# SIMULTANEOUS DUAL SENSOR WAVEFIELD SEPARATION AND SEISMIC DATA COMPRESSION USING PARABOLIC DICTIONARY LEARNING

M.O. Faouzi Zizi[1,2], P. Turquais[1], A. Day[1], M.W. Pedersen[1]

[1] PGS; [2] University of Oslo (UiO)

## Summary

Areas covered by seismic exploration surveys are continuously increasing and the data recorded are becoming of very large size. In addition, conventional seismic processing and imaging is a long meticulous workflow which is time consuming and very expensive in terms of human and computational resources. Consequently, allowing seismic processing steps in a compressed domain can play a key role in the marine seismic industry as it would be a faster and cheaper alternative to the standard seismic processing sequence and would save cost on storage and data transfer. Parabolic dictionary learning has the ability to compress the seismic data by transforming them into an appropriate sparse domain, and also to extract local parameters which relate to the kinematics of the wavefield. In this paper, we use these kinematic parameters to correct for the geophone obliquity problem and thus enable the dual-sensor wavefield separation processing step in the compressed domain. Without the need for data interpolation, our method succeeds in reconstructing aliased events and shows comparable results to an industry-standard FK based algorithm in terms of up- and down-going pressure fields reconstruction. It also comes with the advantage of compressing the data by a factor higher than 10.

# Simultaneous dual sensor wavefield separation and seismic data compression using parabolic dictionary learning

## Introduction

During the past decades, both the area covered by typical seismic exploration surveys and the number of sensors used to acquire the data has continuously increased. Consequently, seismic datasets have become very large, and conventional seismic processing and imaging demands a large amount of human and computational resources. Compressing the seismic data at an early stage of the seismic processing sequence would save cost on storage and data transfer. However, the seismic data would still need to be decompressed by transforming them back into their original time-space domain before seismic processing and imaging steps are carried out. Therefore, enabling seismic processing steps in the compressed domain would overcome the requirement for decompression of the data, and thus provide a faster and cheaper alternative to the standard seismic processing sequence.

Many seismic compression algorithms have been shown to provide relatively concise expressions of the data by transforming the data into an appropriate mathematical domain (e.g., Elad, 2010). These algorithms are generally based on so-called fixed sparse transforms such as wavelets, discrete cosines, and others (e.g., Duval and Rosten, 2000). Dictionary Learning (DL) methods such as the K-mean singular value decomposition (K-SVD) (Aharon et al., 2006) are alternatives to predefined transforms. These methods have recently been shown to provide state-of-the-art results for seismic data compression (Faouzi Zizi and Turquais, 2021). They capture similar events from the seismic data and store them once in a dictionary of atoms that represents these data in a sparse manner. Different modifications of those methods have been shown to be suited to various processing tasks such as noise suppression (Beckouche and Ma, 2014) or interpolation (Turquais et al., 2018). For example, Turquais et al. (2018) propose a Parabolic Dictionary Learning (PDL) method where the learned atoms represent elementary waveforms of constant amplitude along parabolic travel time moveout. Hence, each atom can be characterized by a set of parameters such as the slope and the curvature, which relate to the kinematics of the wavefield (Bortfeld, 1989). These kinematic parameters are then used to interpolate the atoms along their respective slopes, thereby reconstructing the interpolated data.

Such local parameters can be used not only in interpolation but also in other data processing methods, such as wavefield separation. Wavefield separation is generally applied early in the seismic processing sequence. It is used in dual-sensor streamer data processing to decompose the data recorded by pressure and particle velocity sensors into upward and downward travelling waves (Day et al., 2013). The velocity sensor records only the vertical component. Hence, for emergence angles greater than zero, the amplitudes recorded by the velocity sensor must be scaled by an obliquity factor (Söllner et al., 2008). It is most convenient to apply the obliquity scaling after plane wave decomposition. Transforming the data into such a domain requires preconditioning (e.g., data interpolation, zero padding), which comes at a significant computational cost. In this paper, we propose a method which uses PDL to simultaneously compress and extract kinematic parameters from parabolic atoms. These parameters will facilitate the derivation of the obliquity factor for local events in the time domain thus enabling wavefield separation in the compressed domain.

## Method

To describe the different steps of our method, we need to consider two sets of data: geophone, and hydrophone datasets. The geophone property is to measure the vertical component of particle velocity, while the hydrophone is a pressure sensor. The different steps to apply wavefield separation in the compressed domain are as follows:

– Parabolic Dictionary Learning (PDL): First, many time-space 2D patches are independently extracted from both seismic datasets, then vectorized to obtain the training sets matrices $\mathbf{Y_G}$ and $\mathbf{Y_H}$, where the indices $\mathbf{G}$ and $\mathbf{H}$ will refer to the geophone and hydrophone sets, respectively. For each training set PDL is applied as described by Turquais et al., (2018) to obtain the two parabolic dictionaries $\mathbf{D_G} = [\mathbf{d}_1^G, \ldots, \mathbf{d}_K^G]$ and $\mathbf{D_H} = [\mathbf{d}_1^H, \ldots, \mathbf{d}_K^H]$ that are optimal to represent the training data in a sparse manner. Hence, each atom $\mathbf{d}_k$ is now constant along a parabolic moveout of reference position $o_k^{ref}$, of slope $s_k$, and of curvature $c_k$. Turquais et al. (2018) describe in detail the PDL problem and the method to find an approximate solution to it. Finally, the sparse optimization problem that is represented in equation 1 below is solved for 2D overlapping patches of each seismic dataset (**G** and **H**). This process is the same as in conventional dictionary learning methods (Aharon et al., 2006), where $\mathbf{y}_1, \ldots, \mathbf{y}_M$ are the training data patches, $\mathbf{D}$ is the parabolic dictionary, and $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are the sparse coefficient vectors. Each patch can now be described as a linear combination of the parabolic dictionary atoms.

$$\min_{\{\mathbf{x_i}\}_{i=1}^{M}, \mathbf{D}} \sum_{i=1}^{M} \| \mathbf{x_i} \|_0 \text{ subject to } \| \mathbf{y_i} - \mathbf{D}\mathbf{x_i} \|_2 \leq \epsilon, \ i = 1, \ldots M . \quad (1)$$

– Obliquity correction: The amplitude recorded by the geophone needs to be scaled by an obliquity factor $f = \frac{1}{\cos\theta}$ to correct for the fact that only the vertical component of particle velocity is recorded (Söllner et al., 2008), where $\theta$ is the emergence angle of each single event. The equations for the up-going and down-going pressure fields can be written as:

$$P^{up} = \frac{1}{2}(P - \rho v f V_z) \quad (2)$$

$$P^{down} = \frac{1}{2}(P + \rho v f V_z) , \quad (3)$$

where $\rho$ is the water density, $v$ is the propagation velocity in water, $P$ is the recorded pressure, and $V_z$ is the recorded vertical particle velocity. Bortfeld (1989) relates the parameters of the parabolic moveout to the kinematics of the wavefield. In the common shot domain, the slope $s_k$ of an atom $k$ is related to the emergence angle of the trace at the reference position $o_k^{ref}$ as follows:

$$s_k = \frac{\sin\theta_k}{v} \quad (4)$$

From equation 4 and the expression for the obliquity factor, we can write:

$$f_k = \frac{1}{\cos\theta_k} = \frac{1}{\sqrt{1-(v s_k)^2}} . \quad (5)$$

Hence, we have found the obliquity factor to apply to each reference trace of the atoms in the geophone dictionary $\mathbf{D_G}$. Further, we need to take the derivative of the parabolic time moveout function $\Delta t = s_k \Delta o + c_k \Delta o^2$, where $\Delta t$ is the time moveout, and $\Delta o$ is the displacement of a receiver location relative to the reference receiver, to find the right obliquity factor not only for the reference trace of a given geophone atom but also for other traces of that atom. Hence, the obliquity factor at a receiver location $i$ in an atom $k$ can be written as:

$$f_k^i = \frac{1}{\sqrt{1-(v s_k^i)^2}} , \qquad \text{where} \quad s_k^i = 2c_k \left(o_k^i - o_k^{ref}\right) + s_k . \quad (6)$$

– Data compression: Once the atoms of the geophone dictionary $\mathbf{D_G}$ have been scaled by the different obliquity factors following equation 6, we obtain the corrected geophone dictionary denoted $\mathbf{D_{CG}}$. Hence, we end up with two dictionaries $\mathbf{D_{CG}}$ and $\mathbf{D_H}$ and two sets of sparse representations denoted $\mathbf{X_G}$ and $\mathbf{X_H}$. Here, the sparse representation corresponding to the geophone data remains the same after obtaining the corrected dictionary $\mathbf{D_{CG}}$. Further, $\mathbf{D_{CG}}$, $\mathbf{D_H}$, $\mathbf{X_G}$, and $\mathbf{X_H}$ are compressed using the quantization and coding stages described in detail by Faouzi Zizi and Turquais (2021).

– Data reconstruction: First, we decompress the data as described in Faouzi Zizi and Turquais (2021) to obtain the reconstructed dictionaries and sparse representations denoted as $\widehat{\mathbf{D}}_{\mathbf{CG}}$, $\widehat{\mathbf{D}}_{\mathbf{H}}$, $\widehat{\mathbf{X}}_{\mathbf{G}}$, and $\widehat{\mathbf{X}}_{\mathbf{H}}$. Then, one can reconstruct either the hydrophone, the geophone, the up-going pressure, or the down-going pressure shot gathers. Simply multiplying $\widehat{\mathbf{D}}_{\mathbf{H}}$ and $\widehat{\mathbf{X}}_{\mathbf{H}}$ would reconstruct the hydrophone data. Multiplying $\widehat{\mathbf{D}}_{\mathbf{CG}}$ and $\widehat{\mathbf{X}}_{\mathbf{G}}$ would reconstruct the corrected geophone data. The up- and down-going pressure data can be reconstructed by applying equations 2 and 3 to each patch, where $P$ is obtained from $\widehat{\mathbf{D}}_{\mathbf{H}}$ and $\widehat{\mathbf{X}}_{\mathbf{H}}$, and $f V_z$ from $\widehat{\mathbf{D}}_{\mathbf{CG}}$ and $\widehat{\mathbf{X}}_{\mathbf{G}}$.

## Data application

Our wavefield separation method in the PDL compressed domain (WSPDL) was compared with an industry-standard FK based method to assess its capability to correct for the geophone obliquity problem and to reconstruct up-going and down-going pressure fields. To do so, we have modelled synthetic geophone and hydrophone datasets, in addition to the corresponding up-going and down-going pressure fields, each of 100 shots. The data were sampled at 2 ms in time and 3.125 m in space. The modelled up-going and down-going pressure datasets were used as references. The four datasets were pre-processed as follows. First, a low-cut filter was applied at 20 Hz as is typically done in real data cases because the geophone data are in general heavily contaminated by noise at the lowest frequencies (Day et al., 2013). Then, a dip filter was applied to remove high dipping events with apparent propagation velocity values lower than 1550 m/s. This was done because tapering is generally applied when applying FK methods at such dips to avoid instability. Finally, the data were decimated to a sampling of 12.5 m in space which is more similar to real field data cases.

The industry-standard FK based method has been applied a first time after interpolating the data to 3.125 m spatial sampling using an industry-standard interpolation algorithm, and a second time without interpolation. Figure 1 a) shows a shot gather of the up-going reference dataset. Figures 1 b), c), and d) show the difference between the reference up-going dataset and the up-going pressure data output from our WSPDL method, the FK method with interpolation and the FK method without interpolation, respectively for the same shot gather. Note that the upper and lower limits of the colour scale were chosen to best visualize the residuals. The computational costs of the FK method with interpolation and WSPDL were comparable. We have also computed the signal-to-residual ratio (SRR), to assess the quality of the wavefield separation results (Table 1). The SRR was computed as follows

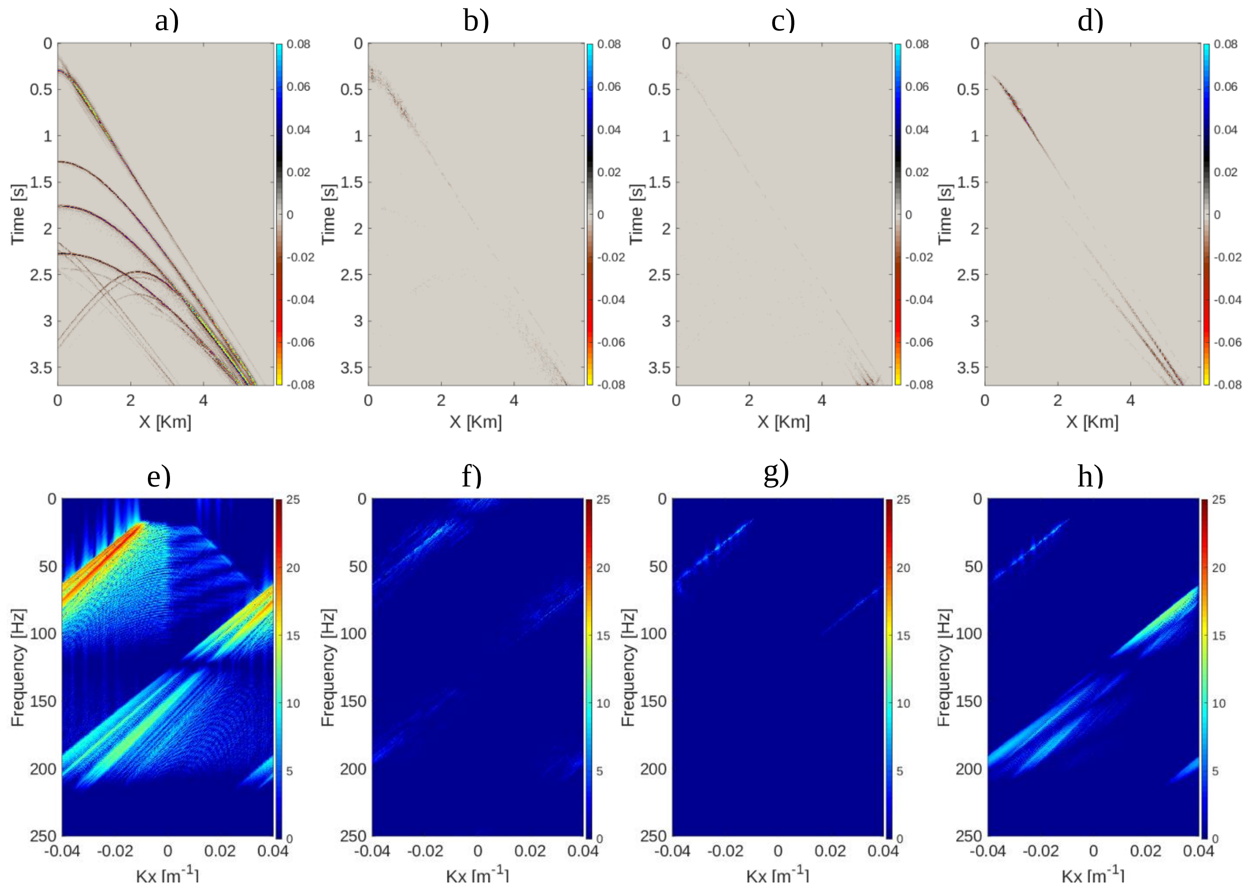


***Figure 1.*** *Wavefield separation results for one shot gather. a) reference up-going pressure field, b) residuals after applying our WSPDL method, c) residuals after applying the FK method with interpolation, d) residuals after applying the FK method without interpolation. (e-h) The f-k spectra of (a-d), respectively.*

$$\mathrm{SRR} = 20\log_{10}\frac{\|\mathbf{d}_{\mathrm{ref}}\|_2}{\|\mathbf{d}_{\mathrm{ref}} - \mathbf{d}_{\mathrm{rec}}\|_2}\,, \tag{7}$$

where $\mathbf{d}_{\mathrm{ref}}$ refers to the up-going reference data and $\mathbf{d}_{\mathrm{rec}}$ the reconstructed up-going data using one of the wavefield separation methods. From Figure 1 and Table 1, we can see that our method performs well in terms of up-going pressure field reconstruction as it reaches a high level of SRR. Also, Figure 1 shows residuals coming only from very high amplitude events. Although, the FK based algorithm achieves slightly better results in terms of SRR when the interpolation is used, our method has shown the ability to reconstruct the aliased events without the need for interpolation. Indeed, the f-k spectra in Figure 1 (e-h) clearly demonstrate that our method succeeds better in reconstructing such events. In addition, our method reconstructed the up-going pressure dataset from compressed data which are almost 13 times smaller than the input geophone and hydrophone datasets. The compression ratio (CR) was computed as follows

$$\mathrm{CR} = \frac{\text{Number of bits of input data}}{\text{Number of bits of output data}}\,. \tag{8}$$

| METHOD | WSPDL | Interpolation + FK | FK |
|---|---|---|---|
| **SRR (dB)** | 24.36 | 27.64 | 16.54 |
| **Aliasing** | + | + | - |
| **CR** | 12.88 | 0.25 | 1 |

***Table 1.*** *Comparison between the different wavefield separation methods in terms of SRR, reconstruction of aliased data parts, and compression ratio.*

## Conclusions

We have developed a novel method which uses kinematic parameters to enable the dual-sensor wavefield separation process in a compressed domain. The method achieves similar results to an industry-standard FK based algorithm with the advantage of being robust to aliasing without the need for data preconditioning such as interpolation. In addition, the method comes with the advantage of significant compression of the data. Finally, our WSPDL method is also automated and thus requires a lesser amount of time and human resources.

## References


Aharon, M., Elad, M. and Bruckstein, A. [2006] K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation: *IEEE Transactions on Signal Processing*, **54**, 4311–4322.

Beckouche, S., and Ma, J. [2014] Simultaneous dictionary learning and denoising for seismic data: *Geophysics*, **79**, A27–A31.

Bortfeld, R. [1989] Geometrical ray theory: Rays and traveltimes in seismic systems (second-order approximations of the traveltimes): *Geophysics*, **54**, 342–349.

Day, A., Klüver, T., Söllner, W., Tabti, H. and Carlson, D. [2013] Wavefield-separation methods for dual-sensor towed-streamer data. *Geophysics*, **78**, WA55–WA70.

Duval, L. C. and Rosten, T. [2000] Filter bank decomposition of seismic data with application to compression and denoising: *SEG Technical Program Expanded Abstracts 2000*, 2055–2058.

Elad, M. [2010] Sparse and redundant representations: From theory to applications in signal and image processing, 1st ed.: *Springer Publishing Company*.

Faouzi Zizi, M. O. and Turquais, P. [2021] A dictionary learning method for seismic data compression: *Geophysics*, **87**.

Söllner, W., Day, A. and Tabti, H. [2008] Space frequency domain processing of irregular dual-sensor towed streamer data. *SEG Technical Program Expanded Abstracts 2008*, 1078-1083.

Turquais, P., Asgedom, E. G., Söllner, W. and Gelius, L.-J. [2018] Parabolic dictionary learning for seismic wavefield reconstruction across the streamers: *Geophysics*, **83**, V263–V282.