

# Structured Dictionary Learning for Interpolation of Aliased Seismic Data

Pierre Turquais<sup>1,2\*</sup>, Endrias G. Asgedom<sup>1</sup>, Walter Söllner<sup>1</sup>

<sup>1</sup> PGS, <sup>2</sup> University of Oslo

## SUMMARY

We present a new dictionary learning method for seismic data interpolation. Dictionary learning methods train a set of basis vectors on the data to capture the morphology of the redundant signal. The basis vectors are called atoms, and the set is referred to as the dictionary. Learned dictionaries are very effective for representing the data as sparse linear combinations of their atoms. In conventional dictionary learning, the atoms are unstructured and do not have an analytic expression. In the proposed method, the atoms are constrained to represent linear events of known slopes. Using the slope information, the atoms can be easily interpolated. Hence, a regularly sampled data can be interpolated over a finer grid by learning a dictionary on the data, finding a sparse representation of the data in the dictionary domain, interpolating the dictionary, and finally taking the sparse representation of the data in the interpolated dictionary domain. The sparsity constraint ensures that atoms with well-fitting slopes are chosen to represent the data, and it hence prevents from aliasing and noise representation. On synthetic and field data, we observe that the proposed method performs a near to exact interpolation of linear events and an accurate interpolation of curved events, and that it is robust to noise.

## INTRODUCTION

In seismic surveys, the data are typically coarsely recorded in space. Interpolating the seismic data over a denser grid is crucial as many seismic processing steps, such as designature, demultiple, wavefield separation, or full waveform inversion, require or benefit from a fine spatial sampling. In the case of regularly sampled data, this interpolation task is challenging because strong spectral aliasing occurs in the data set.

Interpolating beyond aliasing can be seen as an underdetermined problem as the true frequency content of the data is unknown. It requires a priori information to be solved. Some methods assume local linearity of the events and interpolate in the frequency-space domain using error prediction filters (Spitz, 1991; Crawley, 2001). Other methods integrate a priori information about the seismic signal morphology to the problem via sparsity in a transform domain. These methods rely on the following assumption: because the densely sampled seismic wavefield follows the morphology described by the transform, it lies on a small sub-part of the transform domain, and it can be recovered using a sparse optimization process. For interpolating aliased regularly sampled data, several transforms have been proposed, including Fourier (Schonewille et al., 2009; Gao et al., 2013), Radon (Ibrahim et al., 2015), curvelet (Naghizadeh and Sacchi, 2010), seislet (Gan et al., 2015), and

focal transform (Kutscha et al., 2010).

In the quest for sparsity, dictionary learning (DL) methods (Engan et al., 1999; Aharon et al., 2006) are alternatives to predefining a transform. They capture the morphology of the redundant signal present in the data, to provide a dictionary that is optimal to represent the given data in a sparse manner. The resulting dictionary is similar to and has the same roll as a transform, but it cannot be analytically expressed. If a data is represented in an orthogonal analytic transform, for instance in Fourier, the data can be interpolated in another grid as an analytic expression of the data is known. Such interpolation is not possible with a conventionally learned dictionary because it is only physically defined.

In this paper, we propose a structured dictionary learning method that can be used to interpolate regularly sampled seismic data. This method, that we refer to as slope-DL, learns a dictionary in which the atoms are constrained to represent linear events of known slope. Using the slope information, each atom can be easily interpolated over a finer grid, and so the sparse representation of the data. We present the slope-DL method in the next section, and then we show successful synthetic and field data applications.

## METHOD

### The sparse optimization problem

ens We first recall the sparse optimization problem. Consider a recording  $\mathbf{y} \in \mathbb{R}^N$  containing a signal of interest. This recording can be expressed in another domain and recovered via matrix multiplication with a dictionary. This dictionary is a matrix containing an atom, i.e., a unit vector of length  $N$ , in each of its columns ( $\mathbf{D} = [\mathbf{a}_1 \dots \mathbf{a}_K] \in \mathbb{R}^{N \times K}$ ). Given that the dictionary is well chosen, the signal of interest is sparse in the dictionary domain. It exists a sparse vector  $\mathbf{x} \in \mathbb{R}^K$ , containing a small number of non-zero coefficients, such that

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{r}, \quad (1)$$

where  $\mathbf{r}$  is a residual signal of small energy. The sparse optimization problem consists in finding this vector  $\mathbf{x}$ . One possibility is to find  $\mathbf{x}$  that has the least number of non-zero coefficients, and such that the norm of the residual vector is below a small threshold  $\varepsilon$  (Donoho et al., 2006). This minimization problem is formally expressed as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon. \quad (2)$$

In presence of noise, the error threshold  $\varepsilon$  is dictated by the noise variance  $\sigma^2$ ; a natural choice is  $\varepsilon = \lambda \sigma \sqrt{N}$ , where  $N$  is the length of the recording, and  $\lambda$  is a gain factor that controls the strength of the denoising (Elad, 2010). The problem in equation 2 is not tractable if solved exactly, but an approximate

solution can be found using matching pursuit algorithms, e.g., orthogonal matching pursuit (OMP) (Pati et al., 1993).

### The slope-DL method

The dictionary is the key element of the sparse optimization problem. Its atoms need to describe the morphology of the data to be able to compute a representation which is both sparse and accurate. A dictionary representing the morphology of a data set can be obtained by applying a DL algorithm on this data set. For seismic data application, DL is often applied in 2D on a gather. In that case,  $M$  small-sized patches are extracted from the gather and vectorized to obtain the set of vectors  $\mathbf{y}_1, \dots, \mathbf{y}_M$  called the training set. The conventional DL problem consists in finding the dictionary  $\mathbf{D} \in \mathbb{R}^{N \times K}$ , with  $K \ll M$ , and the set of sparse coefficient vectors  $\mathbf{x}_1, \dots, \mathbf{x}_M$  which minimize the representation error given a sparsity constraint  $L$  placed on the sparse coefficient vectors. There is no other constraint on the dictionary. Consequently, the learned atoms are unstructured.

In the proposed method, we constrain the problem to learn atoms that represent linear events of known slope. This structured DL problem may be mathematically expressed as finding the dictionary  $\mathbf{D} = [\mathbf{a}_1 \dots \mathbf{a}_K]$  that satisfies

$$\begin{aligned} \min_{\{\mathbf{x}_i\}_{i=1}^M, \{\mathbf{a}_j\}_{j=1}^K, \{s_j\}_{j=1}^K} & \sum_{i=1}^M \|\mathbf{y}_i - [\mathbf{a}_1 \dots \mathbf{a}_K] \mathbf{x}_i\|_2^2 \\ \text{subject to} & \begin{cases} \|\mathbf{x}_i\|_0 \leq L, i = 1, \dots, M \\ \mathbf{a}_j \text{ is linear of slope } s_j, j = 1, \dots, K, \end{cases} \end{aligned} \quad (3)$$

where by " $\mathbf{a}_j$  is linear of slope  $s_j$ ," we mean that if it is rearranged as a patch, it represents an event that is constant along straight lines of slope  $s_j$ . The problem in equation 3 is very complex, and hence cannot be solved exactly. Similar to conventional DL, we propose to find an approximate solution. The proposed algorithm is a modified version of the K-SVD algorithm presented by Rubinstein et al. (2008). It can be summarized as presented in Algorithm 1. In this algorithm, brackets are used to refer to an index of a vector or a matrix; for instance,  $\mathbf{D}[i, j]$  is the sample at the  $i$ th line and  $j$ th column of the matrix  $\mathbf{D}$ . Columns inside the brackets refer to all the indexes in a dimension;  $\mathbf{D}[i, :]$  would be the  $i$ th line of  $\mathbf{D}$ . The letter T in upper position of a vector stands for the transpose of this vector. In line 9, the function `semblance(a, s)` rearranges the atom  $\mathbf{a}$  as a patch, aligns the traces according to the slope  $s$ , and computes the semblance, which is defined as  $\sum_i (\sum_j A[i, j])^2 / \sum_i \sum_j A[i, j]^2$  for a 2D array  $\mathbf{A}$ . In line 10, the function `mean_along_slope(a, s)` rearranges the atom  $\mathbf{a}$  as a patch, aligns the traces according to the slope  $s$ , averages the traces, and places back the traces in the atom.

### Interpolation using slope-DL

The slope-DL method can be used to interpolate regularly sampled data. The process for interpolating a window  $\mathbf{W}$  of a 2D gather is as follows:

1) A large number  $M$  of patches of size  $O \times P$  are extracted from  $\mathbf{W}$  and are stored in the columns of a matrix to obtain a training set  $\mathbf{Y}$ . The slope-DL algorithm is used to solve the problem in equation 3 and learn a dictionary  $\mathbf{D}$  of  $K$  atoms rep-

---

### Algorithm 1 Slope-DL

---

- 1: **Input:** Training set  $\mathbf{Y} = [\mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_M] \in \mathbb{R}^{N \times M}$
- 2: **Parameters:** Number of dictionary atoms:  $K$ , number of iterations:  $I$ , and sparsity constraint:  $L$
- 3: **Initialization:** Initialize the dictionary  $\mathbf{D} = [\mathbf{a}_1 \dots \mathbf{a}_K]$ , allocate space for the sparse coefficients  $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_M]$  and the slopes  $\mathbf{s} = [s_1 \dots s_K]$
- 4: Repeat  $I$  times,
- 5:   • **Sparse coding:** for  $i = 1, \dots, M$ , use OMP to find a sparse representation of the recording  $\mathbf{y}_i$ :

$$\mathbf{x}_i \leftarrow \arg \min_{\mathbf{x}} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2 \quad \text{subject to } \|\mathbf{x}\|_0 \leq L$$

- 6:   • **Dictionary update:** for  $j = 1, \dots, K$ ,
- 7:   - find the indexes of the recordings that use the atom

$$\Omega \leftarrow \{k | \mathbf{X}[j, k] \neq 0\}$$

- 8:   - find the principal component
- 9:   - find slope that maximizes the semblance of the atom

$$\mathbf{s}[j] \leftarrow \arg \max_{\mathbf{s}} \text{semblance}(\mathbf{a}, \mathbf{s})$$

- 10:   - average the atom in the slope direction

$$\mathbf{a} \leftarrow \text{mean\_along\_slope}(\mathbf{a}, \mathbf{s}[j])$$

- 11:   - normalize the atom

$$\mathbf{a} \leftarrow \mathbf{a} / \|\mathbf{a}\|_2$$

- 12:   - update the coefficients of the atom

$$\mathbf{X}[j, \Omega] \leftarrow \mathbf{a}^T (\mathbf{Y}[:, \Omega] - \mathbf{D}\mathbf{X}[:, \Omega] + \mathbf{D}[:, j] \mathbf{X}[j, \Omega])$$

- 13:   - update the atom in the dictionary

$$\mathbf{D}[:, j] \leftarrow \mathbf{a}$$

- 14: **Output:** Dictionary  $\mathbf{D}$ , sparse coefficient matrix  $\mathbf{X}$ , and slope vector  $\mathbf{s}$
- 

resenting linear events of known slope.

2) The sparse optimization problem that is presented in equation 2 is solved for overlapping patches of the window using the dictionary  $\mathbf{D}$ . It results in a sparse coefficient vector for each overlapping patch.

3) Each atom of the dictionary  $\mathbf{D}$  is interpolated over a finer grid. The  $j$ th atom of the dictionary, denoted with  $\mathbf{a}_j$ , represents a linear event of slope  $s_j$  when it is rearranged as a patch of size  $O \times P$ . Hence, using the slope information, additional traces can be derived from the known traces by time shifts.

4) The interpolated dictionary found in step 3 is multiplied with the sparse representation vectors found in step 2, which results in the interpolation of the overlapping patches extracted from the gather. These overlapping patches are assembled according to their original location to obtain an interpolated version of the window  $\mathbf{W}$ .

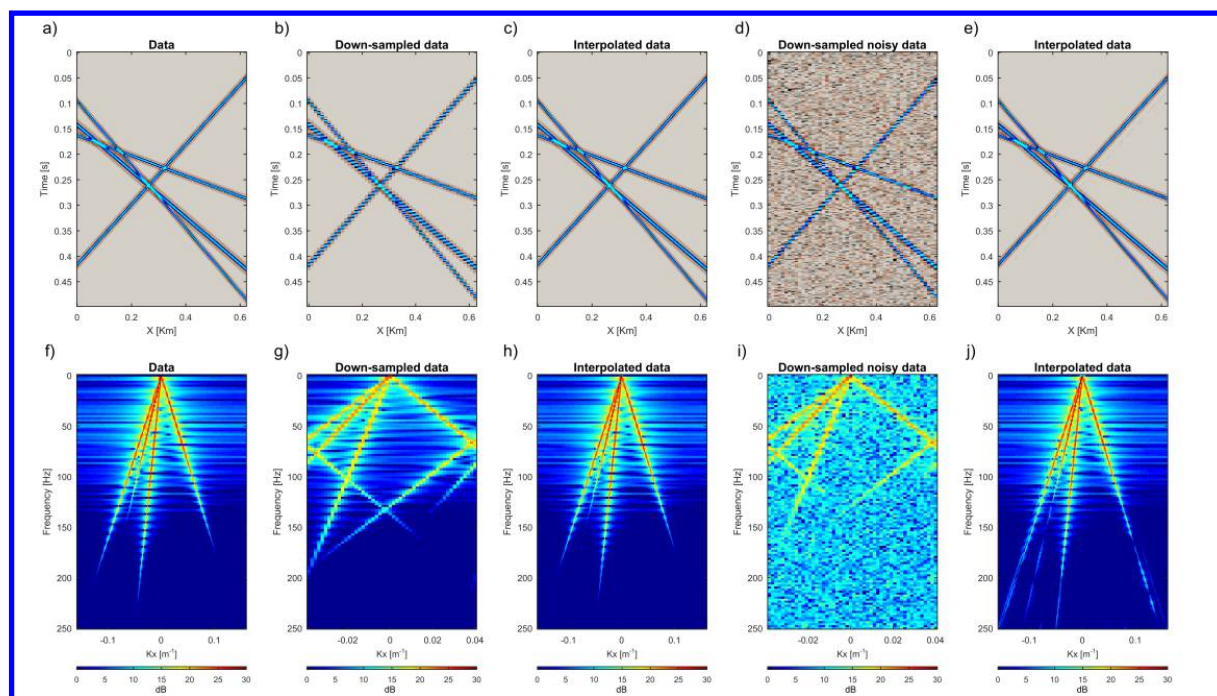


Figure 1: Interpolation of linear events using slope-DL. The original data (a,f), the down-sampled data (b,g) and its corresponding interpolated data (c,h), the noisy down-sampled data (d,i) and its corresponding interpolated data (e,j). In the first row, the data are shown in the time-space domain, and in the second row, their FK spectra are shown.

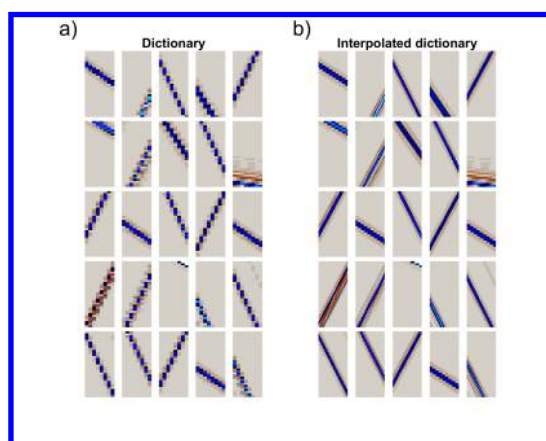


Figure 2: Interpolation of the learned dictionary (a) to a finer sampled dictionary (b).

## SYNTHETIC EXAMPLES

In a first experiment, we assessed the effectiveness of slope-DL when it comes to interpolate aliased linear events. We down-sampled data containing linear events and interpolated back the data using slope-DL. The original data is shown in the time-space domain in Figure 1a. It consists of four linear events sampled at 2 ms in time and at 3.25 m in space. Its FK spectrum, presented in Figure 1f, attests that the original data is not aliased. The down-sampled data had a spatial sampling of 12.5 m and aliasing occurred from 60 Hz (see Figures 1b

and 1g). The slope-DL method was used to interpolate the data as described in the section "Interpolation with slope-DL". In step 1, the slope-DL algorithm was applied with the number of recordings in the training set, the number of iterations, the patch size, the number of atoms, and the sparsity threshold, fixed such as,  $M = 8000$ ,  $I = 10$ ,  $O \times P = 32 \times 8$ ,  $K = 400$ , and  $L = 4$ . The 25 first atoms of the learned dictionary are presented in Figure 2a. In step 2, the sparse optimization problem was solved for patches overlapping on 31 samples in time and 7 samples in space. The error threshold  $\epsilon$  was fixed at a very small value in order to represent any significant signal. In step 3, three traces were interpolated between each two traces of the atom patterns. The 25 first atoms of the interpolated dictionary are presented in Figure 2b. In step 4, the data were interpolated using the sparse representation of the overlapping patches in the interpolated dictionary domain. The interpolated data and its FK spectrum are shown in Figures 1c and 1h. The signal-to-noise ratio (S/R) of the result is 32.0 dB.

In a second experiment, we tested the robustness of the proposed process face to noise. We added a gaussian white noise to the down-sampled data. The down-sampled noisy data, which is presented in Figures 1d and 1f, had an S/R of -1.0 dB. As in the first experiment, we used slope-DL to interpolate the data. The parameters were the same except for the error threshold  $\epsilon$ . Due to the presence of noise, we increased it to  $\lambda \sqrt{N} \sigma$  where  $\lambda$  was a gain factor set at 1.15,  $N$  was the number of samples in a patch, and  $\sigma$  was the standard deviation of the additive noise. The interpolated data is presented in Figures 1e and 1j. The S/R of the interpolated data is 23.4 dB.

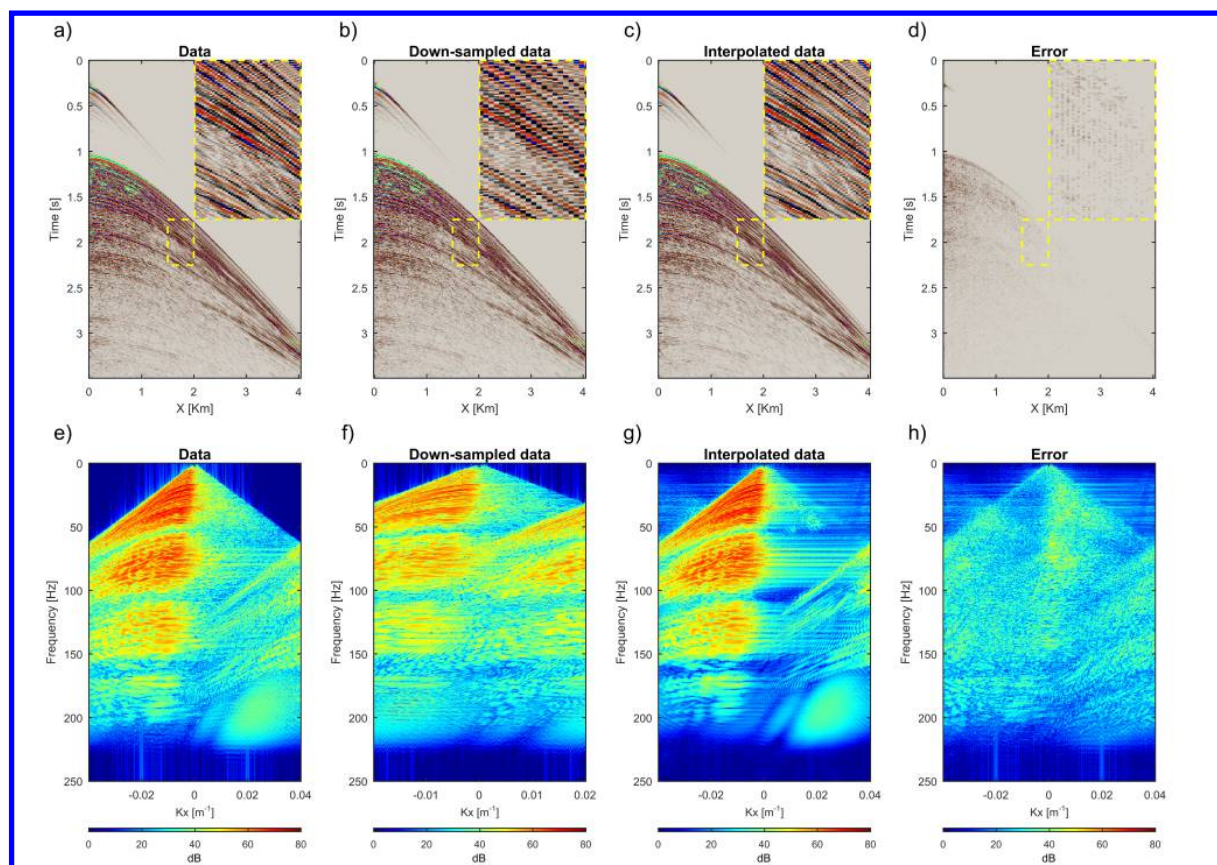


Figure 3: Interpolation of a shot gather using slope-DL. The recorded data (a,e), the down-sampled data (b,f), the interpolated data (c,g), and the error (d,h). In the first row, the data are shown in the time-space domain, and in the second row, their FK spectra are shown.

## FIELD DATA APPLICATION

We selected the pressure recording of a raw shot gather from a marine seismic data set. The sampling was at 2ms in the time dimension and at 12.5m in the offset dimension. The frequencies below 2Hz, and the data outside the signal cone were filtered out to remove the noise. This resulting gather is shown in the time-space domain in Figure 3a and its FK spectrum is shown in Figure 3e. The data were down-sampled at 25 m in the offset dimension (See Figures 3b and 3d). The slope-DL based interpolation process was applied on windows of size  $400 \times 70$  samples. The process for interpolating each window was similar to the one used in the synthetic examples. The interpolated data are presented in Figures 3c and 3g; the S/R is 16.5 dB. The interpolation error, e.g., difference between the interpolated data and the original data, is presented in Figures 3d and 3h. On the spectra, we observe that the interpolated data are aliased only from 60Hz, as are the original data; interpolation was hence beyond aliasing.

## CONCLUSION

We proposed a new dictionary learning method in which the learned atoms are constrained to represent linear events of

known slope. Thanks to this structure, the dictionary can be used to interpolate regularly sampled data beyond aliasing. It consist in finding a sparse representation of the data in the dictionary domain, interpolating the dictionary, and taking the sparse representation of the data in the interpolated dictionary domain. On the presented examples, we observe that this interpolation process performs a near to exact interpolation of linear events and an accurate interpolation of the curved events, and it is robust in the presence of noise. Further work includes constraining the DL problem to learn atoms representing curved events defined by slope and curvature parameters. This would improve the interpolation of the curved events.

## ACKNOWLEDGMENTS

We would like to express our gratitude to the PGS research team of Oslo, as well as to Valérie Maupin and Leiv Gellius, for the fruitful discussions on the subject, and we thank PGS for giving permission to publish this work.

## EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 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, *rmK*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation: IEEE Transactions on Signal Processing, **54**, 4311–4322, <https://doi.org/10.1109/TSP.2006.881199>.
- Crawley, S., 2001, Seismic trace interpolation with nonstationary prediction-error filters: Ph.D. thesis, Stanford University.
- Donoho, D., M. Elad, and V. Temlyakov, 2006, Stable recovery of sparse overcomplete representations in the presence of noise: IEEE Transactions on Information Theory, **52**, 6–18, <https://doi.org/10.1109/TIT.2005.860430>.
- Elad, M., 2010, Sparse and redundant representations: From theory to applications in signal and image processing, 1st ed.: Springer Publishing Company Inc.
- Engan, K., S. Aase, and J. Hakon Husoy, 1999, Method of optimal directions for frame design: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2443–2446.
- Gan, S., S. Wang, Y. Chen, Y. Zhang, and Z. Jin, 2015, Dealised seismic data interpolation using seislet transform with low-frequency constraint: 85th Annual International Meeting, SEG, Expanded Abstracts, 3900–3904, <https://doi.org/10.1109/LGRS.2015.2453119>.
- Gao, J., A. Stanton, M. Naghizadeh, M. D. Sacchi, and X. Chen, 2013, Convergence improvement and noise attenuation considerations for beyond alias projection onto convex sets reconstruction: Geophysical Prospecting, **61**, 138–151, <https://doi.org/10.1111/j.1365-2478.2012.01103.x>.
- Ibrahim, A., M. D. Sacchi, and P. Terenghi, 2015, Wavefield reconstruction using a stolt-based asymptote and apex shifted hyperbolic radon transform: 85th Annual International Meeting, SEG, Expanded Abstracts, 3836–3841, <https://doi.org/10.1190/segam2015-5873567.1>.
- Kutscha, H., D. E. Verschuur, and A. G. Berkhout, 2010, High resolution double focal transformation and its application to data reconstruction: 80th Annual International Meeting, SEG, Expanded Abstracts, 3589–3593, <https://doi.org/10.1190/1.3513595>.
- Naghizadeh, M., and M. Sacchi, 2010, Hierarchical scale curvelet interpolation of aliased seismic data: 80th Annual International Meeting, SEG, Expanded Abstracts, 3656–3661, <https://doi.org/10.1190/1.3513610>.
- Pati, Y. C., R. Rezaifar, 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.
- Rubinstein, R., M. Zibulevsky, and M. Elad, 2008, Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit.
- Schonewille, M., A. Klaedtke, and A. Vigner, 2009, Antialias antileakage fourier transform: 79th Annual International Meeting, SEG, Expanded Abstracts, 3249–3253, <https://doi.org/10.1190/1.3255533>.
- Spitz, S., 1991, Seismic trace interpolation in the *F-X* domain: Geophysics, **56**, 785–794, <https://doi.org/10.1190/1.1443096>.