Appendix C

Least-squares inverses for time-invariant transforms

C.1 Introduction

Stacking along straight lines or hyperbolas was repeatedly applied in previous chapters as part of velocity filtering and moveout estimation. In these applications, the stacking trajectories had the particular property of time-invariance; that is, independence from the time-origin. The reason for stacking along time-invariant trajectories is that the drill-bit source operates continuously in time, rather than being activated at known times as other conventional seismic sources.

Stacking along straight lines, or also slant stacking, is an example of a linear time-invariant transform that is well known in the field of seismic exploration (Claerbout, 1985b). Another example is the parabolic transform (Hampson, 1986). On the other hand, the Normal Moveout (NMO) transform is not time-invariant.

Since the early applications of the slant-stack transform, it was recognized that the finite aperture of the transform — limited range of offsets and ray parameters — introduces undesirable artifacts and loss of resolution (Schultz and Claerbout, 1978). By formulating the computation of a transform as a least-squares estimation problem, Thorson and Claerbout (1985) demonstrated a method for overcoming the limitations due to finite aperture for a class of linear transforms including the slant-stack transform. Beylkin (1987) formulated efficient algorithms for the inversion of time-invariant linear transforms in the frequency domain.

In this appendix, I combine the previous approach of least-squares inversion in the frequency domain with the new observation that the inversion operators have a Toeplitz structure. The Toeplitz structure is a consequence of applying time-delays that depend linearly on the ray parameter. Therefore the Toeplitz structure is present also when the data are *irregularly* sampled in the offset domain, or when the stacking trajectories are parabolas instead of straight lines.

The Toeplitz property allows considerable savings in computational time and storage of the inverse operators both for under-determined and over-determined least-squares problems. Further, I show how the numerical stability of the inverse transforms is related to sampling rates in ray parameter and frequency, and apply these results to the design of transforms with frequency-dependent apertures.

This appendix is organized in three parts: first, I review definitions of least-squares inverses; then I show how to use the Toeplitz structure of the matrix of normal equations in the computations of the least-squares inverses; third, I derive conditions on the optimal sampling in ray-parameter space.

C.2 Linear time-invariant transform pairs

C.2.1 Definitions of transforms in the time domain

Slant-stack transform

The discrete slant-stack transform is defined in the time-offset domain as

$$m(\tau, p) \;=\; \sum_x d(x, t = \tau + px), \tag{C.1}$$

where the data $d(x,t)$, a function of offset and time, are summed along the straight line $t = \tau + px$, parametrized by ray parameter p and intercept traveltime τ (Thorson, 1985).

More generally, the data could be multiplied by a function of offset $w(x)$ before summation:

$$m(\tau, p) \;=\; \sum_x w(x) d(x, \tau = t + px). \tag{C.2}$$

The weights $w(x)$ could be proportional to the distance between channels and thus account for irregular recording geometry, or they could be a window function applied to reduce truncation effects along offset (Thorson, 1985).

Parabolic and hyperbolic transforms

In the parabolic transform, the straight-line stacking trajectories of the slant stack transform are replaced by parabolas centered at zero offset and defined by the following expression:

$$t = \tau + px^2,$$

where τ is the intercept time at zero offset and p is a parameter related to the curvature at zero offset. The formal definition of the parabolic transform, analogous to Equation C.1, is then:

$$m(\tau, p) \;=\; \sum_x w(x) d(x, t = \tau + px^2). \tag{C.3}$$

The hyperbolic transform — a stack along hyperbolas — was defined in Chapter 3 and referred to as a velocity transform (Equation 3.1):

$$m(\tau, z, s) \;=\; \sum_x w(x) d\left(x, t = \tau + \frac{\sqrt{x^2 + z^2}}{v(z)}\right). \tag{C.4}$$

because the parameters of z and $v(z)$ have a physical interpretation, respectively as the depth to a point source and the RMS velocity in a horizontally layered medium.

Dimensions of the slant-stack operators in the time domain

Denoting the slant-stack operator by $\mathcal{L}$, Equation C.1 can be written also as

$$m \;=\; \mathcal{L} d.$$

The data vector d has dimensions $n_x \times n_t$, while the model vector m has dimensions $n_\tau \times n_p$. Thus a matrix representing the operator $\mathcal{L}$ would have $n_\tau \times n_p$ rows and $n_x \times n_t$ columns.

The conjugate-transpose $\mathcal{L}^{\mathrm{H}}$ of this operator maps data from the $\tau - p$ domain to the

original $x - t$ domain,

$$d = \mathcal{L}^{\mathrm{H}} m. \tag{C.5}$$

Figure C.1 illustrates the impulse responses of the operators $\mathcal{L}$ and $\mathcal{L}^{\mathrm{H}}\mathcal{L}$. The operator $\mathcal{L}^{\mathrm{H}}\mathcal{L}$ approximately reconstructs the original spike. The artifacts along straight lines in the time-offset domain are due to the limited range of ray parameters of the transform.

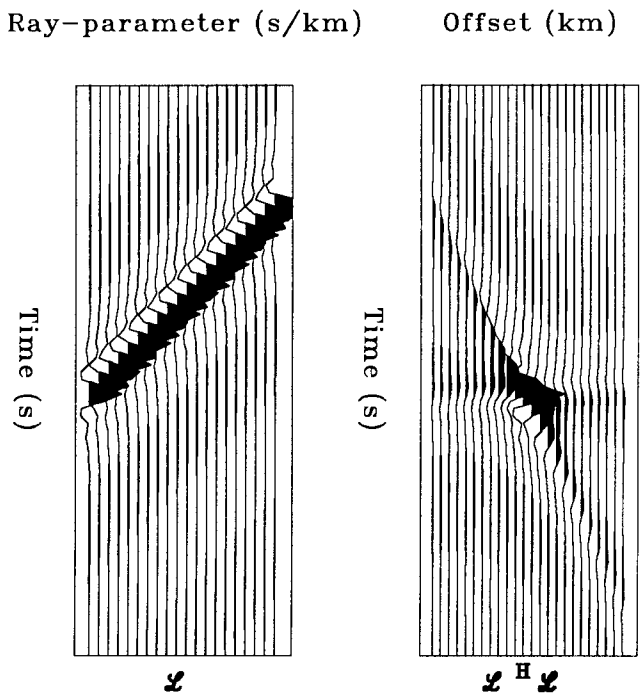

FIG. C.1. Impulse responses of the operators $\mathcal{L}$ and $\mathcal{L}^{\mathrm{H}}\mathcal{L}$. Both, the slant-stack transform operator $\mathcal{L}$ and its complex-conjugate $\mathcal{L}^{\mathrm{H}}$ map impulses into straight lines. Thus, the operator $\mathcal{L}^{\mathrm{H}}\mathcal{L}$ reconstructs approximately an impulse as a superposition of straight lines.

C.2.2 Slant-stack transforms in the $x - \omega$ and $p - \omega$ domains

Applying the Fourier Transform to Equation (C.1) leads to

$$m(\omega, p) = \sum_x d(x, \omega) e^{j\omega p x}, \tag{C.6}$$

where now a time-shift is expressed simply as a multiplication by a complex number. After the Fourier Transform has been applied, different frequency components of the data can be transformed to the $\omega - p$ domain independently from each other. The transform operator

$\mathcal{L}$ is represented at each frequency by a matrix $\mathbf{L}$ of dimensions $n_p \times n_x$, whose elements are:

$$\mathbf{L} = \begin{pmatrix} e^{j\omega p_1 x_1} & \dots & e^{j\omega p_1 x_{n_x}} \\ \vdots & \vdots & \vdots \\ e^{j\omega p_{n_p} x_1} & \dots & e^{j\omega p_{n_p} x_{n_x}} \end{pmatrix}. \tag{C.7}$$

Similarly, the complex-conjugate operator $\mathcal{L}^{\mathrm{H}}$ is represented at each frequency by the matrix $\mathbf{L}^{\mathrm{H}}$, that is the conjugate-transpose of the matrix $\mathbf{L}$ in Equation C.7.

C.2.3 Two forms of the least-squares inverse

When the sampling is uniform in ray parameter, the following identities hold for the elements of matrix $\mathbf{L}$ (Equation C.7):

$$\mathbf{L}_{p,x} = e^{j\omega x k \Delta p} = (e^{j\omega x \Delta p})^k.$$

Thus the entries along each column of the matrix $\mathbf{L}$ are increasing powers of the same complex number. A square matrix with this property is called a Vandermonde matrix (Golub and Van Loan, 1984), and such matrices have full rank, unless two of their columns are identical. Assuming that $\mathbf{L}$ is rectangular with more rows n_p than columns n_x, necessary and sufficient conditions for $\mathbf{L}$ to be full rank equal to n_x are that the complex numbers α_m,

$$\alpha_m = e^{j\omega x_m \Delta p},$$

be distinct from each other. Two particular cases when $\mathbf{L}$ will be *singular* are when $\omega \Delta p$ is zero, or when the transform is aliased. Similar results hold when the sampling in offset, instead of ray parameter, is uniform.

The least-squares inverse of the operator $\mathbf{L}^{\mathrm{H}}$ (equation C.5) can be computed by either one of two operators (Beylkin, 1987),

$$(\mathbf{L}\mathbf{L}^{\mathrm{H}})^{-1}\mathbf{L}, \tag{C.8}$$

or

$$\mathbf{L}(\mathbf{L}^{\mathrm{H}}\mathbf{L})^{-1}. \tag{C.9}$$

Identity of the two last-squares inverses

These expressions for the operators assume that the least-squares residuals are uncorrelated. When there are events in the data that cannot be well explained by the model — either because of errors in the observations, or because the range of model parameters is limited, then the least-squares inverses need to be modified to account for the noise (Tarantola, 1987). I will return to the discussion of such a model later, and continue now with properties of the noise-free least-squares inverses.

The two operators for the noise-free case are identical, as can be seen when we introduce the singular value decomposition (SVD) of the matrix $\mathbf{L}$. The SVD factors the matrix $\mathbf{L}$ into the product of three matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$ of dimensions respectively $n_p \times n_p$, $n_p \times n_x$, and $n_x \times n_x$ (Golub and Van Loan, 1983):

$$\mathbf{L} \;=\; \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}.$$

The matrices $\mathbf{U}$ and $\mathbf{\Sigma}$ are the matrices of left and right eigenvectors, while the matrix $\mathbf{\Sigma}$ is the diagonal matrix of singular values.

The identity of the two least-squares inverses,

$$(\mathbf{L}\mathbf{L}^{\mathrm{H}})^{-1}\mathbf{L} \;=\; \mathbf{L}(\mathbf{L}^{\mathrm{H}}\mathbf{L})^{-1} \;=\; \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{V}^{\mathrm{H}},$$

follows from a straightforward substitution of the SVD decomposition into Equation C.8:

$$(\mathbf{L}\mathbf{L}^{\mathrm{H}})^{-1} \;=\; \mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^{\mathrm{H}};$$

and into Equation C.9:

$$(\mathbf{L}^{\mathrm{H}}\mathbf{L})^{-1} \;=\; \mathbf{V}\mathbf{\Sigma}^{-2}\mathbf{V}^{\mathrm{H}}.$$

Properties of the two least-squares inverses

The first operator (Equation C.8) is the familiar product of the conjugate-transpose and the inverse of the matrix of normal equations. This form of the least-squares inverse operator is well known; it is used for instance in prediction problems in time-series analysis, where prediction filters are obtained as solutions to overdetermined systems of linear equations.

The second form of the inverse operator (Equation C.9) has found applications mainly in problems of interpolation and spline fitting (Mallet, 1989) where the number of unknowns — values of a function at a grid node — far exceeds the available data and requires the solution of an underdetermined linear problem.

To understand how this second operator can be used as an interpolator, consider the expression of the model vector $\mathbf{m}$ (n_p elements for a particular frequency ω) in terms of the data vector $\mathbf{d}$ (n_x elements at the same frequency):

$$\mathbf{m} = \mathbf{L}(\mathbf{L}^{\mathrm{H}}\mathbf{L})^{-1}\mathbf{d}.$$

The elements of the matrix $\mathbf{L}$ can be thought of as providing a set of basis functions $h_p(x)$, which for the slant stack transform are

$$h_p(x) = e^{j\omega px}.$$

The vector of coefficients $\mathbf{b}$ of the decomposition of the model vector $\mathbf{m}$ on that basis are obtained from the equation

$$\mathbf{m} = \mathbf{L}\mathbf{b},$$

by solving the system of linear equations:

$$\mathbf{d} = (\mathbf{L}^{\mathrm{H}}\mathbf{L})\mathbf{b}.$$

The dimensions of the operator $\mathbf{L}^{\mathrm{H}}\mathbf{L}$ are $n_x \times n_x$, independent of the number of values for the ray parameter p. Thus the size of the system of linear equations remains constant as the sampling rate in ray parameter p increases. In contrast, the dimensions of the

operator $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ are $n_p \times n_p$.

Solving a linear system of equations with a full-rank left-hand side matrix is attractive from a numerical point of view. Therefore the operator $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ should be used for over-determined linear systems ($n_p \leq n_x$), and the operator $\mathbf{L}^{\mathrm{H}}\mathbf{L}$ for under-determined systems ($n_x \leq n_p$).

However, the rank of the operator is not the only criterion in choosing between the two forms of the inversion operator. In a later section, I will show that taking advantage of the different structures of these operators may justify solving a rank-deficient linear system.

C.2.4 Finite-aperture versus infinite-aperture inverse

An analytical expression for the least-squares inverse of the slant-stack transform was derived in the frequency-wavenumber domain by Thorson (1985). Thorson's results indicate that the least-squares inverse for the infinite-aperture transform $\mathbf{L}^{\mathrm{H}}$ — that is a transform with infinite range of offsets and dips — is the operator $\mathcal{L}$, followed by a one-dimensional rho-filter, whose transfer function is $|\omega|$. According to these results, the finite-aperture operator $(\mathcal{L}\mathcal{L}^{\mathrm{H}})^{-1}$ should tend to a rho-filter as the aperture increases.

Intuitively, the effect of limited aperture will be strongest for the lowest wavenumbers, that is also at low frequencies. At high wavenumbers however, aliasing will occur. In between, there will be a range of frequencies and wavenumbers in which the finite-aperture and the infinite-aperture inverses should be equivalent. This intuition is confirmed both by Thorson's analysis, and by results given later in this paper.

C.2.5 Inverse transform in the presence of noise

In theory, when the noise on the data is modeled as a Gaussian random process and described by a covariance matrix $\mathbf{C}_D$, it is possible to write an expression for the maximum-likelihood estimator of the model parameters (Tarantola, 1987). The least-squares inverse operators in Equations (C.8) and (C.9) become then

$$(\mathbf{L}\mathbf{C}_D{}^{-1}\mathbf{L}^{\mathrm{H}})^{-1}\mathbf{L}\mathbf{C}_D{}^{-1}, \tag{C.10}$$

and

$$\mathbf{L}(\mathbf{L}^{\mathrm{H}}\mathbf{L} + \mathbf{C}_D)^{-1}. \tag{C.11}$$

In practice, the estimation of signal or noise statistics is a challenging task. The approach often suggested is to introduce a parametric model for the signal, such as an autoregressive model for time series (Burg, 1975), or the superposition of a few plane waves in array processing (Schmidt, 1981; Bresler and al., 1988; Biondi and Kostov, 1989).

An alternative approach, the one developed in this paper, is to increase the number of model parameters, until the residuals become uncorrelated. This second approach is more robust than the first one, because it does not require that the data conform to a particular parametric model. On the other hand, its resolution may be inferior when the model assumptions are appropriate.

C.3 Toeplitz structure of the matrix of normal equations

C.3.1 Toeplitz structure

From the definition of the matrix $\mathbf{L}$ in Equation C.7, I obtain an expression for the elements of the matrix $\mathbf{L}\mathbf{L}^{\mathrm{H}}$:

$$G(p,q) \;=\; \sum_{x=0}^{X_{\max}} e^{j\omega p x} e^{-j\omega q x} \;=\; \sum_{x=0}^{X_{\max}} e^{j\omega (p-q) x}, \tag{C.12}$$

where p and q are ray parameters, constant along rows of the matrix $\mathbf{L}$, and the offsets range from zero to $X_{\max}$. For uniform sampling in the ray parameter, $(p-q)$ is constant along diagonals, and hence the entries of the matrix are also constant along diagonals.

A further property of $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ is that symmetric elements with respect to the diagonal are complex-conjugates, and therefore the matrix is Hermitian-Toeplitz. The Hermitian-Toeplitz property holds also for irregular sampling along offset and for offset-dependent weights applied to the data.

For a parabolic transform (Equation C.3) with weights $w(x)$ along offsets, the elements

of the matrix corresponding to $\mathbf{LL}^{\mathrm{H}}$ become

$$H(p,q) = \sum_{x=0}^{X_{\max}} e^{j\omega(p\,-\,q)x^2} w^2(x).$$

In the case of a hyperbolic transform (Equation C.4) the Toeplitz property holds only if there is only one depth parameter and the sampling in slowness (inverse of velocity) is uniform.

When the sampling in offset x is also uniform, the entries of the matrix $\mathbf{LL}^{\mathrm{H}}$ can be expressed as

$$G(p,q) = \sum_{x} e^{j\omega px} e^{-j\omega qx} = \frac{\sin(\omega X_{\max}(p-q))}{\sin(\omega \delta x(p-q))} e^{j\omega(X_{\max}-\delta x)(p-q)/2}. \qquad \text{(C.13)}$$

Similarly, the matrix $\mathbf{L}^{\mathrm{H}}\mathbf{L}$ is Hermitian-Toeplitz when the sampling in offset x is regular.

To illustrate graphically the Toeplitz property, Figure C.2 displays the magnitude of the elements of the matrix $\mathbf{LL}^{\mathrm{H}}$ for three different frequencies. At low frequencies the matrix is most different from a diagonal matrix; this property confirms that effects of finite aperture are most strongly felt at low frequencies. At intermediate frequencies, the matrix is diagonally dominant; the results of the finite- and infinite-aperture filters should be similar. At high frequencies aliasing introduces identical rows and columns and the matrix becomes singular. Because of their Toeplitz property, each of the matrices $\mathbf{LL}^{\mathrm{H}}$ can be specified by a single column or row. The first rows of these matrices are shown as a function of frequency in Figure C.3, where points with the same wavenumber lay along hyperbolas. The hyperbolas corresponding to wavenumbers that are integer multiples of the Nyquist wavenumber are clearly apparent, because of their high amplitude.

C.3.2 Properties resulting from the Toeplitz structure

Storage

The elements of a Toeplitz matrix are constant along diagonals. Thus, in general, $2n$ complex numbers define a $n \times n$ Toeplitz matrix. When the matrix is Hermitian, only n

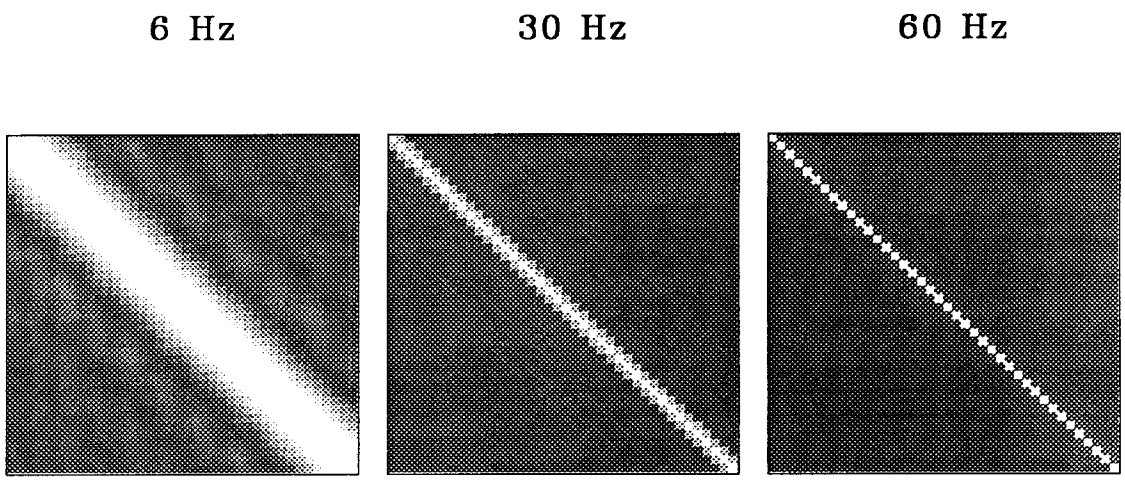

FIG. C.2. Magnitude of the elements in the matrices $\mathbf{LL}^{\mathrm{H}}$ for frequencies equal to 6, 30, and 60 Hz. The matrices are rectangular, with as many rows and columns as there are ray parameters in the transform (Equation C.12). The matrices $\mathbf{LL}^{\mathrm{H}}$ differ most from a diagonal matrix at low frequencies and become singular because of aliasing at high frequencies.

numbers are needed.

Levinson recursion

The Levinson recursion is a fast algorithm for solving systems of linear equations with a left-hand side Toeplitz matrix and a canonical right-hand side vector of the form $(1, 0, \ldots, 0)$. Solving for an arbitrary right-hand side vector requires only a minor modification to the algorithm (Claerbout, 1985).

The recursive solution of a system of Toeplitz linear equations of order n by the Levinson algorithm requires about $2 \times n^2$ floating point operations, only twice the cost of a "matrix times vector" multiplication. The Levinson recursion is computationally efficient, because the solution to a system of order n is expressed simply in terms of the solution to a subsystem of equations, also Toeplitz, of order $(n-1)$. At each order of the recursion, the Levinson algorithm computes a reflection coefficient[1] solution to the canonical system of equations defined above. Given any one of the three sequences — a column of the Toeplitz matrix, the reflection coefficients, or the solutions at each order — the two other sequences can be computed by the Levinson recursion (Marple, 1987).

[1]Cosine of the angle between a vector, and its copy shifted by one element, (Claerbout, 1985a)

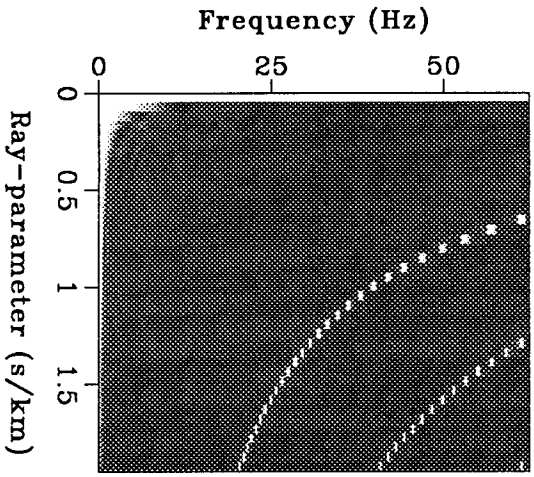

FIG. C.3. First column of the Toeplitz matrices $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ as a function of ray parameter p and frequency ω. Points with the same wavenumber, product of frequency times ray parameter, lay along hyperbolas. The two hyperbolas of high values seen in the bottom right corner of the panel correspond to the Nyquist and to twice the Nyquist wavenumbers. Aliasing of the transform occurs when wavenumbers span an interval larger than twice the Nyquist wavenumber.

Precomputing the data-independent inverse transforms

The least-squares inverses to the slant-stack transforms (Equations C.8 and C.9) are data-independent. Therefore the inverse operators could be precomputed and applied to several data sets, such as common midpoint (CMP) gathers of a seismic survey.

When precomputing the least-squares inverses storage requirements may be a concern. Computations are reduced to a minimum when the inverse matrix is precomputed and stored. However, the inverse matrix of a Toeplitz matrix is only persymmetric (symmetric about the secondary diagonal); therefore storing the inverse of a Hermitian-Toeplitz matrix would require about $n^2/4$ elements.

One alternative, achieving a compromise between storage space and computational savings would be to store the n reflection coefficients, and recompute from them the Cholesky decomposition of the inverse of the Toeplitz matrix (Matrix, 1987).

Finite-aperture unitary transforms

The rows of the matrix $\mathbf{L}$ (Equation C.7) represent vectors – *steering vectors* – whose elements are complex exponentials as a function of the offset. These vectors belong to the data space and have all the same norm.

The matrix $\mathbf{LL}^{\mathrm{H}}$ represents the complex dot products of the steering vectors. When this matrix is diagonal, the steering vectors form an orthogonal basis in the data space. Further, when data and model spaces have the same dimensions, the orthogonality of the rows of $\mathbf{L}$ implies also orthogonality of the columns vectors, and the matrix of the linear transform is a unitary transform.

Working with a unitary transform has several advantages: repeated least-squares transforms between data and model space can be done fast and accurately, and there is least-redundancy between the model parameters and therefore the resolution is improved.

The Cholesky factorization of the inverse of a Toeplitz matrix (Marple, 1987) can be applied in order to obtain finite-aperture unitary transforms. Such unitary transform is perhaps most helpful for working with hyperbolic or parabolic trajectories where choosing orthogonal steering vector is not obvious from considerations of sampling.

C.3.3 Stability of the inverse, sampling, and resolution

The elements of the matrix $\mathbf{LL}^{\mathrm{H}}$ were defined in Equation (A.12) as a function of the ray parameter. Because the product of frequency times ray parameter is a wavenumber, an alternative expression for the elements of $\mathbf{LL}^{\mathrm{H}}$ in terms of wavenumber k is

$$g(p,q) \;=\; \sum_{x=0}^{X_{\max}} e^{jkx}, \tag{C.14}$$

where $k \;=\; (p-q)\omega$. The magnitude of this last expression (Equation A.14) is also the amplitude response to a plane wave incident on a linear array of length $X_{\max}$. Similarly to Equation A.13 and for regular sampling in offset, this amplitude response is a discrete sinc function.

Two remarks about the sampling in wavenumber follow from this observation. First, the array response is periodic in wavenumber, with period equal twice the Nyquist wavenumber, that is

$$2k_{\mathrm{Nyq}} \;=\; 1/\Delta x.$$

Thus sampling over an interval larger than $2K_{\mathrm{Nyq}}$ introduces aliasing. Second, for a linear

array of length $X_{\max}$ and the stacking method, the lower bound on resolution is Δk, where

$$\Delta k \;=\; 1/X_{\max}.$$

The lower bound on resolution is also a lower bound on the sampling rate in wavenumber. For sampling rates in wavenumber finer than Δk, not only is there no gain of resolution, but the system of normal equations $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ becomes ill-conditioned (Golub, 1983), as rows and columns of the matrix of normal equations $\mathbf{L}\mathbf{L}^{\mathrm{H}}$ become nearly identical. Another simple argument confirming the ill-conditioning of the matrix is that the reflection coefficient computed by the first iteration of the Levinson recursion tends to one as the sampling interval in ray parameter tends to zero.

In terms of ray parameters the conditions for aliasing and resolution become:

- Aliasing

  Aliasing occurs for ray parameters p such that

  $$p \;\geq\; P_{\min} \;+\; \frac{2K_{\mathrm{Nyq}}}{\omega} \;=\; P_{\min} \;+\; \frac{1}{\omega \Delta x}.$$

  The number of ray parameters at which the transform is computed increases with frequency in the following way

  $$N_p(\omega) \;=\; \frac{(P_{\max} - P_{\min})}{\Delta p} \;=\; \frac{\omega (P_{\max} - P_{\min}) X_{\max}}{2\pi}.$$

- Sampling

  At a given frequency ω the optimal sampling rate in ray parameter is

  $$\Delta p \;=\; \frac{\Delta k}{\omega} \;=\; \frac{1}{X_{\max}\omega}.$$

Thus, considerations about sampling and resolution recommend regular sampling of the transform in the *wavenumber* domain. When the data are uniformly sampled in offset, the slant-stack transform can be efficiently implemented via a 2-D Fast Fourier Transform (2-D FFT). The orthogonal properties of the FFT imply that the matrices $(\mathbf{L}\mathbf{L}^{\mathrm{H}})^{-1}$ and $(\mathbf{L}^{\mathrm{H}}\mathbf{L})^{-1}$ will be equal to the identity. The inversion of Toeplitz matrices will therefore be

necessary only when the sampling in offset is irregular, or when the traveltime delays are not linearly dependent on offset, or when oversampling in wavenumber.

C.3.4 Aliasing

When there is aliasing, the matrices of normal equations for the inverse operators become singular. The numerical singularity arises from the impossibility of separating sampled signals whose wavenumbers differ by an integer factor of twice the Nyquist wavenumber.

Still, algorithms that compute the least-squares inverses must return an answer even when the transform is aliased for some ray parameters. I have considered two options — either set to zero the aliased points of the transform, or else keep the largest non-singular matrix and then set to zero terms off the diagonal.

Aliasing can be overcome by using prior information, either in the form of an interval of wavenumbers of length smaller than twice the Nyquist wavenumber, or by introducing "a priori" information in Equations C.10 and C.11 via a model-covariance matrix $\mathbf{C}_M$ (Tarantola, 1987):

$$(\mathbf{L}\mathbf{C}_D{}^{-1}\mathbf{L}^{\mathrm{H}} + \mathbf{C}_M{}^{-1})^{-1}\mathbf{L}\mathbf{C}_D{}^{-1}, \tag{C.15}$$

and

$$\mathbf{C}_M\mathbf{L}(\mathbf{L}^{\mathrm{H}}\mathbf{C}_M\mathbf{L} + \mathbf{C}_D)^{-1}. \tag{C.16}$$

The role of the model-covariance matrix is to make the matrices of normal equations (Equations C.15 and C.16) non-singular, by assigning different likelihood values to the ray parameters.

Synthetic examples

Figure C.4 compares the reconstruction of a synthetic bandlimited spike with the conjugate $\mathcal{L}^{\mathrm{H}}$ of the slant-stack operator, and with its inverse $(\mathcal{L}\mathcal{L}^{\mathrm{H}})^{-1}$. The input data consist of a zero-phase bandlimited wavelet in the range of frequencies from 5 to 25 Hz, present at a single trace in the middle of the gather. The least-squares inverse produces a sharper output, with weaker sidelobes than the conjugate operator. The range of dips for both operators is limited, and therefore neither is an exact inverse. However, the least-squares

inverse produces a sharper, higher-frequency output than the conjugate operator.

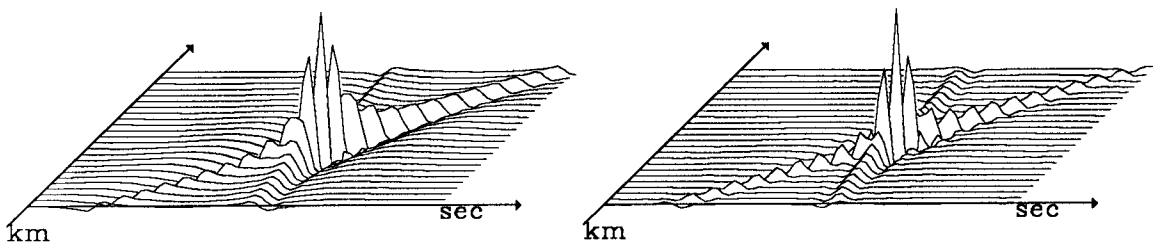

FIG. C.4. Comparison of the inversion of a slant-stack transform with a conjugate (left) and with a least-squares inverse (right) operator. Both inversion operators are dip-limited; however, the least-squares inverse produces a sharper peak with fewer sidelobes.