CWP-1007

Multi-channel compressive sensing for seismic data reconstruction using joint sparsity

Iga Pawelec*, Michael Wakin† & Paul Sava*
*Center for Wave Phenomena and Dept. of Geophysics, Colorado School of Mines, Golden CO 80401
email ipawelec@mines.edu
† Dept. of Electrical Engineering, Colorado School of Mines, Golden CO 80401

ABSTRACT
Despite the many advances in data reconstruction technology and significant increase in the number of channels available for recording the particle motion, acquiring non-aliased seismic data at high signal-to-noise ratio remains a challenge. Areas with restricted access, difficult terrain, and slow velocities in the near surface can prevent the acquisition of properly sampled wavefields which in turn significantly complicates the suppression of the near surface related noise.
There are two popular strategies available for reducing the sampling requirement without loss of information. Recording wavefield and its derivatives yields multiple pieces of information at each sampling point, creating a multi-channel signal and allowing for an increased distance between samples. Compressive sensing (CS) is an alternative way of data acquisition relying on randomized sampling and known data patterns in some domain to reconstruct a fully sampled seismic data volume from reduced measurements. In this report, we use simple 1D examples to build the intuition about the probabilistic nature of compressive sensing and its multi-channel extension. We show that providing additional information (such as signal derivative), yields higher probability of successful signal recovery and allows for more drastic reduction of sampling than a single-channel CS. We test the developed multi-channel approach on a synthetic 3D seismic shot and show that multi-channel reconstruction achieves higher SNR and yields fewer coherent artifacts compared to its single channel counterpart.

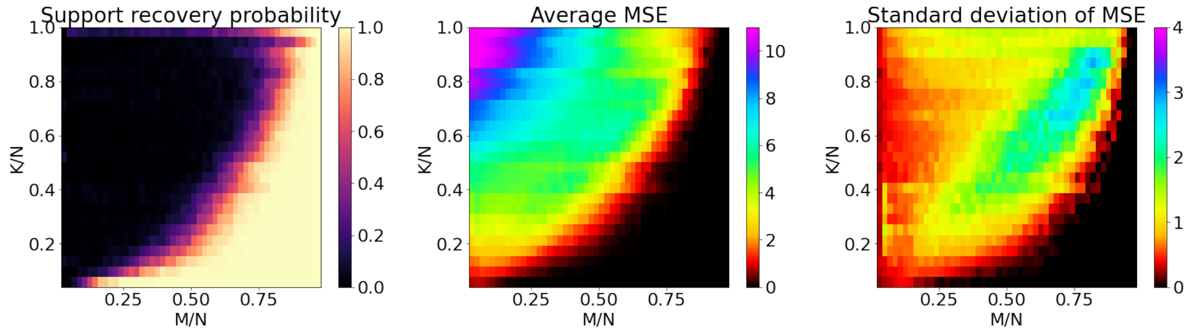
Key words: compressive sensing, generalized sampling, data reconstruction

1 INTRODUCTION

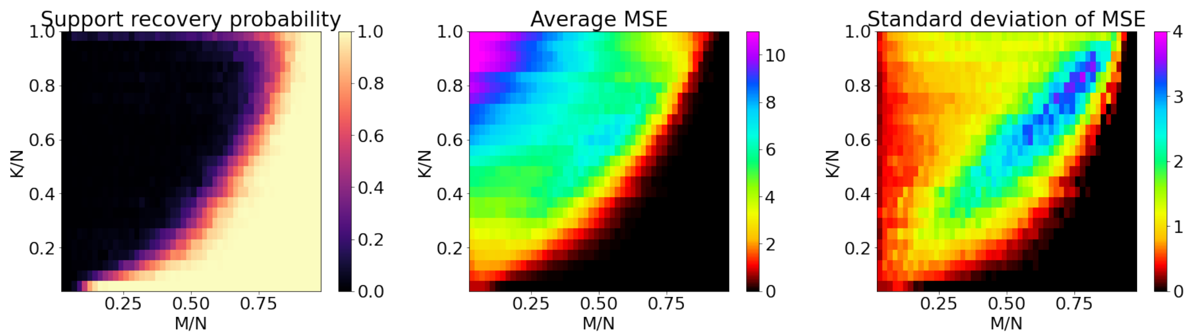
Obtaining non-aliased seismic data with high SNR can be a challenge, particularly in difficult terrain or in areas with access restrictions. The conventional approach to seismic data acquisition relies on recording the particle motion on a regular grid, with spacing between stations dictated by the maximum frequency of the signal and the slowest velocity (water velocity for marine acquisition, or near surface velocities for land acquisition). However, due to terrain obstacles and slow near surface velocities (sometimes $<$ 200 m/s), regular sampling on land can be economically infeasible or even impossible.

In light of the challenges associated with regularly sampled data, there are many available techniques to correct for deviations from regular grids and mitigating the aliasing. A large number of these techniques use the Fourier representation of seismic data. For instance, Liu and Sacchi (2004) develop a framework for data recovery based on L_2 norm minimization, using spectral weights bootstrapped from FK representation of data. Similar strategy extended to five dimensions (Trad, 2009) is even more successful because data in higher dimensional spaces tend to have more compact representations (they achieve higher sparsity level) and thus are easier to reconstruct. Duijndam et al. (1999) tackle the problem of arbitrarily irregular sampling and leverage a weighting scheme based on adjacent sample distances to reconstruct data with one varying spatial coordinate, while Xu et al. (2005, 2010) propose an antileakage version of the Fourier transform that can both handle irregular geometry and mitigate aliasing problems.

The advent of wireless nodal acquisition shifts the way we think about land data acquisition. An increasingly popular alternative to regular sampling is deliberately randomized acquisition which exploits the data patterns in some domain (Mosher et al.,



(a)



(b)

Figure 1. The recovery probability, average mean-squared error and its standard deviation for (a) single channel and (b) multi-channel recovery of a 1D signal sparse in the Fourier domain. K/N is the fraction of non-zero coefficients needed to describe the signal, while M/N is the fraction of available samples from the fully sampled signal. The transition between successful and failed recovery is clear for both scenarios, but reconstructions are more consistent for the multi-channel case.

2017; Allegar et al., 2017; Jiang et al., 2019). Assuming that these patterns can be approximated as sparse, the compressive sensing (CS) theory offers many strategies for successful recovery of fully sampled signals on a regular grid from sub-Nyquist number of sampling points. The challenge for compressive sensing land seismic data acquisition is to find a data representation that captures all features of raw records (including amplitudes spanning several orders of magnitude and instantaneous phase) in what can be considered a sparse form. As shown in Figure 1, there is a trade-off between data sparsity level and the likelihood of successful reconstruction. The less sparse the pattern, the more measurements are needed to ensure success. In Pawelec et al. (2021), we demonstrate compressive sensing recovery for a complex raw land seismic record that reiterates the intuition that the gap pattern and the number of measurement points is critical: the success of recovery depends on a specific realization of random sampling geometry.

We propose to use spatial derivatives of particle motion in addition to the particle motion signal within the compressive sensing framework to further reduce the number of measurement points or improve the SNR of the recovered data. Following the terminology from the signal processing community, we use the term ‘multi-channel’ to refer to multiple measurements available at the same location. We develop a framework for simultaneous sparse approximation of wavefields and their derivatives using the iterative hard thresholding algorithm (Blumensath and Davies, 2008). Our preliminary synthetic results show that the multi-channel approach yields reliable reconstructions in terms of event continuity and signal-to-noise ratio and is superior to the single-channel approach in that respect. However, our analysis indicates that the mulit-channel approach does not always guarantee the superior reconstruction of missing data, with the reconstruction quality dependent on specific sampling pattern.

2 WAVEFIELD DERIVATIVES

Following the traditional sampling theorem (Shannon, 1948; Jerri, 1977) and its extension to multidimensional signals (Petersen and Middleton, 1962) we can recover signals without the loss of information provided that all measurement points on the sampling grid are available. This approach to sampling can be extended for multi-channel signals. Linden (1959) provides a reconstruction

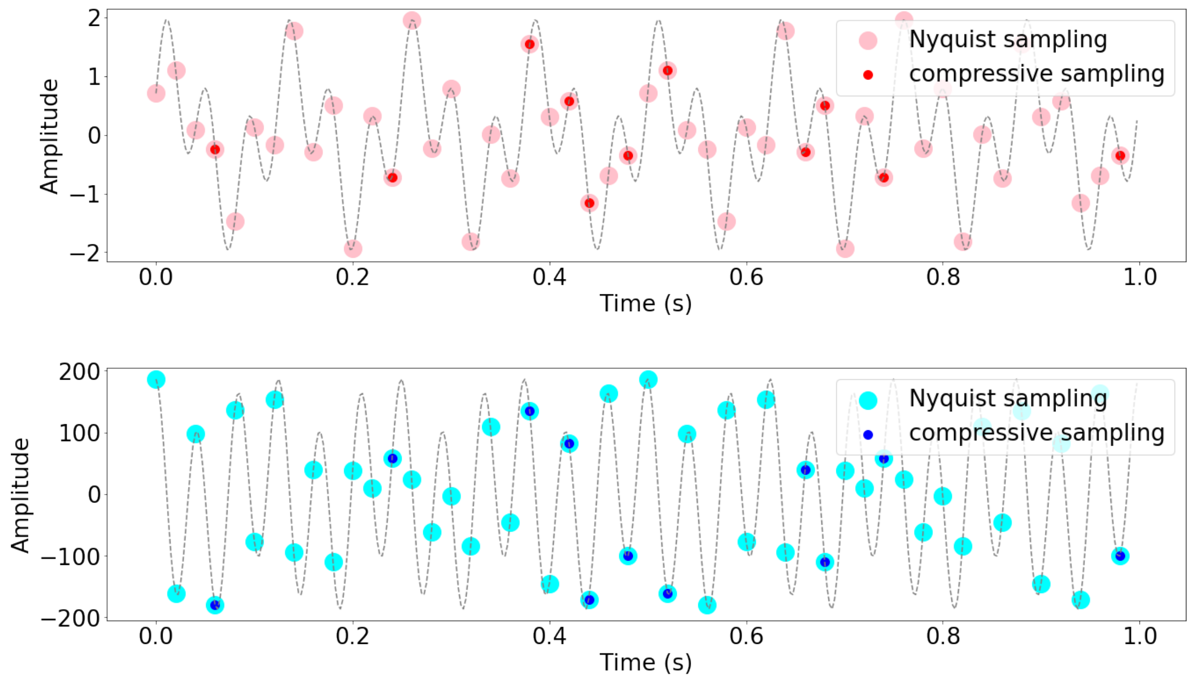


Figure 2. 1D signal composed of two sinusoids recorded with a multi-channel sampling system: by taking discrete samples of a signal (top) and its first derivative (bottom). Large light circles on both panels indicate Nyquist sampling locations while small dark circles show one realization of compressive sampling. Note that compressive samples consist of only only 22% of the Nyquist samples.

formula for simultaneous sampling of a band-limited function and its derivative. Papoulis (1977) generalizes that result showing that a band-limited function $f(t)$ can be uniquely described by samples of m linear systems with input $f(t)$ sampled at $\frac{1}{m}$ the Nyquist rate. Cheung (1993) extends the theory to multidimensional signals. In the Appendix, we show the derivation of a 2D generalized sampling expansion (GSE) for the orthogonal sampling matrix. Note that to increase the distance between the samples by a factor of two, four channels are required.

An interesting application of generalized sampling expansion is presented by Robertsson et al. (2008). They use it to mitigate the issue of coarse crossline sampling in towed streamer data. The pressure gradient is computed from three-component measurements of particle velocity and then combined with independently recorded pressure, providing a two channel measurement at each sampling location. Paired with some simple data processing, such an approach allows one to recover data up to three times the Nyquist wavenumber compared to the single channel measurements.

A natural question arises: is it possible to push the GSE method for seismic data reconstruction even further? Vassallo et al. (2010) show examples of unraveling multiply aliased data relying on the known velocity of the water and predicting aliasing patterns for pressure and its derivative. Their approach, however, cannot be readily applied in land acquisition because the near surface velocities are usually unknown a priori, are rarely constant, and vary as a function of space.

Working with onshore data and with access to spatial derivatives of the wavefield, Muyzert et al. (2019) recover land records with the ground roll aliased up to three times, similar to the results shown by Robertsson et al. (2008). Pushing beyond that result on land may require a different approach. Since we cannot rely on known aliasing patterns and obtaining higher order derivatives can be challenging due to noise considerations, a possible solution is to explore multi-channel recordings in a compressive sensing framework.

Compressive sensing is known for being able to recover sparse signals from significantly reduced measurements, provided that the sampling strategy and sparse domain are carefully selected. Supplementing the compressive sensing framework with the derivative information should, in principle, push the limits of what is possible with one channel only - or with regularly sampled multi-channel signals. Since differentiation is a linear operation, wavefield derivatives have the same bandwidth as the original data, providing an important commonality between different channels, i.e. a shared Fourier domain support. This allows for the use of derivatives in the Fourier-based multi-channel compressive sensing framework by exploiting the idea of joint sparsity.

3 MULTI-CHANNEL COMPRESSIVE SENSING

3.1 Single-channel compressive sensing

Compressive sensing is a sampling paradigm allowing for recovery of signals from incomplete measurements (Candès et al., 2006). The key assumption is that the target signal is or can be approximated as sparse in some representation. In particular, an $N-$length $K-$sparse signal can be expressed using only K non-zero coefficients, usually with $K \ll N$. Recovering that signal means finding the signal support – locations of non-zero coefficients – and values of the coefficients at these locations.

Successful recovery of $K-$sparse or compressible signal depends on three key components: the sampling strategy, the data sparsifying transform, and the sparsity-promoting recovery algorithm. Results from compressive sensing suggest that sparse signals can be recovered without loss of information if the sampling matrix satisfies the restricted isometry property (RIP) (Baraniuk, 2007). RIP is satisfied with high probability for Gaussian matrices (each entry is independent and follows a normal distribution) and random Bernoulli matrices (entries are ± 1 with equal probability) or when sampling non-uniformly Fourier-sparse signals. Depending on the choice of the sampling matrix, the number of measurements to recover a K-sparse signal is $M = \mathcal{O}(K \log(N/K))$.

Let the data in sparse domain be represented as

$$\mathbf{d} = \mathbf{S}\mathbf{\Phi}^H \boldsymbol{\alpha}, \tag{1}$$

where $\mathbf{d}$ is the recorded wavefield, $\mathbf{S}$ is the sampling matrix, $\mathbf{\Phi}$ is the sparsifying transform, and $\boldsymbol{\alpha}$ are the signal coefficients in the sparse domain. From a practical point of view, only the non-uniform sampling (i.e., either placing the receiver and/or source on the grid or skipping the sampling location altogether) can be achieved for seismic acquisition. Although the Fourier domain is not optimal for sparsely representing seismic data, it is possible to obtain a data window that can be approximated as Fourier-sparse with a clever combination of sorting, pre-processing, and windowing. Furthermore, the Fourier domain is natural to consider for extending to the multi-channel case given the relationship between signal and its derivative: $f'(t) = \mathcal{F}^{-1}(i\omega F(\omega))$.

3.2 Extension to the multi-channel case

In the joint sparse recovery problem, also known as simultaneous sparse approximation or multiple measurement vector problem, rather than dealing with one sparse signal, we are attempting to recover an ensemble of p signals:

$$\mathbf{d}_i = \mathbf{S}\mathbf{\Phi}^H \boldsymbol{\alpha}_i \; i = 1, \ldots, p \tag{2}$$

Figure 2 shows an example of a 1D signal composed of two sinusoids and recorded with two channels: by recording samples of signal and its derivative. Each signal is individually sparse in the Fourier domain (Figure 3), but there is also a relationship between the channels. That relationship can be theoretical, as is the case here, or statistical. Formally, the interrelations between channels are described by a joint sparsity model (Duarte et al., 2005). The model we consider in this paper is the common sparse support, where each individual signal coefficient vector has the same support, but the coefficient values can differ.

Consider the matrix $\mathbf{A} = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_p]$. If all $\boldsymbol{\alpha}_i$ share the same support and are K-sparse, then $\mathbf{A}$ contains $N - K$ zero rows. Many algorithms are available for recovering this type of signals. They include simultaneous orthogonal matching pursuit (Tropp et al., 2006), convex relaxation (Tropp, 2006), subspace-based methods (Lee et al., 2012), deep learning (Palangi et al., 2016) and Bayesian approaches (Wipf and Rao, 2007; Chen et al., 2016). Due to the ease of implementation we consider a modification of the iterative hard thresholding (IHT) (Blumensath and Davies, 2008), as described next.

4 STRATEGY FOR JOINT SPARE RECOVERY

In the case of single channel reconstruction, we are interested in solving the following K-sparse optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{d} - \mathbf{S}\mathbf{\Phi}^H \boldsymbol{\alpha}\|_2^2 \text{ subject to } \|\boldsymbol{\alpha}\|_0 \leq K, \tag{3}$$

where $\|\cdot\|_0$ refers to the number of non-zero entries. This problem can be solved with the following iterative algorithm:

$$\boldsymbol{\alpha}^{n+1} = T_K\Big(\boldsymbol{\alpha}^n + \mathbf{\Phi}(\mathbf{d} - \mathbf{S}\mathbf{\Phi}^H \boldsymbol{\alpha}^n)\Big), \tag{4}$$

with the non-linear thresholding operator T_K retaining only K coefficients with the largest magnitude

$$T_K(\alpha_i) = \begin{cases} 0 \text{ if } |\alpha_i| < \lambda_K(\boldsymbol{\alpha}), \\ \alpha_i \text{ if } |\alpha_i| \geq \lambda_K(\boldsymbol{\alpha}). \end{cases} \tag{5}$$

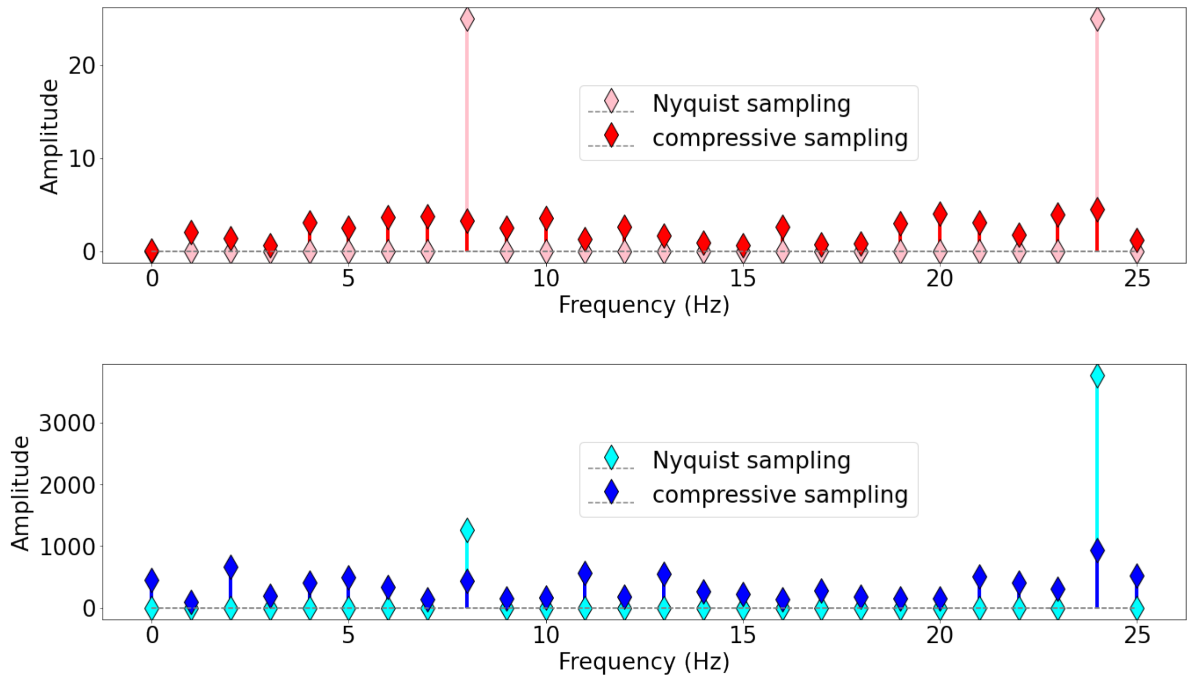


Figure 3. Fourier domain representation of the signal shown in Figure 2. Note that severe decimation in compressive sampling makes it challenging to accurately estimate the true support of the underlying signal (8 and 24 Hz in this example).

The threshold $\lambda_K(\boldsymbol{\alpha})$ is set to K-th largest absolute value of $\boldsymbol{\alpha}^n + \mathbf{\Phi}(\mathbf{d} - \mathbf{S}\mathbf{\Phi}^H\boldsymbol{\alpha}^n)$. For jointly sparse signals with different physical units, we reformulate the problem from equation 3 into

$$\min_{\boldsymbol{\alpha_1}\cdots\boldsymbol{\alpha_p}} \sum_{i=1}^{p} \|\mathbf{d}_i - \mathbf{S}\mathbf{\Phi}^H\boldsymbol{\alpha}_i\|_2^2 \text{ subject to } \|\sum_{i=1}^{p}\mathbf{W}^i\boldsymbol{\alpha}_i\|_0 \leq K. \tag{6}$$

In this formulation, we aim to honor the acquired signal samples while promoting $K-$sparse joint support. To estimate this support, the Fourier representations of the respective channels are combined into one. This is achieved by weighing the respective spectra such that all spectral coefficients corresponding to the first channel (signal samples) remain unchanged ($\mathbf{W}^1$ is an identity matrix). For the remaining channels, $\mathbf{W}$ is a diagonal matrix with entries defined as

$$\mathbf{W}_i^n = \frac{1}{j\omega_i + \epsilon_n},\ \epsilon_n = \frac{\|\mathbf{\Phi}\mathbf{d_n}\|_2}{\|\mathbf{\Phi}\mathbf{d_1}\|_2}, \tag{7}$$

where ω_i represents cycles in appropriate physical units (for example, $\omega_i = 2\pi f_i$ for temporal signals). The purpose of the weighting term is thus to partially 'undo' the derivative operation, using ϵ as a damping factor that also balances the relative contribution of derivative channels with respect to data channel. An example of the weighting function for a 1D time signal is depicted in the top panel of Figure 4. With the weight applied to the derivative channel (dark blue diamonds) its spectral coefficients have similar magnitudes to the coefficients of the signal channel (red diamonds). Adding these two together (black diamonds) yields a joins spectrum whose top two highest energy coefficients are at the same frequencies as for the underlying sparse signal. Correct support estimation is the key in successful signal recovery, as discussed next.

5 NUMERICAL EXPERIMENTS

To develop intuition about the performance of simultaneous sparse approximation compared to single-channel compressive sensing, we design a simple numerical experiment with a two channel recording system. Similar to Figure 2, we define signal as the sum of K sinusoids with different frequencies and random phase shifts:

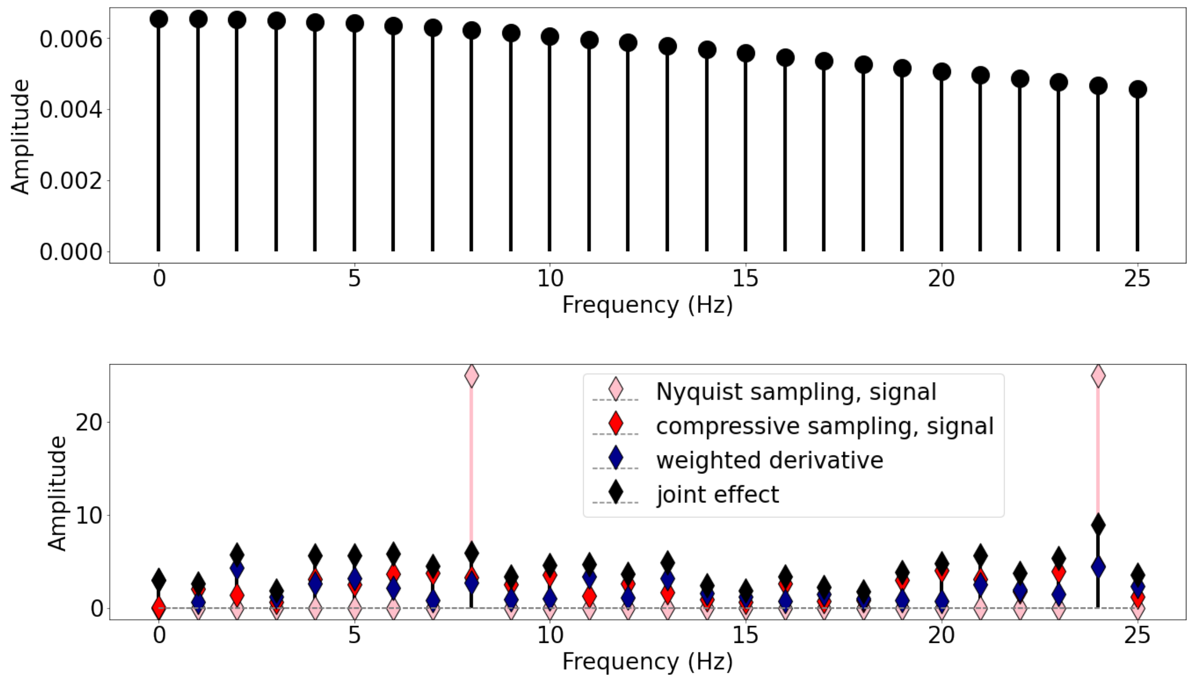


Figure 4. Weights applied to the derivative channel (top) and the effect of combining compressive signal samples with weighed compressive derivative samples. Note that in this instance, the combined signal spectrum yields highest energy coefficients at 8 and 24 Hz, the same as the Nyquist sampled signal. However, this property may not hold for other realizations of signal sampling.

$$f(t) = \sum_{i=1}^{K} \sin(2\pi f_i t + \phi_i) \tag{8}$$

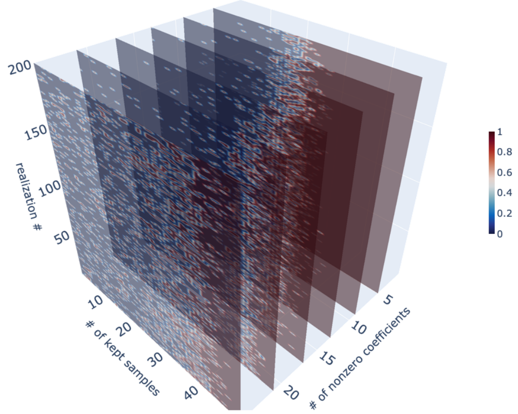
$$f'(t) = \sum_{i=1}^{K} 2\pi f_i \cos(2\pi f_i t + \phi_i), \tag{9}$$

where f_i denotes the temporal frequency for the sinusoid. Note that if f_i is set such that it falls on the DFT grid, its Fourier domain representation is perfectly sparse: the only non-zero coefficients correspond to a bin representing $\pm\omega_i = \pm 2\pi f_i$. Furthermore, because the derivative of sine is a cosine, the Fourier support for the derivative is exactly the same as for the original signal. Thus, it makes for a perfect testing case for multi-channel compressive sensing reconstruction using common sparse support as a joint sparsity model.

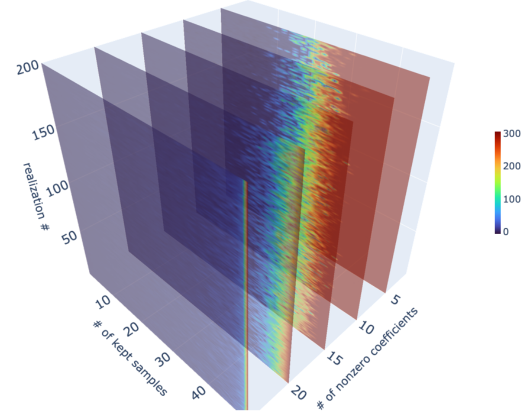
In our test, we assume that all sparse signal components have equal strength (i.e., the amplitude of the sinusoid is always the same). The compressive sampling of both channels is done by selecting a fixed number of samples M from a Nyquist-sampled $N-$length signal (like in Figure 2) uniformly at random. Since there are $\binom{N}{M}$ ways to select sampling geometry meeting these criteria, we run 200 realizations for each M in an effort to capture the probabilistic nature of the problem.

Figure 5 shows the success rate of signal reconstruction between different signal sparsity levels, numbers of kept samples, and realizations. The success is defined as recovering the correct Fourier domain support. Note that the successful recoveries also have very high values of SNR, often as high as 300 dB, indicating signal recovery within the numerical precision. That is only possible when signals are exactly $K-$sparse and are not contaminated by noise. The volumes also highlight the probabilistic nature of compressive sensing. There is a clear and sharp transition between success and failure, that is best captured by summary statistics derived from all realizations. Figure 1 is an example of such: the probability of successful recovery is averaged over the number of realizations, and the mean-squared-error is used to assess the accuracy of each reconstruction. We find that the multi-channel approach is successful with fewer samples than the single-channel approach, but behaves more erratically in the transition zone between 100% success and 100% failure rate.

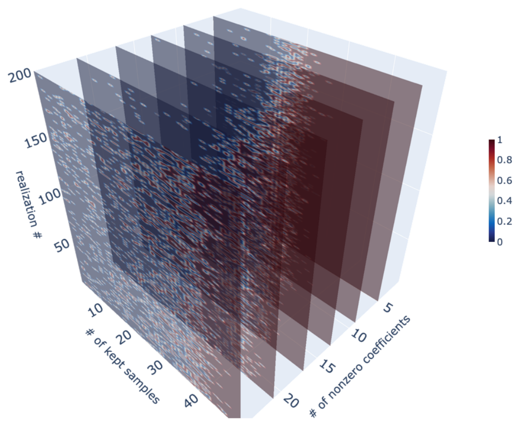
Armed with the developed intuition, we test the applicability of multi-channel compressive sensing to seismic data reconstruction. Unlike the previous experiment, the simulated seismic data are not perfectly sparse in the Fourier domain. However, if



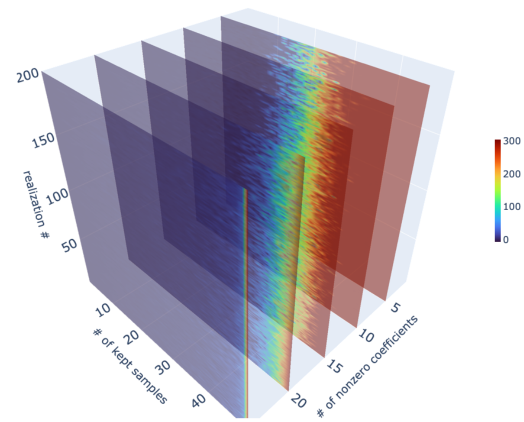
(a)



(b)



(c)



(d)

Figure 5. Quality metrics for single- and multi-channel compressive sensing reconstruction (top and bottom rows, respectively). (a) and (c) are the probabilities of recovering the correct Fourier domain support while (b) and (d) show the quality of the reconstructed signal quantified by SNR.

we consider a 3D volume, we can approximate the data as sparse, thus making them $K-$compressible. Figure 6 shows two slices through a densely sampled 3D seismic volume with all available traces at the top and compressively sampled data at the bottom. The decimation ratio for this experiment is 80%, and due to the computational cost involved, we only consider one realization of the missing trace geometry. Figures 7 and 8 show the reconstruction results for single- and multi-channel approaches, respectively. The multi-channel reconstruction yields results with higher SNR (12.69 dB compared to 11.60 dB) and better event continuity. The near-offset reconstruction artifacts stemming from the difficulty of approximating fast amplitude decay as sparse are also reduced for the multi-channel case. Additionally, the multi-channel aproach reconstructs all channels while enforcing the common sparse support constraint (Figure 9). Maintaining this consistency between the channels is particularly valuable when exploring the use of wavefield derivatives for other applications, such as denoising or mode separation.

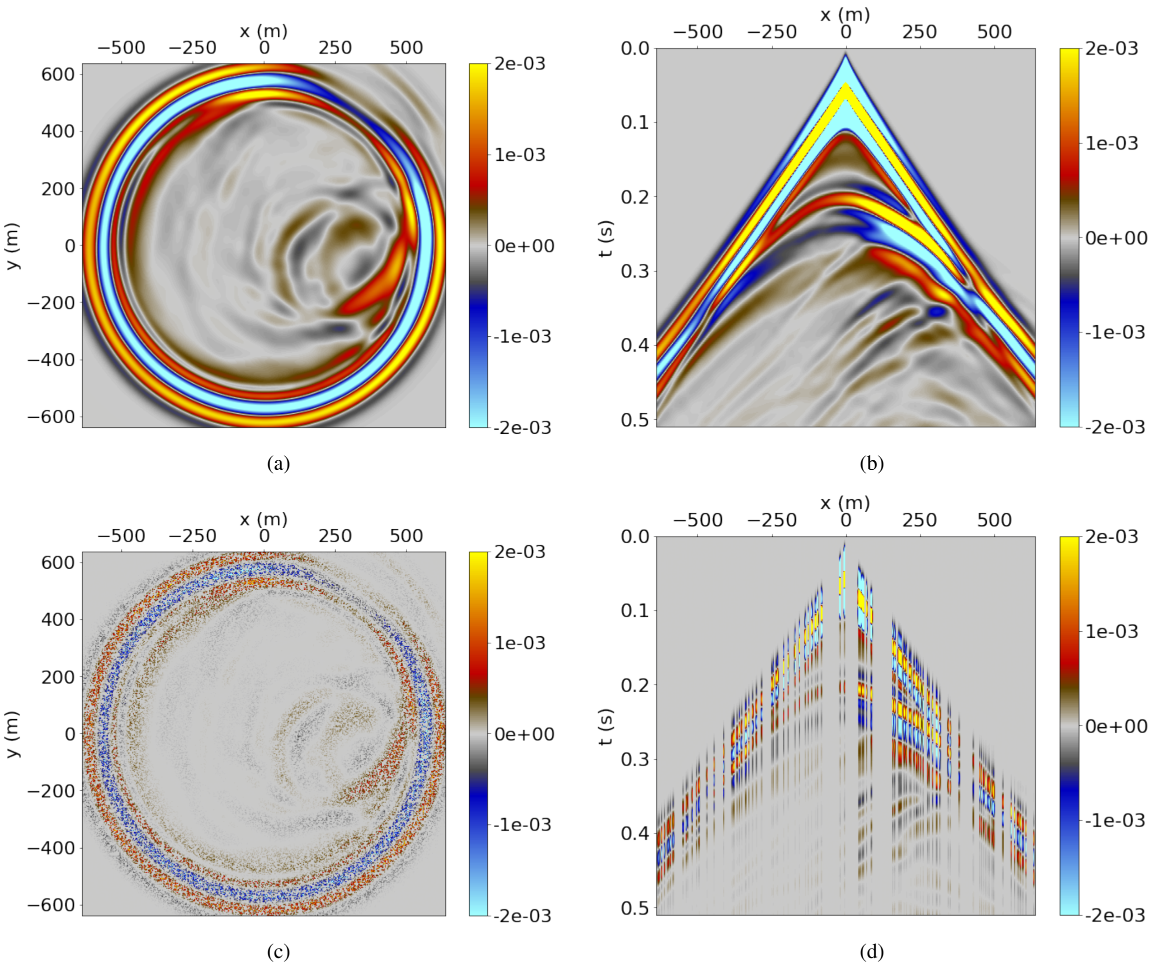


Figure 6. Reconstruction of coarsely sampled seismic data. (a), (b) are the slices from densely sampled 3D volume and (c), (d) are the compressive samples after 80% decimation.

6 DISCUSSION

An important aspect of compressive sensing framework applied to sparse or compressible signals is the probabilistic nature of the recovery success. When the signal sampling is realized as a random selection of a subset from the Nyquist samples, the successful recovery depends on the sparsity ratio and the fraction of preserved samples. As demonstrated in Figure 1, the transition between success and failure is sharp - both for the single- and multi-channel CS. Thus far, we established that for a fixed signal sparsity ratio, the multi-channel approach allows for fewer sampling points than its single-channel counterpart. However, while adding the derivative information tends to help signal recovery most of the time, there are instances where the opposite is true. This is due to the interaction between the sampling strategy and signals recorded on specific channels. If the data decimation with respect to the Nyquist sampling is severe enough, the accurate estimation of signal support may become infeasible. The effect of compressive sampling on the signal and its derivative is not the same. The introduction of the weighting term helps to balance the contributions of respective channels but on rare occasions can steer the solution in the wrong direction.

The performance of the modified IHT algorithm can be improved by preconditioning the input data or introducing a staged recovery approach, starting with a very small sparsity level K and slowly increasing it during the iterations. That would overcome one significant limitation of the current formulation remains: the prior knowledge of the sparsity level. In practice, the more sampling points are missing, the higher the sampling noise introduced to the Fourier domain. Consequently, one may need to opt for a sparser signal approximation than what would be optimal. This in turn means that events with weaker amplitudes are poorly reconstructed or altogether missing. Alternative reconstruction methods which relax the sparsity requirement from $\|\cdot\|_0$ to $\|\cdot\|_1$ may avoid this shortcoming.

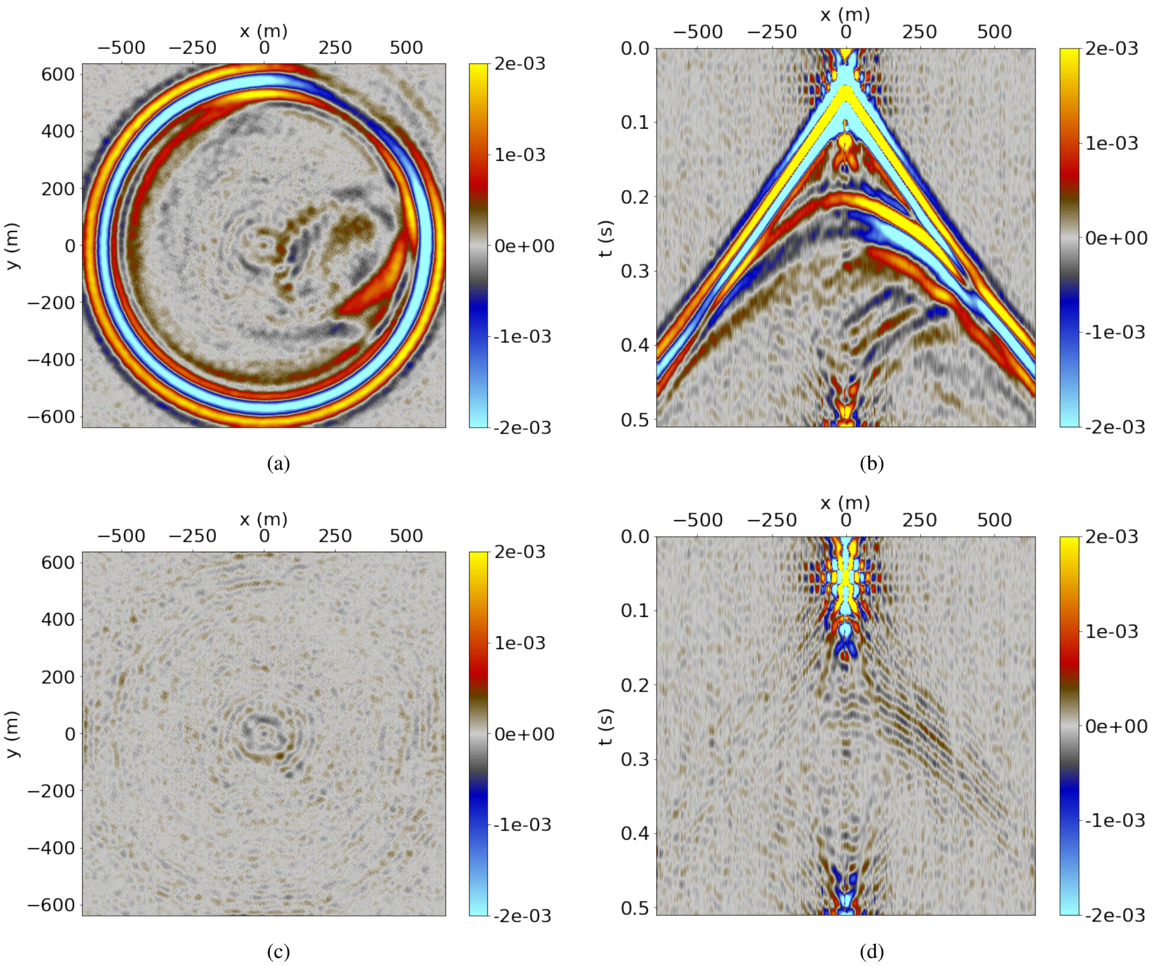


Figure 7. Single-channel reconstruction for the wavefield depicted in Figure 6 and the corresponding differences. The SNR for the recovered volume is 11.60 dB.

7 CONCLUSIONS

Even though seismic data are not necessarily sparse in the Fourier domain, they can be approximated as such and thus considered compressible. Thus, our proposed approach for multi-channel reconstruction can be readily applied in realistic land seismic acquisition scenarios. On average, using wavefield derivative information for compressive sensing reconstruction improves the quality of data reconstruction and lowers the demand on the number of sampling points. A weighted combination of signal vectors with the same support in the sparse domain helps to strengthen the signal from the true support set while decreasing the impact of sampling noise. This allows for more accurate wavefield reconstruction by multi-channel CS with fewer sampling artifacts compared with single-channel CS. Increasing the dimensionality of the problem tends to improve the results because it also increases the sparsity ratio, thus reducing the demand on the requisite number of sampling points.

Additionally, the multi-channel CS simultaneously reconstructs wavefield derivatives on the full grid which may provide interesting opportunities for noise suppression. With more research to establish performance limits and best practices in field data application, multi-channel CS using the joint sparsity is a promising technique that could bring us closer to solving the bad land data challenge.

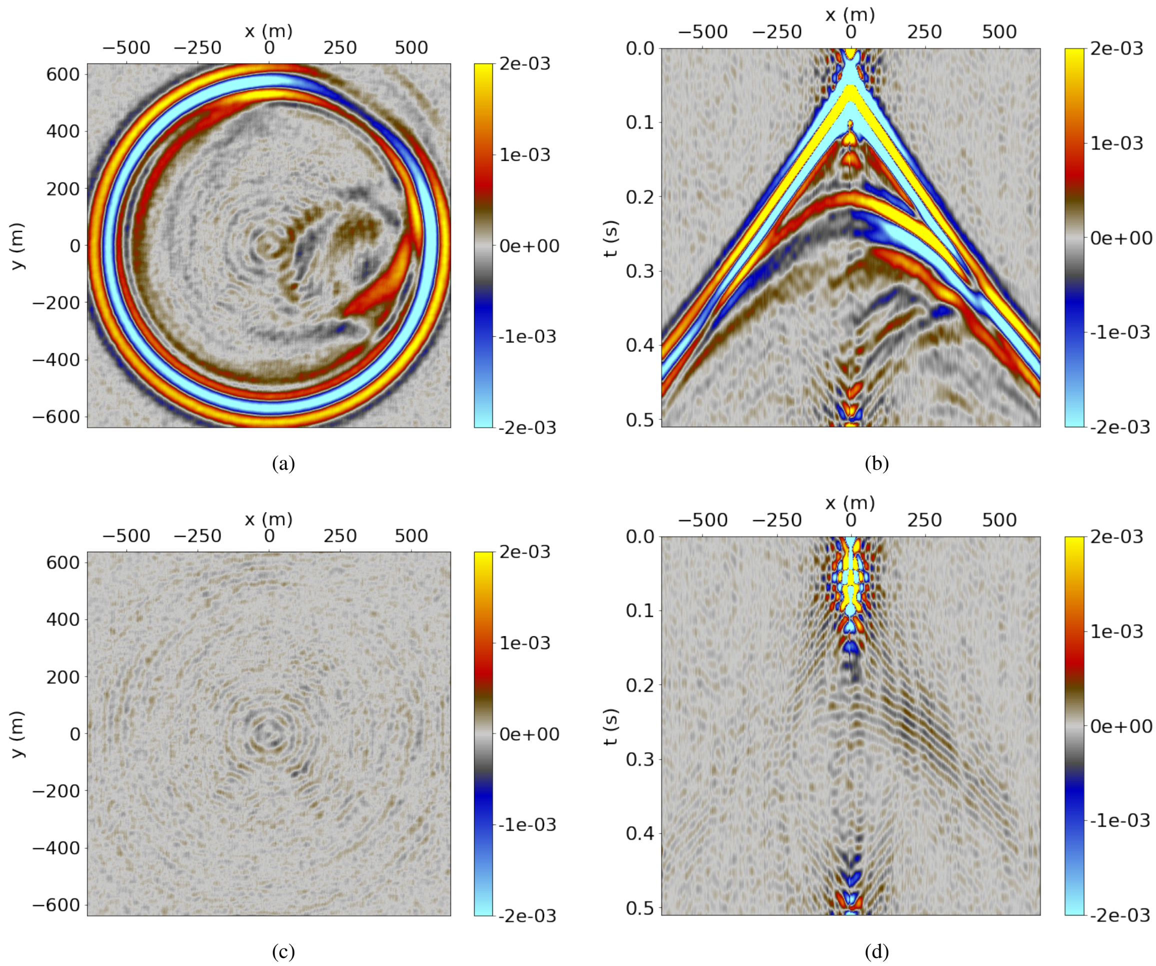


Figure 8. Multi-channel reconstruction for the wavefield depicted in Figure 6 and the corresponding data differences. The SNR for the recovered volume is 12.69 dB.

APPENDIX

Deriving 2D GSE

N-D signal can be represented by samples of itself and its filtered versions. The interpolation formula is

$$f(\mathbf{x}) = \sum_{i=0}^{L-1} \sum_{\mathbf{n}} g_i(\mathbf{V}_g \mathbf{n}) y_i(\mathbf{x} - \mathbf{V}_g \mathbf{n}). \tag{10}$$

L is the number of linear systems. To reduce the sampling of N-dimensional system to $\frac{1}{m}$ Nyquist, $L = m^N$ linear filters are needed. g_i are sample values. y_i are the interpolation functions corresponding to the respective g_i. $\mathbf{V}$ is the ND sampling martix whose columns correspond to sampling vectors in i-th direction. $\|\mathbf{v}_i\|$ is the sampling interval in the i-th direction.

$$\langle \mathbf{v}_j, \mathbf{u}_k \rangle = 2\pi \delta_{jk} \tag{11}$$

Let us define a 2D signal in its native domain, with sampling on rectangular grid such that $\Delta x = \frac{1}{2k_x}$ and $\Delta y = \frac{1}{2k_y}$. Then we have

$$\mathbf{V} = \begin{bmatrix} \frac{1}{2k_x} & 0 \\ 0 & \frac{1}{2k_y} \end{bmatrix} \text{ and } \mathbf{U} = \begin{bmatrix} 4\pi k_x & 0 \\ 0 & 4\pi k_y \end{bmatrix}. \tag{12}$$

Set $m = 2$. Then the sampling and periodicity matrices become:

$$\mathbf{V_g} = \begin{bmatrix} \frac{1}{k_x} & 0 \\ 0 & \frac{1}{k_y} \end{bmatrix} \text{ and } \mathbf{U_g} = \begin{bmatrix} 2\pi k_x & 0 \\ 0 & 2\pi k_y \end{bmatrix}. \tag{13}$$

Now, we need $L = 4$ linear systems to generate 4 sample sets. Let them be:

$$H_0(2\pi v_x, 2\pi v_y) = 1 \tag{14}$$

$$H_1(2\pi v_x, 2\pi v_y) = j2\pi v_x \tag{15}$$

$$H_2(2\pi v_x, 2\pi v_y) = j2\pi v_y \tag{16}$$

$$H_3(2\pi v_x, 2\pi v_y) = -2\pi v_x 2\pi v_y. \tag{17}$$

Sampling density D_g of each sample set is $det(\mathbf{U_g}) = 4\pi^2 k_x k_y$. To obtain interpolation functions y_i, we need to solve the following system:

$$\mathbf{H}^T \mathbf{y} = \mathbf{e},$$

where $H_{k,i} = H_k(\boldsymbol{\omega} + \mathbf{U_g}\mathbf{q}_i)$, with $\mathbf{q}_i$ being $k-$ary representation of integer i, and with the carrier vector entries $e_i = \mathrm{e}^{j\mathbf{q}_i^T \mathbf{U_g}^T \mathbf{x}}$. The interpolation functions are then obtained by

$$y_i(\mathbf{x}) = \frac{1}{D_g} \int_{\mathcal{C}_{g0}} Y_i(\boldsymbol{\omega}, \mathbf{x}) \mathrm{e}^{j\boldsymbol{\omega}^T \mathbf{x}} d\boldsymbol{\omega}. \tag{18}$$

$$(\mathbf{H}^T)^{-1} = \frac{1}{4\pi^2 k_x k_y} \begin{bmatrix} (2\pi v_x + 2\pi k_x)(2\pi v_y + 2\pi k_y) & -2\pi v_x(2\pi k_y + 2\pi v_y) & -2\pi v_y(2\pi k_x + 2\pi v_x) & 2\pi v_x 2\pi v_y \\ j(2\pi k_y + 2\pi v_y) & -j(2\pi k_y + 2\pi v_y) & -j2\pi v_y & j2\pi v_y \\ j(2\pi k_x + 2\pi v_x) & -j2\pi v_x & -j(2\pi k_x + 2\pi v_x) & j2\pi v_x \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

The carrier vector in this instance is

$$\mathbf{e} = \begin{bmatrix} 1 & e^{j2\pi k_x x} & e^{j2\pi k_y y} & e^{j2\pi(k_x x + k_y y)} \end{bmatrix}^T \tag{19}$$

Thus, we can compute Fourier domain representation of interpolation functions $\mathbf{y}$:

$$\begin{aligned}
Y_0 &= \frac{1}{4\pi^2 k_x k_y}\Big((2\pi v_x + 2\pi k_x)(2\pi v_y + 2\pi k_y) - e^{j2\pi k_x x} 2\pi v_x(2\pi k_y + 2\pi v_y) - e^{j2\pi k_y y} 2\pi v_y(2\pi k_x + 2\pi v_x) \\
&\quad + e^{j2\pi(k_x x + k_y y)} 2\pi v_x 2\pi v_y\Big) \\
&= \frac{1}{4\pi^2 k_x k_y}\Big(2\pi v_x 2\pi v_y + 2\pi k_y 2\pi v_x + 2\pi k_x 2\pi v_y + 4\pi^2 k_x k_y - 2\pi k_y e^{j2\pi k_x x} - 2\pi v_y e^{j2\pi k_x x} \\
&\quad - 2\pi k_x e^{j2\pi k_y y} - 2\pi v_x e^{j2\pi k_y y} + 2\pi v_x 2\pi v_y e^{j2\pi(k_x x + k_y y)}\Big) \\
Y_1 &= \frac{j}{4\pi^2 k_x k_y}\Big(2\pi k_y + 2\pi v_y - (2\pi k_y + 2\pi v_y) e^{j2\pi k_x x} - 2\pi v_y e^{j2\pi k_y y} + 2\pi v_y e^{j2\pi(k_x x + k_y y)}\Big) \\
&= \frac{j}{4\pi^2 k_x k_y}\Big(2\pi k_y + 2\pi v_y - 2\pi k_y e^{j2\pi k_x x} - 2\pi v_y e^{j2\pi k_x 2\pi v_x} - 2\pi v_y e^{j2\pi k_y 2\pi v_y} + 2\pi v_y e^{j2\pi(k_x x + k_y y)}\Big) \\
Y_2 &= \frac{j}{4\pi^2 k_x k_y}\Big(2\pi k_x + 2\pi v_x - 2\pi v_x e^{j2\pi k_x x} - (2\pi k_x + 2\pi v_x) e^{j2\pi k_y 2\pi v_y} + 2\pi v_x e^{j2\pi(k_x x + k_y y)}\Big) \\
&= \frac{j}{4\pi^2 k_x k_y}\Big(2\pi k_x + 2\pi v_x - 2\pi v_x e^{j2\pi k_x x} - 2\pi k_x e^{j2\pi k_y y} - 2\pi v_x e^{j2\pi k_y y} + 2\pi v_x e^{j2\pi(k_x x + k_y y)}\Big) \\
Y_3 &= \frac{1}{4\pi^2 k_x k_y}\Big(-1 + e^{j2\pi k_x x} + e^{j2\pi k_y y} - e^{j2\pi(k_x x + k_y y)}\Big)
\end{aligned}$$

To obtain interpolators in the native domain, we can use equation 18, with $\mathcal{C}_{g0}$ defined by a $[\pi k_x, \pi k_y]$ rectangle with a vertex at $[-\pi k_x, -\pi k_y]$. Since integration region is rectangular, we can use Fubini's theorem and split double integral into two cascading integrals. Following that approach, we derive the following 2D interpolation functions:

$$y_0 = \frac{\sin^2(k_x \pi x)\sin^2(k_y \pi y)}{k_x^2 k_y^2 \pi^4 x^2 y^2} = \mathrm{sinc}^2(k_x x)\mathrm{sinc}^2(k_y y) \tag{20}$$

$$y_1 = \frac{\sin^2(k_x \pi x)\sin^2(k_y \pi y)}{k_x^2 k_y^2 \pi^4 x^2 y^2} = \mathrm{sinc}^2(k_x x)\mathrm{sinc}^2(k_y y)x \tag{21}$$

$$y_2 = \frac{\sin^2(k_x \pi x)\sin^2(k_y \pi y)}{k_x^2 k_y^2 \pi^4 x^2 y^2} = \mathrm{sinc}^2(k_x x)\mathrm{sinc}^2(k_y y)y \tag{22}$$

$$y_3 = \frac{\sin^2(k_x \pi x)\sin^2(k_y \pi y)}{k_x^2 k_y^2 \pi^4 x^2 y^2} = \mathrm{sinc}^2(k_x x)\mathrm{sinc}^2(k_y y)xy. \tag{23}$$

REFERENCES

Allegar, N., F. J. Herrmann, and C. C. Mosher, 2017, Introduction to this special section: Impact of compressive sensing on seismic data acquisition and processing: Leading Edge, **36**, 640.

Baraniuk, R. G., 2007, Compressive Sensing [Lecture Notes]: IEEE Signal Processing Magazine, **24**, 118–121.

Blumensath, T., and M. E. Davies, 2008, Iterative thresholding for sparse approximations: Journal of Fourier Analysis and Applications, **14**, 629–654.

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, 489–509.

Chen, W., D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, 2016, Simultaneous Bayesian sparse approximation with structured sparse models: IEEE Transactions on Signal Processing, **64**, 6145–6159.

Cheung, K. F., 1993, A multidimensional extension of Papoulis' generalized sampling expansion with the application in minimum density sampling, *in* Advanced topics in Shannon sampling and interpolation theory: Springer, 85–119.

Duarte, M. F., S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, 2005, Joint sparsity models for distributed compressed sensing: Proceedings of the Workshop on Signal Processing with Adaptative Sparse Structured Representations, 1–4.

Duijndam, A. J. W., M. A. Schonewille, and C. O. H. Hindriks, 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, 524–538.

Jerri, A. J., 1977, The Shannon sampling theorem — Its various extensions and applications: A tutorial review: Proceedings of the IEEE, **65**, 1565–1596.

Jiang, T., P. Eick, Y. Jiang, T. Li, H. Hao, W. Chu, R. Holt, D. Blymyer, K. Koster, and D. Enns, 2019, Compressive Sensing Seismic Processing Tests on A High Density Blended Land Data Set: SEG Technical Program Expanded Abstracts, 4505–4509.

Lee, K., Y. Bresler, and M. Junge, 2012, Subspace methods for joint sparse recovery: IEEE Transactions on Information Theory, **58**, 3613–3641.

Linden, D. A., 1959, A Discussion of Sampling Theorems: Proceedings of the IRE, **47**, 1219–1226.

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568.

Mosher, C. C., C. Li, F. D. Janiszewski, L. S. Williams, T. C. Carey, and Y. Ji, 2017, Operational deployment of compressive sensing systems for seismic data acquisition: The Leading Edge, **36**, 661–669.

Muyzert, E., N. Allouche, P. Edme, and N. Goujon, 2019, A five component land seismic sensor for measuring lateral gradients of the wavefield: Geophysical Prospecting, **67**, 97–113.

Palangi, H., R. Ward, and L. Deng, 2016, Distributed Compressive Sensing: A Deep Learning Approach: IEEE Transactions on Signal Processing, **64**, 4504–4518.

Papoulis, A., 1977, Generalized Sampling Expansion: IEEE Transactions on Circuits and Systems, **24**, 652–654.

Pawelec, I., M. Wakin, and P. Sava, 2021, Missing trace reconstruction for 2D land seismic data with randomized sparse sampling: Geophysics, **86**, P25–P36.

Petersen, D. P., and D. Middleton, 1962, Sampling and reconstruction of wave-number-limited functions in N-dimensional Euclidean spaces: Information and Control, **5**, 279–323.

Robertsson, J. O. A., I. Moore, M. Vassallo, K. Özdemir, D.-j. V. Manen, and A. Özbek, 2008, On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction: Geophysics, **73**, A45–A49.

Shannon, C. E., 1948, A mathematical theory of communication: The Bell system technical journal, **27**, 379–423.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, V123–V132.

Tropp, J. A., 2006, Algorithms for simultaneous sparse approximation. Part II: Convex relaxation: Signal Processing, **86**, 589–602.

Tropp, J. A., A. C. Gilbert, and M. J. Strauss, 2006, Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit: Signal Processing, **86**, 572–588.

Vassallo, M., A. Özbek, K. Özdemir, and K. Eggenberger, 2010, Crossline wavefield reconstruction from multicomponent streamer data: Part 1 — Multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient: Geophysics, **75**, WB53–WB67.

Wipf, D. P., and B. D. Rao, 2007, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem: IEEE Transactions on Signal Processing, **55**, 3704–3716.

Xu, S., Y. Zhang, and G. Lambaré, 2010, Antileakage Fourier transform for seismic data regularization in higher dimensions: Geophysics, **75**, WB113–WB120.

Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87–V95.

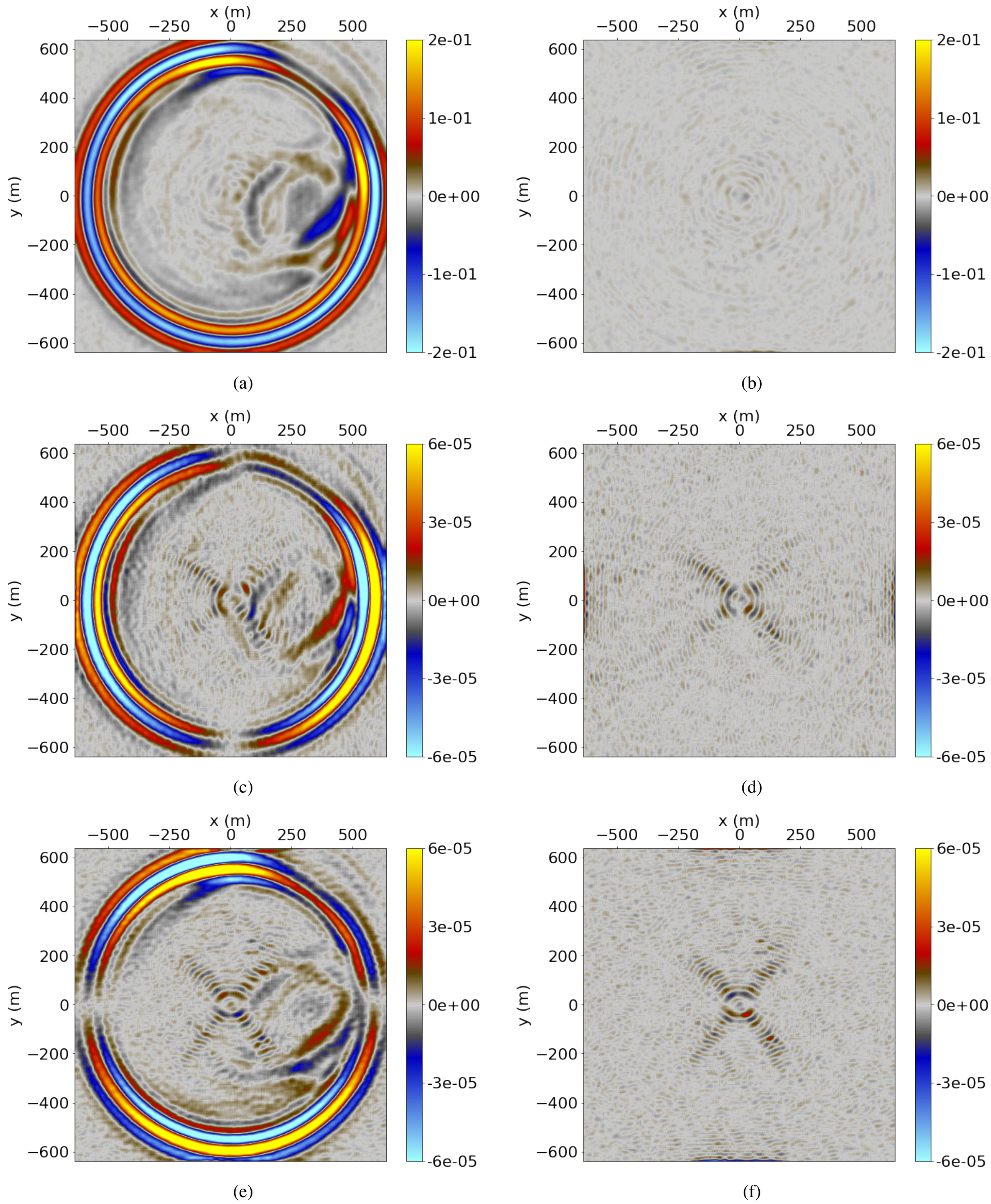


Figure 9. Simultaneous sparse approximation of 4 channels: (a) data samples, (b) time derivative, (c) x derivative, and (d) y derivative. Enforcing the common sparse support constraint helps to preserve the physical relationship between the channels.