows

$$\Phi = \lambda S + Q. \qquad (29)$$

At the MAP solution the following expression must be satisfied

$$\frac{\partial \Phi}{\partial \mathbf{m}} = 0, \qquad (30)$$

where the gradients are given by

$$\frac{\partial S}{\partial \mathbf{m}} = \underset{\sim}{\mathbf{P}}^{-1}\mathbf{m}, \qquad \frac{\partial Q}{\partial \mathbf{m}} = \underset{\sim}{\mathbf{L}}^H \underset{\sim}{\mathbf{C}}^{-1} \underset{\sim}{\mathbf{L}}\mathbf{m} - \underset{\sim}{\mathbf{L}}^H \underset{\sim}{\mathbf{C}}^{-1}\mathbf{d}. \qquad (31)$$

The matrix $\underset{\sim}{\mathbf{P}}(\mathbf{m})$ is diagonal with elements given by

$$p_{ii} = |m_i|^2 + b. \qquad (32)$$

Finally, we write the solution in the following form

$$\mathbf{m} = (\lambda \underset{\sim}{\mathbf{P}}(\mathbf{m})^{-1} + \underset{\sim}{\mathbf{L}}^H \underset{\sim}{\mathbf{C}}^{-1} \underset{\sim}{\mathbf{L}})^{-1} \underset{\sim}{\mathbf{L}}^H \underset{\sim}{\mathbf{C}}^{-1}\mathbf{d}\,. \qquad (33)$$

The last expression can be rewritten as follows (Tarantola, 1987, 158)

$$\mathrm{m} = \underset{\sim}{\mathbf{P}}\underset{\sim}{\mathbf{L}}^H (\lambda \underset{\sim}{\mathbf{C}} + \underset{\sim}{\mathbf{L}}\underset{\sim}{\mathbf{P}}\underset{\sim}{\mathbf{L}}^H)^{-1}\mathbf{d}. \qquad (34)$$

The parameter A controls the ratio between sparseness and fitting. The diagonal matrix P plays a fundamental role in the inversion of sparse models. At each iteration only large powers, $|m_i|^2$, are kept, reducing the powers that are not relevant in fitting the data. The background power b makes the gradient continuous when m_i approaches zero. It also provides a lower bound to the constraint S(m). It is interesting to note that equations (33) and (34) resemble the damped least-squares solution and the minimum model norm solution, respectively.

Our solution has been reduced to a nonquadratic regularized least-squares solution. In damped least-squares, or in the SVD solution, the regularization is accomplished via a constant perturbation to the main diagonal of the pseudoinverse matrix (Lines and Treitel, 1984). This is equivalent to assuming that our prior is Gaussian. This assumption will introduce smoothness in the solution, which is precisely what we do not want. In fact, when b is large, In $(|m_i|^2 + b) \approx |mi|^2/b + \ln$ (b) and equation (26) can be replaced by

$$p(m_i) \propto e^{-|m_i|^2/(b/\lambda_1)}. \quad (35)$$

This is a Gaussian prior that yields the quadratic regularization or damped least-squares solution

$$m = (\mu \mathbf{I} + \mathbf{L}^H \mathbf{C}^{-1} \mathbf{L})^{-1} \mathbf{L}^H \mathbf{C}^{-1} \mathbf{d}, \quad (36)$$

or according to equation (34)

$$m = \mathbf{L}^H (\mu \mathbf{C} + \mathbf{L} \mathbf{L}^H)^{-1} \mathbf{d}, \quad (37)$$

where $\mu = 2\lambda_1/b$. When exact data are considered, C = 0, and equation (37) reduces to the minimum norm solution for underdetermined linear inverse problems (Tarantola, 1987). Another interesting feature of the solution occurs for large values of μ. In this case the pseudoinverse matrix is diagonal dominant, and the model corresponds to the one we might have computed with the conjugate operator.

In the derivation of the prior density, we assumed the existence of only a global constraint S(m). Suppose that we wish to compute the prior probability, then we need to assign a numerical value to the global constraint and manipulate the equations to obtain the Lagrange multipliers. Actually, we do not have a numerical value of the constraint. Our only constraint has been the statement that the solution should be sparse, which is equivalent to finding a minimum of S(m). We can, however, determine the correct prior by means of the misfit function Q. Since the noise has been assumed Gaussian, Q obeys a x^2 statistic and the expected value of the misfit is given by $E(Q) = NP$, where NP is the number of parameters. A line search is performed to find the value of λ which yields the proper misfit target. It is straightforward to see that this single parameter completely defines the priori probability and the numerical value of the constraint. This can be important in problems in which we seek not only a MAP solution but we also wish to describe the inverted model by means of other measures such as mean, median, variance, etc.

The specification of the proper misfit is still a problem since we know the actual structure of the noise covariance matrix only approximately. However, we have found that for reasonable levels of noise the parameter λ can be estimated easily using a tentative misfit target. The results can be visually inspected to find evidence of overfitting or underfitting. When data are over&ted, the algorithm attempts to create hyperbolic energy from the noise. When mapping the model back to the offset space it is easy to recognize new hyperbolas in the original aperture.

The synthetic data plotted in Figure lb were processed with the damped least-squares and our sparse inversion scheme. The results achieved with the damped least-squares technique are shown in Figure 3a (velocity gather) and Figure 3b (reconstructed gather). We emphasize that only offsets ranging from 1000 to 2500 m were used to invert the

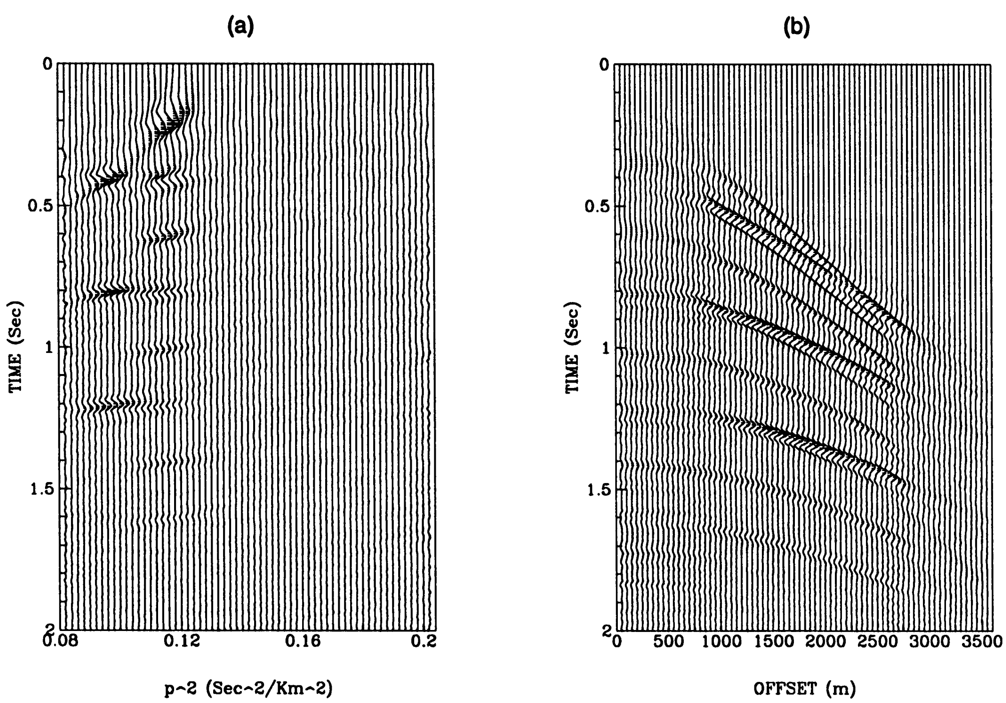

FIG. 3. (a) Velocity gather computed with the least-squares approach. Panel (b) shows the reconstructed CMP gather.

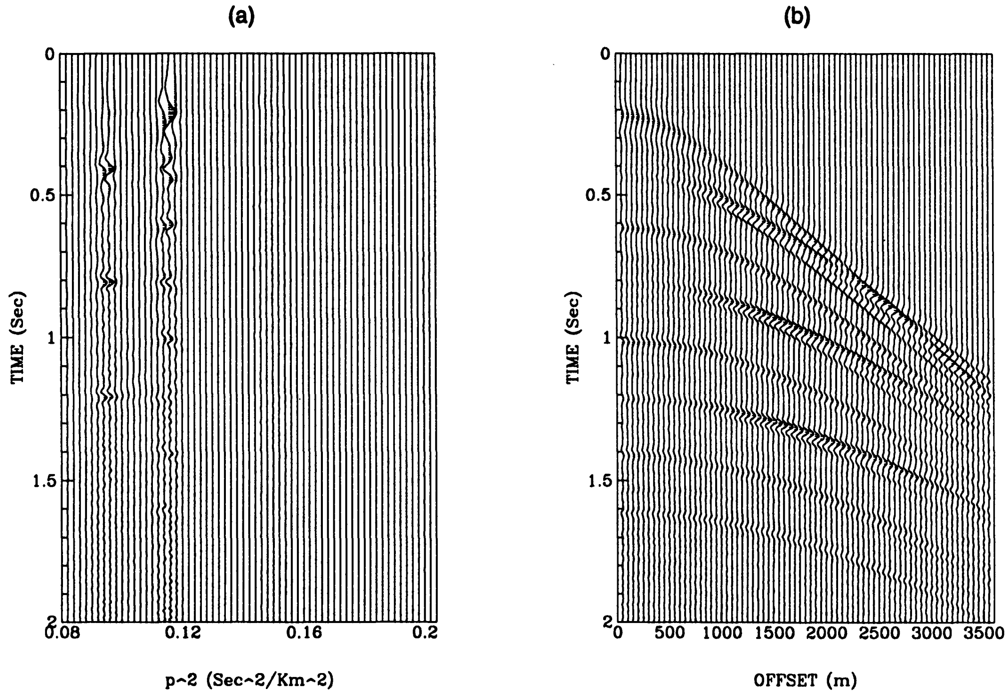

FIG. 4. (a) High-resolution velocity gather computed with the sparse inversion scheme. Panel (b) shows the reconstructed CMP gather.

velocity panel. The resolution in model space is clearly improved. However, near- and far-offset traces are not correctly extrapolated. The same example was then solved with the sparse inversion algorithm. The velocity panel is shown in Figure 4a. Primary and multiple events are perfectly isolated and the noise does not introduce spurious details in the model. Finally, the velocity gather was used to reconstruct the offset space. The result is shown in Figure 4b. Near- and far-offset traces are correctly extrapolated, except for an unavoidable phase shift in the wavelet caused by the t^2 transformation.

DISCUSSION

The high-resolution algorithm described here provides an interesting new approach to limited aperture problems. By defining the problem in the f - h domain, we have been able to decouple discretization and aperture effects. The discretization artifacts are reduced by decreasing the sampling interval in the velocity space. This effect emphasizes the underdetermined nature of the problem.

The sparse inversion technique allows us to construct a solution consisting of a few nonzero velocity events, a feature that is consistent with the expected model. We have shown the importance of the prior information in the overall inversion and how the prior induces a particular regularization of the inverse problem. The resolution that is achieved permits not only the correct reconstruction of the original aperture but also the extrapolation of near- and far-offset traces.

The maximum entropy principle imposes a noncommittal procedure for determining the required a priori densities. The algorithm uses these densities to produce a sparse solution, which according to Burg's definition of entropy, also corresponds to a minimum entropy solution. There is no inconsistency here; the principle of maximum entropy is used to find pdfs. The cost function that is defined in terms of these pdfs may lead to a sparse model.

Finally, we would like to point out that the algorithm can be used efficiently to solve other problems. Two examples that come to mind are 2-D spectral estimation in sonar processing and the separation of down and upgoing waves in VSP. These issues are under investigation.

ACKNOWLEDGMENTS

We wish to thank Oz Yilmaz for stimulating our interest in the quest for higher resolution. We wish to acknowledge the generous support of Imperial Oil Limited through its University Research Grants. MDS appreciates the support of a UBC graduate fellowship. This research was also supported by a NSERC grant to TJU.

REFERENCES

Beylkin, G., 1987, Discrete radon transform, IEEE Trans., Acoust., speech and signal proc., ASSP-35, 162-172.
Biondi, B. L., and Kostov, C., 1989, High-resolution velocity spectra using eigenstructure methods: Geophysics, 54, 832-842.
Burg, J. P., 1975, Maximum entropy spectral analysis: Ph.D. thesis, Stanford University.
Claerbout, J. F., 1992, Earth soundings analysis, Processing versus inversion: Blackwell Scientific Publ.
Hampson, D., 1986, Inverse velocity stacking for multiple estimation: 56th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 422424.
Jaynes, E. T., 1968, Prior probabilities: IEEE Trans., Syst. Sci. Cybern. SSC-4, 227-241.

——— 1985, Where do we go from here?, in Ray Smith, C., and Grandy, W. T. Jr., Eds., Maximum-entropy and Bayesian methods in inverse problems: D. Reidel Publ. Co., 21-58.
Jeffreys, H., 1961, Theory of probability, Oxford Univ. Press (earlier editions 139 and 1948).
Key, S. C., and Smithson, S. B., 1990, New approach to seismic-reflection event detection and velocity determination: Geophysics, 55, 1057-1069.
Lines, L. R., and Treitel, S., 1984, Tutorial: A review of least-squares inversion and its applications to geophysical problems: Geophys. Prosp., 32, 159-186.
Neidell, N. S., and Taner, M. T., 1971, Semblance and other coherency measures for multichannel data: Geophysics, 36, 482-497.
Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science Publ.
Thorson, J. R., and Claerbout, J.F., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, 50, 2727-2741.
Tikhonov, A. N., and Goncharsky, A. V., Eds, 1987, Ill-posed problems in the natural sciences, MIR publisher.
Yilmaz, O., 1989, Velocity-stack processing: Geophys. Prosp., 37, 357-382.