Seismic Data Interpolation by Orthogonal Matching Pursuit

Y. Hollander* (Paradigm Geophysical), D. Kosloff (Paradigm Geophysical), Z. Koren (Paradigm Geophysical) & A. Bartana (Paradigm Geophysical)

SUMMARY

We present a multi dimensional interpolation method for the regularization of seismic data. The method operates in frequency slices where in each slice the data is represented by a Fourier expansion. The algorithm operates iteratively where in each step one Fourier component is selected from an overly redundant space. The coefficients of all previously selected Fourier components are recalculated at every step and in this the present algorithm differs from most algorithms presented in the geophysical literature. The regularization method is tested in two examples of pre stack data.
Introduction

Multi dimensional interpolation of data is a very important topic in exploration seismology, and a large number of publications devoted to the subject have appeared recently. Among the interpolation methods one algorithm which has been particularly successful is the Orthogonal Matching Pursuit (OMP) (Tropp and Gilbert, 2007; Kunis and Rauhut, 2008; Ozbek et al., 2009, 2011; Nguyen and Winnett, 2011; Xu et al., 2010). This method uses an expansion function set to represent the seismic data. The expansion coefficients are calculated one at a time by sequentially selecting the coefficient which minimizes the norm of the data misfit function. This approach usually generates a sparse representation of the seismic data and in certain situations can overcome spatial aliasing.

Although all the published OMP methods are based on the same principle, there have been differences in the actual implementation. Most of the methods described in the exploration geophysics literature sequentially select an additional expansion function and its coefficient while keeping previously calculated expansion coefficients unchanged. Conversely, published methods in the mathematical and electrical engineering literature (Tropp and Gilbert, 2007; Kunis and Rauhut, 2008) recalculate all the expansion coefficients after a new component has been added to the expansion. In the Anti Leakage Fourier Transform (ALFT) the input data are weighted by the area or volume which each input data point occupies (Xu et al., 2010). These differences can have a bearing on the effectiveness of the interpolation algorithm.

In this work we present a data regularization method based on recalculation of all the expansion coefficients at each step as in Tropp and Gilbert (2007), and Kunis and Rauhut (2008), and compare it to existing methods. The basis functions for the expansion are Fourier components which are oversampled in the wavenumber domain. These components form a redundant non-orthogonal expansion set from which the OMP algorithm selects a subset.

Interpolation Method

The interpolation is based on representing the input data by a Fourier expansion and using the expansion coefficients for data reconstruction at any desired spatial location. The method operates in the space-frequency domain in frequency slices and can be applied in 5, 4, 3, and 2 spatial dimensions, respectively. The expansion of offset like coordinates is carried out with the irregular discrete Fourier transform. In 4D and 5D interpolation, the expansion of CMP or shot coordinates uses binning and the regular DFT. For the sake of clarity, the explanation of interpolation in this section is for the 2D case.

Let \( f(x_j) \) \( (j = 0 \ldots N_x - 1) \) represent the temporarily transformed seismic data for a given frequency, where \( N_x \) is the number of seismic traces, and \( 0 \leq x_j < x_{\text{max}} \) is the spatial coordinate of each trace. The expansion of the input data reads,

\[
f(x_j) \approx \sum_l \hat{f}(l) e^{i 2 \pi l x_j / x_{\text{max}}} \tag{1}
\]

The backward transform writes,

\[
\hat{f}(l) = \sum_{j=0}^{N_x-1} f(x_j) e^{-i 2 \pi l x_j / x_{\text{max}}} \tag{2}
\]

The right hand side of these equations can be evaluated via the irregular FFT. The spatial wave numbers are related to the index \( l \) by, \( k_j = \frac{2 \pi l}{x_{\text{max}}} \ (l = -N_k/2 \ldots N_k/2 - 1) \), with \( N_k \) as the number of spatial frequencies. Increasing the value of \( x_{\text{max}} \) results in a finer sampling rate in the wave number domain and in a more redundant expansion space (assuming the maximum wave number is kept fixed). This in turn improves the interpolation (Xu et al., 2010). Typically \( x_{\text{max}} \) is about four time the largest offset in the data.

The OMP algorithm adds each time one component to the expansion (1). The residual after \( m \) steps
writes,

\[ R^m(x_j) = f(x_j) - \sum_{l=0}^{m-1} \tilde{f}(p_l)e^{i2\pi p_l x_j/x_{\text{max}}}, \quad (3) \]

where \( p_l \) is the index of the coefficient selected at the \( l' \)th step. The steps of the algorithm can be summarized as follows,

**Initialization:** Set \( m = 0 \), \( R^0(x_j) = f(x_j) \).

**Iterate:**

1. Calculate \( \tilde{R}^m(l) = \sum_{j=0}^{N_x-1} R^m(x_j)e^{i2\pi l x_j/x_{\text{max}}} \), and find \( p_m \), the index of the coefficient with the largest magnitude.

2. Re-calculate all coefficients by a least square minimization of

\[
\phi(m+1) = \sum_j \left[ f(x_j) - \sum_{l=0}^{m} \tilde{f}(p_l)e^{i2\pi p_l x_j/x_{\text{max}}} \right]^2.
\]

3. Calculate a new residual \( R^{m+1}(x_j) \) according to (3).

4. Stop when \( ||R^{m+1}|| \) becomes sufficiently small or the maximum number of coefficients allowed is reached.

The present scheme differs from most schemes published in the Geophysical literature by the addition of step 2. Step 2 is performed with the conjugate gradient (CG) method. This results in an increase in computational effort since the CG solution requires a number of forward and reverse irregular FFTs. However, the required number of CG iterations has been quite small (about 5-10 iterations). It is possible to reduce the cost by performing step 2 only every certain number of iterations.

**Examples**

**Devon Tempest data set**

The first example of the use of 5D regularization is the reconstruction of missing data in the Tempest data set (Devon Energy Corporation). The input consisted of a group of 3x3 CMP gathers around inline 1003 crossline 2001. Vertical time axis ranges between 0-14s, with a sampling rate of 16msec. The offset range in the gathers was \( \pm 25,000 \)ft in the inline direction and 6,000ft to 13,000ft in the crossline direction. In the test, approximately 30% of the input traces of the middle CMP were removed in a random manner. NMO was applied to the CMP gathers before the reconstruction and then removed after the reconstruction. The interpolation was carried out in vertical windows containing 40 sample where the data in each window was tapered.

Fig 1 plots the original input gather in black. Reconstruction of traces which were not used for calculating the OMP coefficients are plotted in red, on top of the known data at these locations. The insert in Fig 1 magnifies an area with 17 traces, out of which 8 were omitted during estimation, and reconstructed and compared with the known data. As the figure shows the missing trace reconstruction is quite good.

**Field data example**

This example compares 5D regularization of seismic data by the method of the present study, where all the expansion coefficients are updated at every step, to the method where each expansion coefficient is calculated only once.
Reconstruction of traces (red) at offset locations not used for OMP coefficient computations, compared with the known data at these locations (black).

The input contained a group of 3x3 CMP gathers with an offset range of 270-3350m. The distribution of the input offsets in the central CMP gather is plotted in red ‘+’ in Fig 2. The distribution of the output offsets is shown in the figure as green dots. The reconstruction was carried out in vertical windows of 50 samples size. The oversampling factors in the offset and CMP coordinates were 4 and 1, respectively.

Fig 3 compares the reconstruction of the data in the central CMP gather for the two methods. As the figure shows, the reconstruction with the present algorithm appears much better.

Conclusions

We have presented a data regularization algorithm by the Orthogonal Matching Pursuit method. The method operates iteratively, where in each step the largest coefficient from a Fourier expansion of the residual function is selected. After the coefficient selection, all the Fourier coefficients from previously selected components are recalculated by a least squares fit.

The examples presented have shown that the present method is effective in regularizing different types of seismic data. The new method is more expensive than published methods which do not recalculate previously selected coefficients in that it uses a least squares fit at every step. However, the added effort may be justified due to the improvement in the quality of the reconstruction.

Acknowledgements

Tempest original data set courtesy of Devon Energy Corporation. We thank Paradigm for the permission to publish this work.

References

Figure 2 Offset fold map. Input offset fold used for coefficient computation is signed by red '+', whereas regularized output grid is marked by green dots.

Figure 3 Data reconstruction on output grid described in Fig 2 using (a) OMP and (b) ALFT methods. In both methods 15 coefficients were used for every vertical frequency slice.