A global-local-hybrid multi-dimensional seismic data interpolation algorithm

Wenyi Hu, Advanced Geophysical Technology Inc.*

Summary

We report a new algorithm for multi-dimensional (5D) seismic data interpolation. To achieve high efficiency without compromising accuracy, this algorithm employs a novel global-local-hybrid interpolation engine, which is applicable to any arbitrarily distributed shots and receivers. Two key components of this interpolation algorithm are: 1) the global-local-hybrid multi-resolution interpolation scheme; 2) the optimally windowed sinc function serving as the interpolation operator at each resolution level.

The mechanism of this global-local-hybrid interpolation scheme is based on the philosophy of multi-resolution analysis. The multi-dimensional seismic data interpolation problem is posed as a series of inverse problem solved in an iterative manner with gradually increasing spatial resolution. At the first iteration, a low resolution interpolation operator with compact support is applied to derive a continuous function, attempting to match the given dataset with low accuracy but with proper global information. In the second iteration, the same interpolation operator with higher resolution is applied to obtain another continuous function with more local information to match the interpolation residual from the previous iteration. The iteration continues until the interpolation error is within the predefined tolerance. The summation of these constructed functions with different resolutions yields the final broadband interpolation function, which accurately predicts the seismic data at any arbitrarily specified locations.

There are some unique features of this algorithm: 1) both global and local information are fully utilized through a progressive refinement procedure, resulting in a method with global interpolation performance at local interpolation cost; 2) the algorithm is highly efficient because the interpolation operator has compact support with very short stencil length of 4; 3) the interpolation operator is derived using an optimally designed finite impulse response (FIR) filter with precise control over spectral properties; 4) There is no extra regularization parameter required to balance the smoothness and the accuracy of the interpolation function; 5) the interpolation error is quantitatively monitored to ensure the interpolation quality; 6) this hybrid interpolation algorithm is guaranteed to converge.

Introduction

Seismic data acquired have never been uniformly distributed. For marine seismic survey, usually the data are sampled sparsely in one direction and densely in the other direction. Irregularity in data spatial distribution is common mainly due to cable drifting. This problem becomes more serious for land data acquisition. Large quantities of bad traces and big data holes are not uncommon in many land seismic data survey.

The spatial sampling irregularity in seismic data may have severe negative impacts on many seismic processing techniques. For example, this issue may cause strong migration swings and serious aliasing phenomena in the Kirchhoff migration or the reverse time migration (RTM) images. In a velocity tomography algorithm, the irregularly distributed seismic data may cause the ray density imbalance, which reduces the inversion sensitivity in some specific subsurface areas. Therefore, special treatment must be adopted to deal with this issue (Hu et al., 2012). For full waveform inversion (FWI), the data irregularity can directly impair the gradient, a critical component for proper velocity update in a FWI algorithm. Therefore, multi-dimensional seismic data interpolation is very important in modern seismic data processing.

There are several methods developed to solve this challenging problem (Xu et al., 2005; Abma and Kabir, 2006; Poole and Herrmann, 2007; Zwartjes and Sacchi, 2007; Hollander et al., 2012). Xu et al. (2005) proposed an approach called antileakage Fourier transform (ALFT). They realized that the difficulty of irregular data Fourier transform comes from the spectral leakage, caused by the nonorthogonality of the Fourier basis function defined in an irregular grid system. In their approach, the orthogonality is regained by iteratively adjusting each Fourier coefficient to fit the original data. Although the ALFT is accurate, due to the high computational expense of discrete Fourier transform (DFT), it is not efficient for large scale projects. Another category of methods based on fast Fourier transform (FFT) instead of DFT (Trad, 2009) is inherently more efficient. Unfortunately, the seismic data need to be binned during the interpolation procedure because FFT is defined in a regular grid system. The bin intervals can affect the interpolation accuracy, the interpolation stability, and the computation efficiency. The interpolation quality and the robustness of this approach are dependent on the user interaction and experiences.

In this work, we developed another multi-dimensional seismic data interpolation algorithm. This algorithm is based on a global-local-hybrid interpolation strategy (Lee et al., 1997). First, we assign a set of hidden control lattices in a coarse-to-fine hierarchy order. After that, we estimate a continuous interpolation function in the coarsest lattice (i.e., at the lowest resolution) to approximately match the given dataset. Then, the interpolation iteration proceeds to the next level of resolution with the refined control lattice, to compensate the interpolation residual from the previous level. With the progress of the iteration, we are able to obtain a series of interpolation functions with gradually increasing resolution and decreasing interpolation error. The summation of these interpolation functions can be used to predict the seismic data at any arbitrary locations. At each resolution level, the interpolation operator is a Kaiser windowed sinc function (Hicks, 2002) with optimal parameters, a method with precise controlled spectral properties. In summary, this approach includes two steps: 1) interpolation function construction step - use the given dataset (irregularly distributed) to estimate the continuous functions in a series of pre-assigned control lattices (regularly distributed); 2) prediction step - use the obtained interpolation functions in the first step to predict the seismic data at the targeted locations. Note the both steps are implemented in an iterative multi-resolution manner.

Problem Description

Figure 1a) is a diagram showing the shots and receiver distribution geometry of a typical seismic survey, where a huge data coverage gap is observed. Figure 1b) shows the distribution after the interpolation. Both the original seismic data and the interpolated seismic data are irregularly distributed. From a signal processing perspective, this is a scattered data interpolation problem. The goal of scattered data interpolation is to construct a continuous interpolation function in the domain under investigation by matching the given data. With a properly constructed interpolation function, one is able to estimate the data values at any desired locations in the domain. As to seismic data, this function can be represented by $S(x_S, y_S, x_R, y_R, t)$, where x_S, y_S, x_R, *and* y_R are the *x* and *y* coordinates of the shots and the receivers and *t* is the time. Therefore, this is a 5D interpolation. Because the time dimension is always regularly sampled in seismic data, although the function has five dimensions, there are only four spatial dimensions need to be interpolated. Before we consider a 5D interpolation, for ease of explanation, we first investigate a simpler problem, a 2D spatial domain scattered data interpolation. After that, we extend the method to 5D. The 2D spatial domain interpolation problem is: given a set of scattered data points $P_d = \{(x_d, y_d, z_d)\}$ defined in a rectangular domain $\{(x,y) \mid x_{min} < x < x_{max}, y_{min} < y < y_{max}\}$, how to construct a continuous function defined in this rectangular domain to match this given dataset?

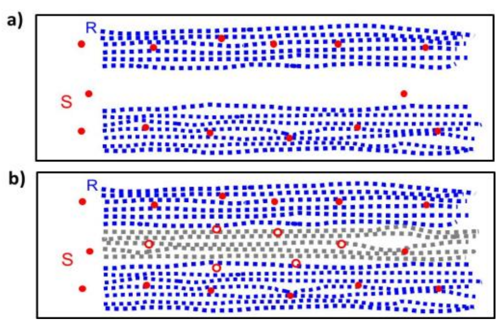

Figure 1: Diagram showing irregular distribution of shots and receivers. a) Before interpolation. Red dots represent shots and blue dots are receivers. b) After interpolation. Red circles represent interpolated shots and gray dots are interploated receivers.

Local Interpolation and Bandwidth Limit

There are two main categories of scattered data interpolation: local methods and global methods. Global methods attempt to use the full dataset simultaneously to build a smooth global interpolation function defined in the whole domain, which implies high computational expense. Both the DFT-based (Xu et al., 2005) and the FFT-based methods (Trad, 2009) fall into this category. On the other hand, a local interpolation method uses only a subset of the given data to build an interpolation defined in the corresponding subdomain. A local interpolation needs to be carried out multiple times until all the subdomains are visited. Local methods are efficient but it has the bandwidth limitation because the subdomain size determines the lower bound of it spatial spectrum.

Our global-local-hybrid method is based on a local interpolation operator while the global information is brought in through an iterative multi-resolution analysis strategy. In other words, our algorithm consists of multiple iterations and each iteration is performed to construct an individual interpolation function with different resolution. Within each iteration, the interpolation is implemented using a local interpolation method to ensure the computational efficiency. For this reason, we call this method a global-local-hybrid interpolation. In this section, we introduce our local interpolation operator to perform the scattered data interpolation at a predefined resolution level.

An example is shown in Figure 2a), where the given data points *P* are irregularly distributed. To conduct the local interpolation, as shown in Figure 2b), we first introduce an underlying control lattice whose grid size determines the resolution of the to-be-constructed interpolation function. The nodes of the control lattice are represented as C_{ij} whose values are φ_{ij}. Using the given data points, the values at the control nodes can be estimated with the following formulations (Lee et al., 1997):

$$\varphi_{ij} = \frac{\sum_{d=1}^{n} w_d{}^2 \varphi_d}{\sum_{d=1}^{n} w_d{}^2} \qquad (1)$$

$$\varphi_d = \frac{w_d z_d}{\sum_{a=0}^{3}\sum_{b=0}^{3} w_{ab}^{\ 2}} \tag{2}$$

$$\begin{aligned} &w_d = w_{kl} = I(s)I(t) \\ &k = (i+1) - \lfloor x_d \rfloor \qquad l = (j+1) - \lfloor y_d \rfloor, \\ &s = x_d - \lfloor x_d \rfloor \qquad t = y_d - \lfloor y_d \rfloor \end{aligned} \tag{3}$$

where n is the number of the given data points. The local interpolation operator is a Kaier-windowed sinc function, defined as

$$I(x) = w(x)\,\mathrm{sinc}(x)\,, \tag{4}$$

$$\mathrm{sinc}(x) \equiv \sin(\pi x)/\pi x\,, \tag{5}$$

$$w(x) = \frac{I_0\left(\beta\sqrt{1-(x/r)^2}\right)}{I_0(\beta)}, \qquad -2 \le x \le 2 \tag{6}$$

where I_0 is the zero-order modified Bessel function of the first kind, and β is a parameter adjusting the shape of the Kaiser window. According to (2), due to the local nature of the interpolation operator $I(x)$, each data point contributes only to its proximity set, the 4 by 4 nodes represented by the blue rectangle in Figure 2b). Consequently, the resolution of the to-be-constructed interpolation function is determined by the grid size of the control lattice.

After the values at all the lattice nodes $\boldsymbol{\varphi_{ij}}$ are obtained, the continuous interpolation function can be evaluated as

$$\begin{aligned} &f(x,y) = \textstyle\sum_{k=0}^{3}\sum_{l=0}^{3} I(s)I(t)\varphi_{(i+k)(j+l)} \\ &i = \lfloor x \rfloor - 1, \quad j = \lfloor y \rfloor - 1, \quad s = x - \lfloor x \rfloor, \quad t = y - \lfloor y \rfloor \end{aligned}, \tag{7}$$

which can be used to predict the values at any specified location over the domain of investigation.

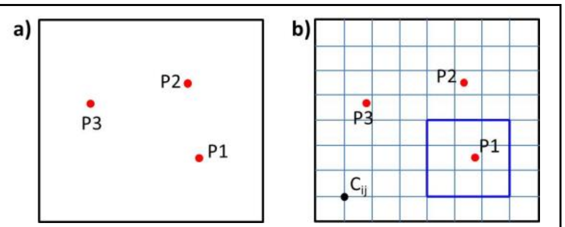

Figure 2. a) three data points for 2D scattered data interpolation; b) underling control lattice and the proximity dataset (blue rectangle).

According to the formulations, this algorithm propagates each data point value to the neighboring lattice nodes. The overall contribution from all the data points automatically constructs the continuous interpolation function. Because the interpolation stencil length is 4, each data point only contributes to the neighboring 16 lattice nodes called the proximity dataset, implying its high efficiency.

The main problem of this local interpolation is the non-uniqueness because scattered data interpolation is usually an ill-conditioned inverse problem. As shown in Figure 3a) and 3b), the interpolation results are highly dependent on the control lattice, although both results fit the given data perfectly. The reason is that the control lattice resolution determines the bandwidth of the interpolation function.

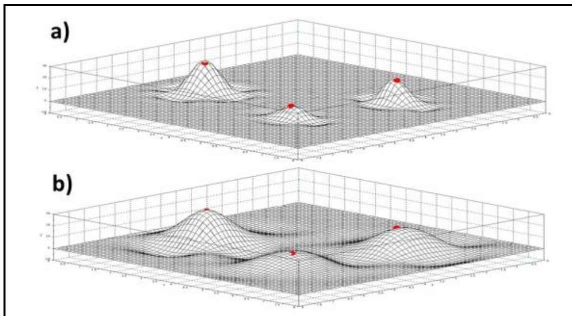

Figure 3. a) interpolation result with high resolution control lattice; b) interpolation result with low resolution control lattice.

Hybrid Interpolation and Extension to 5D

To resolve this resolution ambiguity, we propose the global-local-hybrid iterative interpolation strategy (Lee et al., 1997). With this strategy, the interpolation is implemented in an iterative manner, starting from a low resolution control lattice and gradually increasing the resolution with the iteration proceeds. At the first iteration, using equation (1)–(7), an approximate low resolution continuous function $f^1(x,y)$is obtained, which usually does not fit the data well. The interpolation residual for the first iteration $R^1 = z_d - f(x_d,y_d)$ is calculated. Then, in the second

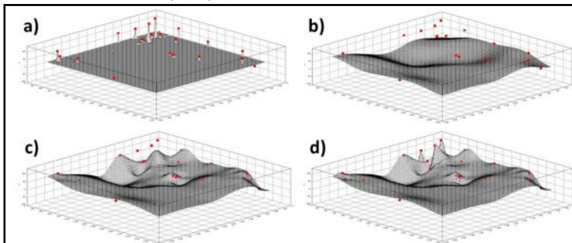

Figure 4. a) local interpolation; b) 1st iteration of hybrid interpolation; c) 3rd iteration of hybrid interpolation; d) final iteration of hybrid interpolation.

iteration, a control lattice with finer grids ($\Delta^2 = 0.5\ \Delta^1$) is employed. Similarly, with equation (1)–(7), we are able to obtain another continuous function $f^2(x,y)$ to approximate the residual R^1. The resolution of this function is doubled due to the finer control lattice grids. With the iteration proceeds, the accumulation of the continuous functions gradually matches the given dataset with higher accuracy, i.e,

$$f(x,y) = \textstyle\sum_{m=1}^{M} f^m(x,y)\,, \tag{8}$$

where M is the number of the iterations.

Figure 4 shows a numerical example to compare the performance of the local interpolation and the global-local-hybrid interpolation algorithm. The investigation domain size is 2560 m by 2560 m. In total, there are 32 data points scattered randomly in the domain. Figure 4a) shows the local interpolation result using a control lattice with the grid size Δ = 10 m. Figure 4b) to Figure 4d) are the evolution of the global-local-hybrid interpolation result. Figure 4b) is the 1st iteration result with Δ = 640 m with the interpolation residual as large as 86.0%; Figure 4c) is the 3rd iteration with Δ = 160 m and the residual of 55.3%. Figure 4d) is the final interpolation result with Δ = 10 m and the residual is 0.1%. Although the results in Figure 4a) and 4d) both successfully interpolate the given dataset with extremely small interpolation error, apparently, the global-local-hybrid interpolation result is more preferable due to its wider spatial spectral bandwidth.

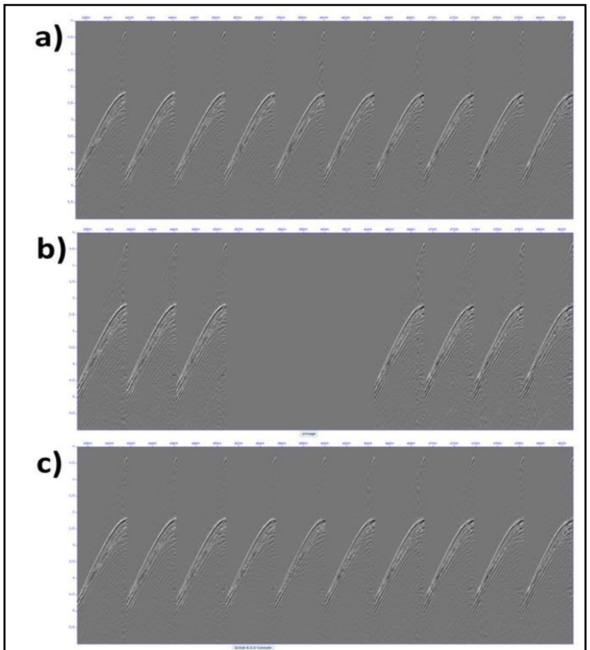

Figure 5. a) Original field data; b) the original data after 3 shots removed; c) reconstructed shot gather by the global-local-hybrid seismic data interpolation method.

The extension of the global-local-hybrid interpolation to 5D seismic data scenarios is straightforward by modifying equation (2) and (3) to

$$\varphi_d = \frac{w_d z_d}{\sum_{a=0}^{3}\sum_{b=0}^{3}\sum_{e=0}^{3}\sum_{f=0}^{3} w_{abef}^{2}}, \tag{9}$$

$$w_d = \prod_{n=1}^{4} I(\varsigma^n), \qquad \varsigma^n = \xi_d^n - \lfloor \xi_d^n \rfloor \tag{10}$$

The four dimensions ξ selected for spatial interpolation can be shot inline, crossline and receiver inline, crossline. Other schemes are also possible, such as inline, crossline, offset, and azimuth (Trad, 2009). It is worth mentioning that the interpolation shows better quality with NMO corrections.

A field data seismic data interpolation example is presented in Figure 5. Figure 5a) shows the original shot gathers. We deliberately remove 3 shots and the remaining shots gathers are shown in Figure 5b). Figure 5c) is the interpolation result showing that nearly full information of the missing shots is reconstructed accurately.

Conclusions and Discussions

We developed an efficient multi-dimensional seismic data interpolation algorithm. This interpolation is based on the idea of multi-resolution analysis and is posed as an iterative inverse problem. At the initial iteration, a low resolution interpolation function is built to roughly approximate the given dataset to resolve the low spatial spectral components without focusing much on accuracy. With the iteration proceeds, more and more fine features of the seismic data are brought into the interpolation procedure through the gradually refined control lattices, while the interpolation error reduces dramatically. With this iterative multi-resolution fashion, the broadband spatial spectral information of the seismic data is reconstructed effciently.

Compared with the widely used FFT-based method, there is no binning procedure needed in this algorithm and the spectral sparsity constraint is not required, which reduces the quality uncertainty caused by user interaction and experiences. Actually, the spectral sparsity constraint is natively embedded in the iterative dyadic control lattice strategy in this hybrid algorithm. Because this algorithm employs a local interpolation operator with the interpolation stencil length of 4, it is much more efficient than the global interpolation methods, especially the DFT-based method. On the other hand, the global information is taken into account through the iterative multi-resolution interpolation strategy.

Furthermore, the local interpolation operator employed by this hybrid interpolation algorithm is an optimally Kaiser-windowed sinc interpolation, whose spectral properties are quantitatively controllable, which ensures its anti-aliasing performance. Another important feature of this algorithm is that the initial control lattice resolution can be determined by the frequency bandwidth of the seismic data and the corresponding velocity model. The final control lattice is determined by the predefined interpolation error. These facts further reduce the computational cost and make this algorithm completely automatic. For these reasons, we believe that the global-local-hybrid interpolation method proposed in this work can be a powerful tool for seismic data processing.

References

Abma, R. and Kabir, N., 2006. 3D interpolation of irregular data with a POCS algorithm. Geophysics, 71(6), pp.E91-E97.

Hicks, G.J., 2002. Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. Geophysics, 67(1), pp.156-165.

Hollander, Y., Kosloff, D., Koren, Z. and Bartana, A., 2012, June. Seismic data interpolation by orthogonal matching pursuit. In 74th EAGE Conference and Exhibition incorporating EUROPEC 2012.

Hu, W. and Marcinkovich, C., 2012. A sensitivity controllable target-oriented tomography algorithm. In SEG Technical Program Expanded Abstracts 2012 (pp. 1-5). Society of Exploration Geophysicists.

Lee, S., Wolberg, G. and Shin, S.Y., 1997. Scattered data interpolation with multilevel B-splines. IEEE transactions on visualization and computer graphics, 3(3), pp.228-244.

Poole, G. and Herrmann, P., 2007, June. Multi-Dimensional Data Regularization for Modern Acquisition Geometries. In 69th EAGE Conference and Exhibition incorporating SPE EUROPEC 2007.

Trad, D., 2009. Five-dimensional interpolation: Recovering from acquisition constraints. Geophysics, 74(6), pp.V123-V132.

Xu, S., Zhang, Y., Pham, D. and Lambaré, G., 2005. Antileakage Fourier transform for seismic data regularization. Geophysics, 70(4), pp.V87-V95.

Zwartjes, P.M. and Sacchi, M.D., 2006. Fourier reconstruction of nonuniformly sampled, aliased seismic data. Geophysics, 72(1), pp.V21-V32.