Seismic compressive sensing beyond aliasing using Bayesian feature learning

Georgios Pilikos*, University of Cambridge, A.C. Faul, University of Cambridge and Neil Philip, BP

SUMMARY

Sampling the seismic wave field and concurrently obtaining the true underlying signal is a challenging task due to environmental, economical and/or equipment limitations involved in seismic surveys. Under-sampling could alias the signal and Compressive Sensing methods use sparse assumptions to reconstruct it on a denser grid. It is assumed that various predefined dictionaries of basis functions provide a sparse representation of the seismic wave field. However, this is very limiting since different signals could contain different structures that require their own sparse representation. We propose to learn dictionaries of basis functions while interpolating with Beta Process Factor Analysis (BPF A). Comparisons with other solvers are undertaken, and learned basis functions are used by Spectral Projected Gradient for L1 (SPGL1) with the performance evaluated. Furthermore, we show that BPF A is able to reconstruct irregular under-sampled seismic signals without any signs of aliasing in the F/K domain. In addition, a feature space is obtained from millions of learned basis functions that could be used to decompose a seismic signal into features for various tasks in seismic data processing.

INTRODUCTION

Coarse spatial sampling of the seismic wave field often occurs in seismic surveys and aliasing is always a potential issue especially for some of the noise types. Because of this, methods that are able to properly interpolate potentially aliased data are of great interest - i.e. methods that will recover the wave field without being significantly affected by aliased events. One category of techniques utilizes the assumption that seismic data can be efficiently represented in some transform domain such as the Fourier (Abma and Kabir, 2006; Zwartjes and Sacchi, 2007), the Radon (Trad et al., 2002), the curvelet (Herrmann and Hennenfent, 2008; Naghizadeh and Sacchi, 2010) and the double focal transformation (Kutscha and Verschuur, 2016). These basis functions are used in conjunction with a solver to obtain a solution fitting the available measurements with SPGL1 (van den Berg and Friedlander, 2008; Naghizadeh and Sacchi, 2010) and the Projection Onto Convex Sets (POCS) (Abma and Kabir, 2006) on the other hand inserts zeros at the location of missing data points and operates in $\mathbb{R}^N$. It calculates iteratively $\hat{\mathbf{w}} = \mathbf{D}^{-1}\hat{\mathbf{x}}$ where $\hat{\mathbf{x}} \in \mathbb{R}^N$ with missing data points set to zero. $\hat{\mathbf{w}}$ is thresholded depending on the criterion used (Abma and Kabir, 2006; Stanton et al., 2015) and a new approximation $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{w}}$ is obtained at each iteration until the specified number of iterations.

Beta Process Factor Analysis (BPF A) also sets missing data to zero. The main difference is that $\mathbf{D}$ is learned by the data as opposed to being predefined. To do this, it divides the signal into many inner patches, $\mathbf{x}^{(l)}$, and uses them as training data. It iteratively finds $\mathbf{w}^{(l)}$ and $\mathbf{D}$ such that $\mathbf{x}^{(l)} = \mathbf{D}\mathbf{w}^{(l)}$. Further information about the optimization procedure, the tuning of parameters, comparisons with POCS and SPGL1 both in accuracy and in computational time along with denoising performances can be found in an upcoming Geophysics paper.

Learning a dictionary while performing interpolation uses training data that are corrupted and one might expect this is only useful for sparsely representing the corrupted signals. Nevertheless, this is not the case as examined in various models for feature learning and in seismic applications (Beckouche and Ma, 2014). In fact, dropping out measurements (Srivastava et al., 2014) or adding noise (Vincent et al., 2008) is recommended as a regularization to avoid overfitting. However, the percentage of measurements necessary for learning basis functions varies depending on the pattern of removal. Large consecutive gaps, in the case of entire traces missing in the x-t (shot record) domain, are a bigger problem than random missing data points in the time slice domain. With that in mind, we decided to operate on the time slice domain, remove receivers corresponding to data points randomly, reconstruct each time slice and then project the data to the x-t domain.

where $\mathbf{x} \in \mathbb{R}^N$ is the original signal from a time slice arranged in a vector, $\mathbf{D} \in \mathbb{R}^{N \times N}$ maps the sparse domain to the acquisition domain and its $l$-th column is the $l$-th basis function, $\mathbf{d}_l \in \mathbb{R}^N$ evaluated at all $N$ points. $\mathbf{w} \in \mathbb{R}^N$ are the sparse coefficients to be estimated. Compressive Sensing (CS) aims to reconstruct this signal using $M$ measurements where $M < N$ given by $\mathbf{y} = \Omega\mathbf{D}\mathbf{w}$. $\mathbf{y} \in \mathbb{R}^M$ is known as the collapsed signal and $\Omega \in \mathbb{R}^{M \times N}$ is the sensing matrix. SPGL1 uses this formulation to obtain an estimate, $\hat{\mathbf{w}}$, and checks $\mathbf{y}$ against $\Omega\hat{\mathbf{D}}\hat{\mathbf{w}}$. It will then refine $\hat{\mathbf{w}}$ until convergence.
SYNTHETIC DATA SET

We use a 3D synthetic data set generated numerically using the SEAM-II model as input. The modeling was carried out by BP in Houston. This data set is composed of one source with a 1281 × 1281 receiver grid and spatial sampling at 6.25 meters. The sampling rate in time is 6ms with a total of 500 time samples per trace. Therefore, 500 time slices were extracted and from each time slice, two sets of 10 sections of 128 × 128 (the last patch is 128 × 129). The sections were extracted near-offset and far-offset as shown in Figure 1 in order to test the reconstruction with different signal structures. The dark gray section illustrates the location of the examples in Figures 2 (127th shot line of section closest to source) and 3. To model irregular undersampling, we created three masks for the given section. The learned basis functions in Figure 3(i) resemble the orientations of the signal’s largest variations. By using these as basis functions, the SPGL1 in Figure 3(h) is able to perform significantly better compared to its DCT configuration irrespective of the patch size. We evaluate the reconstruction accuracy by the quality, Q, defined by

$$Q = 10 \log \frac{\|\hat{x}\|^2_2}{\|x - \hat{x}\|^2_2},$$

where $x$ is the original signal and $\hat{x}$ is the reconstruction.

Each section was processed independently which makes the entire proposed methodology highly suitable for parallel computation. In order to evaluate BPFA, we also performed experiments with POCs and SPGL1 with the Discrete Cosine Transform (DCT) as basis functions. An important clarification to make is that all algorithms could be adapted and their parameters could be tuned to work better. However, an essential aspect of all solvers is the dimension of the space that they work on within the 128 × 128 section. The larger the inner patch size (ie 8 × 8, 16 × 16, etc inside the 128 × 128 domain) the better accuracy but the longer running time. Figure 3(a) shows an example of a section from a time slice and then the same signal using only 30% of measurements randomly in Figure 3(b). It can be seen that when POCs in Figure 3(d) and SPGL1 in Figure 3(e) operate on the entire 128 × 128, they are much better than their 8 × 8 configuration in Figure 3(f) and Figure 3(g) respectively. However, BPFA in Figure 3(c) is better than both even though it operates on 8 × 8 patches. This is because it learns the appropriate dictionary of basis functions for the given section. The learned basis functions in Figure 3(i) resemble the orientations of the signal’s largest variations. By using these as basis functions, the SPGL1 in Figure 3(h) is able to perform significantly better compared to its DCT configuration irrespective of the patch size. We evaluate the reconstruction accuracy by the quality, Q, defined by

$$Q = 10 \log \frac{\|\hat{x}\|^2_2}{\|x - \hat{x}\|^2_2},$$

where $x$ is the original signal and $\hat{x}$ is the reconstruction.

Individual dictionaries for each section were learned near and far offset for all time slices for three different percentages resulting in 7 680 000 basis functions (30 000 sections with 256 basis functions each). To identify the signals’ largest variations over all instances and to obtain a feature space for this data set, we used PCA. We treated each basis function as a feature and then used PCA to reduce this feature space to only 64 basis functions to match our input space. Figure 4(a) shows this feature space as opposed to the DCT shown in Figure 4(b). The learned feature space preserves the curvature of seismic events better as opposed to the DCT which has sharper edges.

In addition, we calculated the mean Q over all 10 000 sections for all percentages and for the eighty inner-most shot records from the 128 × 1281 far-offset sections. Shot records at the ends of the sections are not reconstructed properly due to the lack of data points in their neighborhood. Overlapping reconstruction could help resolve this. Figures 5 and 6 illustrate that BPFA obtains better quality than the rest when they operate on an 8 × 8 patch size. The individual basis functions learned per section when used by SPGL1 provide much better performance as opposed to using DCT. The learned feature space

![Figure 1: Sections from the near-offset and far-offset were extracted over all time steps.](image)

![Figure 2: Reconstruction of the 127th shot line from part of the dark gray section in Figure 1 where 23 traces are used from 80 original traces.](image)
Beyond aliasing with feature learning

![Image of a figure showing various reconstructions and learned basis functions](image)

Figure 3: An example from the dark gray section in Figure 1 with the respective reconstructions and learned basis functions.

![Image of a graph showing mean Q for time slices with all configurations](image)

Figure 5: Mean $Q$ for time slices with all configurations.

![Image of a graph showing mean Q in x-t domain with all configurations](image)

Figure 6: Mean $Q$ in x-t domain with all configurations.

requires further improvement. It provides better quality than DCT however it is not as good as the individual learned basis functions. This is due to the fact that different orientations are not separated. An attempt to overcome this was made by creating 64 different dictionaries where each dictionary corresponds to two angles of the edge orientations $(0^\circ - 180^\circ)$. Edge detection was first performed per section and then the histogram of the orientations of the edges was obtained. The basis functions learned by that section were grouped with others of the same majority bin. PCA was then performed on each group resulting in 64 dictionaries. When reconstructing, the corresponding dictionary was chosen to be used by SPGL1.

Another way to evaluate a method is the ability to not only predict the missing traces but also reconstruct the respective signal without any signs of aliasing in the F/K domain. Figure 7 shows an example of this using BPFA. The F/K domain of the reconstruction in Figure 7(g) from 30% of receivers does not show any signs of aliasing and the reconstruction error in Figure 7(d) is minimal. Near-offset signals exhibit sharper structures which are harder to interpolate beyond aliasing with further research required on larger patch sizes.

### FIELD DATA SET

The Parihaka data set is a 3D seismic image provided for use by New Zealand Petroleum and Minerals (NZPM) and obtained from the SEG wiki website (SEG wiki, 2017). Figure 8(a) shows an entire time slice with BPFA reconstruction in Figure 8(c) from only 30% of receivers. The reconstruction was obtained by splitting the time slice into smaller sections $(128 \times 128, 128 \times 154$ for the upper right part, $230 \times 128$ for the lower part and $230 \times 154$ for the right lower corner). From each section, individual basis functions were learned and used.
Beyond aliasing with feature learning

Figure 7: An example of a shot record (a) with its F/K domain (c). By using only 30% of the traces (b), BPFA is able to reconstruct it (c) and obtain an alias free (g) signal with minimal error (d).

Figure 8: BPFA reconstruction of a time slice at $t = 923$ from the field data set of Parihaka.

Using PCA as before, the feature space in Figure 8(e) was obtained showing similarities with the SEAM synthetic data set.

CONCLUSIONS

Learning basis functions at the same time as interpolating does not limit the solver to a particular basis. By exploring all possible spaces, BPFA is able to learn a sparse representation that captures the signal variations of a seismic signal and provides higher quality of reconstruction compared to other solvers with predefined basis functions. It is also able to reconstruct undersampled seismic signals in denser grids without any signs of aliasing in the F/K domain. Experiments on synthetic and on field data sets illustrate the effectiveness of BPFA. A feature space from millions of basis functions was also learned as well as other dictionaries based on their histogram of edge orientations. Further work is required to improve this for larger patch sizes and near-offset signals. This could be used in various applications where the decomposition of seismic signals is essential, such as automatic classification, compression, denoising, and create new avenues for seismic data processing.

ACKNOWLEDGMENTS

We thank BP for permitting this publication and especially to Ray Abma of BP for his technical input. We also thank SEAM-II for the model and Carl Regone of BP for modelling the seismic data and thank New Zealand Petroleum and Minerals (NZPM) for providing the Parihaka field data set. We acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC) with studentship [1502944].
REFERENCES


