Sparsity Promoting Morphological Decomposition for Coherent Noise Suppression: Application to Streamer Vibration Related Noise

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

SUMMARY

We address the signal and noise separation problem where the noise is coherent. We use a dictionary learning method to learn a dictionary of unit vectors called atoms; each one representing an elementary waveform redundant in the noisy data. In such a learned dictionary, some atoms represent signal waveforms while others represent noise waveforms. Using a multivariate Gaussian classifier trained on a noise recording, the atoms representing noise waveforms are discriminated and separated from the atoms representing seismic waveforms and two sub-dictionaries are created; one describing the morphology of the signal, the other describing the morphology of the noise. Using these sub-dictionaries, a morphological component analysis problem is set to separate the seismic signal and the coherent noise. In contrast to fixing transforms for representing the noise and the signal, our method is entirely adapting to the morphology of the signal and the noise. We present an application for removing streamer vibration related noise and show successful denoising results on synthetic and field data examples.

INTRODUCTION

In marine seismic surveys, the seismic wavefield is generally recorded by sensors attached on streamers lowed by a vessel. Steering devices or barnacles attached along the streamers can cause local vibrations of the streamers. These vibrations are recorded by the motion sensors and appear in the seismic data. Because these vibrations are recorded continuously in time and by several neighboring receivers, they are spatio-temporally coherent in the data, and attenuating them is challenging.

Morphological Component Analysis (MCA) has been developed to separate different components in a data based on their morphology (Starck et al., 2004, 2005). For data containing a signal and a noise component, the data is sparsely represented subject to two dictionaries; each one describing the morphology of one of the components. If the two dictionaries have a low mutual coherence, and noise and signal have highly sparse representations subject to their respective dictionaries, MCA correctly separates the signal and the noise (Bruckstein et al., 2009). For instance, Wang et al. (2010) propose to remove ground-roll noise from land data by solving the MCA problem using a wavelet transform for representing the signal and a discrete cosine transform for representing the noise. However predefining dictionaries for representing the signal and the noise components in a sparse manner is risky, as the signal or the noise component might not have a sparse representation subject to its attributed dictionary. In this case, the signal and noise separation is incomplete, and the quality of the denoising is poor.

Dictionary Learning (DL) methods (Engan et al., 1999; Aharon et al., 2006) are alternatives to predefining a dictionary. They capture the morphology of the redundant signal present in the data to provide a dictionary which is optimal to sparsely represent the given data in a sparse manner. Such methods have proved to perform well on seismic data such that the learned dictionaries enable an accurate sparse representation of the data (Beckouche and Ma, 2014; Turquais et al., 2015).

Here, we propose to tackle the coherent noise removal problem by performing DL before setting an MCA problem. First, a dictionary is learned on the noise contaminated data. In this dictionary, both coherent noise and seismic signal morphologies are described. Then, using a multivariate Gaussian classifier (Anderson and Bahadur, 1962) trained on a noise recording, the learned dictionary is separated into two sub-dictionaries; one describing the morphology of the noise, and the other describing the morphology of the signal. Finally, the MCA problem is set using the two sub-dictionaries and the coherent noise is separated from the seismic signal.

On synthetic and field data, we show a successful application of the combined Dictionary Learning and Morphological Component Analysis (DLMCA) method for removing streamer vibration related noise from marine seismic data.

METHOD

In this section, we present a method to separate coherent noise from signal within a data window, by learning a dictionary on the window, dividing the dictionary into two sub-dictionaries, and solving the MCA problem.

Dictionary Learning

Dictionaries are used by sparsity promoting methods as tools for performing sparse representations of data. Practically, a dictionary is a set of unit vectors called atoms stored as columns in a matrix. Hence computing a sparse representation of a recording $\mathbf{z} \in \mathbb{R}^N$ subject to a dictionary of atoms $\mathbf{D} = [\mathbf{a}_1 \ldots \mathbf{a}_K] \in \mathbb{R}^{N \times K}$ consists of finding a sparse coefficient vector $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{Dx}$ equal or approximate $\mathbf{z}$. The dictionary is the key element of the problem because its atoms should describe the morphology of the data in order to enable a sparse representation. A dictionary optimally adapted to represent recordings from a given dataset can be learned by capturing the morphology present in the dataset with a DL algorithm. For a 2D window, 2D small patches containing N samples are extracted from the window and rearranged as vectors $\mathbf{z}_1$, ..., $\mathbf{z}_M$. Then, one possibility for learning the dictionary is to find the dictionary $\mathbf{D}$ and the set of sparse coefficient vectors $\{\mathbf{x}_i\}_{i=1}^M$ which minimize the representation error given a sparsity constraint m placed on the sparse coefficient vectors. This minimization problem is mathematically expressed as

$$\min_{\{\mathbf{x}_i\}_{i=1}^M,\mathbf{D}} \|\mathbf{z}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \text{ subject to } ||\mathbf{x}_i||_0 \leq m \, , \, i = 1,...,M \, . \quad (1)$$

The problem in Equation 1 is very complex and highly undetermined. However, several algorithms, such as MOD (Engan et al., 1999) or K-SVD (Aharon et al., 2006), approximate the problem and find a good approximate solution.

Atom classification

A dictionary $\mathbf{D}$ learned on a data window contaminated by a coherent noise describes both the morphology of the noise and the signal. If the noise and the signal are independently generated, they are also independently distributed in the window. If, in addition, the morphology of the noise is different from the morphology of the signal, the two morphologies will be described by different atoms of the dictionary. Hence, the atoms can be classified to create two sub-dictionaries; one signal sub-dictionary $\mathbf{D}_s$ containing the atoms describing the morphology of the signal and one noise sub-dictionary $\mathbf{D}_n$ containing the atoms describing the morphology of the noise. Such signal and noise sub-dictionaries would be optimal to represent the morphology of the signal and the noise, respectively.

The atoms $\mathbf{a}_1$,..., $\mathbf{a}_K$ learned on a 2D window of a seismic data describe 2D patterns when rearranged as 2D patches. The morphology of such patterns can be characterized with second order statistics features. That is, features characterizing the statistical distribution of observed combinations of amplitude values from a pair of samples at specified relative positions in the pattern. For instance, the inertia is a textural feature that reflects the homogeneity of the pattern when moving in a given direction (Δt, Δx) (Haralick et al., 1973). The larger the probability that two samples separated by Δt samples in time and Δx samples in spatial dimension have close amplitude values, the lower the inertia is. The inertia feature of a pattern is mathematically given by

$$I(\Delta t, \Delta x) = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-j)^2\mathbf{P}[i,j] \, , \quad (2)$$

where $\mathbf{P}$ is the Gray-Level Co-occurrence Matrix (GLCM) computed in the direction (Δt, Δx) and G is its length in both dimensions. For the given direction (Δt, Δx), each element $\mathbf{P}[i,j]$ of the GLCM is computed as the probability of changing from the amplitude i to j when moving Δt samples in time and Δx samples in spatial dimension. Haralick et al. (1973) presents the different textural features and how to compute the GLCM with more details.

Given a vector of textural features $\mathbf{f}$ that characterizes its pattern, an atom $\mathbf{a}$ can be classified as signal atom (i.e., atom describing the morphology of the signal) or noise atom (i.e., atom describing the morphology of the noise) using a multivariate Gaussian classifier trained on the noise morphology (Anderson and Bahadur, 1962). That can consist to compute the probability of an atom to be a noise atom as a function of its feature vector such that

$$p(\mathbf{f}) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_c|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{f}-\mu_c)^T\Sigma_c^{-1}(\mathbf{f}-\mu_c)\right] \, , \quad (3)$$

where D is the length of $\mathbf{f}$, and μ_c and Σ_c are respectively the mean vector and covariance matrix that characterize the multivariate Gaussian distribution of the noise feature vectors. Both μ_c and Σ_c can be obtained by training on a noise window. Here, it can consist of obtaining a test dataset by learning a dictionary on the noise window and computing the feature vectors of its atoms. Then, μ_c and Σ_c are computed as the mean vector and covariance matrix of the test dataset. The atom $\mathbf{a}$ is classified as a noise atom if the probability $p(\mathbf{f})$ is above a probability threshold. This threshold can be fixed as the maximum probability threshold such that all the feature vectors from the test dataset are classified as noise atoms. Or, to minimize the impact of eventual outliers in the test dataset, one would prefer to fix the threshold as the maximum probability threshold such that a given percentage of the feature vectors (e.g. 95%) from the test dataset are classified as noise. Finally, if $p(\mathbf{f})$ is under the fixed threshold, $\mathbf{a}$ is classified as a signal atom.

Morphological Component Analysis

Consider a recording $\mathbf{z}$, containing a signal component that can be sparsely represented subject to a dictionary $\mathbf{D}_s$, and a noise component that can be sparsely represented subject to a dictionary $\mathbf{D}_n$. MCA can be used to separate the noise and signal components. The separation of the two components is exact if the sparsity of the recording subject to the two dictionaries is below a threshold dictated by the mutual coherence between the dictionaries (Bruckstein et al., 2009). This implies that, for obtaining a high quality signal and noise separation, the signal and noise dictionaries should describe different morphologies, and the signal and noise components should both have a very sparse representation subject to their corresponding dictionary. The MCA problem consists of finding a sparse representation of the recording subject to two dictionaries. One possibility of setting such a problem is finding the sparse vectors $\mathbf{x}_s$ and $\mathbf{x}_n$ that are the solution of the following minimization problem

$$\min_{\mathbf{x}_s,\mathbf{x}_n}||\mathbf{z}-\mathbf{D}_s\mathbf{x}_s-\mathbf{D}_n\mathbf{x}_n||_2 \text{ subject to } ||\mathbf{x}_s||_0+||\mathbf{x}_n||_0 \leq m \, , \quad (4)$$

where m is a threshold that constrains the solution to be sparse. Then the signal and noise components can be reconstructed by computing the sparse approximations $\mathbf{D}_s\mathbf{x}_s$ and $\mathbf{D}_n\mathbf{x}_n$, respectively. The residual vector (i.e., $\mathbf{z}-\mathbf{D}_s\mathbf{x}_s-\mathbf{D}_n\mathbf{x}_n$) could contain random noise if present in the recording or signal and coherent noise if the recording is not strictly sparse subject to the two dictionaries. In the case where the residual vector is suspected to contain signal, the signal can be retrieved by subtracting the noise component from the recording. The problem in Equation 4 is NP-hard, and therefore not tractable for realistic seismic data sizes. However, several techniques (e.g., Orthogonal Matching Pursuit (OMP) (Pati et al., 1993), basis pursuit (Chen et al., 1998)) solve an approximate problem to find the correct or a good approximate solution to the problem in Equation 4.

Practically, if the two dictionaries $\mathbf{D}_s$ and $\mathbf{D}_n$ contain atoms

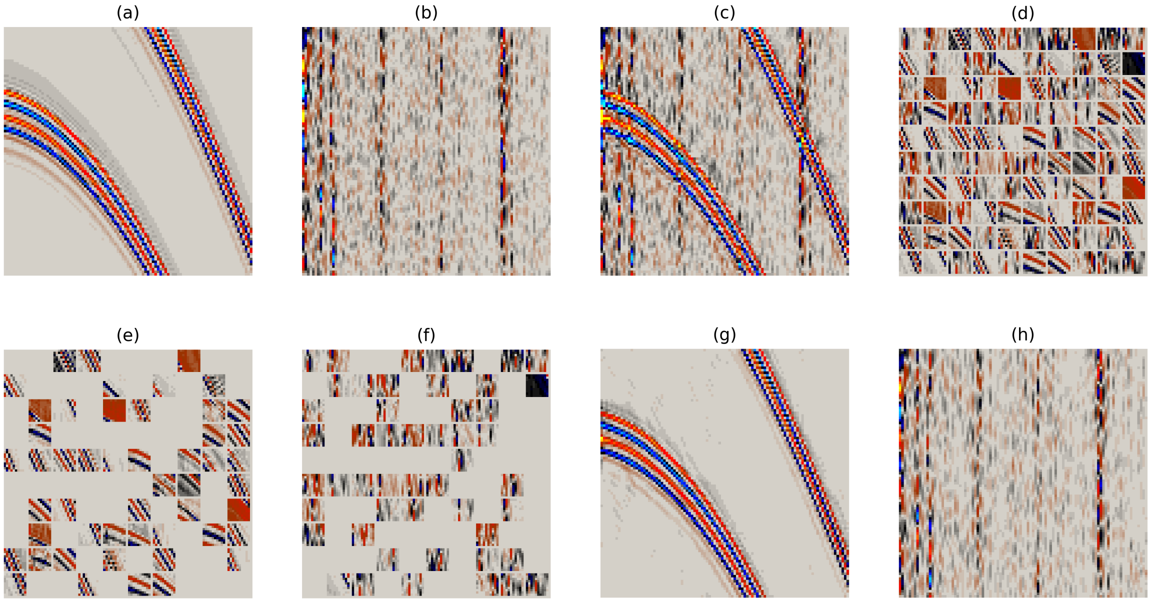

Figure 1: DLMCA for removing streamer vibration related noise from a synthetic signal. (a) Signal window from a noise free synthetic shot gather. (b) Noise window from a shot gather recorded during a marine survey when no seismic source is fired. (c) Noisy data window resulting from the addition of the noise to the signal (SNR=2.86dB). (d) Dictionary learned on the noisy data window. (e) Signal sub-dictionary and (f) noise sub-dictionary consequent to the atom classification. (g) Signal window (SNR=15.43dB) and (h) noise window retrieved by solving the MCA problem.

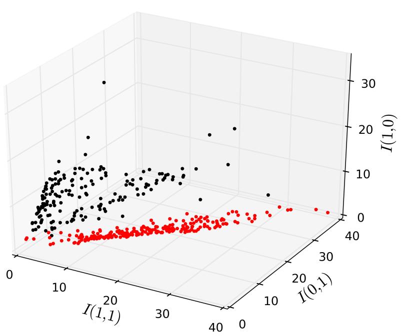

Figure 2: Location of the classified signal (black points) and noise (red points) atoms in the feature space defined by the inertia in the three directions (1,1), (0,1), and (1,0).

representing small 2D patches, the signal and noise components within a 2D data window are separated by solving the MCA problem for overlapping patches of the size of the atoms. The signal components obtained for all the patches are assembled given their location in the window and averaged to obtain the signal component in the 2D data window. A noise window can similarly be obtained by assembling the noise component of the patches.

SYNTHETIC EXAMPLE

An experiment was designed to illustrate the separation of coherent noise from seismic signal using DLMCA. The signal of interest (cf. Figure 1(a)) was a window of size 100×100 samples from a noise free synthetic shot gather. The coherent noise was a window from a shot gather acquired during a marine survey when no seismic source is fired (cf. Figure 1(b)). The noise in this window is caused by vibrations of the streamer. The frequencies below 10Hz were prior removed from both windows. The noisy data window (cf. Figure 1(c)) obtained by adding the noise to the signal had a Signal-to-Noise Ratio (SNR) of 2.86dB. Using the K-SVD algorithm to solve the problem in Equation 1, a dictionary of 400 atoms was learned from the noisy data window. The atoms have been learned such that they represent patterns of size 10×10 samples. A subset of 100 patterns described by the atoms is pictured in Figure 1(d). Considering the small size of the patterns, low frequency noise patterns cannot be distinguished from low frequency signal patterns. Hence, for filtering low frequency noise, one should use larger size patterns. The atoms of this dictionary were characterized with three features: the inertia computed in the directions (0,1), (1,0), and (1,1). A multivariate Gaussian classifier was trained on an window that has been extracted from the same shot gather as the window presented in Figure 1(b). This classifier was used to discriminate the noise atoms from the signal atoms. In Figure 2, the 400 atoms of the learned dictionary are located in the space formed by the three chosen features. The noise atoms are concentrated

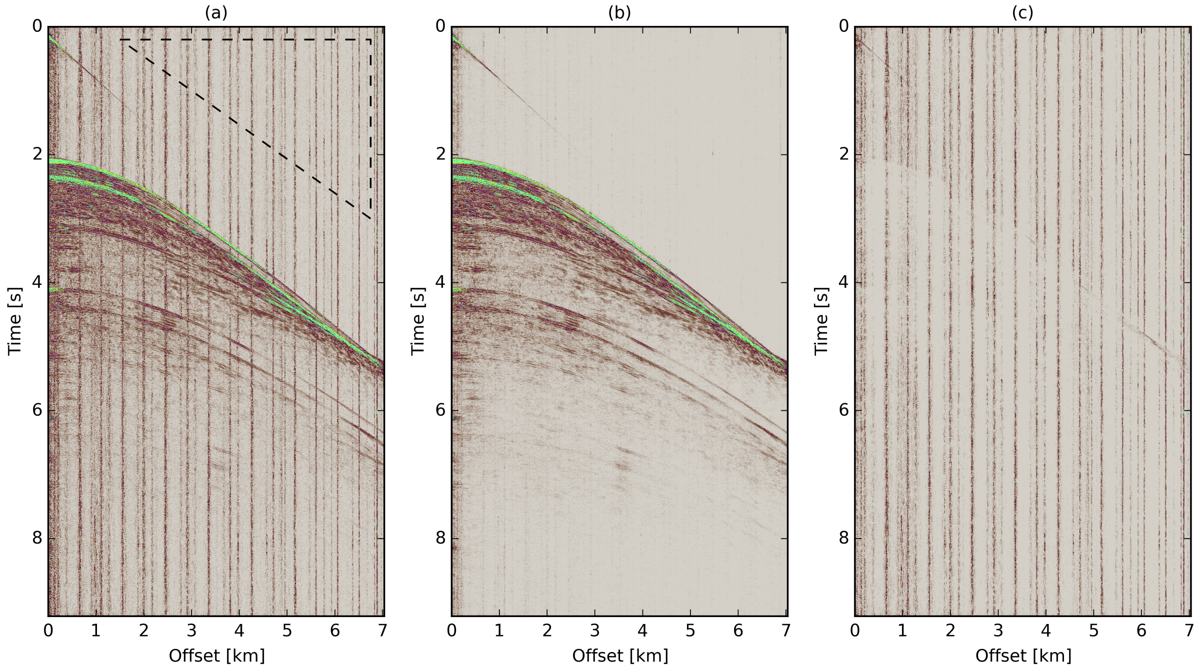

Figure 3: DLMCA based denoising for removing streamer vibration related noise from a common shot gather. Vertical particle velocity of the (a) input common shot gather, (b) denoised result, and (c) removed noise.

in a small part of the feature space which attests of the quality of the features for discriminating the noise patterns. The results of the classification for the 100 atoms subset are shown in Figure 1(e)-(f) and attest of a correct classification. Using two sub-dictionaries resulting from the classification, the MCA problem was solved for all overlapping patches of the window with a sparsity constraint $m = 10$. The signal components in the patches were assembled and averaged to obtain the signal within the window (cf. Figure 1(g)). Similarly, the noise components in the patches were assembled and averaged to obtain the noise within the window (cf. Figure 1(h)). The results show that the signal and the noise are accurately separated. It is confirmed by the SNR of the reconstructed signal reaching 15.43dB.

FIELD DATA APPLICATION

A raw shot gather, acquired during a marine survey was selected. The data was sampled at 2ms in the temporal dimension and 12.5m in the offset dimension. The frequencies below 10Hz were filtered. This shot gather is presented in Figure 1(a). DLMCA was performed on windows containing 25 complete neighboring traces. For each window, a dictionary was learned using 10000 patches of size 10×10. A signal and a noise sub-dictionary were obtained using a multivariate Gaussian classifier trained on the part of the shot gather framed with dotted line in Figure 1(a). This data part cannot contain signal from the seismic source. The MCA based signal and noise separation was performed for all overlapping patches of the window. The denoised shot gather shown in Figure 1(b) has been obtained by removing the retrieved noise component windows from the data. We can observe that the majority of the noise has been removed leaving a high quality data. The removed noise shown in Figure 3(c) attests that the signal has been preserved to large extend by the denoising process with minor differences in the direct arrival and refracted waves.

CONCLUSION

We proposed a new sparsity promoting method to remove coherent noise from seismic data. In this method, a dictionary is learned from the data and separated into two sub-dictionaries; one describing the morphology of the signal and the other describing the morphology of the noise. Then, the two sub-dictionaries are used to separate the noise from the signal via MCA. In contrast to predefining dictionaries, the proposed method entirely adapts to the morphology of the signal and the noise present in the data. In addition, since the sub-dictionaries are specifically learned to provide a very sparse representation of the data, MCA performs an optimal signal and noise separation. For removing streamer vibration related noise from seismic data, the proposed method performs state of the art denoising results on a synthetic and a field data example.

ACKNOWLEDGMENTS

We would like to express our gratitude to the PGS research team of Oslo for the fruitful discussions on the subject, and thank PGS for giving the permission to publish this work.

http://dx.doi.org/10.1190/AM2016-13856592.1

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016 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**, 4311–4322, http://dx.doi.org/10.1109/TSP.2006.881199.

Anderson, T. W., and R. R. Bahadur, 1962, Classification into two multivariate normal distributions with different covariance matrices: Annals of Mathematical Statistics, **33**, 420–431, http://dx.doi.org/10.1214/aoms/1177704568.

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.

Bruckstein, A. M., D. L. Donoho, and M. Elad, 2009, From sparse solutions of systems of equations to sparse modeling of signals and images: SIAM Review, **51**, 34–81, http://dx.doi.org/10.1137/060657704.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 1998, Atomic decomposition by basis pursuit: SIAM Journal on Scientific Computing, **20**, 33–61, http://dx.doi.org/10.1137/S1064827596304010.

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

Haralick, R. M., K. Shanmugam, and I. Dinstein, 1973, Textural features for image classification: IEEE Transactions on Systems, Man, and Cybernetics, **SMC-3**, 610–621, http://dx.doi.org/10.1109/TSMC.1973.4309314.

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

Starck, J.-L., M. Elad, and D. Donoho, 2004, Redundant multiscale transforms and their application for morphological component separation: Advances in Imaging and Electron Physics, **132**, 287–348, http://dx.doi.org/10.1016/S1076-5670(04)32006-9.

Starck, J.-L., Y. Moudden, J. Bobin, M. Elad, and D. Donoho, 2005, Morphological component analysis: Wavelets, **XI**, 209–223.

Turquais, P., E. G. Asgedom, W. Sllner, and E. Otnes, 2015, Dictionary learning for signal-to-noise ratio enhancement: 85th Annual International Meeting, SEG, Expanded Abstracts, 4698–4702, http://dx.doi.org/10.1190/segam2015-5846080.1.

Wang, W., W. Chen, J. Lei, and J. Gao, 2010, Ground roll separation by sparsity and morphological diversity promotion: 80th Annual International Meeting, SEG, Expanded Abstracts, 3705–3710, http://dx.doi.org/10.1190/1.3513621.