Stanford Exploration Project, Report 72, December 18, 1997, pages 1–84

Short Note

Interpolation and Fourier transform of irregularly sampled data

Lin Zhang[1]

keywords: *Fourier transform, interpolation, algorithm*

INTRODUCTION

Many algorithms in seismic data processing require that input data have uniform sampling intervals both in time and in space. However, in reality, the difficulties in positioning the sources and receivers may cause the spatial sampling intervals to vary from place to place. If the variations of the sampling intervals are too large to be acceptable, an interpolation process must be done. Because seismic data are almost always sampled uniformly in time, the spatial interpolation can be done along each constant time slice or each constant frequency slice. Therefore, the problem can be reduced to one dimension.

Although the interpolation of seismic data beyond aliasing is still under development, the theories and techniques for interpolating unaliased data are well understood and routinely used in industry. In the 1950s, Yen (1956 and 1957) developed an algorithm for interpolating or resampling irregularly sampled, one-dimensional data. In his algorithm, the irregularly sampled data are modeled as the interpolation of unknown uniformly sampled data:

$$s(t_i) = \sum_j s(j\Delta t)\mathrm{sinc}(t_i - j\Delta t) \qquad i = 1, \ldots, N. \tag{1}$$

Yen solved this linear system for $s(j\Delta t)$, and derived an exact, explicit formula for evaluating the continuous signal $s(t)$.

Hale (1980) tested Yen's algorithm and pointed out that, although the method is exact for noise-free data, it could give erroneous results for noisy data. With Yen's method, a single noisy sample, even one far away from the correct sampling positions, may have an undesirable influence on the values of the interpolated samples. Moreover, Yen's algorithm is computationally expensive. Resampling N points requires an order of N^3 operations. To solve these two problems, Hale replaced the sinc function

[1]**email:** not available

in equation (1) with a truncated, tapered sinc function. Although this procedure sacrifices accuracy at high frequencies, it performs better and at much less computational cost than Yen's method for data contaminated with truncations, aliasing data or noisy data.

This paper describes a new, one-dimensional interpolation method. In this method, a continuous signal is modeled with its Fourier series expansion. The coefficients of this expansion are discrete spectra of the signal, and can be estimated by fitting the modeled data to the irregularly sampled data using the standard least squares technique. It turns out that the normal matrix in this least squares problem is a Hermitian Toeplitz matrix, and its inverse requires only an order of N^2 operations. Therefore, the algorithm is quite efficient. Assuming that the irregularly sampled signal is not aliased and that it has a finite duration, the continuous signal can be recovered from the irregular samples exactly. This algorithm can also be used to compute the discrete Fourier transform of the irregularly sampled data.

In the following sections, I first formulate the problem and observe the Toeplitz structure of the matrix to be inverted. Then, I show two examples of one-dimensional interpolation and one example of computing the discrete Fourier transform. Finally, I offer conclusions and discuss some potential problems of the method.

CONTINUOUS SIGNALS WITH DISCRETE SPECTRA

Let us suppose we have a one dimensional, continuous signal $s(t)$ that has a finite time duration T. Within the time interval T, this signal can be represented by a Fourier series expansion as follows:

$$s(t) = \sum_k S_k e^{j\omega_k t}, \tag{2}$$

where S_k is the discrete spectrum of the signal at the discrete frequency

$$\omega_k = k\Delta\omega = \frac{2\pi k}{T}.$$

Let $\{s_i = s(t_i); \quad i = 1, 2, \ldots, N\}$ be a sequence of samples of this signal. If one can reconstruct the original continuous signal from this finite number of samples, then one can interpolate or resample this signal arbitrarily. This reconstruction is possible if the spectrum of the signal is band-limited, in other words, if the Fourier series expansion in equation (2) has a finite number of terms. With this assumption, equation (2) tells us that the continuous signal can be fully described by a finite number of samples of its spectrum. Thus, the key step of the reconstruction is to estimate the spectrum.

Substituting the known samples into equation (2) gives

$$s_i = \sum_k S_k e^{j\omega_k t_i} \qquad i = 1, 2, \ldots, N. \tag{3}$$

If we define

$$s^T = (\, s_1 \quad s_2 \quad \ldots \quad s_N \,)\,,$$
$$S^T = (\, S_1 \quad S_2 \quad \ldots \quad S_K \,)$$

and

$$\mathbf{A} = \begin{pmatrix} e^{j\omega_1 t_1} & e^{j\omega_2 t_1} & \ldots & e^{j\omega_K t_1} \\ e^{j\omega_1 t_2} & e^{j\omega_2 t_2} & \ldots & e^{j\omega_K t_2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j\omega_1 t_N} & e^{j\omega_2 t_N} & \ldots & e^{j\omega_K t_N} \end{pmatrix},$$

where K is the number of frequencies required to model the signal, we can rewrite equation (3) into the corresponding matrix form

$$\mathbf{A}S = s. \tag{4}$$

The least squares solution of this system equation is defined by the following normal equation:

$$\mathbf{R}S = \mathbf{A}^H s, \tag{5}$$

where $\mathbf{R} = \mathbf{A}^H\mathbf{A}$ is a $K \times K$ matrix with complex elements and $\mathbf{A}^H$ is the conjugate transpose of $\mathbf{A}$.

There exist many numerical algorithms for solving a general normal equation. But the computation required is generally in an order of K^3. Fortunately, as we will see later, the matrix $\mathbf{R}$ has a special structure that makes it possible to solve equation with an order of K^2 operations.

TOEPLITZ STRUCTURE

The elements in the matrix $\mathbf{R}$ are

$$r_{lk} = \sum_i e^{-j\omega_l t_i} e^{j\omega_k t_i} = \sum_i e^{j(\omega_k - \omega_l)t_i}. \tag{6}$$

Because the frequency is uniformly sampled, $\omega_k - \omega_l$ is constant along diagonals of the matrix; hence the elements of the matrix $\mathbf{R}$ are also constant along its diagonals. Matrices with this kind of structure are called Toeplitz matrices. Since one can further show that symmetric elements with respect to the main diagonal of the matrix are complex conjugates, the matrix is Hermitian-Toeplitz. Figure **??** shows an example of the real and imaginary parts of $\mathbf{R}$. With a left-side Hermitian-Toeplitz matrix, equation (5) can be efficiently solved using the generalized Levinson recursion (Golub and Van Loan, 1989; Kostov, 1989). The number of operations required is $2K^2$, just twice the cost of multiplying a vector by a matrix. Once the discrete Fourier spectra are found, equation (2) can be used to evaluate the continuous signal $s(t)$ at arbitrary positions. If the evaluations desired are at uniformly spaced positions, the fast Fourier transform should be used. When the input signal is uniformly sampled, matrix $\mathbf{R}$ becomes an identity matrix, and the whole interpolation process is equivalent to the ordinary sinc interpolation.

The elements of matrix **R** depend only on the sampling positions, not on the values of samples. Therefore, to interpolate missing traces, one can pre-compute the inverse of matrix **R**, and then apply it to each time or frequency slice.

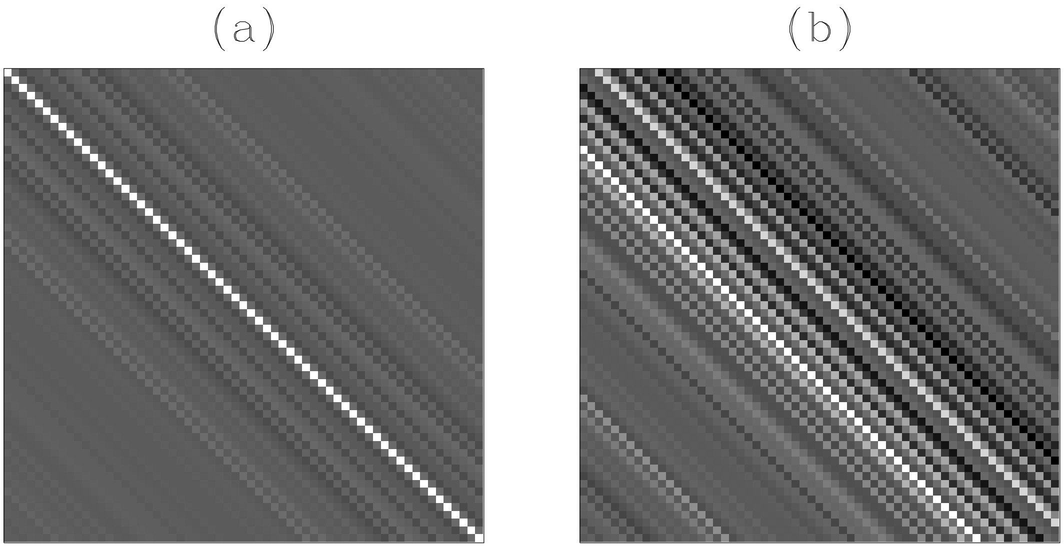

Figure 1: Elements in the matrix **R** of a general normal equation set: (a) real part, (b) imaginary part. lin2-ata [ER]

EXAMPLES

For the sake of simplicity, I demonstrate the method by interpolating one dimensional time sequences. The data sequence shown in the following examples is a segment of a trace from a field dataset. The original trace has a uniform time sampling interval of 2 ms and a maximum frequency of 160 Hz. To prevent the signal from aliasing after random sub-sampling, the signal is low-passed to 100 Hz. Figure **??** shows two examples of interpolating the randomly sub-sampled signal. Both examples have 32% randomly missing samples, but the distributions of the missing samples are different. The distribution of the missing samples in the first example is close to uniform, while the missing samples in the second example are clustered into a few groups. From Figure **??**, we see that the results of the interpolation are satisfactory. Figure **??** compares the amplitude spectra computed from the original signal and from the signal with random missing samples, respectively. The results indicate that the algorithm can accurately estimate the spectrum of an irregularly sampled signal.

CONCLUSIONS

I have presented an efficient algorithm for interpolating irregularly sampled, one dimensional data. This algorithm can also be used to compute the discrete Fourier

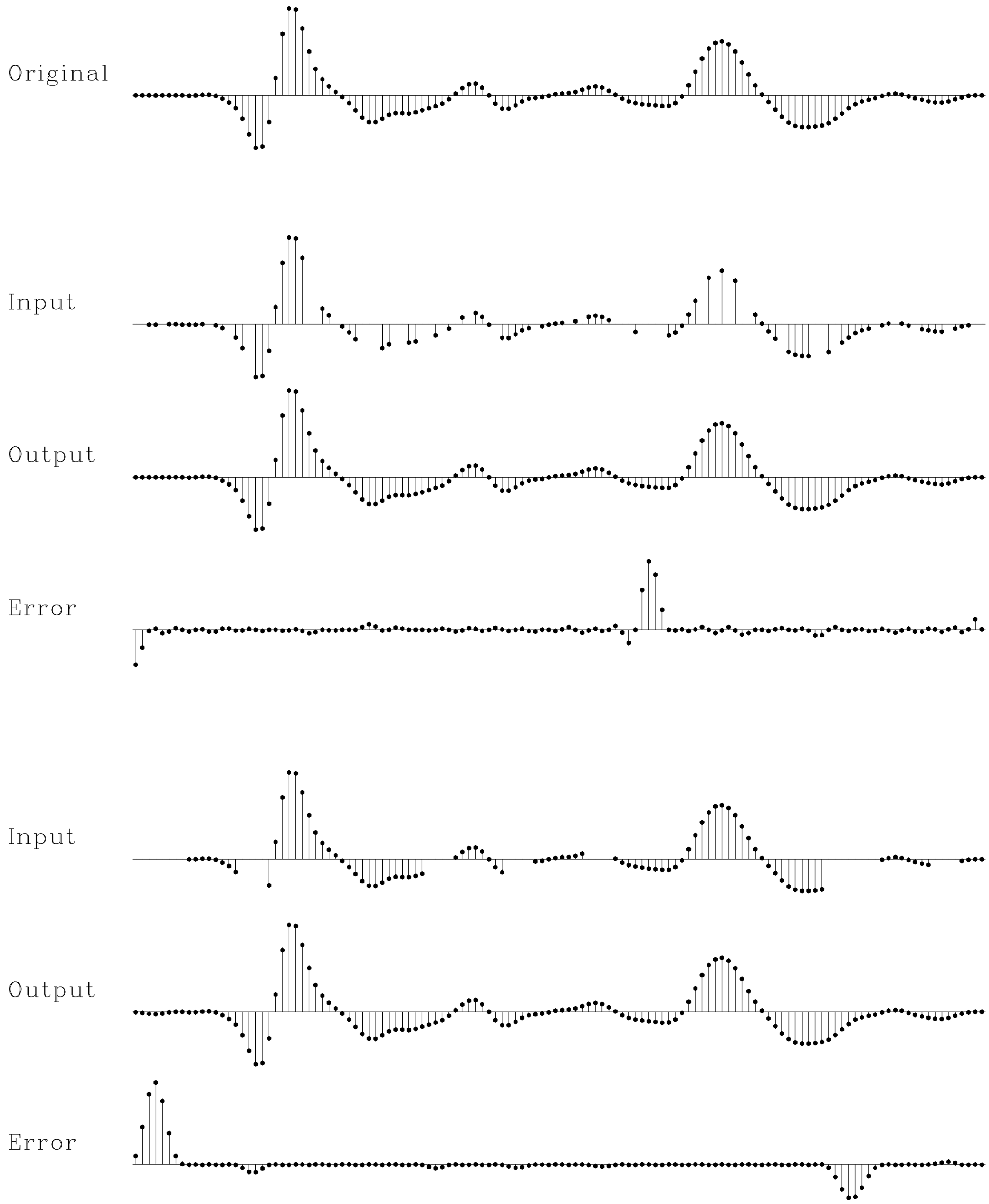

Figure 2: Two examples of interpolating irregularly sampled data. The top trace shows a segment of seismic data. The three traces in the middle show data before and after interpolation, and interpolation error, which is magnified by a factor of 500. The three traces at the bottom show another example in which the missing samples are clustered. The error in this second example is magnified by a factor of 40. lin2-interp [ER]

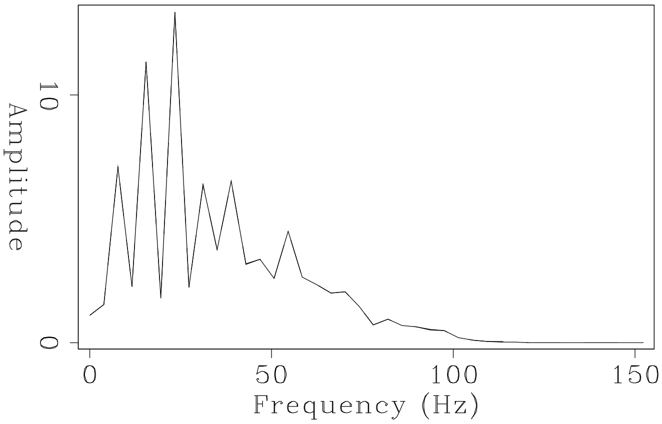

Figure 3: Amplitude spectra: The solid curve is computed from the original signal; The dashed curve is computed from the signal with random missing samples. lin2-specirre [ER]

transform of irregularly sampled data. Examples show that, if the given data samples are not aliased and are of finite length, the algorithm gives satisfactory results. Two potential problems are not addressed in this paper. The first one is the truncation effects at the beginning and ending points of the interpolated signal. Because the signal is modeled with its Fourier series expansion, it behaves periodically. The ending point of one period of the signal is connected with the starting point of the next period. Any discontinuity at the juncture cannot be modeled by a finite number of frequencies. This problem results in interpolation errors near the beginning and ending points of the signal. This problem can be partially solved by padding free samples at the end of the signal to be interpolated. The second problem is related to noises and bad samples. To prevent isolated noisy samples from affecting far away interpolated samples, Hale (1980) suggested that a local interpolation operator should be used. The operator defined in equation (2) contains periodic functions; hence it is a global operator. It is not obvious how one can localize this operator. Nevertheless, the output of the algorithm is insensitive to random noises outside of the signal band.

ACKNOWLEDGMENTS

I thank Francis Muir for providing inspiration for this work.

REFERENCES

Golub, G.H., and Van Loan, C.F., 1989, Matrix Computations: The Johns Hopkins University Press, 186–187.

Hale, D., 1980, Resampling irregularly sampled data: SEP–**25**, 39–57.

Kostov, C., 1989, Finite-aperture slant-stack transforms: SEP–**61**, 261–279.

Yen, J.L., 1956, On the nonuniform sampling of bandwidth-limited signals: IRE Trans. Circuit Theory, v. CT–**3**, 251–257.

Yen, J.L., 1957, On the synthesis of line-sources and infinite strip-sources: IRE Trans. on Antennas and Propagation, v. AP–**5**, 40–46.