Dictionary learning for signal-to-noise ratio enhancement

Pierre Turquais[1*], *Endrias G. Asgedom*[1], *Einar Otnes*[2], *Walter Söllner*[1]
[1] *Petroleum Geo-Services ASA,* [2] *Formerly Petroleum Geo-Services ASA*

SUMMARY

We investigate dictionary learning (DL) method for enhancing the signal-to-noise ratio (SNR) of a seismic data. The method is applied in windows, where each of these windows is subdivided into overlapping small two dimensional (2D) patches. Each patch is approximated using a linear combination of elementary signals (or atoms) from a set called dictionary. Denoising is performed by assuming the signal in each patch can be represented as a linear combination of a few of the atoms in the dictionary. This dictionary can be constructed either assuming a mathematical model for the signal (e.g. wavelets and curvelets) or can be learned to perform best on a training set. Unlike predefined general purpose dictionaries, learned dictionaries avoid any assumption about the morphology of the seismic data. Hence, denoising by DL ought to provide state-of-the-art results. In this paper, we demonstrate the performance of the K-means singular value decomposition (K-SVD) based DL denoising both on synthetic and field datasets. The significantly high signal preservation and SNR enhancement ability of DL denoising is illustrated with a comparison with that of conventional FX deconvolution.

INTRODUCTION

Raw seismic data are often contaminated with random noise over the entire time and frequency band. This noise obscures details and hinders seismic imaging from revealing the real subsurface structures. Random noise in seismic signal processing is a well-known problem and there are many approaches that have been proposed to attenuate such a noise (Yilmaz, 2001). The classic random noise attenuation method is the FX deconvolution (Canales, 1984) which mainly relays on spatial linearity of seismic signals. However, over the past decade, sparse and redundant representations for denoising have received a lot of attention in image processing by providing the state-of-the-art denoising results (Elad, 2010). When used for denoising, sparse and redundant representations assume that the desired signal can be reconstructed with few bases (or atoms) in a dictionary. The dictionary could be predefined (e.g. wavelets (Mallat, 1999), curvelets (Ma and Plonka, 2010), seislet (Fomel and Liu, 2010), etc) or learned from a training dataset. When using a predefined dictionary, we fix the representation space and assume that it can efficiently describe the data. However, the DL methods directly capture the morphology of the data and provide the atoms that can sparsely represent the signal. Thus, DL methods can overcome the limitation of needing prior information about the morphology of the data. In this paper, we utilized the K-means singular value decomposition (K-SVD) based DL method to perform denoising (Aharon et al., 2006). The results of K-SVD based DL method and FX deconvolution are presented for both synthetic and field data examples.

METHOD

The seismic section is first divided into windows. Considering each window as a training set, it is subdivided into M overlapping patches, ordered lexicographically as column vectors $\{\mathbf{y}_i\}_{i=1}^{M}$. DL based denoising aims at learning the dictionary $\hat{\mathbf{D}}$ of size $n \times K$ adapted to represent all the patches within the training set and finding the optimal sparse vectors $\hat{\mathbf{x}}_i$ such that the product $\hat{\mathbf{D}}\hat{\mathbf{x}}_i$ is the sparse approximation of the patch i. This problem can mathematically be expressed as

$$(\{\hat{\mathbf{x}}_i\}_{i=1}^{M}, \hat{\mathbf{D}}) = \underset{\{\mathbf{x}_i\}_{i=1}^{M}, \mathbf{D}}{\arg\min} \; ||\mathbf{x}_i||_0 \;\text{ subject to } \; \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \le \varepsilon \,, i = 1, ..., M \,, \tag{1}$$

where $||.||_0$ denotes the ℓ_0-norm and ε is the representation error threshold. To not represent the noise in the sparse approximation, ε is fixed to the noise energy present in each patch.

The minimization problem presented in Eqn. (1) is non-convex in regard to $(\{\mathbf{x}_i\}_{i=1}^{M}, \mathbf{D})$ and hence very complex to solve. Thus, a common approach for solving such a problem is to decompose it into two minimization subproblems and alternatively solve each one of them. For the k-th iteration, this two-step procedure can be summarized as:

I Use the dictionary from the previous iteration $\mathbf{D}_{\text{(k-1)}}$ and solve the sparse approximation problem

$$\min_{\mathbf{x}_i} ||\mathbf{x}_i||_0 \;\text{ subject to }\; \|\mathbf{y}_i - \mathbf{D}_{\text{(k-1)}}\mathbf{x}_i\|_2^2 \le \varepsilon \,, \tag{2}$$

for each patch i of the learning set. For the first iteration, $\mathbf{D}_{(0)}$ is initialized with patches randomly chosen from the learning set.

II Find the optimal dictionary by minimizing the errors between the dataset and its sparse representation

$$\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}_{\text{(k)}}\|_F^2 \,, \tag{3}$$

where the columns of the matrices $\mathbf{X}_{\text{(k)}}$ and $\mathbf{Y}$ are the solutions $\{\hat{\mathbf{x}}_i\}_{i=1}^{M}$ from (I) and the learning set $\{\mathbf{y}_i\}_{i=1}^{M}$ respectively.

Eqn. (2) is also a non-convex problem and hence it is very complex to find its exact solution. However, the matching pursuit algorithms (e.g. orthogonal matching pursuit (OMP) (Pati et al., 1993)) can provide quite an effective approximate solution. Eqn. (2) can also be relaxed using ℓ_1-norm rather than ℓ_0-norm, which makes the problem convex and possible to solve

DOI http://dx.doi.org/10.1190/segam2015-5846080.1

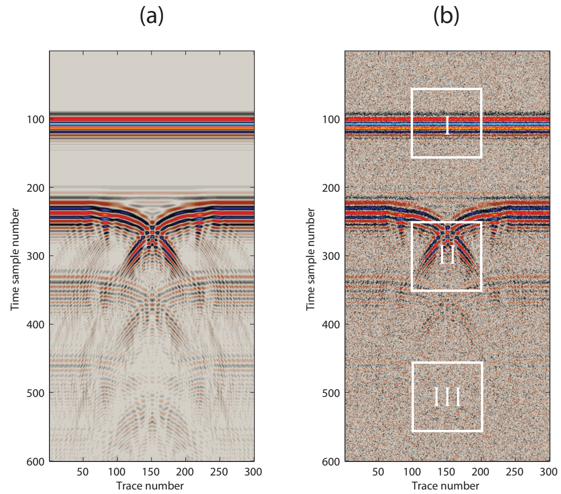

Figure 1: (a) Noise free and (b) noise contaminated synthetic common offset sections. Here, I, II, and III denote the three windows selected for detailed analysis.

using the basis pursuit algorithm (Chen et al., 1998). In this work, we choose to use the OMP algorithm for its simplicity and robustness.

Eqn. (3) can be simply solved by minimizing the average residual error using the method of optimal direction (MOD) (Engan et al., 1999). This method has been utilized by Beckouche and Ma (2014) for denoising a seismic section and they have showed its superiority in comparison with different predefined dictionary methods. Here we use the K-SVD method for dictionary learning, where the dictionary update is more efficient and converges faster than MOD. In K-SVD, K atoms of the dictionary are updated sequentially using singular value decomposition (SVD).

By applying iteratively the two steps specified in the procedure above, the algorithm converges to $\hat{\mathbf{D}}$ and $\{\hat{\mathbf{x}}_i\}_{i=1}^{M}$. The learned dictionary $\hat{\mathbf{D}}$ contains the most efficient bases for representing the dataset. Features present in several data patches are obviously more useful to represent efficiently the entire dataset and will consequently constitute the dictionary. The random noise present in the data is different from one patch to another and therefore will not be part of the dictionary. Finally, each approximation $\hat{\mathbf{D}}\hat{\mathbf{x}}_i$ is the sparse linear combination of bases describing seismic features repeated over the data, which minimizes the representation of noise. Thus, DL uses the redundancy of the seismic features over the dataset to attenuate the random noise.

SYNTHETIC DATA EXAMPLE

A synthetic data was generated using finite difference for an earth model consisting of plane and syncline reflectors. Since DL exploits the redundancy of the features over the dataset, we selected the common-offset domain, where we expect higher

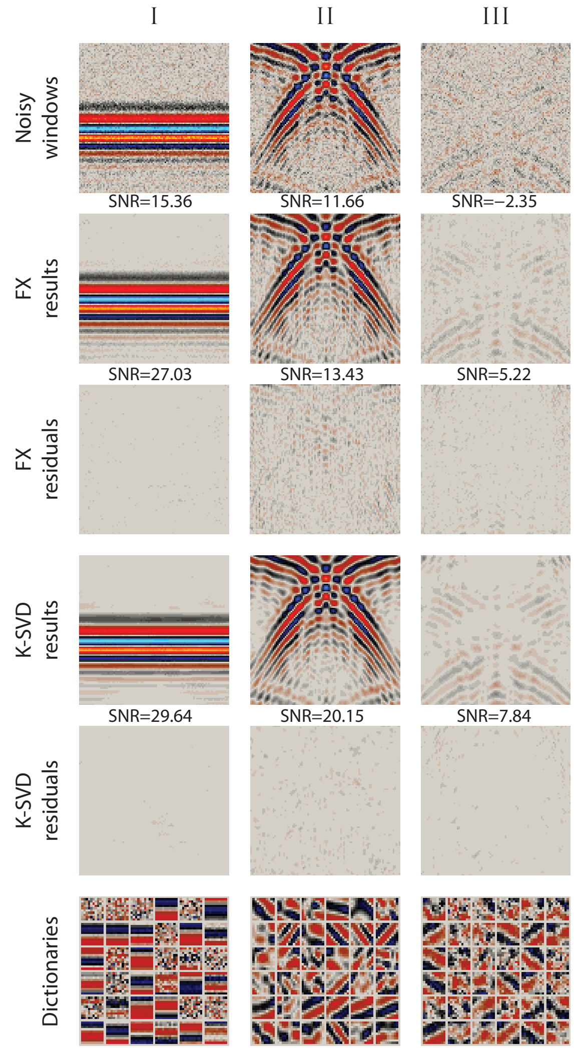

Figure 2: FX deconvolution and K-SVD based denoising results for the three windows selected in Fig. 1(b). Note that the same color scale is used for all the plots.

redundancy to be present, to apply the denoising algorithm. Fig. 1 (a) and (b) respectively show a common-offset section before and after the addition of random Gaussian noise with a standard deviation of 5% of the maximum amplitude in the noise free data. To present the denoising results, we selected three windows of size 100×100 from the common-offset section (cf. Fig. 1(b)). For each one of these windows, we applied the FX deconvolution and K-SVD denoising. For FX deconvolution we have used a filter length of 10 time samples. To apply the K-SVD method, we extracted from each window fully overlapping patches of size 10×10 samples in time and space. The size of the patches has been chosen as a compromise between a good quality of denoising and a small computation time. Since learning the dictionary on the entire dataset is expensive, we first learned the optimal dictionary on a subpart

DOI http://dx.doi.org/10.1190/segam2015-5846080.1

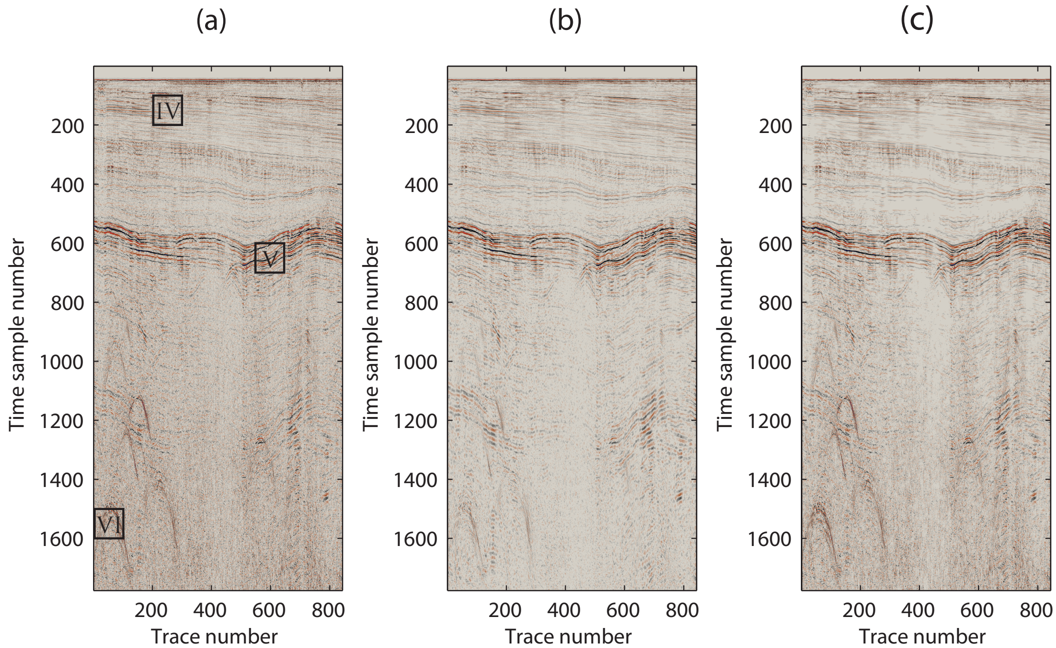

Figure 3: Raw common-offset field data (a), FX deconvolution result (b) and K-SVD method result (c). IV, V, and VI indicate the windows selected for detailed analysis.

of the dataset (learning set) and then used this dictionary to sparsely represent the entire dataset.

For learning the optimal dictionary, the two-step procedure specified in the methodology was iterated 25 times. In the first step, the patches of the learning set were sparsely coded using Eqn. (2). The representation error threshold was estimated by $\varepsilon = \sigma\sqrt{n}$, where σ is the known standard deviation of the noise present in the data and n is the length of the patch. The first iteration dictionary was initialized with 36 randomly selected patches from the learning set. Given that the patches contain 100 samples (i.e. 10×10), a dictionary with only 36 atoms creates an undercomplete system of equations. However, given the simplicity of the synthetic signal, this choice provides better denoising results. In the second step, Eqn. (3) was used to update the dictionary. Then, we used the learned optimal dictionary and Eqn. (3) to sparsely represent the entire dataset and finally averaged the overlapping patches to obtain the denoising results for a given window.

The learning and denoising results of the three selected windows (cf. Fig. 1(b)) are shown in Fig. 2. Note that the 3 dictionaries are quite different and constitute features that are redundant in their respective windows. This shows the efficiency and advantage of learning a dictionary rather than using a predefined one. The quality of the denoising is assessed quantitatively using SNR calculated for the denoised data $\hat{\mathbf{y}}$ as

$$\mathrm{SNR}(\hat{\mathbf{y}}, \mathbf{y}_{\mathrm{ref}}) = 10\,\log_{10} \frac{\|\mathbf{y}_{\mathrm{ref}}\|_2^2}{\|\mathbf{y}_{\mathrm{ref}} - \hat{\mathbf{y}}\|_2^2}\,, \qquad (4)$$

where $\mathbf{y}_{\mathrm{ref}}$ is the noise free data. The signal preservation ability of the denoising algorithms was assessed by calculating the residuals (i.e. which are obtained by subtracting the denoised data from the noise free data) and observing the presence of any noise or signal left.

For window I, where we have a linear flat event, both FX deconvolution and K-SVD method provides a good SNR enhancement with K-SVD method being slightly better. Windows II and III respectively are characterized by having nonlinear events and poor SNR. For these two windows (i.e. II and III), denoising based on K-SVD clearly shows a much better signal preservation and SNR enhancement (cf. Fig. 2).

FIELD DATA EXAMPLE

A raw common offset section (cf. Fig. 3(a)) was selected to validate the performance of K-SVD based denoising. To denoise the entire common-offset section, we first divided the section into windows of size 100×100 samples in time and space overlapping on half of their sizes in both dimensions. Then, each window has been filtered with the K-SVD and FX deconvolution methods. The resulting overlapping windows have been weighted with hamming filters and averaged to obtain the denoising results of the entire section (cf. Fig. 3(b) and (c)). A 10 time samples long filter was used for FX deconvolution.

For the K-SVD method, the same process as presented in the synthetic data example has been applied. Fully overlapping patches of size 10×10 both in space and time were extracted for each window. Since the data contains complex features, we made overcomplete dictionaries by setting the number of atoms in the dictionaries to 196. Since the characteristics of the noise are unknown, the error threshold needed for Eqn. (2) was estimated using the median absolute deviation (MAD) of data selected from the noisy part of the seismic section. Then, the standard deviation σ is given by $\sigma = \mathrm{MAD}(\mathbf{d}_n)/0.6745$, where $\mathbf{d}_n$ is part of the data containing only noise.

Fig. 4 shows the detailed analysis for the three windows (i.e. IV, V and VI) selected in Fig. 3(a). For each dictionary, only 36 of the 196 atoms have been pictured. Note that the dictionary parts constitute again features that are redundant in their respective windows. Here, the denoising performance of FX deconvolution and K-SVD methods are assessed using the residuals after subtraction of the denoised results from the noisy data. For window IV, where we have high-frequency flat events, and for window V, where we have linear high-amplitude dipping events, we observe that the residuals of K-SVD method are random while the FX deconvolution residuals show some removed signal. For window III, we observe that the K-SVD method preserves the complex non-linear dipping events due to diffraction in contrary to the FX deconvolution method (cf. Fig. 4). Here, it is pertinent to note that the results of both FX deconvolution and K-SVD method could be made better by reducing the size of the window, which however increases the computational cost.

If the standard deviation σ of the noise present in the data is correctly estimated, the performance of K-SVD based denoising remarkably preserves the seismic signal while reducing the noise. Here, we estimated σ using only one noisy window in the data and assumed it can represent the noise level for the entire section. This option was possible because the standard deviation of the noise was relatively constant for the entire common offset section. However, this might not always be the case and one possible extension to this technique is to estimate the noise standard deviation for every window containing both signal and noise and adapt the algorithm to the local standard deviation of the noise (Donoho and Johnstone, 1994).

K-SVD method for denoising is computationally expensive compared to FX deconvolution. As an example, filtering the field data window IV requires 37.09s for the K-SVD method and 0.04s for the FX deconvolution method. However, the computation time of the K-SVD method is highly dependent of the filtering parameters, such as the size or the overlapping rate of the patches, and can be improved in many aspects by adapting these parameters. Here, we focused mainly on the quality of the results but not much on the computation time, and consequently implemented an expensive algorithm. Adaptation can also be realized on the algorithm side. For example, it is possible to switch the K-SVD algorithm for faster algorithms such as the Online DL (Mairal et al., 2009).

CONCLUSION

Seismic data denoising was performed using both K-SVD based DL and FX deconvolution methods. The two methods were compared on both synthetic and field datasets. For a fixed window size the K-SVD method out performed FX deconvolution in terms of signal preservation and SNR enhancement. Though, K-SVD method is computationally expensive compared to FX deconvolution, there are quite a lot of improvements that can be realized.

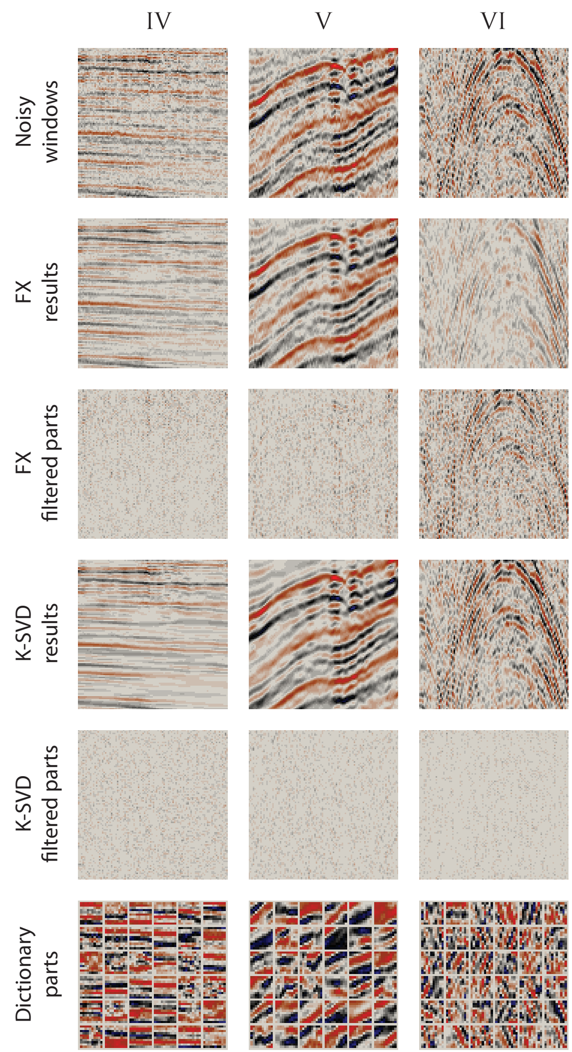

Figure 4: FX deconvolution and K-SVD method denoising results for the three data windows IV, V, VI shown in Fig. 3. Note that the same color scale is used for all the plots.

ACKNOWLEDGMENTS

We would like to thank PGS for providing the field data and giving the permission to publish this work. We also would like to express our deepest gratitude to Leiv Gellius and our PGS office team for the fruitful discussions on the subject. We finally give thanks to the internal reviewer team of PGS.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2015 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Aharon, M., M. Elad, and A. Bruckstein, 2006, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation: IEEE Transactions on Signal Processing, **54**, no. 11, 4311–4322, http://dx.doi.org/10.1109/TSP.2006.881199.

Beckouche, S., and J. Ma, 2014, Simultaneous dictionary learning and denoising for seismic data: Geophysics, **79**, no. 3, A27–A31, http://dx.doi.org/10.1190/geo2013-0382.1.

Canales, L. L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 1998, Atomic decomposition by basis pursuit: SIAM Journal on Scientific Computing, **20**, no. 1, 33–61, http://dx.doi.org/10.1137/S1064827596304010.

Donoho, D. L., and I. M. Johnstone, 1994, Ideal spatial adaptation by wavelet shrinkage: Biometrika, **81**, no. 3, 425–455, http://dx.doi.org/10.1093/biomet/81.3.425.

Elad, M., 2010, Sparse and redundant representations: Springer.

Engan, K., S. Aase, and J. Hakon Husoy, 1999, Method of optimal directions for frame design: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, IEEE, 2443–2446, doi:10.1109/ICASSP.1999.760624.

Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, no. 3, V25–V38, http://dx.doi.org/10.1190/1.3380591.

Ma, J., and G. Plonka, 2010, The curvelet transform: IEEE Signal Processing Magazine, **27**, no. 2, 118–133, http://dx.doi.org/10.1109/MSP.2009.935453.

Mairal, J., F. Bach, J. Ponce, and G. Sapiro, 2009, Online dictionary learning for sparse coding: Proceedings of the 26th Annual International Conference on Machine Learning, Association for Computing Machinery (ACM), 689–696, http://www.di.ens.fr/willow/pdfs/icml09.pdf.

Mallat, S., 1999, A wavelet tour of signal processing: Academic Press.

Pati, Y. C., R. Rezaiifar, and P. S. Krishnaprasad, 1993, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition: Proceedings of the 27th Annual Conference on Signals, Systems, and Computers, IEEE, 40–44, doi:10.1109/ACSSC.1993.342465.

Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: SEG.