Data interpolation in pyramid domain

Xukai Shen

ABSTRACT

Pyramid domain is defined as a frequency-space domain with different spatial grids for different frequencies. Data interpolation in pyramid domain is preferable since for stationary events only one prediction error filter (PEF) is needed for estimating all offsets and frequencies. However, when it is necessary to estimate both missing data and the PEF, it becomes a nonlinear problem. By solving both iteratively, we can linearize the problem and frame it as a least-squares problem. Initial results show that the iterative method will recover the missing data and PEF quite well.

INTRODUCTION

Data interpolation is an important step in seismic data processing that can greatly affect the results of later processing such as multiple removal, migration and inversion. There are many ways to interpolate data, including PEF-based approaches (e.g., Spitz, 1991; Crawley, 2000) and Fourier transform-based approaches (e.g., Xu et al., 2005). The first paper interpolate data using prediction filters in the frequency-space domain while the second one does it by using prediction-error filters (PEFs) in the time-space domain. On the other hand, the third one uses Fourier transform on irregularly sampled data to perform interpolation. The first two belong to the class of methods which is PEF-based data interpolation. A PEF is a filter that predicts one data sample from n previous samples, where n is the length of the PEF. One important feature of a PEF is that when convolved with known data, it minimizes the convolution result in the least-square sense, thus making it a good representation of correlations between samples of known data. Previous work includes Claerbout (1999), Curry (2007) and Crawley (2000). However, one disadvantage is that if PEF estimation is done in frequency-space (f-x) domain, one PEF is needed for every frequency.

Pyramid domain is a resampled representation of ordinary f-x domain. Although it has the axes of frequency and space, the spatial sampling is different for different frequencies. This is attractive because we can use sparser sampling to adequately sample the data at lower frequencies, which makes uniform sampling for all frequencies unnecessary. Therefore in pyramid domain, coarser grid spacing is used for lower frequencies, while finer spacing is used for higher frequencies. This makes it possible to capture the character of all frequency components of stationary events with only one PEF, making data interpolation more efficient.

In this paper, I present a way to interpolate data in pyramid domain based on PEF estimation. The paper is organized as follows: I first show how to transform data into pyramid domain and describe the method for missing-data interpolation and PEF estimation. I then show synthetic data examples, and discuss possible future work.

METHODOLOGY

There are two important parts of the pyramid-based interpolation algorithm, the first one is forward and inverse pyramid transform. The more accurate these transforms are performed, the better the result we can get for missing data interpolation. The second step is data interpolation and PEF estimation in pyramid domain, which are done alternately in a iterative way.

Forward and inverse pyramid Transform

A pyramid domain is derived from f-x domain, where instead of a uniform spatial grid spacing for all frequencies, larger grid spacing is used for lower frequencies. As a result, low-frequency data components are represented by fewer data points in pyramid domain, while higher-frequency data components are represented by more data points. However the grid spacing is not allowed to be smaller than the given grid spacing in f-x, since our data does not contain information at finer scales. This means that even for high frequencies, the maximum spatial samples that can be used is equal to the number in original f-x data.

Significant work has been done on this topic, a good example of which can be found in Sun and Ronen (1996), who transform the data from t-x space to f-x space, where f denotes frequency; the spatial grid spacing is then calculated for each frequency f using the equation

$$\Delta x(f) = \max\left(\Delta x_0, v/f\right), \tag{1}$$

where Δx_0 is the uniform spatial grid spacing in the original f-x data and v is the velocity that controls the slope of the pyramid. The equation gives the maximum grid spacing for a certain frequency without aliasing stationary events in the t-x domain. Next, an inversion scheme is employed to transform data in f-x space to pyramid domain as follows.

$$\mathbf{Lm} - \mathbf{d} \approx \mathbf{0}, \tag{2}$$

where $\mathbf{m}$ is the model to be solved, which is data in pyramid domain, $\mathbf{d}$ is the known data in f-x space, and $\mathbf{L}$ is the interpolation operator. In this paper, I use simple linear interpolation; however, a more accurate interpolation would improve the PEF estimation. Sun and Ronen (1996) suggests using 8-point sinc interpolation to produce a relatively accurate result. The inverse transform uses a similar equation to the one mentioned above.

$$\mathbf{d} = \mathbf{Lm}. \tag{3}$$

Instead of solving an inverse problem, it is now a forward modeling problem where $\mathbf{m}$ is known and $\mathbf{d}$ is unknown data in f-x space.

Missing data estimation and PEF estimation

When transforming data from the f-x domain to the pyramid domain, there will be missing data in the pyramid domain if there are missing data in the f-x domain. We must interpolate the missing data and estimate the PEF in pyramid domain. If they are done simultaneously, the problem becomes non-linear. Instead of directly solving this non-linear problem, we alternately solve two linear problems at each iteration, which are PEF estimation and missing data interpolation.

For missing data estimation, we start with a known PEF $\mathbf{A}$, and try to solve the following least-squares problem (Claerbout, 1999):

$$\begin{aligned} \mathbf{W}\left(\mathbf{d}_{interp} - \mathbf{d}_{known}\right) &\approx \mathbf{0} \\ \epsilon \mathbf{K}\mathbf{A}\mathbf{d}_{interp} &\approx \mathbf{0}, \end{aligned} \tag{4}$$

where $\mathbf{W}$ is a diagonal masking matrix that is 1 where data is known and 0 elsewhere, ϵ is a weight coefficient that reflects our confidence in the PEF, $\mathbf{K}$ is a diagonal masking matrix that is 1 where the convolution can be used and 0 elsewhere.

For PEF estimation, I try to solve the following problem assuming training data are $\mathbf{D}_{known}$, and W is the same as explained above while $\mathbf{a}$ is the unknown PEF (Claerbout, 1999):

$$\mathbf{K}\mathbf{D}_{known}\mathbf{a} \approx \mathbf{0}, \tag{5}$$

EXAMPLE

Here I show the result of the missing data interpolation using PEF estimated on a synthetic data set consisting of 3 plane waves of different frequency bands, locations and dips (Figure 1a). After transforming data in f-x domain (real part shown in Figure 1b) into pyramid domain (real part shown in Figure 1c), I selected a subset of the data, keeping only one trace out of every twenty in the radial direction (Figure 2a). This sub-sampling is done by first creating a missing data mask in f-x domain, then transform this mask into pyramid domain, then apply the mask onto known data in pyramid domain, thus creating missing data in pyramid domain. This is different from how it should be done normally, which is decimate data in f-x domain before go into pyramid domain. But it is a good start to understand how things work in pyramid domain. Notice that the interpolated data (Figure 2b) looks like the original data (Figure 1c). The different between interpolated data and original data at known locations (real part shown in Figure 2c) may be further minimized by employing more accurate interpolation scheme to go from f-x domain to pyramid domain.

FUTURE WORK

Currently a linear interpolation operator is used for the forward and inverse pyramid transform, and the nonlinear iteration have not yet been performed. A better understanding of how data behaves in pyramid domain will allow me to perform the nonlinear iterative method which estimates PEF and interpolates missing data at the same time. I find that linear interpolation operator introduces artifacts that affect PEF estimation, in future tests, I will try an 8-point sinc interpolation operator, which may yield a better result for PEF estimation and data interpolation. Adding a weighting function may also improve PEF estimation by enabling increased reliance on known data, and less on interpolated data. Also, I plan to test the algorithm on more realistic data set. Since in the Pyramid domain only one PEF is needed for all frequencies, interpolation should be faster than the f-x domain PEF estimation scheme described in Curry (2007).

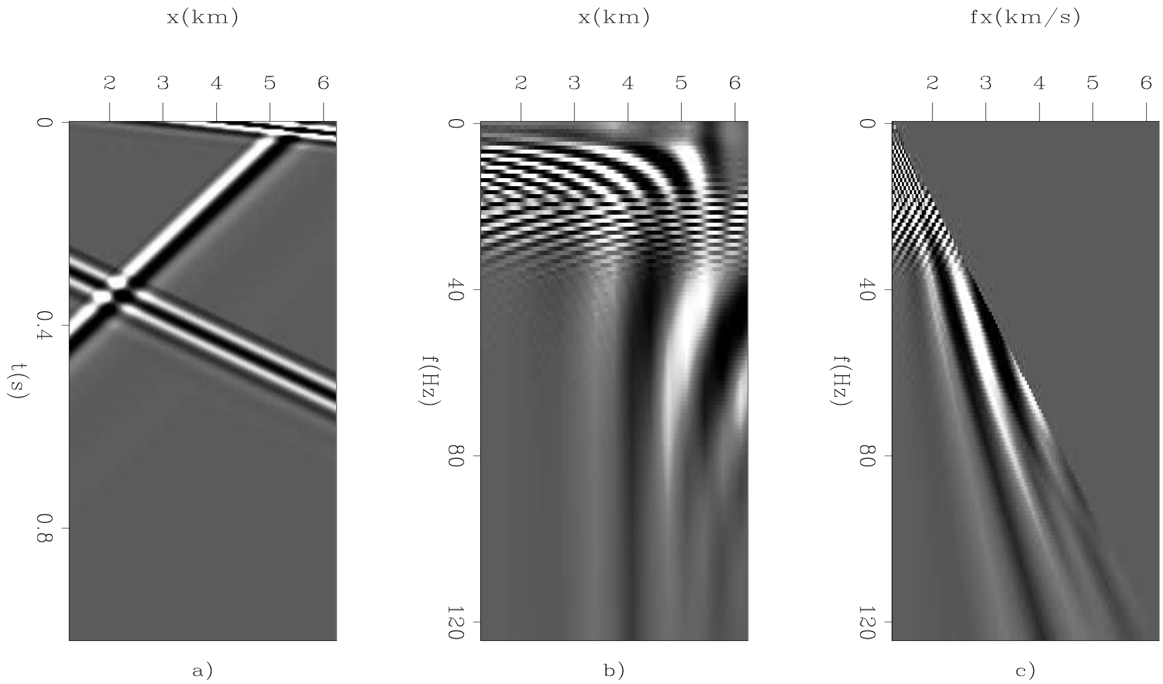

Figure 1: a) Input data, consisting of three plane waves with different spatial extents and frequency contents. b) real part of the input data in the f-x domain. c) real part of input data in pyramid domain. [**ER**]

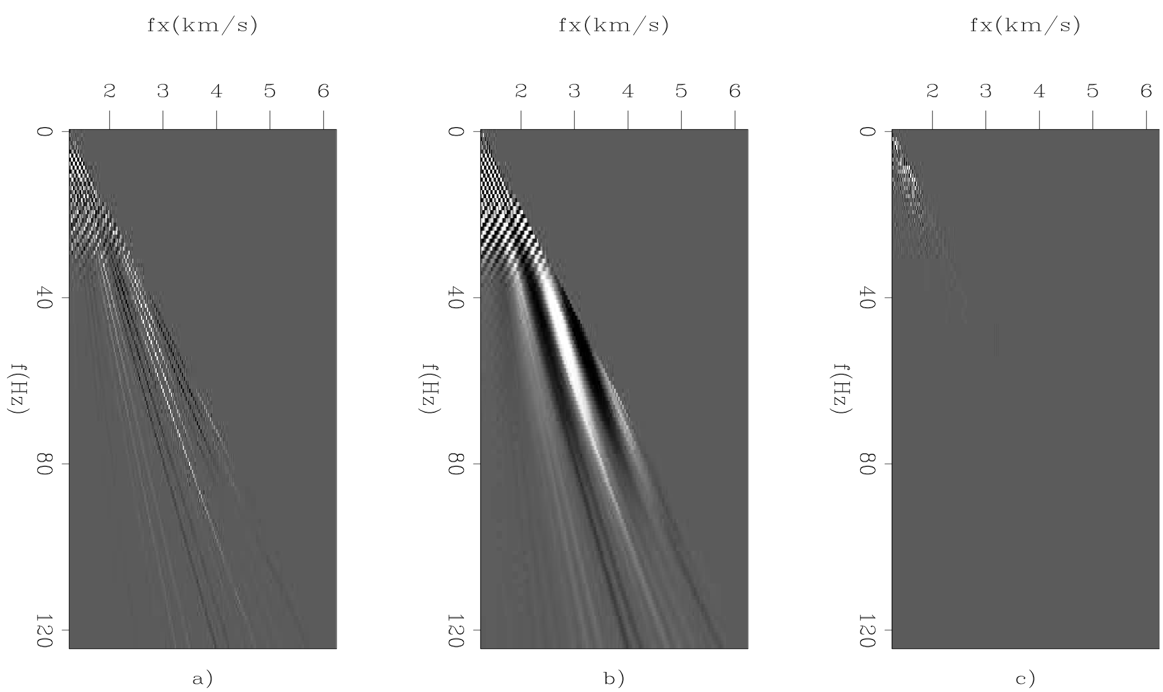

Figure 2: a) real part of known data in pyramid domain. b) real part of the interpolated data in pyramid domain. c) real part of weighted difference between original data and interpolated data in pyramid domain, clipped at 0.01 for display purpose, with a) and b) clipped at 0.2. [**ER**]

CONCLUSIONS

Pyramid domain is a very promising domain for missing data interpolation. The synthetic example demonstrates that with good initial PEF estimate, we can interpolate the missing data relatively accurately. With only one PEF needed for the whole data set, the efficiency of the algorithm is competitive.

ACKNOWLEDGMENTS

I would like to thank Jon for ideas about data interpolation in pyramid domain, and also thank Antoine and Bill for numerous debugging suggestions.

REFERENCES

Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project.

Crawley, S., 2000, Seismic trace interpolation with nonstationary prediction-error filters: PhD thesis, Stanford University.

Curry, W., 2007, f-x domain non-stationary interpolation in patches: SEP-Report, **131**.

Spitz, S., 1991, Seismic trace interpolation in the F-X domain: Geophysics, **56**, 785–794.

Sun, Y. and S. Ronen, 1996, The pyramid transform and its application to signal/noise separation: SEP-Report, **93**, 161–176.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87–V95.