Chapter I

Invertibility of Velocity Stacks

1.1 Preface

Performing normal moveout and stacking is a standard way of transforming data into what can be called velocity space. Under what conditions is this operation invertible? In other words, to what extent can an inverse operator be found, that will transform data from velocity space back into the original space of the common midpoint gather? This question will be addressed in detail in the following chapters, and there it will be seen to be closely related to the problems of velocity analysis, velocity filtering, and, probably most importantly, to the undersampling of seismic field data. The results of later chapters will be briefly described below with the help of a real-data example. By illustrating the central points to be made with field data, this chapter serves as both an outline and a summary to the remaining chapters of the thesis.

Terms like *common midpoint gather*, *stack*, and *normal moveout (NMO)* are common terms in the literature of reflection seismology. Definitions for most of these terms, except those that are explicitly defined in mathematical expressions, will not always be given. For a comprehensive glossary of reflection seismology terminology, see Sheriff (1973).

1.2 Velocity stacks

Velocity stacks are commonly used to estimate velocities during the course of processing reflection seismic data. A velocity stack is a collection of traces, each

trace being the result of applying normal moveout (NMO), consistent with a constant velocity for that trace, to the traces of a common midpoint gather and summing. Figure 1.1 is an example of a velocity stack of real data (as opposed to synthetic data). The stack was performed by discrete summation, but for our purposes a *velocity* stack will be defined as an integral over the finite range of offsets h_1 to h_2:

$$u(v,\tau) \equiv \int_{h_1}^{h_2} d(\, h,\, t=\sqrt{\tau^2 + h^2/v^2}\,)\, dh \qquad (1.1)$$

where $d(h,t)$ denotes the input to the velocity stack and $u(p,\tau)$ denotes the output. The velocity-time pair (v,τ) are the coordinate axes of the velocity stack, and the offset-time pair (h,t) are the coordinate axes of the *common midpoint* (CMP) gather, also referred to as a *common depth point* (CDP) gather. Offset is a spatial axis, with units of length. When (v,τ) and (h,t) are considered to be linear vector spaces (they can easily be defined so), equation (1.1) defines a linear transformation $\mathbf{L}^T$ from the (h,t) domain into the (v,τ) domain. The (h,t) domain will subsequently be referred to as the *offset space* or *data space*, and the (v,τ) domain as the *velocity space*. A typical seismic survey is comprised of 48 or 96 offset traces, and approximately half that number of discrete velocities might be used to make a velocity stack. Because a typical trace consists of 1500 sampled time points, the dimensions of the discretized offset and velocity spaces can reach beyond 10^5. The fact that the operator (1.1) can be applied at all in practice, is because it is a very large, very sparse linear system.

For our purposes it is more appropriate to parameterize velocity stacks by slowness p, the inverse of velocity, instead of the velocity v. To standardize nomenclature, the transformed panel $u(p,\tau)$ from now on will be called a *velocity stack*, and in this chapter the illustrations show velocity scales at the top, though the traces in the illustrations are actually evenly sampled in slowness p. A velocity

stack is now defined to be

$$u(p,\tau) = \int_{h_1}^{h_2} d(\, h,\, t=\sqrt{\tau^2 + p^2 h^2}\,)\, dh \qquad (1.2)$$

1.3 Velocity space

We illustrate operation (1.2) with figure 1.1, a velocity stack of a CMP gather from the offshore Texas Gulf Coast. Panel D, the common midpoint gather, displays the attributes of a typical gather. Pre-critical reflections from sedimentary layers demonstrate *hyperbolic moveout*: the arrival times fit a hyperbola parameterized by, first, the average velocity of waves through the sediments and, second, the zero-offset travel time.

A velocity stack can be used to determine velocities, because, with the correct velocity, the values of a reflector will add constructively when the sum is taken over the correct hyperbolic path. Panel LTD in figure 1.1 demonstrates the typical appearance of a velocity stack. The deeper the reflector is on the section, the more diffuse its response is; this degradation in velocity from shallow to deep on the section is a consequence of the finite recording aperture laid out on the earth's surface: thus velocity resolution degrades with depth. The velocity curve implied by figure 1.1(LTD) does not necessarily have to remain a single-valued function of time, and the only characteristic that allows us to distinguish between two reflectors with the same zero-offset travel time is their relative moveout. For use in velocity analysis, the transformation into velocity space should have the desirable property of focusing hyperbolic events into points; this property obviously would aid in the resolution of two events with different velocities. Velocity stacks are able to resolve different events well in velocity space, but they do have the undesired side effect of adding horizontal tails (*sidelobes*) to each event in velocity

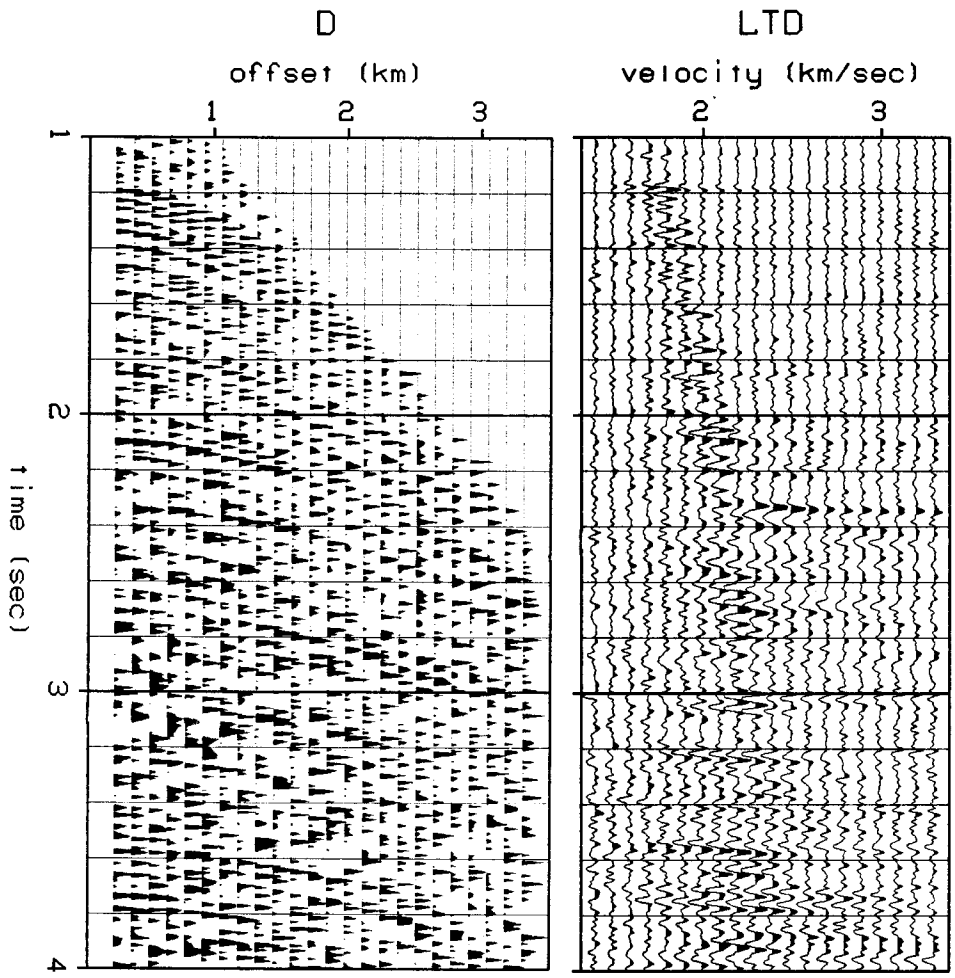

FIG. 1.1. Velocity stacking. Panel **D** is a portion of a common midpoint gather from the Offshore Texas Gulf area (courtesy of Western Geophysical Co.). Refractions have been removed with a triangular mute. For clarity, only positive polarities have been plotted. Panel **LTD** is a velocity stack of **D**. It was made using equation (1.2) with the integration over offset replaced by a uniformly weighted summation. Though the stacking velocity curve may be picked off the panel, it lacks the resolution in velocity that a semblance velocity analysis would give.

space. In comparison, a standard semblance velocity analysis (figure 1.5) seems to be able to suppress the energy in the tails. The main reason for the strength of the tails on the velocity stack is that the moveout curve for any event, regardless of its velocity, has a slope that converges to zero at zero offset. When summing over a hyperbolic moveout path, we cannot avoid contributions from events at other velocities as zero offset is approached, as long as we continue to weight each

trace of the common midpoint gather equally in the summation.

1.4 Designing a velocity transform

It is desirable to have an alternate choice for the velocity stacking operator $\mathbf{L}^T$ that does not suffer the shortcomings of $\mathbf{L}^T$ mentioned above. One obvious approach is to selectively weight the traces of the common midpoint gather before summing: one choice is to set the weighting to be proportional to offset h. In this way, the far-offset traces are allowed to contribute more to the stack than the near-offset traces, where there is no moveout discrimination between events with different velocities. If we denote this offset-weighting in operator notation by **S**, an alternate definition for a velocity stack is the product of the two operators $\mathbf{L}^T$ and **S**:

$\mathbf{L}^T\mathbf{S}$:
$$u(p,\tau) = \int_{h_1}^{h_2} h\; d(\, h,\, t=\sqrt{\tau^2 + p^2h^2}\,)\; dh \qquad (1.3)$$

Figure 1.2 compares the two velocity stacks, $\mathbf{L}^T$ and $\mathbf{L}^T\mathbf{S}$, applied to a window of the data of figure 1.1. The main advantage that $\mathbf{L}^T\mathbf{S}$ has over $\mathbf{L}^T$ is that it reduces the horizontal "smear" of the events on the stack. The remaining streaks, or sidelobes, on the panel are artifacts resulting from the truncation of events at the far offset of the common-midpoint gather. One of our main goals will be to alleviate this offset truncation problem on the velocity stack. The velocity stack of equation (1.3) may be generalized to allow an arbitrary weighting function of offset $f(h)$ inside the integral, whose purpose is to reduce the truncation effects at far offsets (Larner, 1979).

There is however a better approach to designing a replacement for the velocity stacking operator $\mathbf{L}^T$, via least squares. Assume that a common-midpoint gather $d(h,t)$ is the result of some transformation on a function $u(p,\tau)$ in velocity space. Perhaps $d(h,t)$ is also corrupted with additive noise:

$$\mathbf{d} = \mathbf{L}\mathbf{u} + \mathbf{n} \tag{1.4}$$

A straightforward definition for the operator L is the adjoint of the operator L^T defined by equation (1.2). Through the suitable definition of an inner product space, L turns out to be simply the process of reverse NMO and stacking:

$$d(h,t) = \int_{p_1}^{p_2} u(p, \tau=\sqrt{t^2 - p^2h^2}\,)\, dp \tag{1.5}$$

The adjoint relationship between operators L^T and L will be derived in chapter 2. L is identical in operation to L^T except that normal moveout "stretch" is replaced by an inverse-moveout compression of each trace.

Equation (1.5) makes sense as a modeler of events on a common-midpoint gather: an impulse in velocity space gives rise to a hyperbolic event with a uniform amplitude across all offsets on the gather (data space). To what extent are the operations L and L^T inverses? Applying the least squares method to (1.4), in order to minimize the energy in the noise term of (1.4), yields

$$\mathbf{u} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{d} \tag{1.6}$$

To zeroth order, the diagonally dominant L^TL approximates the identity operator; the amplitudes of the diagonal elements of L^TL may be orders of magnitude greater than those of the off-diagonal elements. But this zero-order assumption is inadequate, as illustrated by the wide spread in velocity exhibited by the events in figure 1.1(LTD). Even more basic is the question of whether $(L^TL)^{-1}$ exists, and how easily it may be found. Both the finite sampling interval and finite range in offset contribute to the ill-conditioning of L^TL and increase the likelihood of it being singular. Probably the greatest contribution to ill-conditioning is the nonuniqueness of the hyperbolic summation paths near zero offset.

It was seen above that the offset-weighted velocity stack L^TS had properties that favored it over the uniformly weighted velocity stack L^T. We may say that L^TS

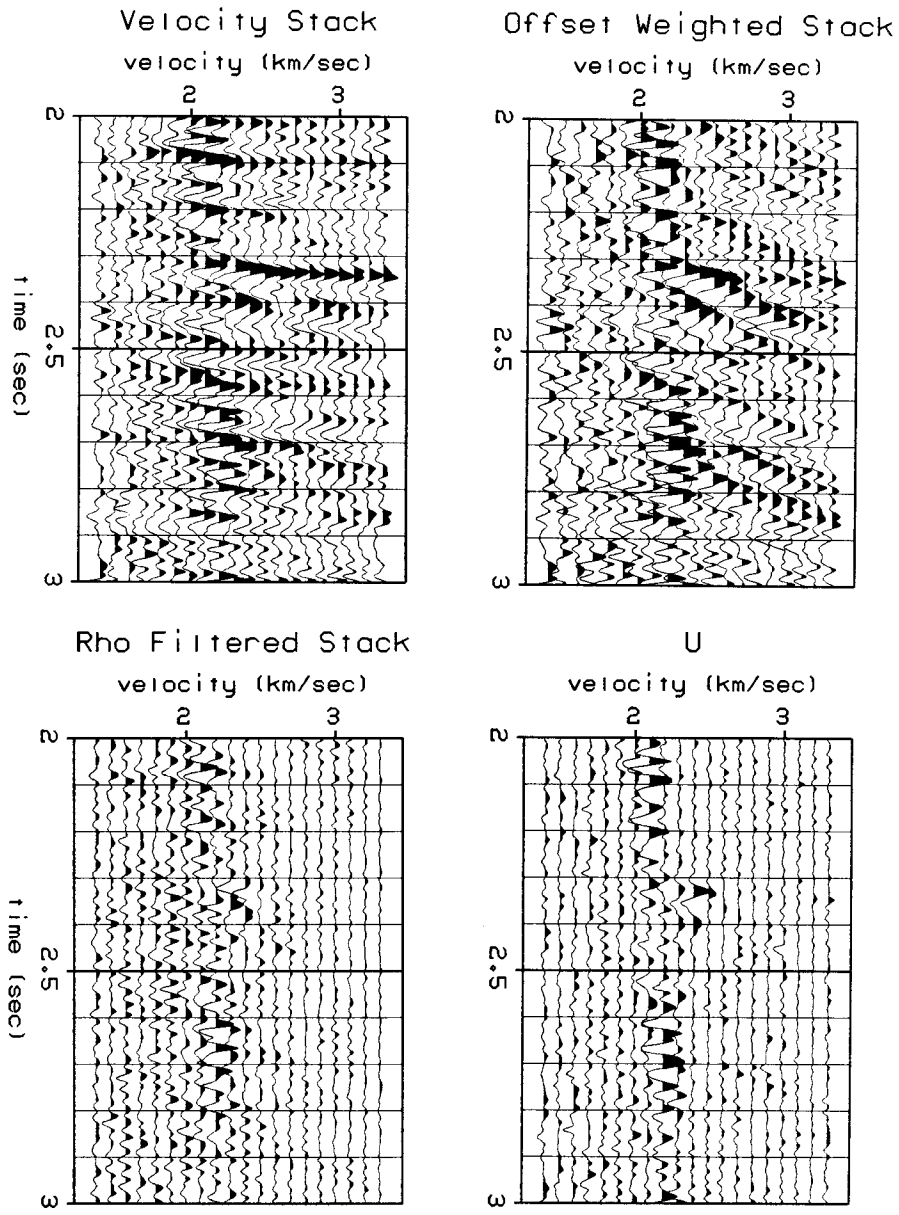

FIG. 1.2. Comparison of the pseudoinverse and stochastic inverse to the velocity stack. The upper left panel is a portion of the velocity stack of figure 1.1 from 2 to 3 seconds. In text notation, it is $\mathbf{L}^T\mathbf{d}$. The upper right panel is the offset-weighted velocity stack over the same data, $\mathbf{L}^T\mathbf{S}\mathbf{d}$. The lower left panel is an estimate of the pseudoinverse $(\mathbf{L}^T\mathbf{L})^+\mathbf{L}^T\mathbf{d}$ of the data. The "rho filter" (a term borrowed from algebraic reconstruction theory) refers to the filter $(\mathbf{L}^T\mathbf{L})^+$. The lower right panel is an approximate solution to the stochastic inverse equations (1.7). The resolution of velocities in velocity space gets progressively better from upper left to lower right.

is an example of a transformation that is somehow closer to the inverse of L than L^T is. But the approach which we shall take will deal with the estimation of $(L^TL)^{-1}$. The main reason for preferring L^T over L^TS as the definition of the velocity stack lies in the important assumption (1.4): that the gather is the result of L (the "reverse" velocity stack of equation 1.5) applied to some function u in velocity space. Once the form of L is fixed by equation (1.5), the definition of its adjoint L^T is likewise fixed, and from now on will be the preferred definition for the standard velocity stack.

We have two basic routes open to us to design a substitute for the ill-conditioned (or nonexistent) inverse $(L^TL)^{-1}$. The first approach is to use the pseudoinverse, or generalized inverse $(L^TL)^+$ (Lanczos, 1961), which by definition inverts the nonzero singular values of L^TL (chapter 2). In this approach, the operator L^TL may have a large null space, which is related to the offset limitations on the original gather (chapter 4); by use of the pseudoinverse the null space components are effectively constrained to be zero in the inversion. The pseudoinverse corresponding to L is $(L^TL)^+L^T$. Figure 1.2 illustrates the steps taken in applying the pseudoinverse filter to a window of the data from figure 1.1. First apply the velocity stack L^T of equation (1.2) to the data. Next, apply the filter $(L^TL)^+$ to give the "rho filtered" stack (the lower left panel of figure 1.2). This filter, which is derived in chapter 4, is relatively localized in time and space. It also happens to be nonstationary. Recall that our objective is to design an operator that is able to transform a common-midpoint gather, like that of figure 1.1(D), into a sharply focused velocity panel. The rho filter panel of figure 1.2 shows that the pseudoinverse $(L^TL)^+$ does this fairly well. False events at high velocities have been eliminated, but there is still some problem with the discrimination between real events and artifacts at low velocities.

The second approach to designing an approximate inverse to L, is by solving

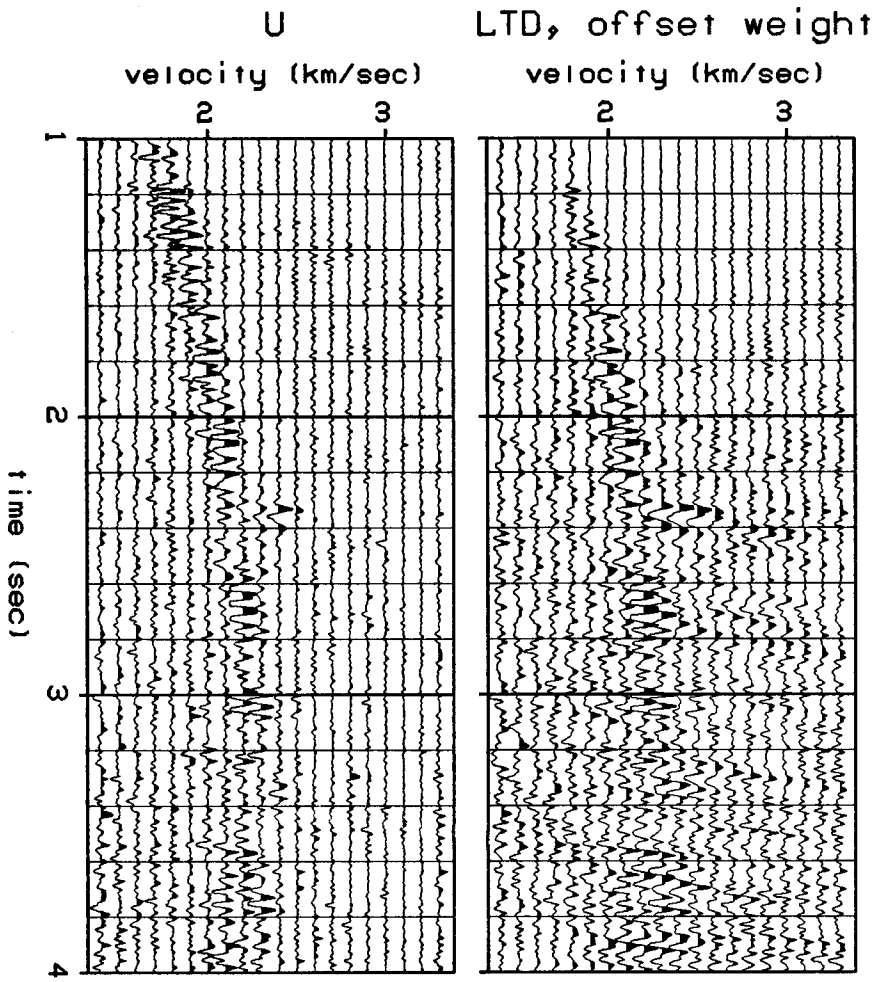

FIG. 1.3. Comparison between the stochastic inverse **u** and the velocity stack $\mathbf{L}^T\mathbf{Sd}$ over a time window of 1 to 4 seconds.

equation (1.6) directly, after perturbing $\mathbf{L}^T\mathbf{L}$ toward an operator that is guaranteed to have a stable inverse. *A priori* knowledge of the variance of the solution may be incorporated by adding the term $\mathbf{u}^T\mathbf{Du}$ to the least squares functional to be minimized (chapter 5). The term **D** is defined as the ratio of noise variance to variance of the model in velocity space. Adding this diagonal term to the least squares linear system $\mathbf{L}^T\mathbf{L}$ converts it into what Aki and Richards (1980) call the *stochastic inverse*:

$$\mathbf{u} = (\mathbf{L}^T\mathbf{L} + \mathbf{D})^{-1}\mathbf{L}^T\mathbf{d} \qquad (1.7)$$

When the variance of the solution is not precisely known, the **D** term may be

"bootstrapped" or iteratively refined from estimates of **u**. A mathematical justification for this bootstrapping is derived in chapter 5 by relating the variance estimate to a nonlinear *parsimony* (sparseness) measure. In this case, when **D** depends explicitly on **u**, equation (1.7) is converted into a nonlinear system of equations. The lower right panel of figure 1.2 shows the results obtained when the stochastic inverse in (1.7) is applied to the Gulf Coast CMP gather. Because the system of equations ($\mathbf{L}^T\mathbf{L} + \mathbf{D}$) is nonlinear and possesses a vast dimensionality, it is impossible to attain exact convergence with an iterative algorithm, much less solve equations (1.7) directly. Yet in practice, substantial convergence can be made within a few iterations. For example, only five iterations were necessary to attain the solution **u** in figure 1.2. In the lower right panel of figure 1.2 the resolution of individual velocity events is significantly enhanced over that of the two alternatives, standard velocity stacking and pseudoinversion. Another comparison of the stochastic result **u** to the velocity stack $\mathbf{L}^T\mathbf{Sd}$, over a greater range in time, is shown in figure 1.3.

It should be noted that the definition of stochastic inverse used here is a generalization of the definition given in Aki and Richards (1980, sec. 12.3.5). There, the diagonal term **D** is constrained to be constant.

The least squares equations (1.7) can now be identified as a velocity transformation from data space to velocity space. Once a parsimonious solution **u** has been found, the process of "inversion" consists simply of an application of the operator **L** (figure 1.4) to **u**. Coherent events remaining on the residual $\mathbf{d} - \mathbf{Lu}$ of figure 1.4 are for the most part nonhyperbolic. The low-frequency coherent events with linear moveout happen to be normal modes confined to the shallow water layer.

Figure 1.5 compares the envelope of the stochastic inverse **u** with a standard semblance velocity analysis from the same gather, and the envelope of the velocity stack $\mathbf{L}^T\mathbf{Sd}$. The velocity resolution is virtually the same on the first two plots,

though the semblance plot was created by scanning twice the number of velocities. The envelope of the standard velocity stack also shows good resolution of the strong events, but suffers from having a higher level of background noise, due, as always, to truncation artifacts. The main objective of the stochastic inversion procedure may be summarized as an attempt to drive down the level of the artifacts on a standard velocity stack, in order that weak events, which may have been obscured by the artifacts, can be seen. Notice a high-velocity event at 2.4 seconds on figure 1.5 which is distinct from the general trend of the velocity function. It is a reflection from a steeply dipping fault (Hale, 1983); that its apparent velocity is higher than that of the surrounding events, is due to dip effects.

As a final motivation for developing approximate inverses to the velocity stacking operator L, consider figure 1.6. The various approximations to the inverse of L described in the text may be used to reconstruct the original CMP gather of figure 1.1. The left-hand panel is a reconstruction of the common midpoint gather employing $\mathbf{L}^T$ as an estimate to the inverse of L. In the center panel the generalized inverse $(\mathbf{L}^T\mathbf{L})^+$ (Lanczos, 1961) was used to estimate the inverse. Finally, the stochastic inverse $(\mathbf{L}^T\mathbf{L} + \mathbf{D})^{-1}\mathbf{L}^T$ was used in the right panel to reconstruct the data. There is an obvious disparity between the data set itself (figure 1.1) and the reconstructions of it using the approximate forward- and inverse-transform pairs. The stochastic inverse in this case best restores the original data.

1.5 Limitations

It must be emphasized that L functions as a common-midpoint gather modeler, albeit a crude one. The inverse transformation of a common-midpoint gather into velocity space is meaningful only when the implicit assumptions that justify employing L as a modeler are true. What are these assumptions? A reflector on a common-midpoint gather must have uniform amplitude from trace to trace; it must

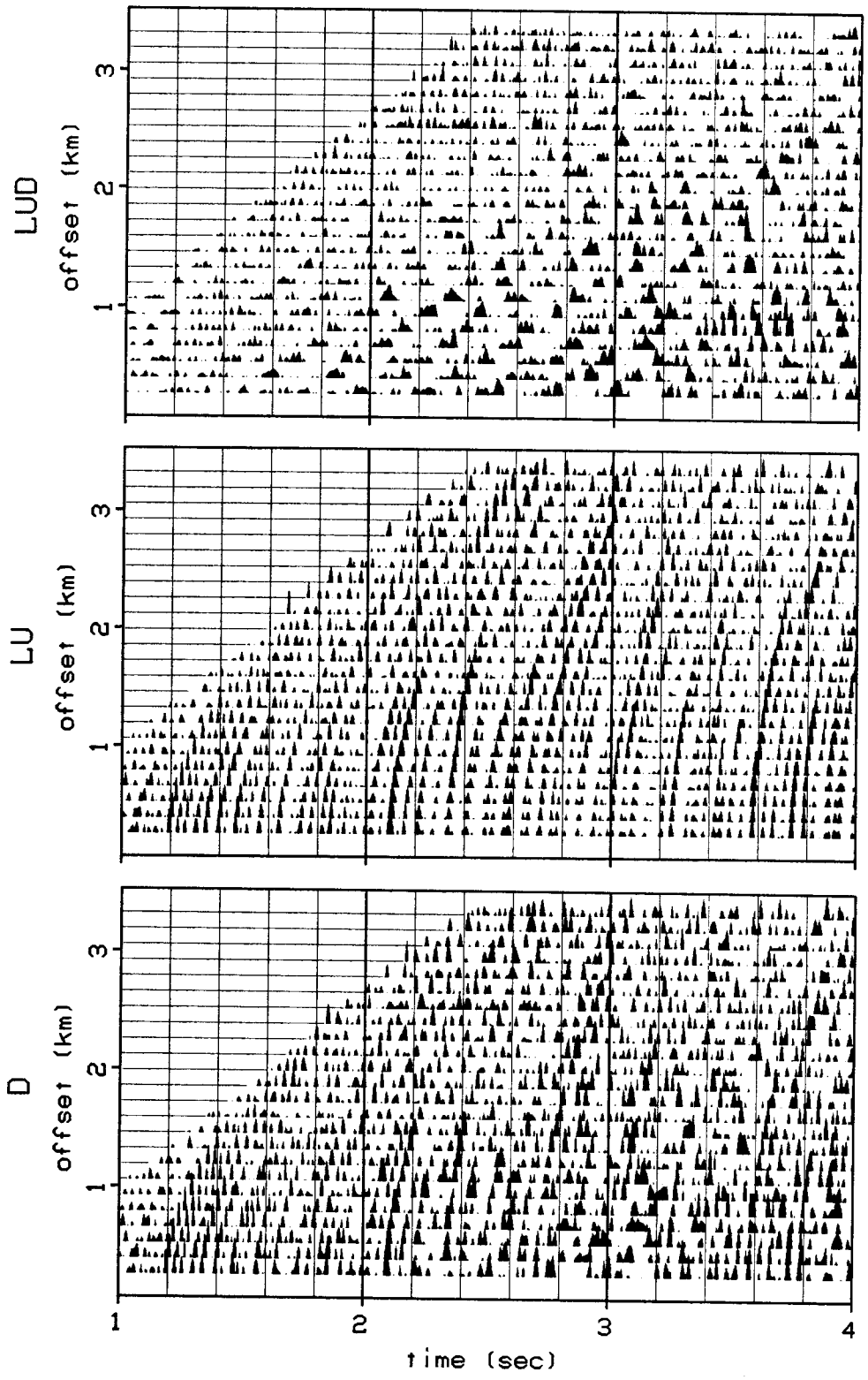

FIG. 1.4. Reconstructing the original data. Panel **LU**, an estimate of the original data (Panel **D**), is made by applying the velocity stacking operator **L** (equation 1.5) to the stochastic inverse **u** of figure 1.3. Panel **LUD** is the difference **d** – **Lu** between the original and the estimated common-midpoint gathers.

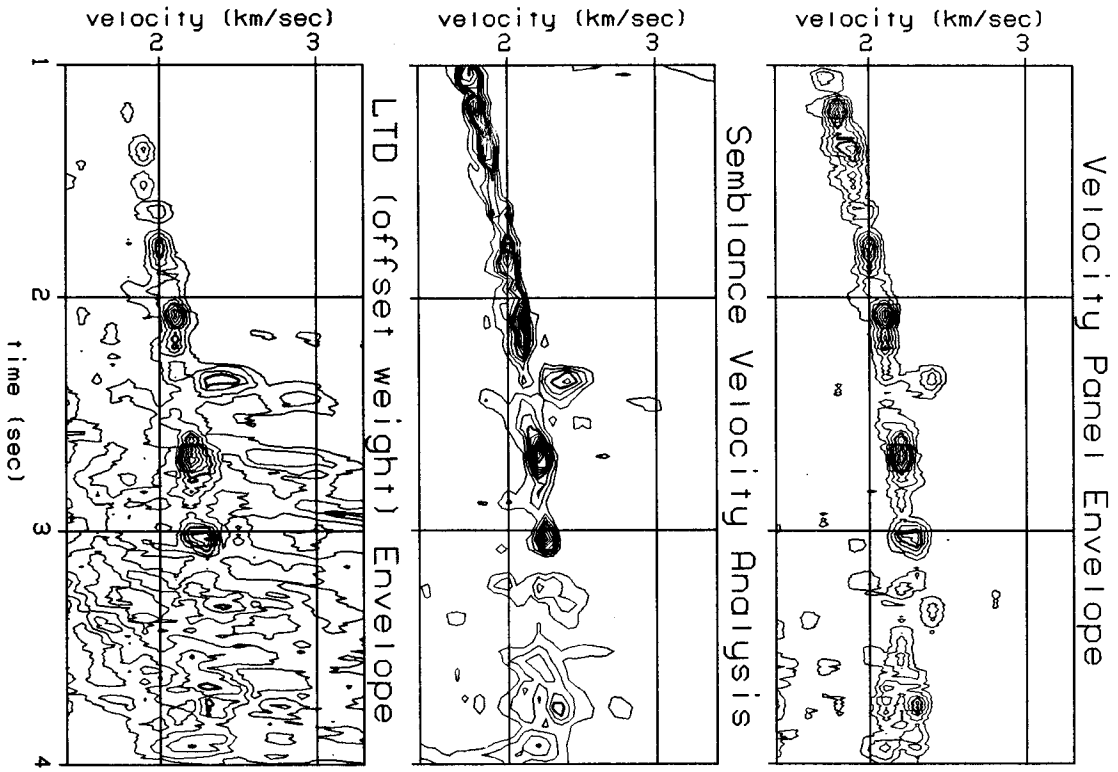

FIG. 1.5. Velocity analyses. The right panel was made by taking a time averaged, normalized envelope of the stochastic inverse **u** of figure 1.3. For comparison the middle panel is a standard semblance velocity analysis on the common midpoint gather. Each point of the panel is a measure of the similarity of waveforms over the corresponding moveout path in data space. Perfect semblance equals 1. The left panel is an envelope of the velocity stack $\mathbf{L}^T\mathbf{Sd}$ of figure 1.3. The fine contour interval on the plots is 0.1. The resolution of velocities in each panel is good. 40 velocities were scanned to make the semblance plot, 20 velocities to make **u** and $\mathbf{L}^T\mathbf{Sd}$. The events in the velocity panel envelope closely track the events of the semblance velocity analysis; for this example the horizontal sampling used in velocity space (20 traces) is adequate.

smoothly vary in moveout from trace to trace; traces must be balanced in amplitude, and no amplitude residuals may be present. Actual data sets satisfy none of these assumptions precisely, but the better the assumptions hold with the seismic data at hand, the better the resultant velocity panel can resolve true earth velocities.

On the brighter side, the results that will be developed in the following

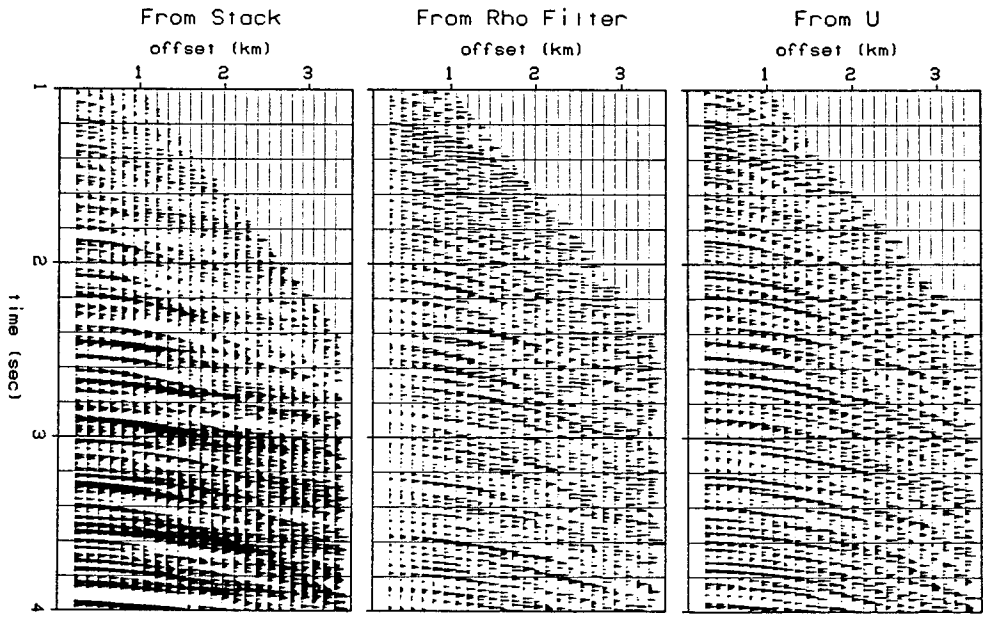

FIG. 1.6. A comparison of various data reconstructions. The left panel is a reconstruction using $\mathbf{L}^T$ in place of the true inverse: in the notation of the text it is $\mathbf{L}\mathbf{L}^T\mathbf{d}$. The middle panel is a reconstruction from the rho-filtered velocity stack, or $\mathbf{L}(\mathbf{L}^T\mathbf{L})^{+}\mathbf{L}^T\mathbf{d}$. The right panel is a reconstruction from the stochastic inverse $\mathbf{u}$, or $\mathbf{L}(\mathbf{L}^T\mathbf{L} + \mathbf{D})^{-1}\mathbf{L}^T\mathbf{d}$. By comparing these panels with the data of figure 1.1 we see that the generalized inverse and stochastic inverse better approximate the original data, and so are "closer" in a sense to the inverse of $\mathbf{L}$ than $\mathbf{L}^T$ is.

chapters may be easily adapted to a better, physically more realistic choice for the modeling operator $\mathbf{L}$. This applies as long as the chosen operator is linear and has an input space not restricted to a single velocity function. For example, a wave equation modeler would eliminate the waveform distortion that accompanies NMO operations.

1.6 Overview of the remaining chapters

The velocity-stacking operator defined in this chapter has a close relationship to slant stacking. They are both back-projection techniques; only in the shapes of the summation paths taken over the data do they differ. Slant stacking involves

summing over paths with linear moveout; that is, the operator can be defined to be

$$u(p,\tau) = \int_{h_1}^{h_2} d(\, h,\, t = \tau + ph \,)\, dh \qquad \text{[slant stack]}$$

Recall that the definition of a velocity stack in (1.2) is

$$u(p,\tau) = \int_{h_1}^{h_2} d(\, h,\, t = \sqrt{\tau^2 + p^2 h^2}\,)\, dh \qquad \text{[velocity stack]}$$

Redefining L as a slant stack raises the new issue of slant stack invertibility. Two effects, aliasing and truncation, definitely preclude the existence of an exact slant stack inverse pair, as will be seen in chapter 3.

Chapter 2 presents greater detail on the issue of invertibility of the operator L, in particular for the above two choices of L: velocity stacking and slant stacking. Estimates of the generalized inverse for each transformation are derived in chapters 3 and 4 and are applied to real data examples. The term "real" will be used from now on to distinguish measured field data from synthetically generated data. Chapters 5 and 6 describe the stochastic inverse to the velocity stack and to the slant stack (respectively) as choices for L. A number of real-data examples there illustrate the performance of the stochastic inverse in resolving events in velocity space.

1.7 Motivation

This thesis is an outgrowth of a proposal of Jon Claerbout's to study the problem of missing data restoration. Seismic data sets are often irregularly or incompletely sampled spatially, and the outcome of subsequent processing of a data set will most likely be influenced by the incomplete sampling. But interpolation of the missing parts of the data always involves some a priori assumptions about the data: what the interpolated traces should look like.

The focus of this study eventually shifted from missing data restoration, to the effect missing parts of the data have on the important step of stacking, and finally to the design of a stacking operator that is relatively insensitive to missing data. As described above, it is basically a least squares approach to fitting of the recorded data. Interpolation of the data is then a byproduct of the least squares inversion step, because a simple operation takes the inversion estimate back into the data domain, yielding an arbitrarily finely sampled estimate of the data. Figure 5.16 in chapter 5 is an example of such an interpolation.

The main a priori assumption made is that of parsimony in the model domain. The incorporation of parsimony into the least squares formulation was motivated by the successful work of Gray (1979) and Godfrey (1979) on parsimonious (or minimum entropy) deconvolution, which in turn was inspired by the work of Jon Claerbout and Francis Muir at the Stanford Exploration Project. Chapters 5 and 6 are an attempt to extend the use of parsimony to a two-dimensional problem, that of imaging velocity space, in comparison to the one-dimensional deconvolution problems of Godfrey and Gray.