geoconvention
Calgary • Canada • May 13-17 2019

A Non-Parametric Bayesian dictionary learning Method for SNR enhancement

Chao Zhang and Mirko van der Baan
Department of Physics, University of Alberta

Summary

The signal-to-noise ratio (SNR) enhancement plays the key role for subsequent seismic signal processing and interpretation. In this article, we propose an efficient and appropriate dictionary learning strategy to realize denoising. It learns features (basis functions) for seismic signals from a smaller portion data by an unsupervised non-parametric Bayesian method. In this dictionary learning strategy, an Indian Buffet Process (IBP) prior is used to promote sparsity and infer an adaptive number of atoms. So, the size of dictionary is no more set a prior. There is no parameter adjustment in the process of dictionary learning thanks to the non-parametric Bayesian method, which makes it easier to be applied. Tests on synthetic and real datasets demonstrate that this proposed strategy achieves superior denoising performance in terms of SNR enhancement.

Introduction

Sparse representation has shown to be a very powerful model for signal denoising, recently it has drawn considerable attention in seismic data processing (Ma et al., 2010; Zhu et al., 2015; Chen, et al., 2016). The basic assumption of this model is that seismic signals can be represented if combined with an appropriate dictionary. Specifically, solely a few columns (also referred to as atoms) of the dictionary capture most signal features. Most algorithms formulate the dictionary learning as an optimization problem (Engan et al., 1999; Aharon et al., 2006; Siahsar et al., 2017; Chen, 2017; Hou, et al., 2017). They seek to match the dictionary to the signal of interest and simultaneously encourage a sparse representation. These methods have achieved good performance for denoising, but they still have several limitations. They may require the knowledge of the noise level, the size of the dictionary or the sparsity level. To mitigate these limitations, non-parametric Bayesian approaches have been introduced (Zhou et al., 2012; Zhang and Van der Baan, 2018). All involved parameters and their distributions are estimated from the data using a Bayesian framework. In this article, we use a non-parametric Bayesian dictionary learning method by employing the Indian Buffet Process (IBP, Griffiths and Ghahramani, 2006) for sparse representation of seismic signals. An IBP prior is adopted to infer the size of dictionary. The IBP defines a probability distribution of sparse binary matrices with a potentially infinite number of rows, which is the key to learning a dictionary for sparse representation with adaptive size. An approximation to the posterior distributions is manifested via Gibbs sampling (Albert and Chib, 1993; Dang and Chainais, 2018), yielding an ensemble of dictionary representations. In this non-parametric Bayesian method, the dictionary is directly learned from underlying noisy data. Furthermore, the noise variance need not be known, and the noise can be non-stationary. These merits are deemed an important advantage over other dictionary learning methods.

Method

Dictionary learning consists of estimating features (basis functions) from observed data by invoking sparsity. In dictionary learning, the noisy data $\mathbf{Y}$ are modeled by:

$$\begin{cases} \mathbf{Y} = \mathbf{X} + \varepsilon \\ \mathbf{X} = \mathbf{D}\,\mathbf{W} \end{cases}, \tag{1}$$

where $\mathbf{Y} = \left[\mathbf{y}_1, ..., \mathbf{y}_N\right] \in \mathbb{R}^{P \times N}$ is a set of N observations $\mathbf{y}_i$, $\mathbf{y}_i$ is a column vector of dimension P ($\mathbf{y}_i \in \mathbb{R}^P$), whose elements are taken column-wise from a square patch (image) of size $\sqrt{P} \times \sqrt{P}$. The patches are extracted at regular interval from observed data in the horizontal and vertical directions. $\mathbf{X} = \left[\mathbf{x}_1, ..., \mathbf{x}_N\right] \in \mathbb{R}^{P \times N}$ is a set of $\mathbf{x}_i$ ($\mathbf{x}_i \in \mathbb{R}^P$), the elements of $\mathbf{x}_i$ are taken column-wise from a square patch of the original pure 2D signal and $\boldsymbol{\varepsilon} \in \mathbb{R}^{P \times N}$ represent some additive noise. $\mathbf{W} = \left[\mathbf{w}_1, ..., \mathbf{w}_N\right] \in \mathbb{R}^{K \times N}$ are the encoding coefficients and $\mathbf{D} = \left[\mathbf{d}_1, ..., \mathbf{d}_K\right] \in \mathbb{R}^{P \times K}$ is the dictionary of K atoms (K is the size of the dictionary).

The IBP based non-parametric Bayesian model is expressed as (Dang and Chainais, 2018):

$$\mathbf{y}_i = \mathbf{D}\mathbf{w}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

$$\mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i, \tag{3}$$

$$\mathbf{d}_k \sim \mathcal{N}(\mathbf{0}, P^{-1}\mathbf{I}_P), \tag{4}$$

$$\mathbf{z}_i \sim IBP(\alpha), \tag{5}$$

$$\mathbf{s}_i \sim \mathcal{N}(\mathbf{0}, \gamma_s^{-1}\mathbf{I}_K), \tag{6}$$

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \gamma_\varepsilon^{-1}\mathbf{I}_P), \tag{7}$$

where $\mathbf{y}_i$ is a column vector of dimension P, $i = 1, ..., N$, $\odot$ represents the element-wise or Hadamard product. $\mathbf{I}_P(\mathbf{I}_K)$ represents a $P \times P(K \times K)$ identity matrix. The representation coefficients are defined as $\mathbf{w}_{k,i} = \mathbf{z}_{k,i}\mathbf{s}_{k,i}$, the *K*-dimensional vector $\mathbf{z}_i$ consisting of 0 and 1 ($\mathbf{z}_i \in \{0,1\}^K$) encodes which of the K columns of $\mathbf{D}$ are used to represent $\mathbf{y}_i$; $\mathbf{s}_i \in \mathbb{R}^K$ are the values of the coefficients. The sparsity properties of $\mathbf{W}$ are induced by the sparsity of $\mathbf{Z} = \left[\mathbf{z}_1, ..., \mathbf{z}_N\right] \in \mathbb{R}^{K \times N}$ thanks to the IBP prior which will be discussed later. α is modeled by a prior with a Gamma distribution (α ~Gamma(a, b)), a and b are set to 1 in this paper. $\mathbf{d}_k$, $\mathbf{s}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to follow Gaussian distributions as priors for modeling convenience. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the Gaussian distribution with expectation μ and covariance Σ. Both γ_s and γ_ε follow a gamma distribution, that is $\gamma_s \sim \Gamma(c,d)$, $\gamma_\varepsilon \sim \Gamma(e,f)$ (Zhou *et al.*, 2012), $\{c,d,e,f\}$ are parameters which describe the Gamma distribution. They are all set to 10^{-6} as is done usually to make them non-informative (Tipping, 2001).

The posterior distributions (after observing the data) of the model variables are estimated by a Gibbs sampling strategy (Dang and Chainais, 2018). The final values of the variables $\mathbf{Z}, \mathbf{D}, \mathbf{S}, \gamma_s, \gamma_\varepsilon, \text{and } \alpha$ are obtained after a number of iterations on sampling. Then, the denoised patch matrices can be obtained by $\hat{\mathbf{X}} = \mathbf{D}(\mathbf{Z} \odot \mathbf{S})$.

Results

We test the method first on a 2-D synthetic example, as shown in Figure 1. The sample interval of this data set is 1ms. The dominant frequencies of the three events are 35 Hz, 30 Hz, and 25 Hz, respectively. We add strong white Gaussian noise to the synthetic model. The signal-to-noise ratio (SNR) of noisy record is -10 dB as shown in Figure 1b. The signal-to-noise ratio is defined by the power ratio. We compare the proposed method with K-Singular Value Decomposition (KSVD) which has shown success in seismic denoising (Zhu et al., 2015). In the proposed method, P is 64 and the total number of overlapping patches N is 14761. The number of iterations is 20. The variables are initialized randomly from their respective prior distributions. The patch size and the number of iterations for the KSVD are the same as the ones used in the proposed method. The size of dictionary K used in the KSVD is fixed as 100. The results of two methods are shown in Figure 2. The laterally coherent features are better enhanced by the proposed method, and there is less residual noise compared with the KSVD method.

Next we verify the performance of the proposed method on a real surface microseismic record with a sample interval of 4 ms, which is shown in Figure 3(a). In this example, we also compare the proposed method with the KSVD method. P is 64 and the total number of overlapping patches N is 12325. The number of iterations is 40. The patch size and the number of iterations for the KSVD are the same as the ones used in the proposed method. The size of dictionary K used in the KSVD is fixed as 100. The results of two methods are shown in Figure 3(b) and 3(c), respectively. The result of the proposed method is much better than using the KSVD method in terms of signal recovery and noise suppression.

Finally, the proposed method is applied to a real seismic section with a sample interval of 4 ms, which is shown in Figure 4(a). Again the KSVD method is applied for comparison. The results of two methods are shown in Figure 4(b) and 4(c), respectively. Both the KSVD and proposed methods enhance the continuity of the reflectors but the proposed method better eliminated residual high-frequency noise (red rectangle).

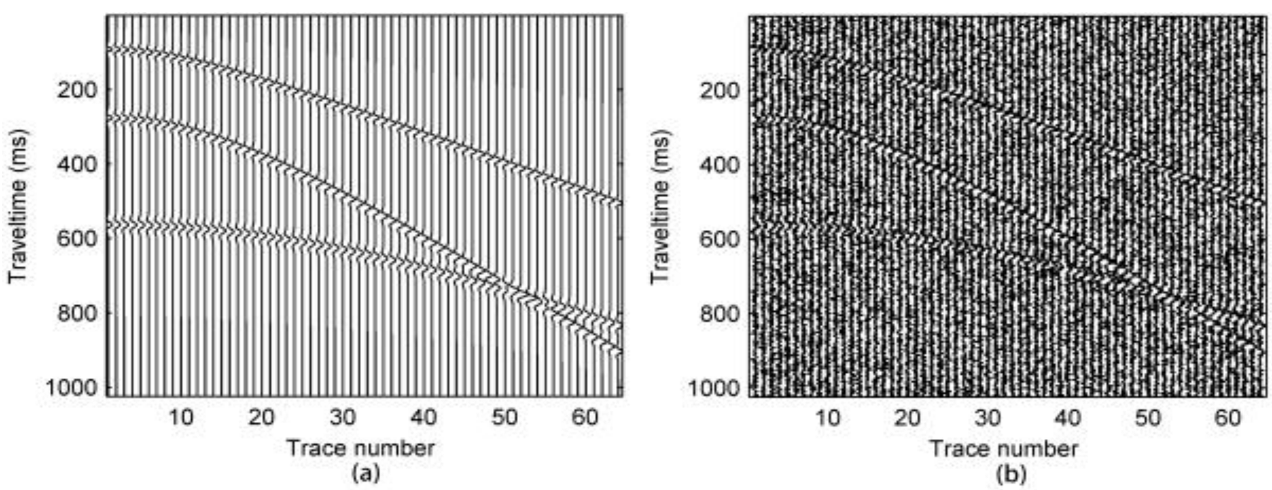

Figure 1: The synthetic model. (a) Noise-free data. (b) Noisy data (SNR=-10 dB) contaminated with strong white Gaussian noise.

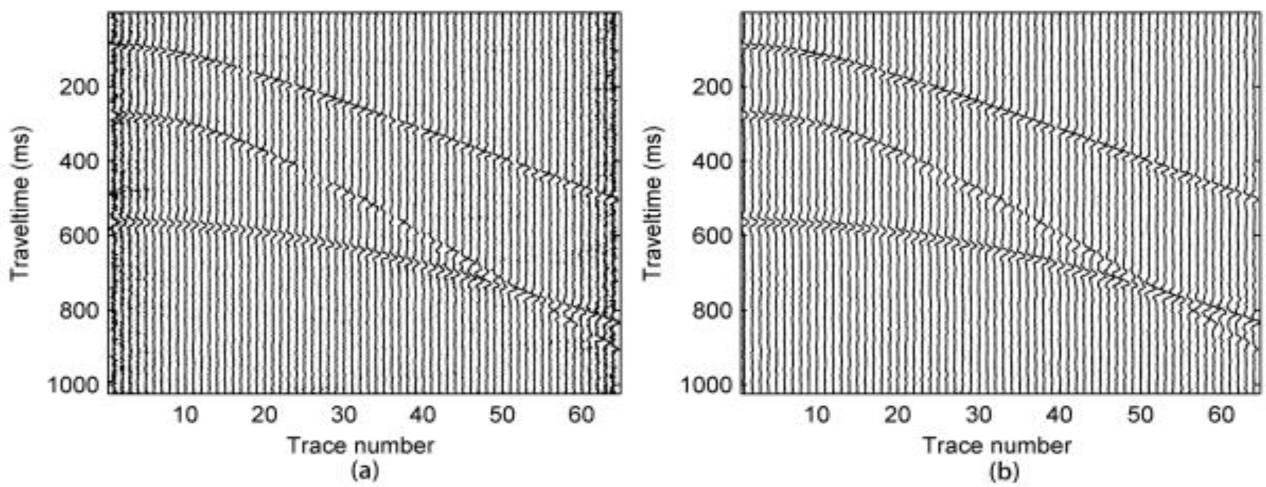

Figure 2: Result after (a) the KSVD, (b) the proposed method.

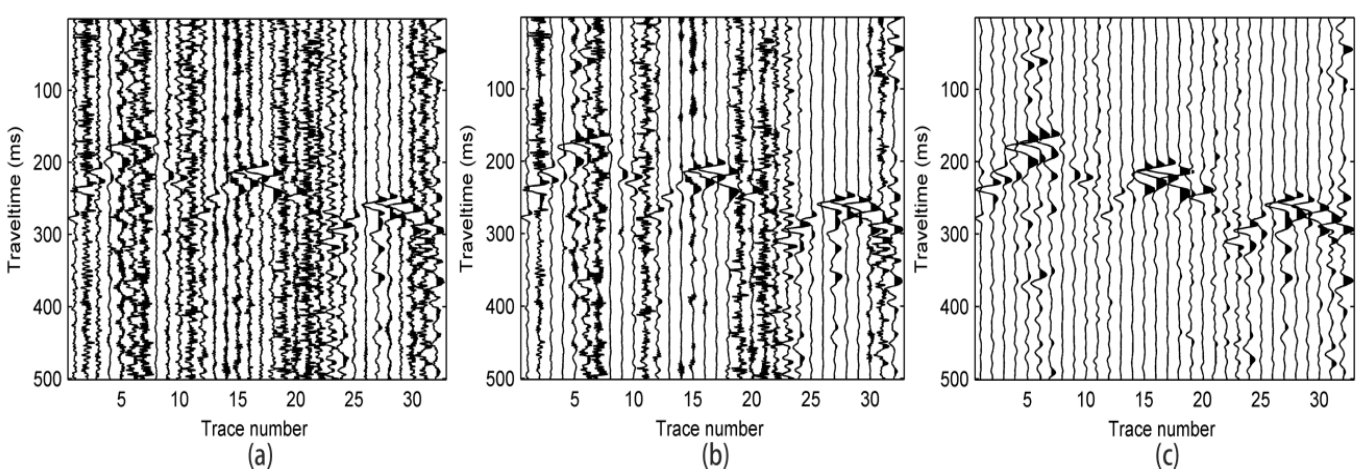

Figure 3: Real surface microseismic data example. (a) Real surface microseismic data . Result after (b) the KSVD method (the size of dictionary is 100), (c) the proposed method (the size of dictionary is 12).

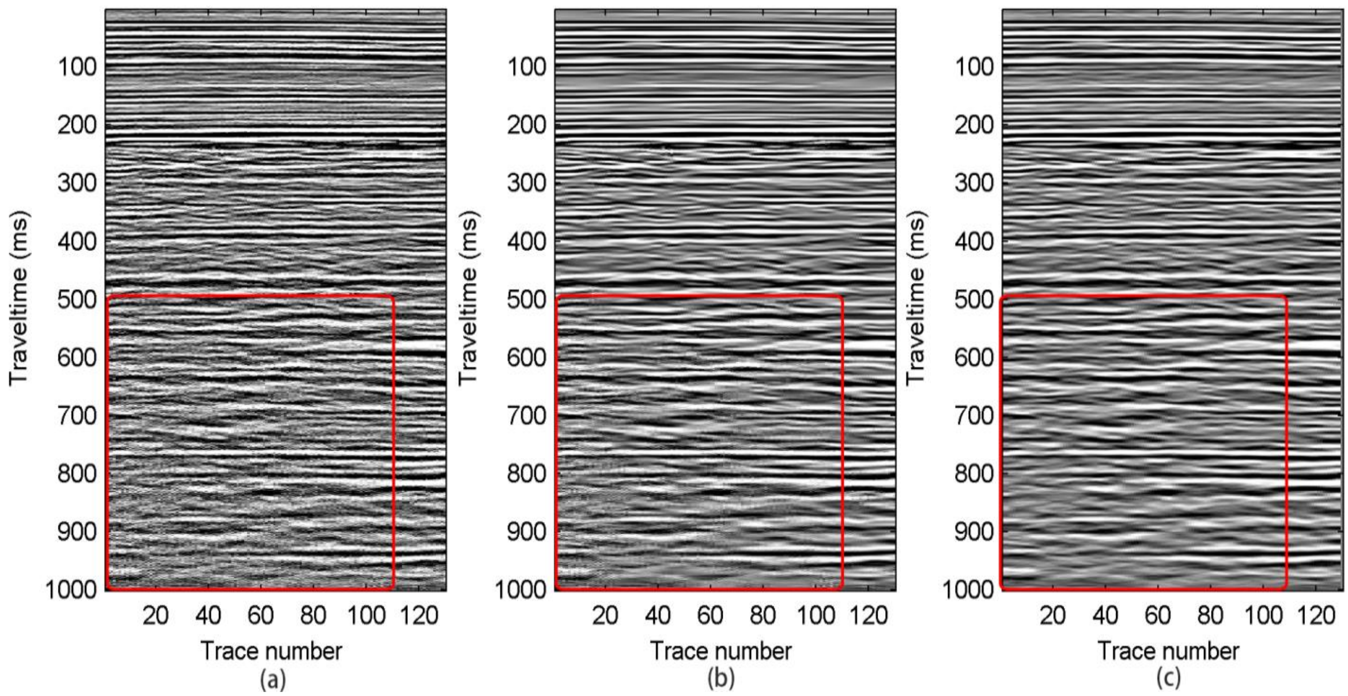

Figure 4: Real seismic section example. (a) Real stacked section . Result after (b) the KSVD method (the size of dictionary is 100), (c) the proposed method (the size of dictionary is 15).

Conclusions

The Indian Buffet Process based Bayesian non-parametric model learns an appropriate dictionary from noisy data. It provides a very useful tool for effective signal extraction. The size of the dictionary as well as the other parameters of this model are simultaneously inferred by the Gibbs sampling strategy. The results of synthetic and real data examples verify the validity of the proposed method for both noise reduction and signal recovery.

Acknowledgements

The authors thank the financial support provided by the sponsors of the Microseismic Industry Consortium.

References

Albert, J. H., and S. Chib (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *J. Bus. Econ. Stat*. **11**, no. 1, 1-15.

Aharon, M., M. Elad, and A. Bruckstein (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.***54**, no. 11, 4311-4322.

Chen, Y., J. Ma, and S. Fomel (2016). Double-sparsity dictionary for seismic noise attenuation, *Geophysics*, **81**, no.2, V103-V116.

Chen, Y. (2017). Fast dictionary learning for noise attenuation of multidimensional seismic data, *Geophys. J. Int*. **209**, no. 1, 21-31.

Dang, H. P., and P. Chainais (2018). Towards Dictionaries of Optimal Size: A Bayesian Non Parametric Approach, *Journal of*

Signal Processing Systems, **90**, no. 2, 221-232.

Engan, K., S. O. Aase, and J. H. Hakon-Husoy (1999). Method of optimal directions for frame design, *IEEE International Conference on Acoustics, Speech and Signal Processing*, **5**, 2443-2446.

Griffiths, T. L., and Z. Ghahramani (2006). Infinite latent feature models and the indian buffet process, *Advances In Neural Information Processing Systems*, 475-482.

Hou, S., F. Zhang, X. Li, et al. (2017). Simultaneous multicomponent seismic denoising and reconstruction via K-SVD, *87th Annual International Meeting, SEG, Expanded Abstracts*, 2503-2507.

Ma J., G. Plonka, and H. Chauris (2010). A new sparse representation of seismic data using adaptive easy-path wavelet transform, *IEEE Geosci. Rem. Sens. Lett.* **7**, no.3, 540-544.

Polatkan, G., M. Zhou, L. Carin, et al. (2015). A Bayesian non-parametric approach to image super-resolution, *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, no. 2, 346-358.

Siahsar, M. A. N., S. Gholtashi, V. Abolghasemi, et al. (2017). Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning, *Signal Process*. **141**, 309-321.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine, *J. Mach. Learn. Res.* **1**, 211-244.

Zhou, M., H. Chen, J. Paisley, et al. (2012). Non-parametric Bayesian dictionary learning for analysis of noisy and incomplete images, *IEEE Trans. Image Process.* **21**, no. 1, 130-144.

Zhu, L., E. Liu, and J. H. McClellan (2015). Seismic data denoising through multiscale and sparsity-promoting dictionary learning, *Geophysics*, **80**, no. 6, WD45-WD57.

Zhang C. and Van der Baan M. (2018). Microseismic and seismic denoising by a Bayesian non-parametric method with optimal dictionary size, *88th Annual International Meeting, SEG, Expanded Abstracts*.