Beta Process Factor Analysis for efficient seismic Compressive Sensing with uncertainty quantification

Georgios Pilikos
Laboratory for Scientific Computing, Maxwell Centre
Department of Physics, University of Cambridge
Cambridge, United Kingdom, CB3 0HE
Email: ggp29@cam.ac.uk

Neil Philip
BP Exploration
Chertsey Road, Sunbury-on-Thames
United Kingdom, TW16 7LN
Email: Neil.Philip@uk.bp.com

***Abstract*—Compressive Sensing for seismic surveys uses sparse signal assumptions to reconstruct the reflected wave field. In the past, most methods utilised dictionaries of fixed basis functions for sparse representation. Recently, algorithms that learn the basis from data are being used with better reconstruction accuracy but longer computational times. One of these is the Beta Process Factor Analysis (BPFA). We propose faster inference for BPFA using Gibbs sampling analysis and illustrate that the reduced computational time does not severely affect the reconstruction accuracy with no aliasing in the frequency spectrum. In addition, associating each prediction with a level of uncertainty is essential but very challenging. Using the Gibbs samples, we create uncertainty maps that are highly correlated with the reconstruction error and are not affected by the faster BPFA inference. Experiments on synthetic and field data illustrate the effectiveness of our proposed methodology for both reconstruction and uncertainty quantification in seismic surveys.**

***Index Terms*—seismic compressive sensing, dictionary learning, sparse representation, Bayesian machine learning**

I. Introduction

Seismic surveys involve an artificial source at the Earth's surface creating a wave field that is reflected by changes in impedance. Receivers at the surface are placed in a certain spatial arrangement and record the reflected wave field. This has the effect of sampling an analogue into a digital signal and care is needed to avoid aliasing. In addition, restrictions on the surface, malfunction of receivers or other environmental reasons can cause gaps in the acquired seismic data.

Compressive Sensing (CS) [1], [2] is a framework that tackles missing data using signal reconstruction algorithms. An important assumption for CS is that the signal of interest is sparse [3], [4] in the acquired domain or in some basis. Recently, there has been a great interest in CS methods [5]-[7]. For seismic CS, Fourier [8], discrete cosine [9], curvelet [10], Radon [11] or focal dictionaries [12] are used for sparse representation. Algorithms that use dictionaries of bases such as Projection Onto Convex Sets (POCS) [13], [14], Spectral Projected Gradient for L1 (SPGL1) [15] and the Relevance Vector Machine (RVM) [16] are limited by the pre-selection of the basis. To overcome this limitation, dictionary learning methods [17]-[21] have been proposed to learn a sparse representation.

However, most algorithms are ad hoc and their only aim is to fill in the gaps or denoise a signal with their predictions. They do not provide any degree of uncertainty or confidence for the predicted receivers' values. From the methods mentioned earlier, the RVM is able to provide uncertainty information [9] due to the fact that is built around the Bayesian framework for CS [22]-[25]. Another Bayesian technique, the Beta Process Factor Analysis (BPFA) [26] has been proposed for seismic CS and denoising [27]. It is able to obtain state-of-the-art results by learning dictionaries of bases without any aliasing in the Frequency-Wavenumber (FK) domain [28]. Nevertheless, it has been shown that its computational time is much slower compared to others in the literature [27]. Since it is also built around the Bayesian framework, it is possible to create uncertainty maps for its predictions.

In this paper, our contributions are the following: we propose faster BPFA inference using Gibbs sampling analysis. We apply this for seismic Compressive Sensing (CS) and uncertainty quantification. Using this, we create state-of-the-art uncertainty maps with high correlation between uncertainties and reconstruction error and no aliasing in the FK domain. We compare against other techniques in the literature obtaining more informative uncertainty maps. We also investigate the effect that the reduced computational time has on accuracy and uncertainty. All these are applied on synthetic data provided by BP and field data provided by New Zealand Petroleum and Minerals (NZPM). The rest of the paper is organised as follows: First we introduce the BPFA model, then we describe how we can obtain uncertainty maps and after that we describe the Gibbs sampling analysis using various initialisations. Then, the fast BPFA inference is described along with results and comparisons on various data sets.

II. The BPFA model

BPFA is a hierarchical Bayesian model that constitutes a finite approximation of the Indian Buffet Process (IBP) [29]. It is a truncated beta-Bernoulli process with a fixed, large number of features, L, and shrinks if there is redundancy. To introduce this model, we assume that a data matrix $\mathbf{X}$ is generated by an underlying process with its columns $\mathbf{x}^{(i)} \in \mathbb{R}^K$, $i = 1, ..., T$. Each $\mathbf{x}^{(i)}$ can be considered a patch from a signal and is assumed to be generated by,

$$\mathbf{x}^{(i)} = \mathbf{D}\mathbf{w}^{(i)} + \boldsymbol{\epsilon}^{(i)}, \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{K \times L}$ denotes the dictionary of bases. Prior distributions are defined for each model variable. We start with the columns $\{\mathbf{d}_l\}_{l=1}^{L}$ of $\mathbf{D}$ which are modelled by

$$p(\mathbf{d}_l) = \mathcal{N}(\mathbf{d}_l; 0, K^{-1}\mathbf{I}_K) \tag{2}$$

where $\mathbf{I}_K \in \mathbb{R}^{K \times K}$ is the identity matrix. Then, we explicitly separate the value of a coefficient in $\mathbf{w}^{(i)}$ from the fact whether it is non-zero or zero. In particular, we introduce $\mathbf{z}^{(i)}$ and $\mathbf{s}^{(i)}$,

$$\mathbf{w}^{(i)} = \mathbf{z}^{(i)} \odot \mathbf{s}^{(i)}, \tag{3}$$

with $\odot$ the elementwise vector product, $\mathbf{z}^{(i)} \in \{0,1\}^L$ signifies whether a basis is used and $\mathbf{s}^{(i)} \in \mathbb{R}^L$ are the values of the coefficients. The prior distribution for $\mathbf{z}^{(i)}$ is given by,

$$p(\mathbf{z}^{(i)}) = \prod_{l=1}^{L} \text{Bernoulli}(z_l^{(i)}; \pi_l), \tag{4}$$

where π_l is the probability that the l-th basis is used when $\mathbf{x}^{(i)}$ is generated. This means that the l-th component of $\mathbf{z}^{(i)}$ is generated by a Bernoulli distribution with probability π_l. The probabilities $\boldsymbol{\pi} = [\pi_1, ..., \pi_L]^T$ themselves are a priori distributed by a hyper-prior defined by,

$$p(\boldsymbol{\pi}) = \prod_{l=1}^{L} \text{Beta}(\pi_l; a/L, b(L-1)/L), \tag{5}$$

where a, b are parameters characterising the Beta distribution.

The latent variable, $\mathbf{s}^{(i)}$, on the other hand models the value of the coefficients and is assumed to be generated by,

$$p(\mathbf{s}^{(i)}) = \mathcal{N}(\mathbf{s}^{(i)}; 0, \gamma_s^{-1}\mathbf{I}_L), \tag{6}$$

where $\mathbf{I}_L$ is the $L \times L$ identity matrix and γ_s is modelled by a Gamma distribution. The noise in equation 1 is also modelled with a normal distribution. Using the above, we can obtain a joint distribution for the BPFA model over all the available data. This is given by,

$$P(\mathbf{D}, \mathbf{Z}, \mathbf{S}, \boldsymbol{\pi}, \gamma_\epsilon, \gamma_s, \mathbf{Y}, \boldsymbol{\Delta}) = \prod_{i=1}^{T} \mathcal{N}(\mathbf{y}^{(i)}; \boldsymbol{\Delta}^{(i)}\mathbf{D}(\mathbf{s}^{(i)} \odot \mathbf{z}^{(i)}), \gamma_\epsilon^{-1}\mathbf{I}_{\|\boldsymbol{\Delta}^{(i)}\|_0}) \mathcal{N}(\mathbf{s}^{(i)}; 0, \gamma_s^{-1}\mathbf{I}_L) \tag{7}$$

$$\prod_{l=1}^{L} \mathcal{N}(\mathbf{d}_l; 0, K^{-1}\mathbf{I}_K) \text{Beta}(\pi_l; a_0, b_0) \prod_{i=1}^{T}\prod_{l=1}^{L} \text{Bernoulli}(z_l^{(i)}; \pi_l)$$

$$\Gamma(\gamma_s; c_0, d_0)\Gamma(\gamma_\epsilon; e_0, f_0).$$

where we define $\mathbf{Y} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(T)}]$ and $\boldsymbol{\Delta} = [\boldsymbol{\Delta}^{(1)}\boldsymbol{\Delta}^{(2)}, ..., \boldsymbol{\Delta}^{(T)}]$ the available training data and the sampling matrix per patch respectively. $\mathbf{Z} = [\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, ..., \mathbf{z}^{(T)}]$, $\mathbf{S} = [\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, ..., \mathbf{s}^{(T)}]$ are the respective latent variables.

Posterior distributions for each variable are obtained using the Bayes' rule. Using these, it is possible to create an algorithm for inference by estimating the model's variables considering the rest fixed for a given iteration (with the posterior conditional distributions). This is called a Gibbs iteration and the entire procedure of obtaining an estimate for one variable given the others is called Gibbs sampling. Further information can be found in the Appendix of [26].

III. Producing uncertainty maps with BPFA

In order to obtain training data for the BPFA, we split a signal into smaller overlapping patches, $\mathbf{x}^{(i)}$, and use them in a sequential manner. We chose a patch size of 8×8 in a grid of 128×128 receivers. First, we extract all 256 patches with $\frac{128}{8} = 16$ along the vertical axis and $\frac{128}{8} = 16$ patches along the horizontal axis. This is the first Gibbs round and each round can have many Gibbs iterations. In the second round, we shift the starting point of extraction by one receiver down which results in the extraction of 240 patches. This continues for all receiver locations in a given patch resulting in 64 such rounds. We use the current extracted training data to perform Gibbs sampling over the unknown variables. More patches are added sequentially until finally all patches are used.

At the last Gibbs iteration, each variable is drawn from its corresponding distribution and used to calculate the receivers' value for all patches. Thus, each value is inferred various times since it is contained in multiple patches (at most 64). The mean (final prediction) of each receiver's value is obtained by averaging over all its estimated values. The uncertainty of the prediction at a receiver's location is obtained by calculating the variance of all its estimated values. [30] contains more information regarding the patch processing procedure.

IV. Proposed analysis of the Gibbs sampler

Before starting the inference for the Gibbs sampling, we need to initialise all unknown variables. By doing so, we place the sampler in a location at the variable's space. One of the most important variables in the BPFA model is $\mathbf{D}$ as defined in equation 1 and we will investigate its initialisation. We will evaluate six different options. The first three are popular dictionaries of basis functions: the Discrete Cosine Transform (DCT), the Haar wavelets transform and the radial (Gaussian) basis functions. Another option for initialisation is the Singular Value Decomposition (SVD) which decomposes the available data into singular vectors that capture the largest variances in the data. A further option is a dictionary that was inferred by using 7 680 000 bases, from 30 000 seismic signals with 256 bases each, learned during previous reconstructions [28]. Finally, random initialisation is also investigated.

We would like to investigate how the Gibbs samplers are affected and we thus track how the reconstruction accuracy changes over time using Q defined by

$$Q = 10 \log \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}, \tag{8}$$

where $\mathbf{x}$ is the original signal and $\hat{\mathbf{x}}$ is the current reconstruction. On the other hand, the uncertainty maps are judged by their correlation with the reconstruction error. A suitable metric for this is the Spearman's correlation coefficient, S, defined by,

$$\text{S} = \frac{\sum_{i=1}^{n}(\text{error}_i - \mu_{\text{error}})(\text{uncer}_i - \mu_{\text{uncer}})}{\sqrt{\sum_{i=1}^{n}(\text{error}_i - \mu_{\text{error}})^2}\sqrt{\sum_{i=1}^{n}(\text{uncer}_i - \mu_{\text{uncer}})^2}} \tag{9}$$

where n are the number of receiver locations used, $\{\text{error}_i\}_{i=1}^{n}$ and $\{\text{uncer}_i\}_{i=1}^{n}$ are the reconstruction error and uncertainty

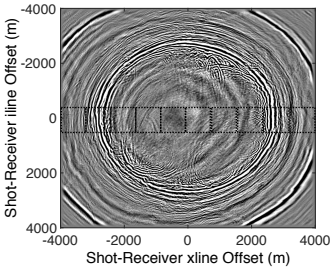


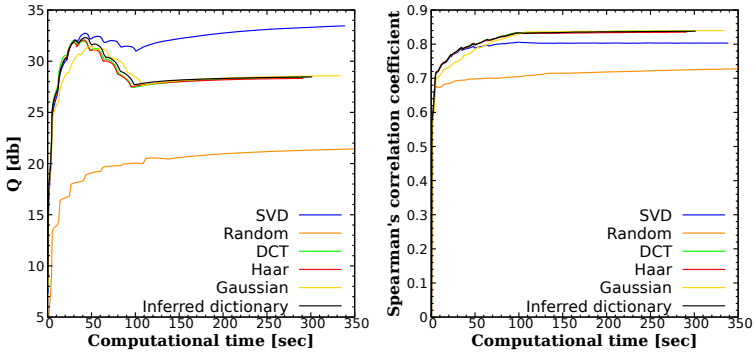
Fig. 1: A time slice illustration - Extraction of 10 sections.

at the respective receiver locations. μ_{error} and μ_{uncer} are their respective means. Note that each variable is ranked from lowest to largest in magnitude and S ranges from -1 to $+1$. Positive correlation means that as one grows or decreases so does the other. We would like a value close to $+1$.

To test our proposed system, we extracted seismic data from a 3D synthetic data set which was generated numerically using the SEAM-II [31] model as input. The modelling was carried out by BP in Houston. We process seismic data in sections of time slices as it was done in [28]. Figure 1 shows a time slice where all receivers in a grid are used at a particular time step. Then, from a time slice, ten sections are extracted of 128×128 receivers in order to operate on. The spatial sampling is at 6.25 metres and the time sampling is at 0.006 seconds. To test aliasing, the x-t domain is transformed into its Frequency-Wavenumber (FK) domain. The x-t domain uses a receiver line with all time steps as it can be seen in Figure 3(a). All experiments were performed as single-core jobs on machines with Intel(R) Xeon(R) CPU E5-2650 with 2.00GHz using the software package from [32].

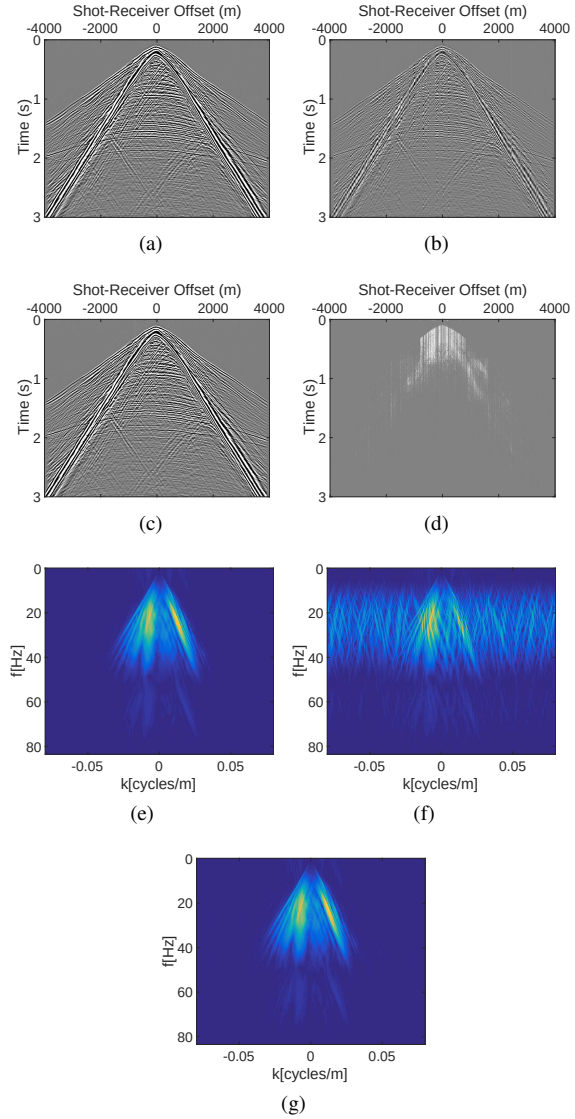
To analyse the Gibbs samplers, we used 21 different seismic signals and plotted the mean Q and mean S at each time for each different initialisation of basis. Figure 2(a) shows the average Q against the average time. All Gibbs samplers exhibit similar behaviour. At the beginning of the inference, Q rapidly increases with the extraction of patches. Every 8 Gibbs rounds, there is a rapid change in Q when a horizontal shift in the extraction of patches occurs. It increases Q at the start of the inference but decreases it afterwards. This is because more training data are being used, adjusting the model. When 63 Gibbs rounds are completed (with one iteration per round), the 64th starts (with 100 iterations) with all available patches as training data. This results in a better estimation and Q gradually increases. From the various initialisations, the Gibbs sampler with SVD performs much better. The inferred dictionary from 30000 sections peaks near the former's performance. The three dictionaries of basis functions perform similarly, but the Gaussian basis functions peak slightly lower. Finally, the random initialisation performs the worst.

Figure 2(b) shows the average S against the average time. There is a rapid change in S every 8 Gibbs rounds but much smaller compared to Q. In addition, S does not worsen with time and slightly improves. The random initialisation provides the worst uncertainty maps compared to the others.



(a) (b)

Fig. 2: Mean reconstruction quality (a) and mean Spearman's correlation coefficient (b) against computational time.



(a) (b) (c) (d) (e) (f) (g)

Fig. 3: Original (a), using 50% (b), the BPFA reconstruction (c), the error (d), the FK domain of original (e), the FK domain for 50% (f) and the FK domain of reconstruction (g).

V. Fast BPFA inference

BPFA is orders of magnitude slower than others in the literature [27]. If we consider that there are potentially thousands of instances of the BPFA that need to be executed in a complete seismic signal, reducing the processing time can have great computational improvements in general.

By using the insight from the Gibbs sampling analysis, we can speed up the BPFA procedure depending on the requirements of reconstruction. From Figure 2, we can see that Q does not improve significantly at the last Gibbs round. Thus, not all 100 iterations are necessary. In addition, S is not affected significantly by increasing the iterations. In the following experiment, we perform slice processing and then re-sort in the x-t domain. We use all 64 Gibbs rounds to reach high Q but stop the iterations at 50 with SVD as initialisation since it obtains the best Q. This allows a speed up of approximately 120 seconds (or 2 minutes) per section and only approximately 0.5 db loss with S only 10^{-5} smaller. We are interested in the reconstruction of entire seismic signals which are composed of numerous sections. In particular, in our experiment we use 5000 sections of time slices (10 sections per time slice with 500 time steps) which results in 10000 minutes or 166.6 hours of speed up. Figure 3 shows a receiver line of a seismic signal in the x-t domain. We used 50% of the receivers and reconstructed the signal with no signs of aliasing in the FK domain illustrating that the significant speed up does not compromise the quality of the reconstruction.

VI. Comparisons for uncertainty maps

We wanted to test our system against others in the literature that produce uncertainty maps. The Relevance Vector Machine (RVM) is one such algorithm that was applied to seismic data and produced uncertainty maps [9]. It was found that the predictive variance of the RVM does not always provide accurate uncertainties [33] and thus an improvement was proposed using the expected change in the model's likelihood [9]. We used 50% of the receivers from 1000 sections of time slices, reconstructed the signals and produced respective uncertainty maps. Then, we calculated the Spearman's correlation coefficient, S, for our system, the RVM's predictive variance and its extension [9]. We found that our system using the fast BPFA obtained an average S$= +0.5248$, the RVM's predictive variance obtained an average S$= +0.2109$ and its extension an average S$= +0.3775$. This illustrates that our system produces more accurate uncertainty maps even with the fast inference proposal. [9] contains illustrative uncertainty maps for the RVM and its extension. A more thorough comparison with more percentages and analysis is not included for brevity.

VII. Field data set

We use a field data set called Parihaka which is a 3D seismic image provided for use by New Zealand Petroleum and Minerals (NZPM) obtained from the SEG wiki website [34]. Figure 4 shows a section of this data set with its respective reconstruction and the learned dictionary of bases. It can be seen that the reconstruction from 50% of receivers matches

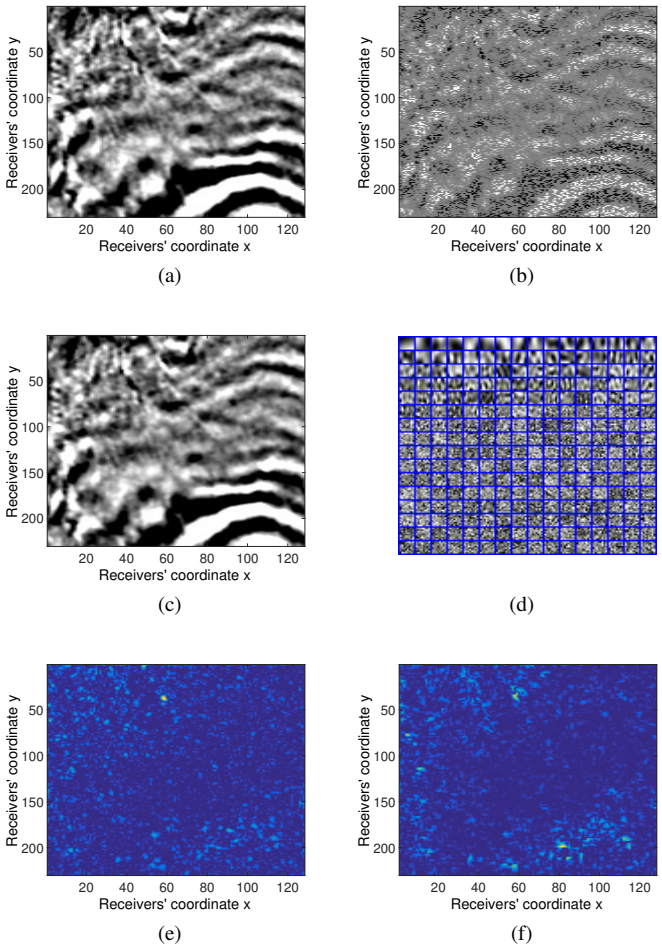


Fig. 4: Section from field data with original (a), using 50% of receivers (b), BPFA reconstruction with Q = 26.1 db (c), the learned dictionary of bases (d) the error (e) and the uncertainty map with $S = +0.82$ (f). Bright colours mean high error and uncertainty in (e) and (f) respectively.

well with the original signal with Q = 26.1 db using the learned bases. Furthermore, we can see that the uncertainty map produced by the BPFA has high correlation with the reconstruction error (S = $+0.82$). Using the uncertainty maps, under certain conditions, it could be possible to place receivers where there is high uncertainty and minimise reconstruction error in future surveys.

VIII. Conclusion

Efficient seismic Compressive Sensing (CS) techniques are crucial in seismic surveys. Fast and accurate reconstructions without any aliasing along with uncertainty information are desirable. These allow the utilisation of seismic images for further processing. We proposed a fast BPFA inference procedure using Gibbs sampling analysis. We illustrated that shorter computational time does not signifcantly compromise the reconstruction accuracy nor the quality of the uncertainty maps. In addition, thousands of sections were reconstructed

using the fast BPFA inference without any aliasing in the FK domain and at the same time saving hundreds of hours of computation. Furthermore, a comprehensive study of the Gibbs sampler initialisation showed that the SVD of the available data provides the best starting place in the basis space. This allows the learning of dictionaries of bases for sparse representation with higher reconstruction accuracy. Initialisation with an inferred dictionary of bases trained on thousands of sections also provided improved performance.

The importance of uncertainty quantification in data acquisition is also growing. We utilised the probabilistic nature of the BPFA and produced accurate uncertainty maps having high correlation with their respective reconstruction error. We compared this against others in the literature and obtained more informative uncertainty maps for the BPFA model illustrating its suitability for both reconstruction accuracy and uncertainty quantification in seismic surveys.

ACKNOWLEDGMENT

We thank BP for permitting this publication. We also thank Anita Faul for her contribution to this work. In addition, we thank Ray Abma of BP and Nikos Nikiforakis of University of Cambridge for 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

[1] E. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[3] H. Liu, Y. Li, Y. Zhou, and T. K. Truong, "Greedy pursuit algorithms for sparse signal reconstruction in the case of impulsive noise," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, Oct 2016, pp. 705–709.

[4] R. Song, L. Chen, and Y. Gu, "Performance estimation of sparse signal recovery under bernoulli random projection with oracle information," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, Oct 2016, pp. 695–699.

[5] X. Li and G. Bi, "Image reconstruction based on the improved compressive sensing algorithm," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 357–360.

[6] Y. Wang, S. Feng, and P. Zhang, "Information estimations and acquisition costs for quantized compressive sensing," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 229–233.

[7] F. Li, T. J. Cornwell, F. de Hoog, L. Xin, and Y. Guo, "Compressive sensing based multi-frequency synthesis," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, Oct 2016, pp. 710–714.

[8] M. Sacchi, T. Ulrych, and C. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 31–38, Jan 1998.

[9] G. Pilikos and A. C. Faul, "Relevance vector machines with uncertainty measure for seismic bayesian compressive sensing and survey design," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2016, pp. 925–930.

[10] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, no. 1, pp. 233–248, 2008.

[11] D. O. Trad, T. J. Ulrych, and M. D. Sacchi, "Accurate interpolation with high-resolution time-variant Radon transforms," *Geophysics*, vol. 67, no. 2, pp. 644–656, 2002.

[12] H. Kutscha and D. Verschuur, "The utilization of the double focal transformation for sparse data representation and data reconstruction," *Geophysical Prospecting*, vol. 64, no. 6, pp. 1498–1515, 2016.

[13] R. Abma and N. Kabir, "3D interpolation of irregular data with a POCS algorithm," *Geophysics*, vol. 71, no. 6, pp. E91–E97, 2006.

[14] A. Stanton, M. D. Sacchi, R. Abma, and J. A. Stein, "Mitigating artifacts in Projection Onto Convex Sets interpolation," *85th Annual International Meeting, SEG, Expanded Abstracts, 3779-3783*, 2015.

[15] E. van den Berg and M. P. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.

[16] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[17] T. Variddhisaï and D. P. Mandic, "Online multilinear dictionary learning for sequential compressive sensing," *CoRR*, vol. abs/1703.02492, 2017. [Online]. Available: http://arxiv.org/abs/1703.02492

[18] S. Beckouche and J. Ma, "Simultaneous dictionary learning and denoising for seismic data," *Geophysics*, vol. 79, no. 3, pp. A27–A31, 2014.

[19] L. Zhu, E. Liu, and J. H. McClellan, "Seismic data denoising through multiscale and sparsity-promoting dictionary learning," *Geophysics*, vol. 80, no. 6, pp. WD45–WD57, 2015.

[20] P. Turquais, E. G. Asgedom, W. Sollner, and E. Otnes, "Dictionary learning for signal-to-noise ratio enhancement," *85th Annual International Meeting, SEG, Expanded Abstracts, 4698-4702*, 2015.

[21] S. K. Sahoo and A. Makur, "Replacing k-svd with sgk: Dictionary training for sparse representation of images," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 614–617.

[22] E. Karseras, K. Leung, and W. Dai, "Bayesian compressed sensing: Improving inference," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, July 2013, pp. 365–369.

[23] Q. Wu, Y. D. Zhang, M. G. Amin, and F. Ahmad, "Autofocus bayesian compressive sensing for multipath exploitation in urban sensing," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 80–84.

[24] Q. Wu, Y. D. Zhang, and M. G. Amin, "Continuous structure based bayesian compressive sensing for sparse reconstruction of time-frequency distributions," in *2014 19th International Conference on Digital Signal Processing*, Aug 2014, pp. 831–836.

[25] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[26] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 130–144, Jan 2012.

[27] G. Pilikos and A. C. Faul, "Bayesian feature learning for seismic compressive sensing and denoising," *GEOPHYSICS*, vol. 82, no. 6, pp. O91–O104, 2017.

[28] G. Pilikos, A. Faul, and N. Philip, "Seismic compressive sensing beyond aliasing using bayesian feature learning," in *SEG Technical Program Expanded Abstracts 2017*, 2017, pp. 4328–4332.

[29] T. L. Griffiths and Z. Ghahramani, "The indian buffet process: An introduction and review," *J. Mach. Learn. Res.*, vol. 12, pp. 1185–1224, Jul. 2011.

[30] M. Zhou, H. Chen, L. Ren, G. Sapiro, L. Carin, and J. W. Paisley, "Non-parametric bayesian dictionary learning for sparse image representations," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., 2009, pp. 2295–2303.

[31] M. Oristaglio, "Seam phase iiland seismic challenges," *The Leading Edge*, vol. 31, no. 3, pp. 264–266, 2012.

[32] M. Zhou, "Beta process factor analysis: software," *last accessed 6 April 2018*. [Online]. Available: http://mingyuanzhou.github.io/Code.html

[33] C. E. Rasmussen and J. Quionero Candela, "Healing the relevance vector machine through augmentation," in *Proceedings of the 22nd International Conference on Machine Learning*, January 2005, pp. 689–696.

[34] SEG, "SEG wiki open access data," *last accessed 6 April 2018*. [Online]. Available: http://wiki.seg.org/wiki/Parihaka-3D