Restoration of Missing Data by Parsimony in the Frequency Domain

Jeff Thorson

Basic Problem

Consider the problem of expanding a two-dimensional data set, such as a seismic reflection profile, by interpolation or extrapolation. Intuitively it should be done in such a way that no new information is added to the expanded data set. "New information", such as the truncation of a bed, high frequency components or an event interpolated at an unreasonable dip, is essentially undesirable information, and can be recognized as such by the eye on the seismic section. However, if further processing (e.g. migration) is performed, the artifacts may be transformed into apparent events that are difficult to distinguish from real events.

In the case of a gather of time traces that are sparsely or unevenly sampled in the spatial (x) direction, padding the gather with zero traces is injecting new information into it in the form of high spatial wavenumber components in the (k_x,ω) domain. Looking at the zero-padded gather back in the (x,t) domain it is obvious where the real information does or does not lie; a particular domain must be selected where the essential information is hypothetically local. For example, if it is claimed that a gather shall consist only of straight dipping continuous events, the Fourier transform domain localizes this information: valid events are aligned on straight lines passing through the origin. Once a domain is defined in which information is local in some sense, filtering may be applied in this domain. Spurious events originating solely from the padding of zero traces in the (x,t) domain may be removed as far as they do not correspond to the model for real data. As a consequence the padded traces are no longer zero, but some smooth continuation of the real data.

Two general steps are involved here. First, a desired field must be estimated by examining data in information space (the selected domain where information is supposedly local). Second, the data itself is extrapolated to approximate this estimation as well as possible,

subject to constraints. We shall always impose the constraints that the original data in the data space (the (x,t) domain) remains unchanged; the variables in the procedure are the samples of the padded traces.

Formulation

The discussion in the previous section has been very general; let us now formulate the problem specifically. Assume first a unitary transformation **F** from data space (x) to model space (y). A vector **x** in the data space can be partitioned into two pieces; $\mathbf{x}_A$ consisting of those elements in the region to be extended, and $\mathbf{x}_B$ consisting of those elements of the original data. Let **S** ("select") be the projection operator that zeros out the $\mathbf{x}_B$ component of any **x**. The transform **F** may also be partitioned into pieces (prime indicates transpose):

Restricted Transforms			domain	range
$\mathbf{F}_A$	$\equiv$	$\mathbf{FS}$	$\mathbf{x}_A$	**y**
$\mathbf{F}_B$	$\equiv$	$\mathbf{F(I-S)}$	$\mathbf{x}_B$	**y**
$\mathbf{F}'_A$	$\equiv$	$\mathbf{SF}'$	**y**	$\mathbf{x}_A$
$\mathbf{F}'_B$	$\equiv$	$\mathbf{(I-S)F}'$	**y**	$\mathbf{x}_B$

From now on we will be concerned with the following problem: minimize $\mathbf{y}'\mathbf{W}\mathbf{y}$ under the constraint that the original data be fixed.

$$\min_{\mathbf{y}} \mathbf{y}'\mathbf{W}\mathbf{y} \quad \textit{with constraints} \quad \mathbf{F}_B\mathbf{y} = \mathbf{x}_B \tag{2}$$

W is considered to be a diagonal matrix, so that the functional $\mathbf{y}'\mathbf{W}\mathbf{y}$ is simply a weighted least squares energy of the vector **y** in information space. The constraints $\mathbf{F}'_B = \mathbf{x}_B$ restrict any solution **y** to honor the original data $\mathbf{x}_B$.

The estimation step is involved with selecting the weighting function **W** in model space. To see this, suppose a desired function **g** has been given (by some estimation) in model space that **y** should be closest to in some sense. Minimizing $\|\mathbf{y}/\mathbf{g}\|_2$ ($\mathbf{y}/\mathbf{g}$ is termwise division) is equivalent to setting $\mathbf{W} = 1/|\mathbf{g}|^2$. Note that in specifying **W** only the amplitude of the desired **g** need be given, not the phase. Two things are crucial for the minimization of (2) to work: the nature of the constraints and the nature of **W**. Equation (2) has a trivial solution **y** = 0 if there are no constraints supplied, and with respect to continuity of the functional $\mathbf{y}'\mathbf{W}\mathbf{y}$ one would imagine that a "good" solution may be obtained only if the constraints provide enough support to the function **y**. For the case of interpolating a new trace in between each original trace, the constraints are fairly dense in (x,t) space.

On the other hand, if **W** is constant, the interpolated traces are identically zero. That is,

$$\min_{\mathbf{y}} \mathbf{y}'\mathbf{W}\mathbf{y} \;=\; w \min_{\mathbf{x}} \mathbf{x}'\mathbf{F}'\mathbf{F}\mathbf{x} \;=\; w \min_{\mathbf{x}} \mathbf{x}'\mathbf{x}$$

with the constraints $\mathbf{x}_B = constant$ gives the obvious solution $\mathbf{x}_A = 0$. Since the estimation $\mathbf{y}'\mathbf{W}\mathbf{y}$ is biased toward smaller total energy, the values of **W** have to have a sufficient amount of dynamic range to make the spectrum $\mathbf{y}'\mathbf{y}$ look like the inverse of **W**, that is, the desired spectrum $\mathbf{g}'\mathbf{g}$.

Implementation

There are various ways to perform the minimization of (2). First, the unknowns to be solved for may be taken to be the vector $\mathbf{x}_A$. Partition (2) into A and B spaces, using definitions in (1):

$$\min_{\mathbf{x}_A} \begin{bmatrix} \mathbf{x}_A' & \mathbf{x}_B' \end{bmatrix} \begin{bmatrix} \mathbf{F}_A' \\ \mathbf{F}_B' \end{bmatrix} \mathbf{W} \begin{bmatrix} \mathbf{F}_A & \mathbf{F}_B \end{bmatrix} \begin{bmatrix} \mathbf{x}_A \\ \mathbf{x}_B \end{bmatrix}$$

Making the variation of this functional over x_A zero yields the linear system

$$\mathbf{F}_A'\mathbf{W}\mathbf{F}_A\mathbf{x}_A \;=\; -\mathbf{F}_A'\mathbf{W}\mathbf{F}_B\mathbf{x}_B \tag{3}$$

(Claerbout, SEP-25, p. 3).

Alternately, one may choose to solve a system whose dimension is as large as the dimension of $\mathbf{x}_B$: the data region. By introducing Lagrange multipliers, the minimization of (2) is equivalent to

$$\min_{\mathbf{y},\,\lambda} \mathbf{y}'\mathbf{W}\mathbf{y} + \lambda'(\mathbf{F}_B'\mathbf{y} - \mathbf{x}_B)$$

where the dimension of the unknown λ vector is the dimension of $\mathbf{x}_B$. Taking variations over **y** yields $\mathbf{W}\mathbf{y} + \mathbf{F}_B\lambda = 0$ while taking variations over λ gives the constraint equation $\mathbf{F}_B'\mathbf{y} = \mathbf{x}_B$. Suppose **W** to be positive definite. Then $\mathbf{y} = -\mathbf{W}^{-1}\mathbf{F}_B'\lambda$, and together with the constraints gives

$$\mathbf{F}_B'\mathbf{W}^{-1}\mathbf{F}_B\lambda = -\mathbf{x}_B \tag{4}$$

Equation (4) allows to solve for lambda, and once lambda is known, **y** and $\mathbf{x}_A$ can be obtained:

$$\mathbf{x}_A = \mathbf{F}_A'\mathbf{y} = -\mathbf{F}_A'\mathbf{W}^{-1}\mathbf{F}_B'\lambda$$

For a two-dimensional data set, the size of the linear systems (3) or (4) are prohibitively large for a direct solution. If the transform **F** is taken for example to be a two-dimensional DFT, the systems (3) and (4) are far from being banded. Two iterative methods may be used to approximately solve (3) or (4), which for example do not require knowledge of the Hessian: steepest descent and the conjugate gradient method. The algorithms are described in figure 1. One iteration of the conjugate gradient method is worth about two iterations of steepest descent, while taking only a minimal amount of extra time and storage over that of steepest descent.

Solve: $\mathbf{Qx} = \mathbf{b}$

Steepest Descent Algorithm	Conjugate Gradient Algorithm
Initial $\mathbf{x}_0 = 0$	Initial $\mathbf{x}_0 = 0$
Gradient $\mathbf{g}_0 = \mathbf{Q}\mathbf{x}_0 - \mathbf{b}$	Gradient $\mathbf{g}_0 = \mathbf{Q}\mathbf{x}_0 - \mathbf{b}$
	$\delta\mathbf{x}_0 = -\mathbf{g}_0, \quad \gamma_0 = \mathbf{g}_0' \mathbf{g}_0$
for $k = 0,1,2, \cdots$	for $k = 0,1,2, \ldots, dim(\mathbf{Q})$
$\delta\mathbf{g}_k = \mathbf{Q}\mathbf{g}_k$	$\delta\mathbf{g}_k = \mathbf{Q}\mathbf{g}_k$
$\alpha_k = \dfrac{\mathbf{g}_k' \mathbf{g}_k}{\mathbf{g}_k' \delta\mathbf{g}'_k}$	$\alpha_k = \dfrac{\gamma_k}{\delta\mathbf{x}_k' \delta\mathbf{g}_k}$
$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{g}_k$	$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \delta\mathbf{x}_k$
$\mathbf{g}_{k+1} = \mathbf{g}_k - \alpha_k \delta\mathbf{g}_k$	$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \delta\mathbf{g}_k$
next k	$\gamma_{k+1} = \mathbf{g}_{k+1}' \mathbf{g}_{k+1}$
	$\beta_k = \gamma_{k+1} / \gamma_k$
	$\delta\mathbf{x}_{k+1} = -\mathbf{g}_k + \beta_k \delta\mathbf{x}_k$
	next k

FIG. 1. Descent algorithms that do not require knowledge of the Hessian of **Q**. Equations (3) and (4) are of the form $\mathbf{Qx} = \mathbf{b}$; $\mathbf{g}_k$, $\delta\mathbf{g}_k$, $\mathbf{x}_k$, and $\delta\mathbf{x}_k$ are vectors, while α_k, β_k, and γ_k are scalars. Primes denote complex conjugate transpose. These algorithms may be further streamlined in the case of equation (3) by incorporating the computations with the right hand side **b** into that of **Qx**.

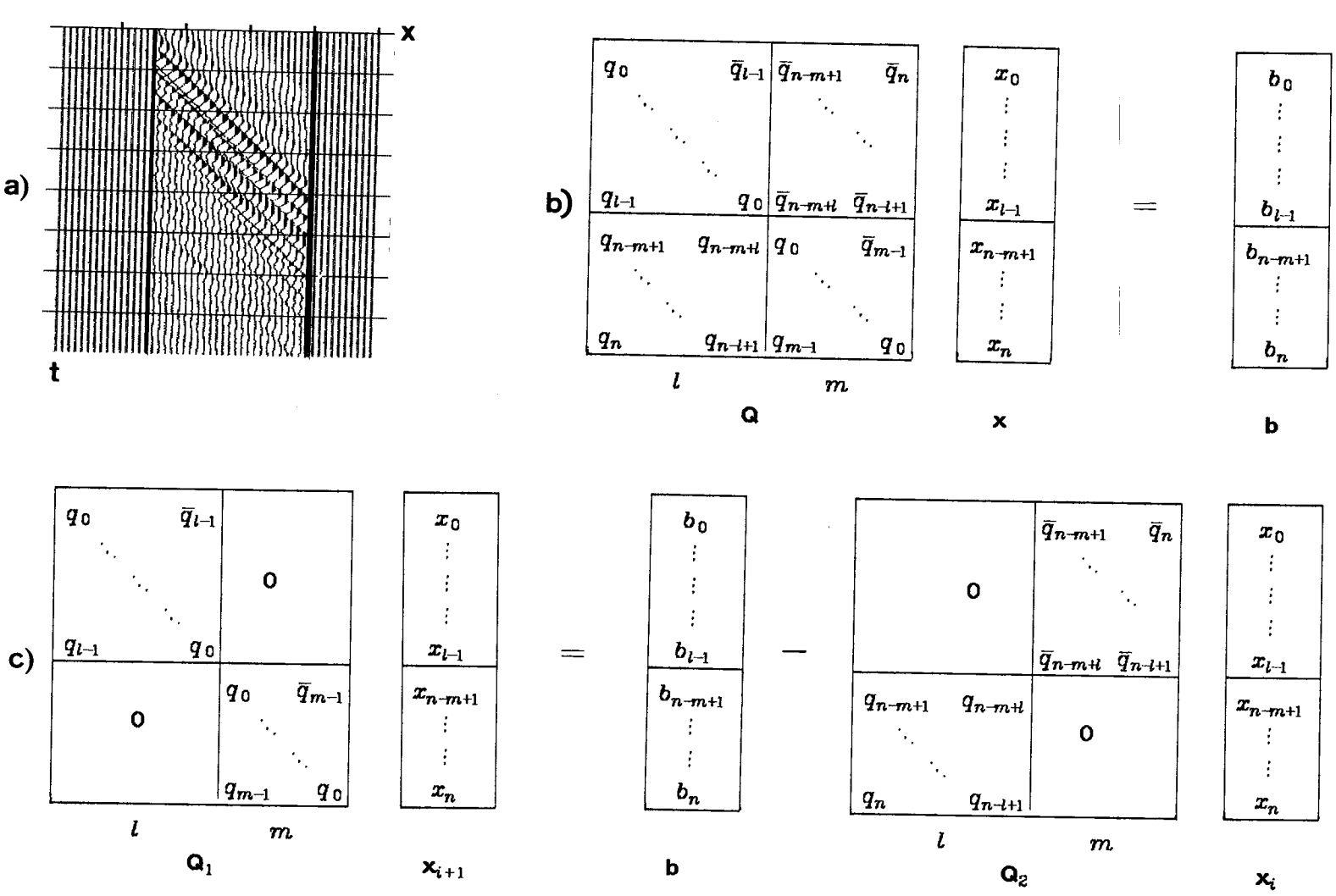

FIG. 2. A partial-Toeplitz system (b) arising from the configuration (a) of traces to be interpolated. **Q** in (b) is actually $\mathbf{F}_A'\mathbf{W}\mathbf{F}_A$ where $\mathbf{F}_A$ is defined in (1) in the text. The projector **S** selects the traces to be interpolated (the zero traces) of (a). The iterative system (c) can be derived from (b) by splitting **Q** into $\mathbf{Q}_1$ and $\mathbf{Q}_2$. $\mathbf{Q}_1$ can be a number of small independent Toeplitz systems. The rate of convergence of (c) is proportional to $\| \mathbf{Q}_1^{-1}\mathbf{Q}_2 \|^{-1}$ (Strang, p. 283).

Toeplitz and Near-Toeplitz systems

If **F** is taken to be a two-dimensional Fourier transform, the resulting systems (3) and (4) are block Toeplitz matrices under certain conditions. The condition for (3) to be Toeplitz is that the samples to be interpolated are either all contiguous or all evenly spaced. The condition for (4) to be Toeplitz is the same, but with respect to the known data space $\mathbf{x}_B$.

There is another assumption that can practically always be made that will simplify the work of solving (3) or (4): the projection operator **S** is defined in units of whole traces; that is, complete traces in time are to be interpolated rather than parts of traces. When given an estimated **W**, the problem separates in temporal frequency ω. Each frequency ω may be solved for independently. This is because the matrices $\mathbf{F}_A'\mathbf{W}\mathbf{F}_A$ and $\mathbf{F}_B'\mathbf{W}^{-1}\mathbf{F}_B$ decouple in ω when the projection operator **S** that carries **x** into $\mathbf{x}_A$ is independent of time t. This remains true for any arbitrary **F** that separates in t and x. Therefore, depending on **F** a two-dimensional problem (2) may degenerate into a number of one-dimensional problems, each of which is Toeplitz.

When the interpolated traces are not evenly spaced or contiguous, the systems are not Toeplitz. Two possible geometries of missing traces to known traces are indicated in figures 2 and 3. Though they are not Toeplitz, the almost-Toeplitz nature may be taken advantage of in an iterative scheme based on splitting the matrices into a Toeplitz and a non-Toeplitz part. Convergence depends on the spectral radius of $\mathbf{Q}_1^{-1}\mathbf{Q}_2$ of figures 2 and 3, and is guaranteed if the spectral radius is less than 1 (Strang, p. 283). See the figures for a description of the splitting. In each of the cases, the matrix is split so that "most" of it resides in the Toeplitz part $\mathbf{Q}_1$, where hopefully $\|\mathbf{Q}_2\| \ll \|\mathbf{Q}_1\|$. Assume that the eigenvalues of $\mathbf{Q}_1$ and the original matrix **Q** are all positive real (i.e. all the weights in **W** are positive). Then the eigenvalues of $\mathbf{Q}_1^{-1}$ are positive and real. Now a sufficient condition that $\|\mathbf{Q}_1^{-1}\mathbf{Q}_2\| < 1$ is that $\|\mathbf{Q}_1^{-1}\| . \|\mathbf{Q}_2\| < 1$, or $\kappa(\mathbf{Q}_1)\|\mathbf{Q}_2\| < \|\mathbf{Q}_1\|$. Therefore $\|\mathbf{Q}_1\|$ has to be greater than $\|\mathbf{Q}_2\|$ by at least the factor $\kappa(\mathbf{Q}_1)$, where $\kappa(\mathbf{Q}_1) \equiv \|\mathbf{Q}_1^{-1}\| . \|\mathbf{Q}_1\|$ is the condition number of $\mathbf{Q}_1$. $\kappa(\mathbf{Q}_1)$ is about the ratio of the largest value of the weight **W** to its smallest. Convergence of the methods of figures 2 and 3 thus depend on the dynamic range of **W**.

A similar statement concerning convergence may be made for the case of the steepest descent algorithm in figure 1. Luenberger (p. 153) shows that the convergence rate of steepest descent depends on the condition number $\kappa(\mathbf{Q})$ of $\mathbf{Q} = \mathbf{F}_A'\mathbf{W}\mathbf{F}_A$ (or $\mathbf{F}_B'\mathbf{W}^{-1}\mathbf{F}_B$ in the case of equation 4). The higher $\kappa(\mathbf{Q})$ is, (i.e. the greater the dynamic range of **W**) the smaller the convergence rate of the steepest descent method. There is therefore a tradeoff between accuracy of the interpolation and the rate of convergence of the algorithms. As we shall see in the examples of the following sections, a good interpolation requires a large ratio

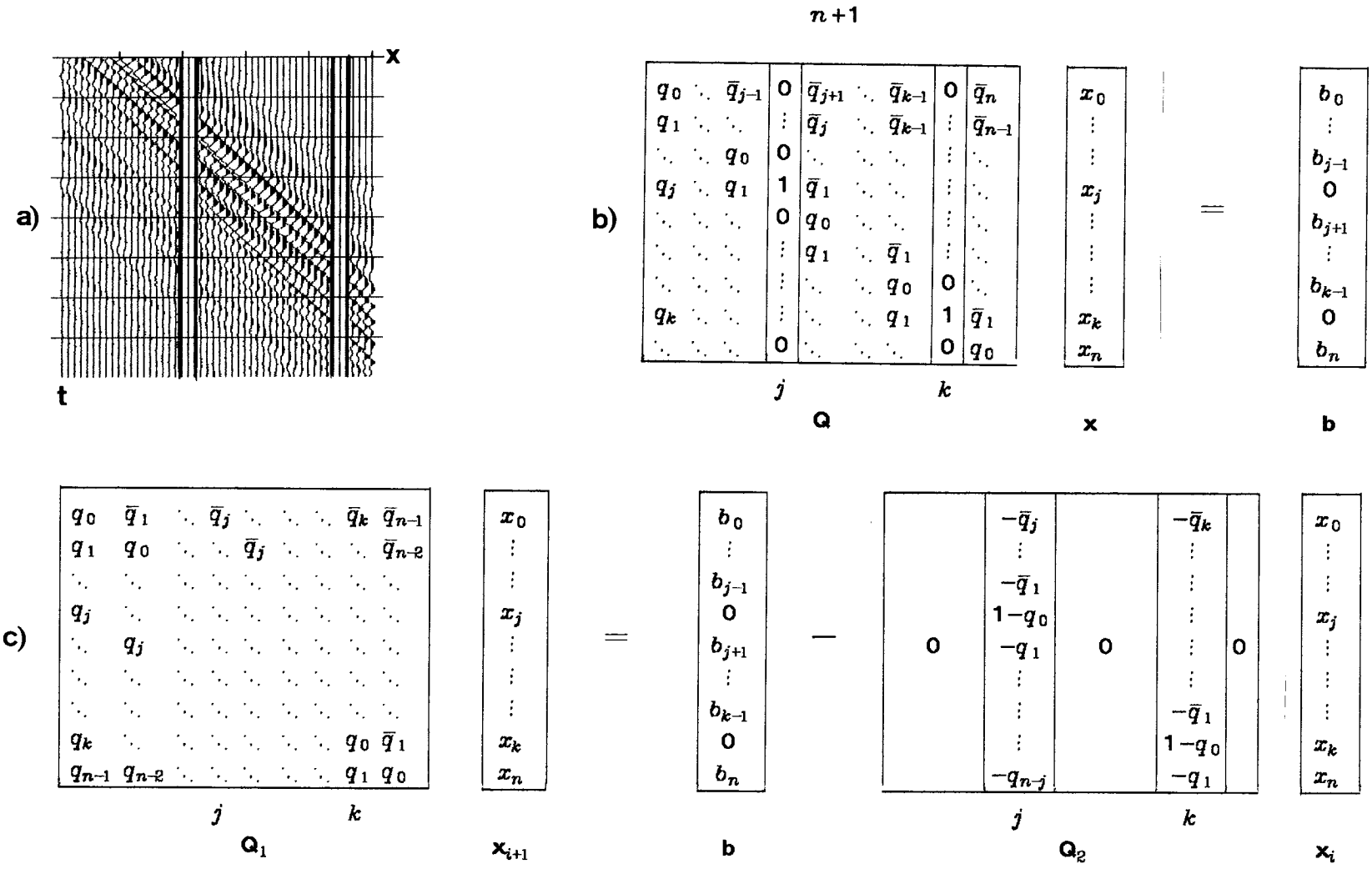

FIG. 3. Another partial-Toeplitz system (b) arising from a geometry (a) of missing traces. In this case, **Q** is $\mathbf{F}_B'\mathbf{W}\mathbf{F}_B$ with extra columns to expand it out to the full dimension of $\mathbf{F}'\mathbf{W}\mathbf{F}$. The jth and kth equations of (b) are superfluous and do not affect the solution **x**, except for the values x_j and x_k respectively. **Q** may be split into $\mathbf{Q}_1$ and $\mathbf{Q}_2$ of (c). Convergence depends, as in figure 2, on the value $\|\mathbf{Q}_1^{-1}\mathbf{Q}_2\|$. The norm of $\mathbf{Q}_2$ can be quite small compared to $\|\mathbf{Q}_1\|$ for general cases of this type.

between the largest and smallest values of **W**, as well as an even distribution of values of **W** such that there are substantial regions of high weight contrasting with regions of low weight in model space.

Imperfect Transformations

So far the transformation to model space **F** has been considered to be a Fourier transform, and thus unitary. If the transformation is not unitary, or even has zero eigenvalues in it, and if the pseudo-inverse of **F** is being used as an approximation to $\mathbf{F}'$ in the application of the operator **Q**, the result may be quite unlike that of using the transpose $\mathbf{F}'$. If $\mathbf{F}'\mathbf{W}\mathbf{F}$ happens to have a null space (some zero eigenvalues in **F**) then any component of $\mathbf{x}_A$ lying in the null space has no contribution to the weighted least squares functional, and may have a tendency due to roundoff to grow and dominate the estimate of $\mathbf{x}_A$.

Take **F**, for example, to be a constant-velocity migration. It has a propagating region with eigenvalues of magnitude 1, and an evanescent region with eigenvalues tapering off to zero. Any evanescent energy that creeps into the update of the extrapolated traces $\mathbf{x}_A$ will remain and contaminate the estimate, because the evanescent component is unconstrained in the minimization.

One solution is to modify the transform **F** so that it is unitary for a particular subspace, the propagating region, and null for the rest of the domain space. To assure the absence of evanescent energy in $\mathbf{x}_A$, it may be filtered out every few iterations. With $\mathbf{x}_A$ thus restricted to lie in a subspace where **F** is unitary, the algorithms may be used as they are.

Estimating the Weights W

Getting a reasonable estimate on a weight function **W** is quite crucial to how good the data extension looks. In case the transformation **F** is designated to be a Fourier transform, estimating a weight (which is diagonal and non-negative) is equivalent to estimating a desired power spectrum from the original data. In this instance we find ourselves involved with the general problem of spectral estimation. For the examples that follow, **F** is taken to be a two-dimensional DFT, and the desired spectrum is estimated in the following manner. First, the original data, padded with the traces to be extrapolated, is transformed to (k_x,ω). The modulus of the transform is first summed over constant dips passing through the origin in (k_x,ω), and second, summed over constant frequencies ω. Two functions are obtained: an average frequency spectrum $\mathbf{G}_1(\omega)$ and an average dip spectrum $\mathbf{G}_2(p)$ where $p = k_x/\omega$. In this way any aliased dips (that do not pass through the origin) are discriminated against.

The desired spectrum $\mathbf{G}_d$ is set to the product of $\mathbf{G}_1$ and $\mathbf{G}_2$:

$$\mathbf{G}_d(\omega,k_x) = \mathbf{G}_1(\omega)\mathbf{G}_2(k_x/\omega).$$

The weighting **W** is the inverse of $\mathbf{G}_d$, normalized and smoothed, and having a specified dynamic range w_{max}/w_{min} where $w_{max} = 1$ and w_{min} is specified.

Experimentally it has been observed that setting **W** to an inverse power of $\mathbf{G}_d$,

$$\mathbf{W}(\omega,k_x) = \mathbf{G}_d{}^{-\alpha}(\omega,k_x)$$

where $\alpha > 1$, gives "better" extrapolations in the sense that the extrapolated traces are at about the same energy as the data traces. Otherwise, as seen in the examples following, the interpolated traces may uniformly have less energy than the original traces. Using a power of α not only increases the dynamic range of **W** but also changes the quantile distributions of the values of **W**. It is not currently clear how to pick an optimal value for α.

Example I: Three Beds

Missing data estimation will now be tried on a relatively simple synthetic model. Figure 4(a) depicts the model -- plane events at three distinct dips, interspersed with and padded with zero traces on either side. The projection operator **S** is zero on the nonzero traces of figure 4(a) and unity elsewhere. Figure 4(b) is a plot of the modulus of the model in (ω,k_x) space. The spectrum is replicated at the nyquist k_x because every other trace in the (x,t) domain is identically zero. A steepest descent procedure was used in the solution of equation (3). One important modification to the procedure was made: the weighting function **W** is re-estimated at each iteration. This changes (3) into essentially a non-linear procedure, combining the steps of estimation and extrapolation in a bootstrap manner. With this modification and for this particular synthetic model, convergence is practically attained in only a few iterations. Figures 5 and 6 are the results of performing two steepest-descent iterations on the "three-beds" model of figure 4. Figure 5 depicts the initial estimated weight **W** in (a) and the final estimated **W** in (b). The weights were estimated by averaging along lines of constant dip and constant ω in the (k_x,ω) domain. Note the final estimated weight shows no tendency to pass any energy in the aliased region, nor the energy due to abrupt truncation of the model. Figure 6 shows the extrapolated data after two iterations, (a) displaying the (x,t) domain and (b) displaying the modulus in the (k_x,ω) domain. For this model the extrapolation has quickly converged to a desirable answer. There is little doubt that this is due to the simplicity of the model: the the aliased energy is clearly separated from the desired spectrum. The next example does not share this attribute.

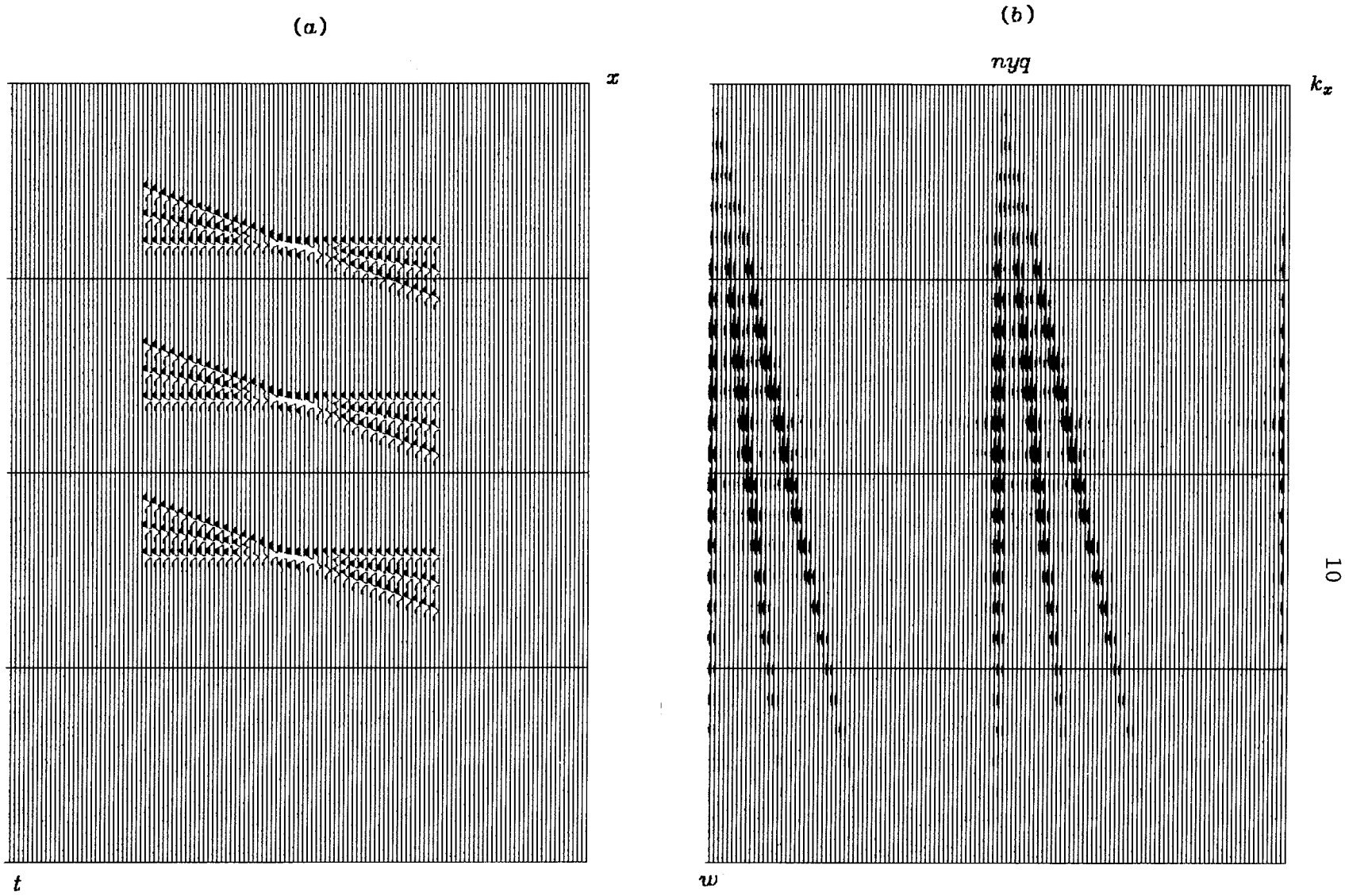

FIG.4. (a) The three-beds synthetic model in (x,t) space. Every other trace is zero and a padded region of 32 traces has been added to either side. All nonzero traces are considered to be the inviolate data traces. Size of panel: 128 x points by 256 t points. (b) The three-beds model in (k_x,ω) space. Modulus is plotted. The vertical axis extends from ω = 0 to ω_{nyq}. The origin k_x = 0 is in the upper left, right corners of the plot. The replication of the three prominent dips is due to the insertion of blank traces at every other trace, and the faint k_x-independent energy is due to the truncation of the beds.

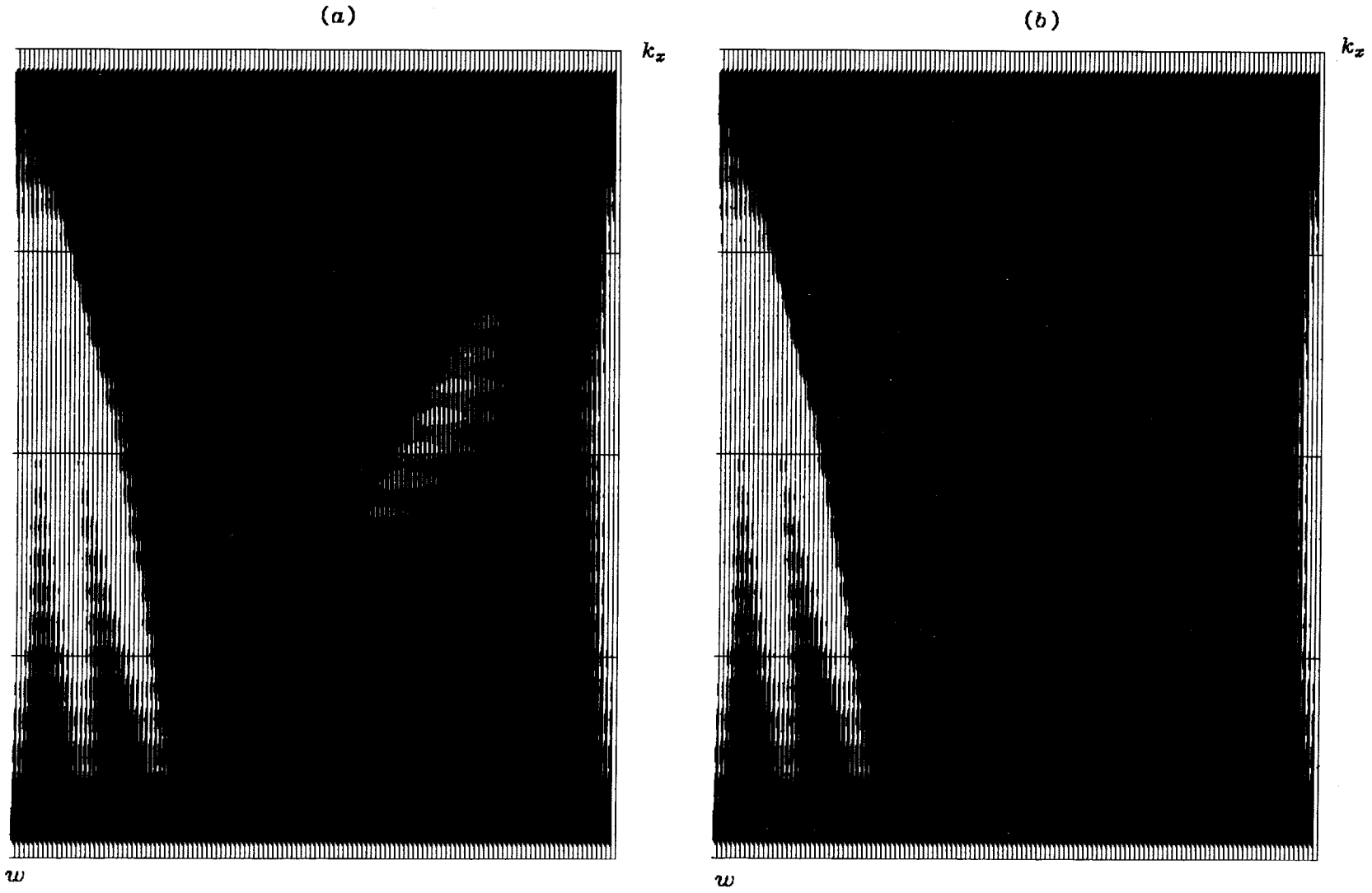

FIG.5. (a) The Initial weighting function W estimated from figure 4(b). Black indicates a weight of 1, while the lightest area (the pass region) corresponds to a weight of 0.05, for a dynamic range of 20 to 1. (b) The final (second) estimated weighting function W. Dynamic range is still 20 to 1. This function exhibits considerable "parsimony".

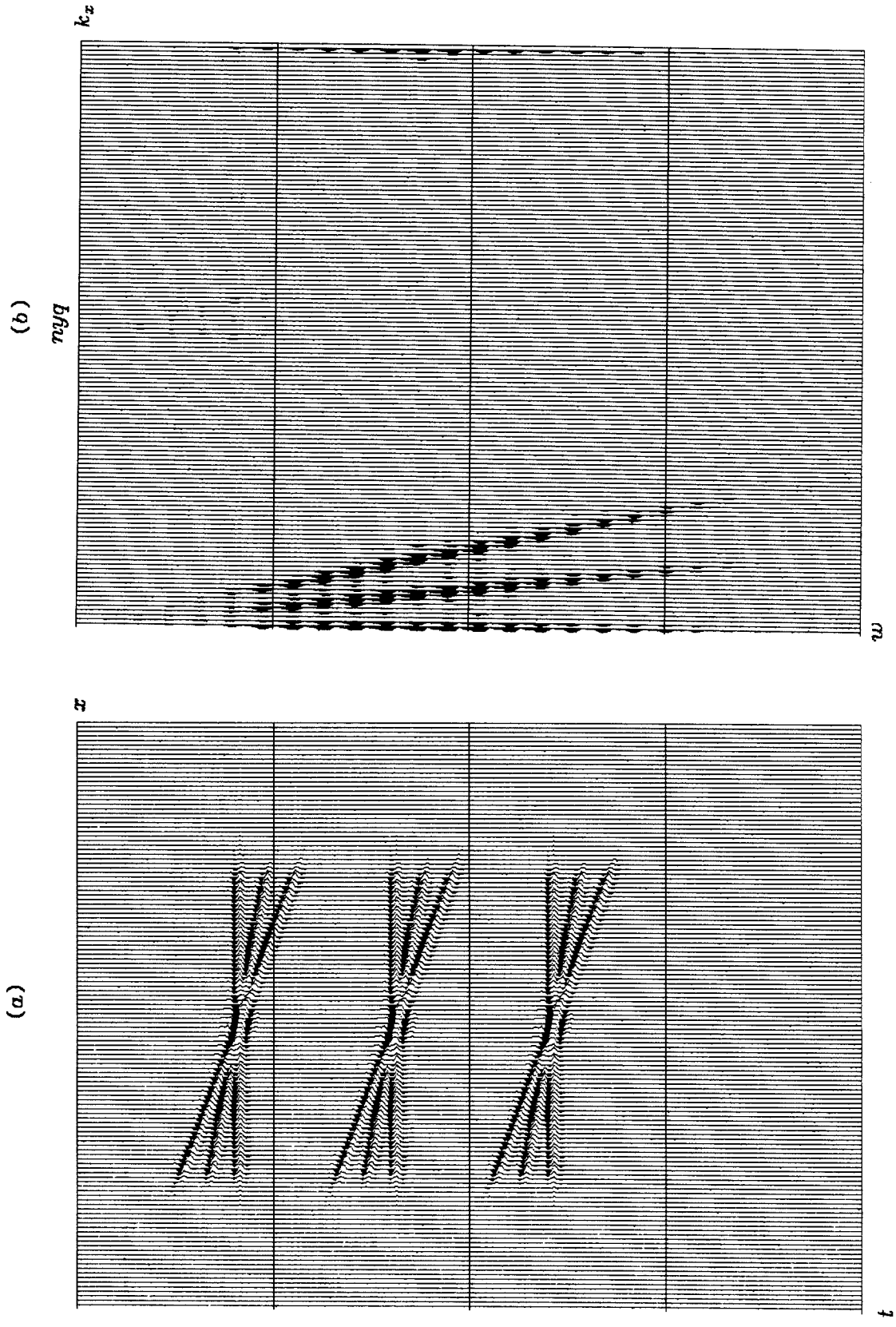

FIG.6. (a) The extended three-beds model in the (x,t) plane. (b) The extended model in the (k_x,ω) domain.

Example 2: Vertical Seismic Profile

An extrapolation procedure identical to that described for the three-beds model was next tried on real data. Figure 7 illustrates a portion of a vertical seismic profile, with depth down the borehole on the horizontal axis and time on the vertical axis. The original data, now padded with three zero traces for each data trace, is to be interpolated. Events at four predominant dips are present on the profile: downgoing and upcoming primary events at a shallow apparent dip, and downgoing and upcoming tube waves at a steeper apparent dip. Figure 8 is a plot of the modulus of the Fourier transform of figure 7, and the four dips can be seen: the downgoing primary and tube wave are in well-defined bands of dip, while reflected primaries and a tube wave of opposite dip are much weaker though still discernible. In figure 8 there is obviously drastic aliasing of the tube wave: aliased tube wave energy is overlapping primary energy at frequencies near 70 Hertz. There are other obvious locations in the (k_x,ω) plane where aliased and unaliased energy overlap each other. The objective of the 3-to-1 interpolation is to suppress energy that does not match our parsimony model: energy that does not line up along lines through the (k_x,ω) origin (the upper right and left corners of figure 8). Figure 9 is the result after two descent steps, and it can be seen that enhancement of the parsimonious model has taken place. Figure 10 is the interpolated profile in the (x,t) domain, and some shortcomings of the procedure are obvious:

First, energy in the interpolated traces is uniformly lower than in the data traces. This is due to the problem with limiting the weights **W** to a certain dynamic range, as described previously. In this case, the ratio of the largest to smallest value of **W** is 20 to 1. A trace-equalized version of figure 10 is shown in figure 11, and the interpolation of the primaries appears somewhat smoother.

Second, the procedure has the limitation of being global: the interpolations are performed at only a few predominant dips. A short portion of the primary event in the upper right corner of figure 8 has a slightly steeper dip than primaries of the rest of the profile, yet this dip is completely missed in the interpolation. This problem could be easily fixed by making the operation more local: interpolate smaller pieces of the profile separately.

The third observation is that aliased energy that happens to intersect a pass-band on figure 9 has a tendency to stick there. This can be seen at various locations on figure 10, where interpolation at dips corresponding to the tube wave are present in a region where only primaries obviously should be, and vice-versa. Making the interpolation more local may cure this problem to some degree, but it is fundamental. The desired answer is declared to be parsimonious in the (k_x,ω) domain, and only a few select dips are going to be used to interpolate the entire section.

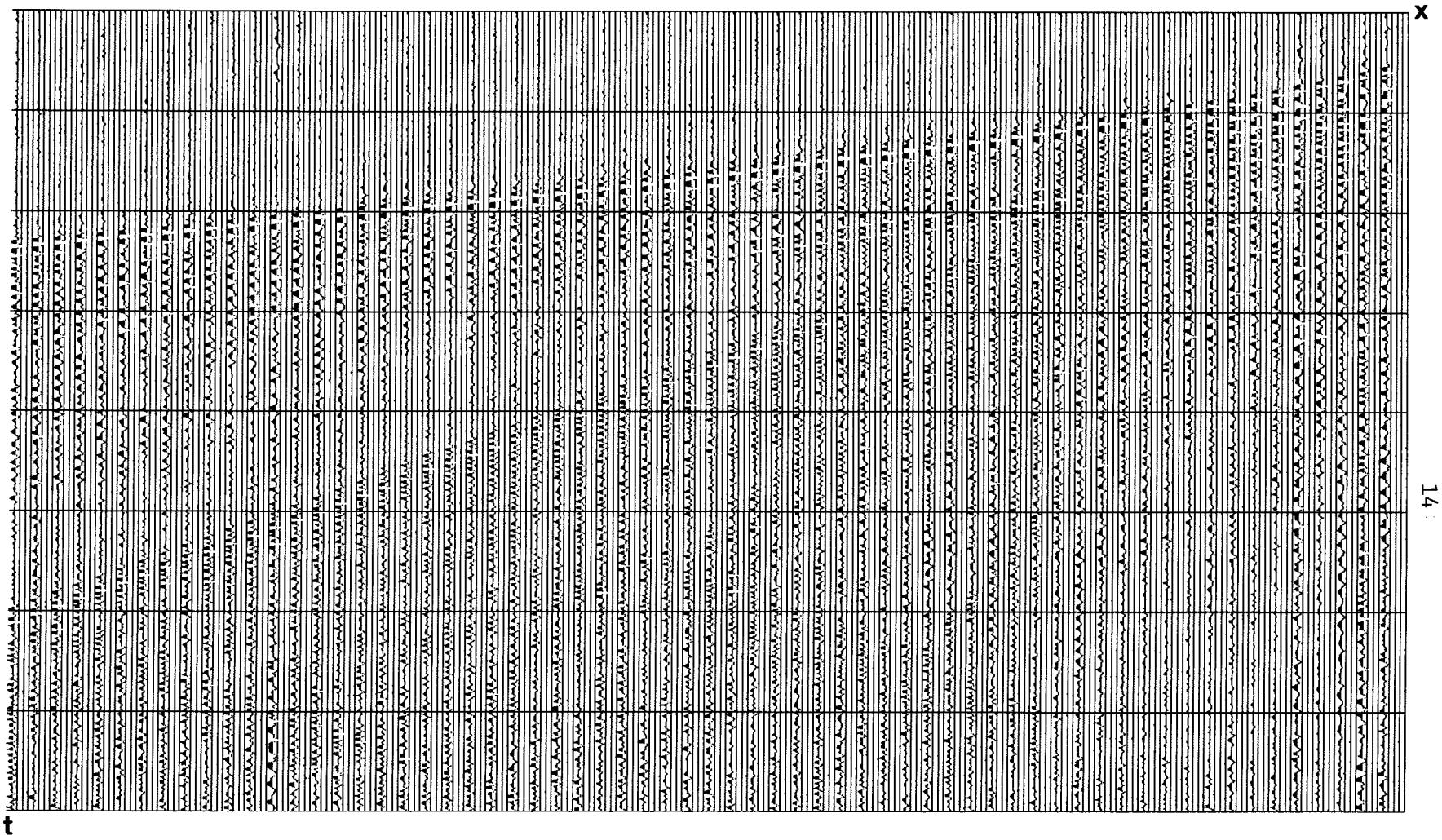

FIG.7. A vertical seismic profile with three zero traces inserted between each data trace. These traces are to be interpolated. The main events on this profile are (1) the downgoing primary (the strong first event in shallow time), (2) the aliased tube wave extending from upper right to lower left corner of the profile, (3) weak upcoming primaries at the opposite dip of the downgoing primaries, and (4) a weak upcoming tube wave in the center of the profile.

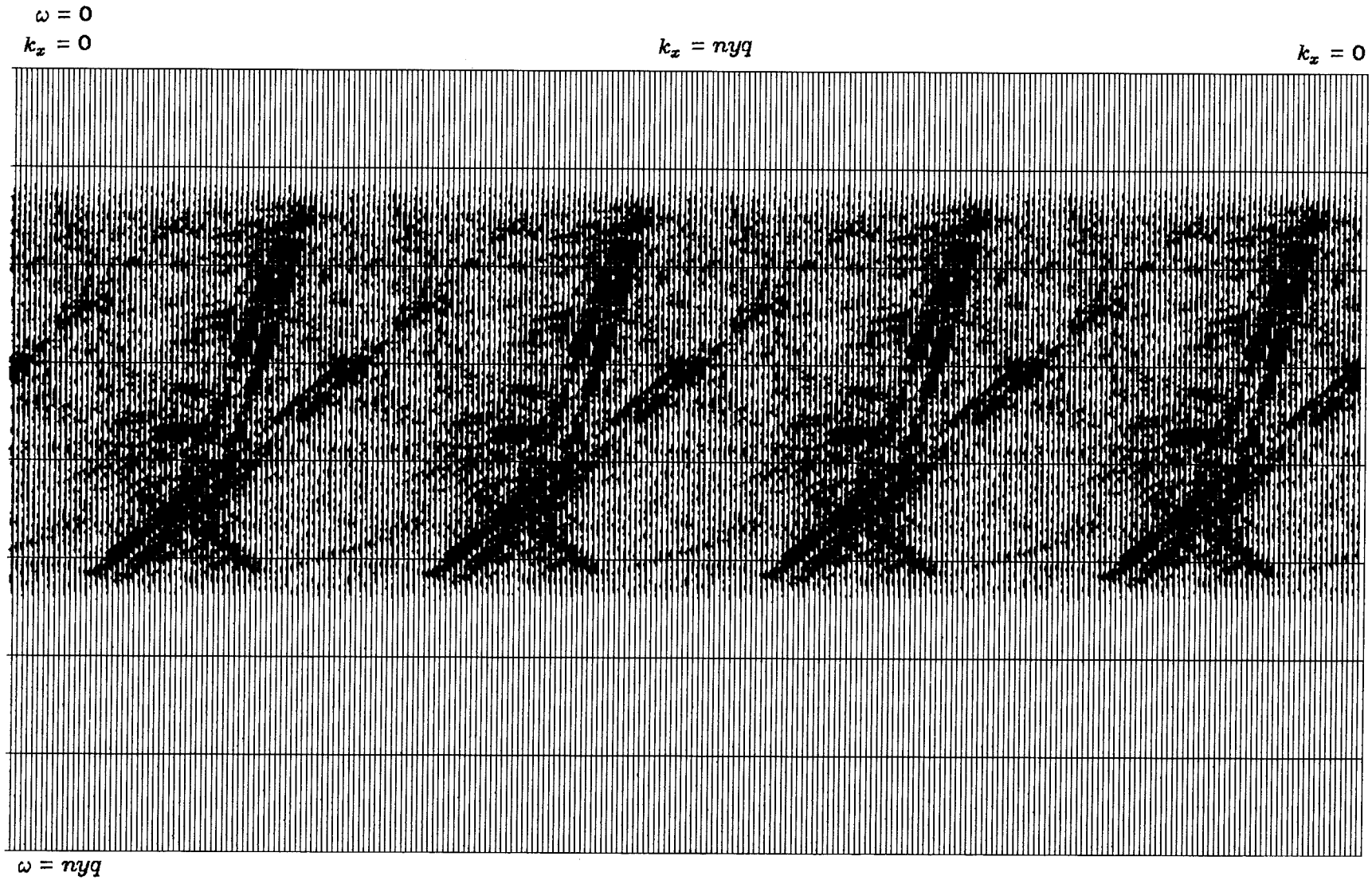

FIG.8. The profile in (k_x,ω) space. ω ranges from 0 to nyquist while k_x ranges from 0 to +nyquist at center, then -nyquist to 0. Modulus is displayed. The spectrum of the original data is replicated four times due to the inserted traces. In this way it can be seen that virtually all the energy in the tube wave is aliased, and overlaps a portion of the strong direct wave. The direct and tube waves exhibit splitting because of a slight change in dip of the events at the center trace of figure 7, possibly due to a well-casing change at that point. Weak reflected events of opposite sign may be seen. The pass band in ω is 20 Hz to 80 Hz.

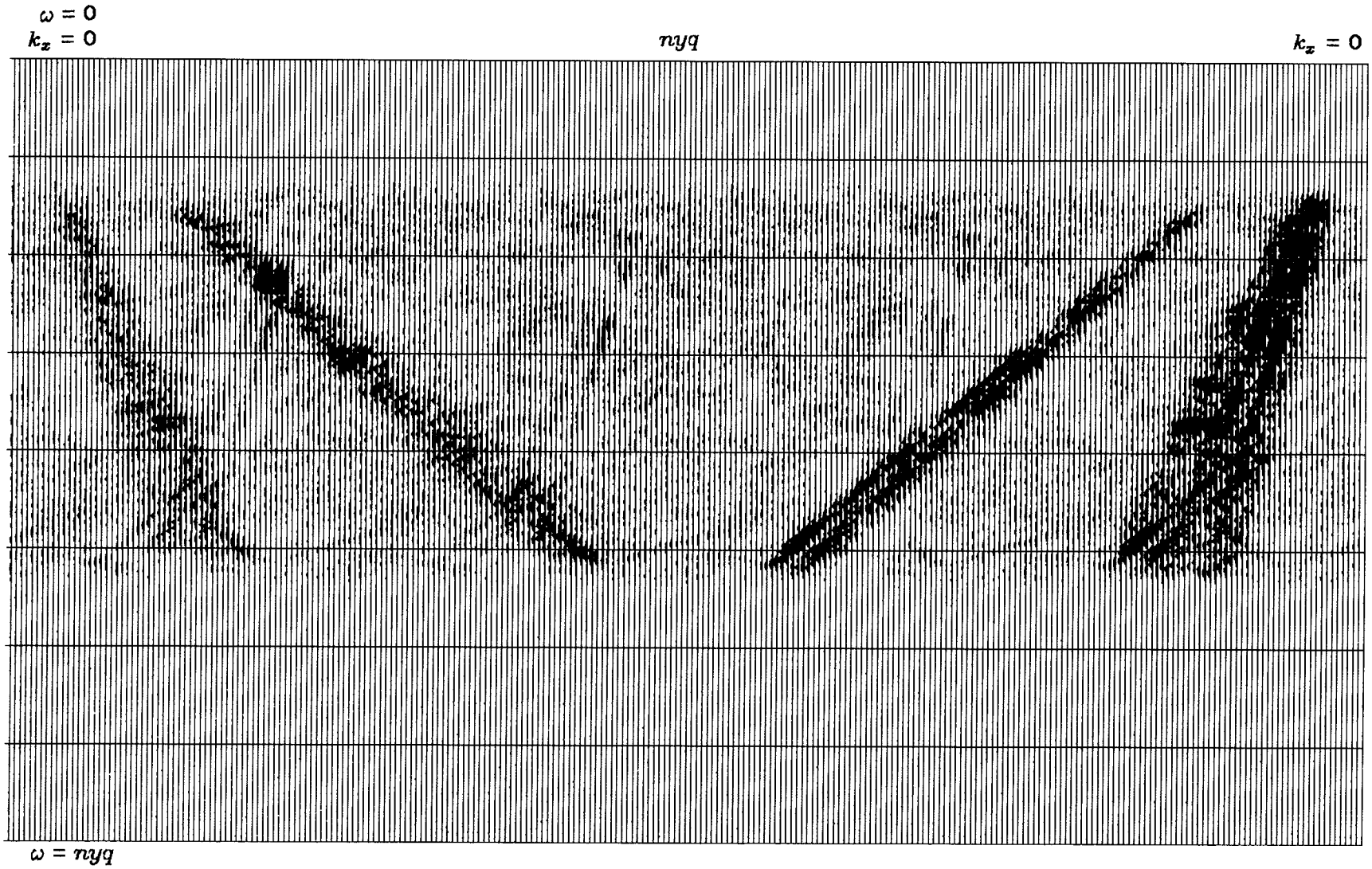

FIG.9. The extended data in (k_x,ω) space after two iterations. Modulus is displayed. Energy in dips not passing through the (k_x,ω) have been discriminated against. Some energy that is obviously aliased and that crosses dip pass-bands remains: the image of the aliased tube wave in the band of the downgoing direct wave. This energy is manifested in figures 10 and 11 as an interpolation of the tube wave with the dip of the direct wave.

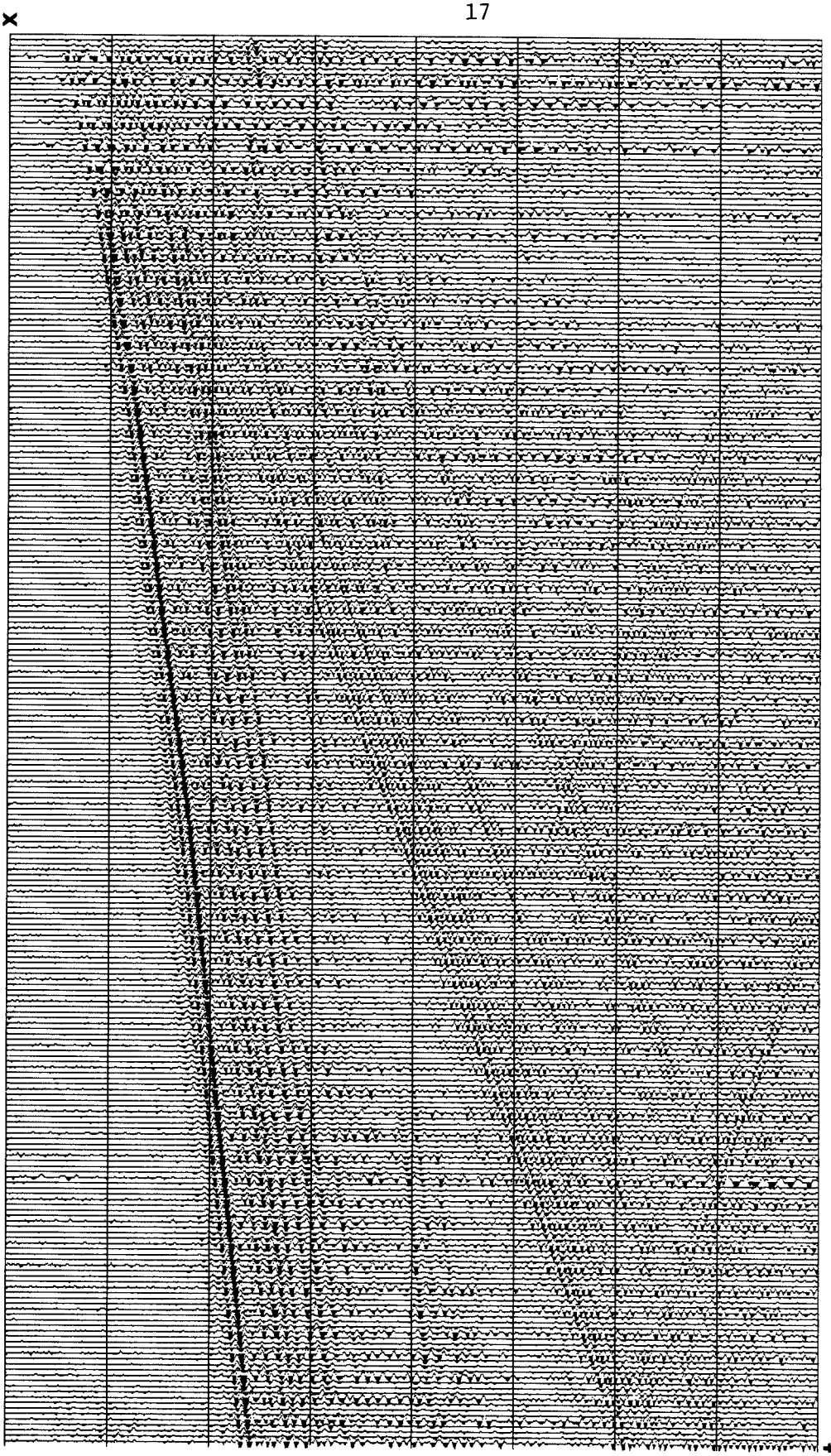

FIG.10. The extended data after two iterations in the (x,t) plane.

FIG.11. The extended data in (x,t) after a trace equalization has been applied. Reflected primaries have not been successfully interpolated and many parts of the profile have been interpolated at the dip of the reflected tube wave.

Example 3: Burg Extrapolation

As an example of a more local type of extrapolation, a completely different procedure than what has up to now been described is now given. It has been first used here by Clayton in extrapolating common-midpoint gathers to zero offset, and is described by Morley and Muir (SEP, this report). The procedure is an elementary application of a prediction-error filter to extrapolate laterally. Consider the three-beds model of figure 12(a). The model may be extrapolated out into the padded traces by performing the following operations:

1) transform to (ω,x) domain;

2) for each ω, calculate a prediction-error filter in the x-direction, by the Burg algorithm;

3) apply the negative of the prediction-error filter to extrapolate in x.

The result of this operation on the model of figure 12(a) is shown in 12(b). Obviously the method works perfectly for models such as that of 12(a), which have sharp line spectra. This extrapolation technique is closely related to the original problem stated early in this paper: the (k_x,ω) domain is still the parsimonious domain. The estimation stage is the calculation of the prediction filter, and the extrapolation stage is simply it's application to the data.

Another extrapolation example is shown in figures 12(c) and 12(d). Here, the data (c) exhibits some hyperbolic moveout. The resulting extrapolation is shown in (d); there is a tendency to extrapolate at the average dip seen by the prediction filter. This is not surprising, for data with curved events is nonstationary in the x direction. The iterative procedure of the previous examples would perform similarly, for it assumes the same parsimonious domain, the frequency domain. A section with hyperbolic moveout is no longer "parsimonious" in this domain, so that if a better extrapolation is desired, some other space must be sought where local-ness of information can be assumed.

ACKNOWLEDGMENTS

SEP would like to thank ARCO for generously supplying the Vertical Seismic data used in this report.

REFERENCES

Claerbout, J., Restoration of Missing Data by Least Squares Optimization: SEP-25, p.1.
Luenberger, D.G., Introduction to Linear and Nonlinear Programming: Addison-Wesley, 1973.
Morley, L. and F. Muir, Space-Time 1/2-Plane Prediction Filtering: SEP-26 (this volume).
Strang, G., Linear Algebra and Its Applications: Academic Press, 1976.

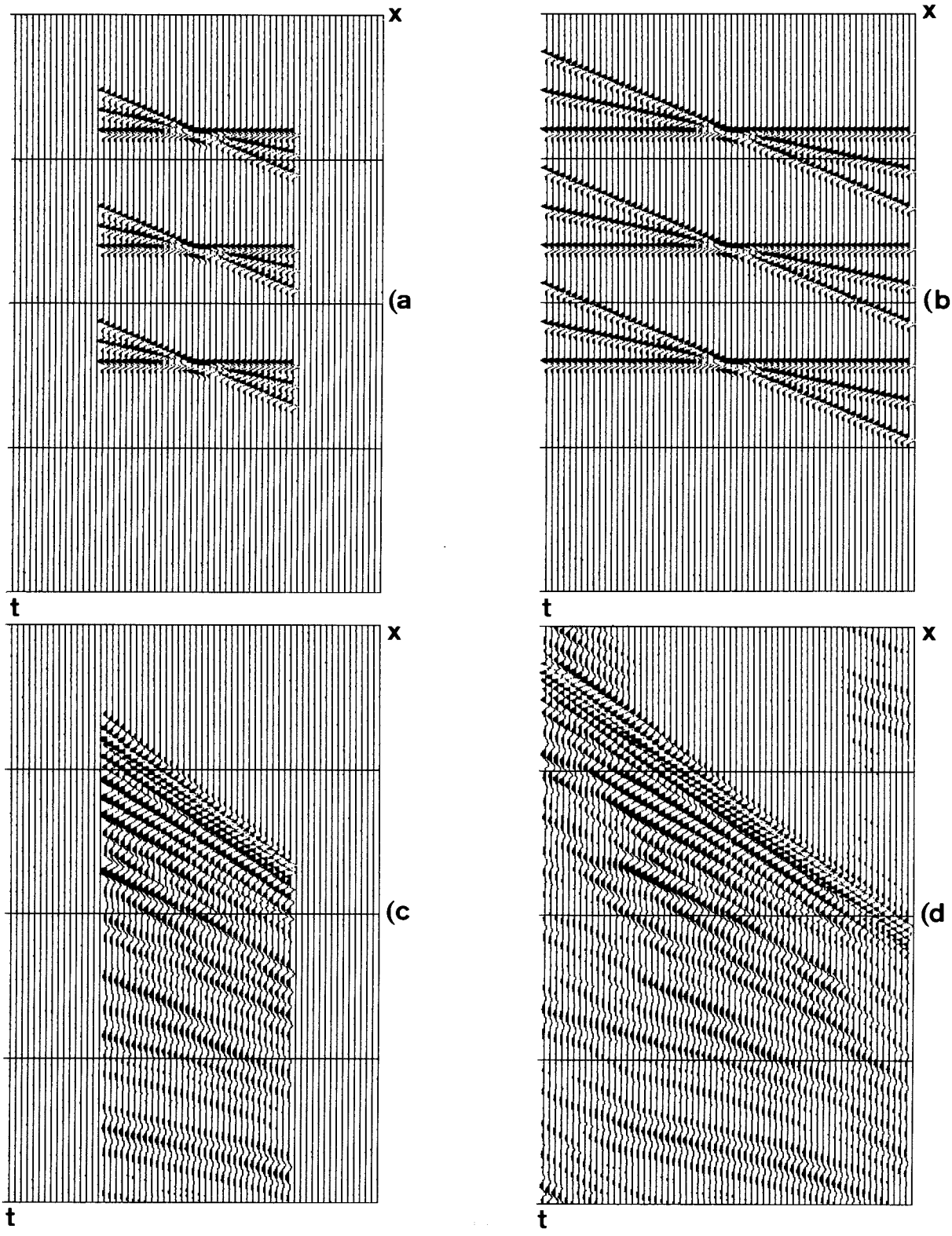

FIG.12. (a) Three-beds model with 16 traces padded on each side. (b) Extended three-beds model, by calculating a prediction filter for each ω and applying it sequentially to the padded traces. Length of filter is 10 samples. (c) A real data set with some hyperbolic moveout. 16 traces padded. (d) Extended data set by prediction filtering. Note the sigmoid shape of the extrapolation for the events with the greatest change in dip across the patch of data: events extrapolate at their average dip.

Solutions to Chess Problems, SEP-24, p. 250.

Kubbel Problem

1 N-K3ch K-N6

If 1...K-R7, 2 Q-KB2ch K-R6
3 Q-N2ch K-R5 4 Q-N4 mate.

2 Q-N4ch K-B7
3 Q-B4ch K-K7 (or K8)

If 3...K-N8, 4 Q-N3ch.

4 Q-B1ch K-Q7

If 4...KxN, 5 Q-K1ch and
Black's queen is lost.

5 Q-Q1ch K-B6
6 Q-B2ch K-N5

If 6...K-Q5, 7 N-B5ch again
wins the queen.

7 Q-N2ch N-N6

If 7...K-R4, 8 N-B4ch K-R3
9 Q-N6 mate.

8 Q-R3ch!! KxQ

If the Black king moves away,
the queen is lost.

9 N-B2 mate!

Gottlieb Problem

1 N-B7ch K-Q5
2 R-B4ch PxR
3 P-B3ch PxP
4 PxRch PxP
5 R-K4ch PxR
6 B-K5ch PxB
7 N/Q6-N5ch RxN

If 7...PxN, 8 Q-Q8ch B-Q4
and 9 QxB mate.

8 Q-Q8ch B-Q4
9 QxBch PxQ
10 NxR mate!