A practical workflow for land seismic wavefield recovery with weighted matrix factorization

Yijun Zhang and Felix J. Herrmann,
Department of Electrical & Computer Engineering, Georgia Institute of Technology

SUMMARY

While wavefield reconstruction through weighted low-rank matrix factorizations has been shown to perform well on marine data, out-of-the-box application of this technology to land data is hampered by ground roll. The presence of these strong surface waves tends to dominate the reconstruction at the expense of the weaker body waves. Because ground roll is slow, it also suffers more from aliasing. To overcome these challenges, we introduce a practical workflow where the ground roll and body wave components are recovered separately and combined. We test the proposed approach blindly on a subset of the 3D SEAM Barrett dataset. With our technique, we recover densely sampled data from 25 percent randomly subsampled receivers. Independent comparisons on a single shot demonstrate significant improvements achievable with the presented workflow.

INTRODUCTION

One of the critical phases in the early stages of oil and gas exploration is seismic data acquisition. Inspired by relatively recent developments encouraged by the field of Compressive Sensing (Candès et al., 2006), seismic data is increasingly collected randomly along the spatial coordinates to shorten the acquisition time and to reduce cost. While random sampling improves acquisition productivity (Mosher et al., 2014), it does shift the burden from field acquisition to data processing (Chiu, 2019) since fully sampled seismic data is a prerequisite to subsequent steps such as multiple removal and migration.

Wavefield recovery is one of the key steps to reconstruct fully seismic data from subsampled data. Recovery methods based on wavefield reconstruction that exploits data sparsity in different transform domains, such as wavelet (Villasenor et al., 1996), Fourier (Sacchi et al., 1998), and curvelet (Herrmann and Hennenfent, 2008), have been proposed. More recently, several seismic studies have investigated wavefield recovery via low-rank matrix factorizations (Kumar et al., 2015), which are relatively simple and computationally cheap. The general idea of these methods is to exploit low-rank structure of fully sampled frequency slices when they are organized in a matrix. Oropeza and Sacchi (2011) and Kumar et al. (2015) showed that the presence of noise or missing traces increases the rank of these matrices, and they used this property to recover the fully sampled frequency slices via low-rank matrix factorization.

While the low-rank matrix factorization method has had some success, especially for low to midrange frequencies, it struggles to recover high frequency slices, which require higher ranks because they cannot be accurately approximated by low-rank factorizations. To solve this problem, Aravkin et al. (2014), Eftekhari et al. (2018), and Zhang et al. (2019) used the wavefield recovery via weighted matrix factorization to reconstruct

seismic data by introducing matrix weights defined in terms of factorizations at neighboring frequencies that live in close-by subspaces. By moving the matrix weights from the constraint to the data-misfit term, Zhang et al. (2020) proposed a computationally more efficient scheme capable of handling high frequencies.

Even though this weighted approach has had success, there remains the challenge that land seismic data contains ground roll, which because of its strong amplitude and high spatial frequency content is known to (Liu, 1999) degrade the wavefield reconstructions based on promoting structure whether this is sparsity or low rank. The reason for this possible degradation is two-fold. First, groud roll corresponds to Rayleigh-type surface waves, which are slow and for this reason often aliased. Second, ground roll has strong amplitudes, which causes the reconstruction to focus on the ground roll at the expense of reconstructing the low-amplitude body waves. While ground roll is typically dominant at the low temporal frequencies, its separation from body waves is complicated by the fact that it is spatially aliased. By reconstructing the wavefield to a fine grid, where the ground roll is no longer aliased, we allow for a separation of ground roll and body waves using $f - k$ filtering (Yilmaz, 2001) or Radon domain techniques (Trad et al., 2003). During this talk, we present a practical workflow aimed at removing the complications of carrying out wavefield reconstruction on land data dominated by ground roll.

We organize this expanded abstract as follows. First, we discuss the seismic wavefield reconstruction via weighted matrix factorization. Next, we discuss the impact of ground roll. And then, we introduce our proposed practical workflow step by step. We conclude by demonstrating our approach on synthetic 3D data simulated from the Barrett model and show improved recovery quality compared to the conventional workflow.

RECONSTRUCTION WITH WEIGHTED MATRIX FACTORIZATIONS

In Aravkin et al. (2014), Eftekhari et al. (2018) and Zhang et al. (2019), the authors proposed a wavefield recovery via weighted matrix factorization. These factorizations are carried out on data organized in monochromatic frequency slices and involve the following optimization problem:

$$
\begin{aligned}
\underset{\mathbf{X}_i}{\text{minimize}} \quad & \|\mathbf{Q}\mathbf{X}_i\mathbf{W}\|_* \\
\text{subject to} \quad & \|\mathscr{A}(\mathbf{X}_i) - \mathbf{B}_i\|_F \leq \eta.
\end{aligned}
\tag{1}
$$

In this expression, the symbol $\|\cdot\|_*$ represents the nuclear norm, given by the sum of the singular values, and $\|\cdot\|_F$ denotes the Frobenius norm, the energy of the matrix entries. The matrix $\mathbf{X}_i$, for $i \in [1, \ldots, N_f]$, represents a fully sampled monochromatic frequency slice at the ith frequency. N_f corresponds to the number of frequencies, and the matrix $\mathscr{A}(\cdot)$ represents a mask

operator used to subsample the fully sampled frequency slice. The matrix $\mathbf{B}_i$ represents the observed input data with missing traces at the ith frequency. The misfit tolerance η depends on the noise level in the observed data.

To exploit the fact that seismic data exhibits low-rank behavior in the so-called non-canonical organization (Kumar et al., 2015), we matricize the frequency slices in the source-x receiver-x organization—i.e., the to-be-recovered monochromatic data is represented by the matrix $\mathbf{X}_i \in \mathbb{C}^{(N_{sx} \times N_{rx}) \times (N_{sy} \times N_{ry})}$ where N_{sx}, N_{sy} are the number of sources along the x and y coordinates, respectively. N_{rx}, N_{ry} are the corresponding numbers of receivers. The $\{\mathbf{Q}, \mathbf{W}\} \in \mathbb{C}^{(N_{sx} \times N_{rx}) \times (N_{sx} \times N_{rx})}$ are the weighting matrices, which include information on the subspaces of a neighboring factorization as we reconstruct the wavefield from low-to-high frequencies (Zhang et al., 2019). These weighting matrices are given by

$$\mathbf{Q} = w_1 \mathbf{U}\mathbf{U}^H + \mathbf{U}^\perp \mathbf{U}^{\perp H} \tag{2}$$

and

$$\mathbf{W} = w_2 \mathbf{V}\mathbf{V}^H + \mathbf{V}^\perp \mathbf{V}^{\perp H}. \tag{3}$$

In these expressions, the symbol ^{H} denotes the Hermitian transpose. The projection matrices $\mathbf{U} \in \mathbb{C}^{(N_{sx} \times N_{rx}) \times r}$, $\mathbf{V} \in \mathbb{C}^{(N_{sy} \times N_{ry}) \times r}$ contain rank r column and row subspaces that derive from neighboring (lower) frequencies. The matrices $\mathbf{U}^\perp$, $\mathbf{V}^\perp$ are the orthogonal complements of $\mathbf{U}$, $\mathbf{V}$. Because factorizations of neighboring (lower) frequencies share information with the current frequency slice, they can serve as prior information aiding the wavefield recovery. The scalars $w_1 \in (0,1]$ and $w_2 \in (0,1]$ quantify the similarity between prior information and the current to-be-recovered frequency slice. Small values for these scalars indicate that we have more confidence in the prior information.

As shown in (Zhang et al., 2019), considerable improvements can be made during the recovery when reliable prior information is available. However, including weighting matrices in the nuclear norm objective function complicates the optimization making the minimization in equation 1 computationally more expensive. To avoid this issue, we follow Zhang et al. (2020) and rewrite equation 1 into

$$\begin{aligned} \underset{\tilde{\mathbf{X}}_i}{\text{minimize}} \quad & \|\tilde{\mathbf{X}}_i\|_* \\ \text{subject to} \quad & \|\mathscr{A}(\mathbf{Q}^{-1}\tilde{\mathbf{X}}_i\mathbf{W}^{-1}) - \mathbf{B}_i\|_F \le \eta. \end{aligned} \tag{4}$$

To arrive at this formulation, we replace the optimization variable with $\tilde{\mathbf{X}}_i = \mathbf{Q}\mathbf{X}_i\mathbf{W}$. After solving equation 4, the original solution $\mathbf{X}_i$ can be recovered by $\mathbf{X}_i = \mathbf{Q}^{-1}\tilde{\mathbf{X}}_i\mathbf{W}^{-1}$. Mathematically, equations 1 and equation 4 are equivalent except that the solution of the second formulation is easier to compute by moving the weighting matrices to the data misfit constraint.

To prevent computationally expensive singular value decompositions (SVDs) part of the nuclear norm computations, we write equation 4 in the following factored form:

$$\begin{aligned} \underset{\tilde{\mathbf{L}}_i, \tilde{\mathbf{R}}_i}{\text{minimize}} \quad & \frac{1}{2}\left\|\begin{bmatrix}\tilde{\mathbf{L}}_i \\ \tilde{\mathbf{R}}_i\end{bmatrix}\right\|_F^2 \\ \text{subject to} \quad & \|\mathscr{A}(\mathbf{Q}^{-1}\tilde{\mathbf{L}}_i\tilde{\mathbf{R}}_i^H\mathbf{W}^{-1}) - \mathbf{B}_i\|_F \le \eta. \end{aligned} \tag{5}$$

In this expression, the $\tilde{\mathbf{L}}_i \in \mathbb{C}^{(N_{sx} \times N_{rx}) \times r}$ and $\tilde{\mathbf{R}}_i \in \mathbb{C}^{(N_{sy} \times N_{ry}) \times r}$ represent the low-rank factorization of $\tilde{\mathbf{X}}_i$ with rank $r \ll \min(N_{sx} \times N_{rx}, N_{sy} \times N_{ry})$ (Zhang et al., 2020).

While wavefield recovery based on weighted matrix factorization has been applied successfully (see e.g., Zhang et al. (2019) and Zhang et al. (2020)), its performance is challenged by data that contains strong-amplitude aliased ground roll. Because of its large-amplitude, ground roll dominates the reconstruction at the expense of body waves that are of prime interest. In the next section, we will introduce a practical workflow addressing this challenge.

IMPACT OF GROUND ROLL

Acquisition and processing of land data are often challenged because it is contaminated by strong ground roll. Because ground roll is slow, it is often spatially aliased, complicating subsequent processing efforts to remove this coherent noise component with $f - k$ or Radon filtering (Yilmaz, 2001, Trad et al. (2003)). Unfortunately, it is financially unfeasible to decrease the periodic receiver sampling interval to avoid aliasing (Bahia et al., 2020), and we have to resort to alternative randomized acquisition methodologies (Mosher et al., 2014, Kumar et al. (2015)) that are in principle conducive to in silico unaliased wavefield reconstruction. While this has proven to work, the presence of strong-amplitude ground roll complicates wavefield reconstruction.

To investigate this issue, we collaborate with Klaas Koster from Occidental to conduct a blind study where we were provided with a subset consisting of 21 source lines, extracted from the synthetic 3D SEAM Barrett dataset (Van De Coevering et al., 2019, Tan et al. (2019)). This dataset is designed to benchmark land data processing. As part of this blind study, with the acquisition geometry plotted in Figure 1, we receive 3D shot records that are randomly subsampled along the receivers. The 8×8km receiver aperture is moving with the source location, which means that between neighboring shots randomly sampled receivers are mostly shared while some drop-off and others are added (cf. red and blue rectangle in Figure 1). Approximately 75 percent of receiver positions are missing from the regular densely sampled periodic grid of 12.5m, yielding an effective average sample interval of 50m, which is well below Nyquist. The data consists of 667 time samples with a sample interval of 0.006 s. The shots are sampled periodically with a sample interval of 25m in the shot-line direction and 100m in the perpendicular direction.

To illustrate, the effects of the strong ground roll on our factorized low-rank wavefield reconstruction scheme, we recover a patch of 4×4 shots with on the dense periodic receiver grid of 641 receivers in each direction sampled at 12.5m. This recovery corresponds to solving a total of 384 monochromatic matrix factorization problems involving data volumes of $667 \times 641 \times 641 \times 4 \times 4$. These volumes are factored into the product of two $(641 \times 4) \times 340$ matrices where 340 is the rank r. After reconstruction, the results of which are plotted in Figure 2 for a single shot record, we recover shot gathers sampled at 12.5m from receivers collected at random at an average receiver spacing of 50m. While we are able to recover this shot record, strong

noisy artifacts remain especially at the long offsets. In addition, important reflection and diffraction information is missing and the ground roll is not well recovered making this wavefield recovery unsuitable for subsequent processing.
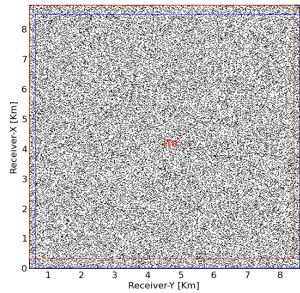

Figure 1: Acquisition geometry for the recovered patch. Black ·'s represent receiver locations, and red ·'s in the middle represent the source locations. The red rectangle is the receiver aperture for the top left source and the blue rectangle is the receiver aperture for the bottom right source.
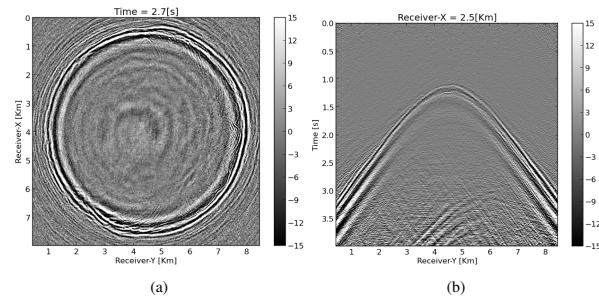
Figure 2: One common-shot record obtained by factorized wavefield reconstruction. *(a)* Time slice at 2.7 s. *(b)* Shot record along the *y* direction.

PROPOSED PRACTICAL WORKFLOW

To mitigate the effects of ground roll, we propose the reconstruction of the body and surface (ground roll) waves separately. In this approach, outlined in Figure 3, we use the fact that ground roll is slow and relatively easily separable by applying a linear shift to the data. Below, we describe the different steps outlined in the dashed boxes in Figure 3.

Ground roll estimation

Because the ground roll is slow, it is steep and therefore at least in an approximate sense, separable from the body waves. This allows us to devise a separate reconstruction scheme to recover the ground roll before adding it back to the reconstruction of the body waves. We obtain an estimate of the ground roll by carrying out the following steps: (1) after zero-padding the input data (Figure 4a), we apply a linear shift aligning the ground roll (Figure 4b); (2) we apply a smooth taper with smooth cutoffs around at $t = 0$s and $t = 1$s designed to extract the ground roll (Figure 4c), followed by (3) undoing the linear shift, yielding an estimate of the randomly subsampled ground
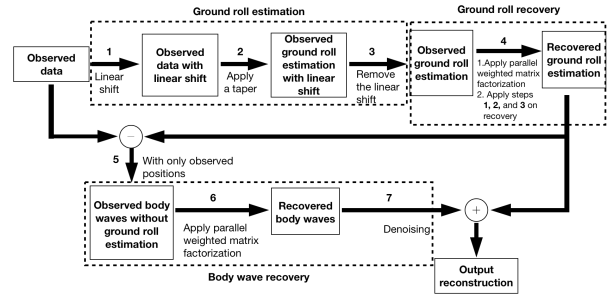

Figure 3: Flowchart of proposed method.

roll plotted in Figure 4d. This estimate for the ground roll serves as input for the reconstruction.

Ground roll recovery

We use the estimated ground roll as input to our wavefield reconstruction based on weighted matrix factorizations for a rank $r = 250$, which we find empirically by observing continuity of signals and limited noise in the reconstructed data. We run the reconstruction over all shots simultaneously for 320 iterations of SPG-ℓ_2 (Lopez et al., 2015) per frequency slice. To avoid reconstruction leakage, we apply steps *(1)-(3)* from the previous section again to get the final ground roll recovery.
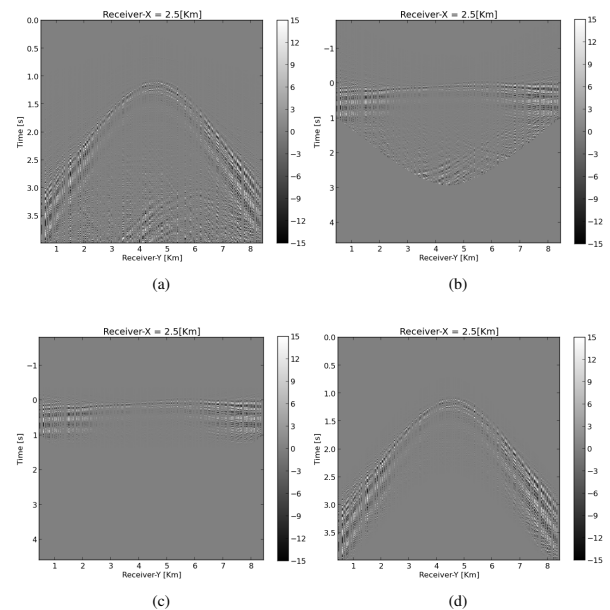

Figure 4: Ground roll estimation. *(a)* Input shot record. *(b)* Input after linear shift. *(c)* Tapered data. *(d)* Estimate subsampled ground roll after undoing linear shift.

Body wave recovery

After reconstruction of the ground roll, we apply the mask $\mathscr{A}$ to restrict the ground roll reconstruction to the observed receiver positions again and subtract it from the original subsampled input data. The resulting "ground roll free" estimate

for the body waves subsequently serves as input to a second wavefield reconstruction now for the body waves. Since these waves are more complex than ground roll, we choose the rank higher ($r = 340$). As we can observe from Figure 5, the reconstructed body waves contain, as expected, some remaining low-amplitude ground roll. Before inverse Fourier transforming the reconstructed body waves, we apply a $f - k$ filter to each shot along both receiver coordinates to remove remnant noise. The resulting recovery for the body waves shown in Figure 5b shows reconstruction of high frequency reflected and diffracted energy. To arrive at the final result, we combine the wavefield reconstructions for the ground roll and body waves. The result of this blind study for a single shot is included in Figure 6.
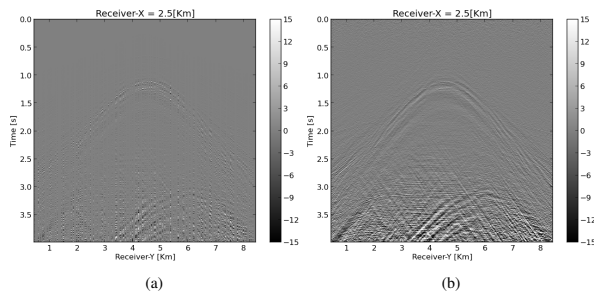

(a)

(b)

Figure 5: Body waves in the time domain. *(a)* Observed body waves. *(b)* Reconstructed body waves
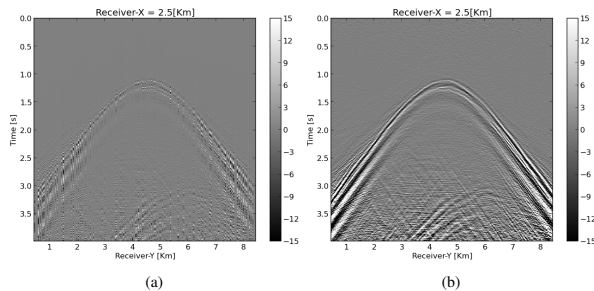

(a)

(b)

Figure 6: Seismic data reconstruction in the time domain. *(a)* Observed shot gather. *(b)* Reconstructed shot gather.

QUALITY CONTROL (QC)

By comparing Figure 6b with Figure 2b, we observe that the proposed method produces results with less artifacts at the long offsets, and the ground roll is well recovered, especially at the near offsets (see Figure 6b for the receiver coordinate *y* between $4 - 6$ km for the time interval $1 - 1.5$ s).

To further verify our proposed practical workflow, we sent one reconstructed shot gather to Occidental and obtained the following plots in return. Figure 7 contains time slices at 2.1s with the ground truth (Figure 7a), reconstructed wavefield (Figure 7b) and difference plot (Figure 7c). From these plots, we observe that the proposed method successfully recovers body waves (reflections and diffractions) despite the presence of strong aliased ground roll. During our presentation, we will further QC our results by working with Occidental to produce stacks.
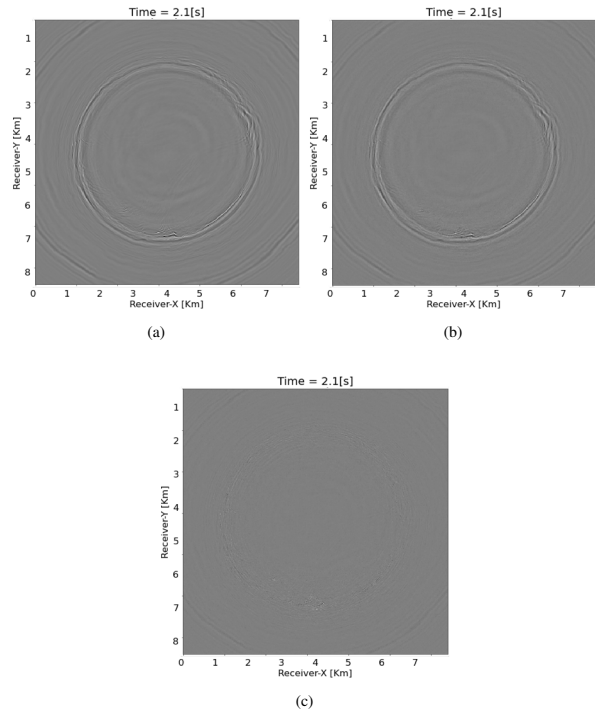

(a)

(b)

(c)

Figure 7: Wavefield recovery of one common shot gather. *(a)* Time slice at 2.1s of ground truth. *(b)* Time slice at 2.1s of reconstructed data. *(c)* Time slice at 2.1s of difference between ground truth and recovery. All the subfigures are plotted on the same scale.

CONCLUSIONS

We presented a practical workflow successfully recovering a subset of the synthetic 3D SEAM Barrett dataset randomly sampled at 25 percent receiver sampling. Our workflow consists of the combination of a weighted matrix factorization scheme and a separation of the subsampled input data into ground roll and body wave components. Thanks to this decomposition, we were able to mitigate the effects induced by the strong aliased ground roll. Initial findings of the blind test we carried out in collaboration with Occidental show that our method is capable of dealing with ground roll while recovering high-frequency body waves.

RELATED MATERIALS

The Julia code for this work is available on the SLIM GitHub page https://github.com/slimgroup/Software.SEG2021.

REFERENCES

Aravkin, A., R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, 2014, Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation: SIAM Journal on Scientific Computing, **36**, S237–S266.

Bahia, B., I. Papathanasaki, and M. D. Sacchi, 2020, Ground-roll attenuation through quaternionic inversion with sparsity constraints, *in* SEG Technical Program Expanded Abstracts 2020: Society of Exploration Geophysicists, 3254–3258.

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.

Chiu, S. K., 2019, 3d attenuation of aliased ground roll on randomly undersampled data, *in* SEG Technical Program Expanded Abstracts 2019: Society of Exploration Geophysicists, 4560–4564.

Eftekhari, A., D. Yang, and M. B. Wakin, 2018, Weighted matrix completion and recovery with prior subspace information: IEEE Transactions on Information Theory, **64**, 4044–4071.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.

Kumar, R., C. Da Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2015, Efficient matrix completion for seismic data reconstruction: Geophysics, **80**, V97–V114.

Liu, X., 1999, Ground roll supression using the karhunen-loeve transform: Geophysics, **64**, 564–566.

Lopez, O., R. Kumar, and F. J. Herrmann, 2015, Rank minimization via alternating optimization-seismic data interpolation: 77th EAGE Conference and Exhibition 2015, European Association of Geoscientists & Engineers, 1–5.

Mosher, C., C. Li, L. Morley, Y. Ji, F. Janiszewski, R. Olson, and J. Brewer, 2014, Increasing the efficiency of seismic data acquisition via compressive sensing: The Leading Edge, **33**, 386–391.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38.

Tan, J., T. Li, F. Jarrah, K. Lee, R. Holt, N. V. D. Coevering, and K. Koster, 2019, Seam phase ii barrett model classic data study: Processing, imaging, and attributes analysis, *in* SEG Technical Program Expanded Abstracts 2019: Society of Exploration Geophysicists, 3840–3844.

Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse radon transform: Geophysics, **68**, 386–399.

Van De Coevering, N., K. Koster, and R. Holt, 2019, A sceptic's view of vvaz and avaz, *in* SEG Technical Program Expanded Abstracts 2019: Society of Exploration Geophysicists, 3230–3234.

Villasenor, J. D., R. Ergas, and P. Donoho, 1996, Seismic data compression using high-dimensional wavelet transforms: Proceedings of Data Compression Conference-DCC'96, IEEE, 396–405.

Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of exploration geophysicists.

Zhang, Y., S. Sharan, and F. J. Herrmann, 2019, High-frequency wavefield recovery with weighted matrix factorizations, *in* SEG Technical Program Expanded Abstracts 2019: Society of Exploration Geophysicists, 3959–3963.

Zhang, Y., S. Sharan, O. Lopez, and F. J. Herrmann, 2020, Wavefield recovery with limited-subspace weighted matrix factorizations, *in* SEG Technical Program Expanded Abstracts 2020: Society of Exploration Geophysicists, 2858–2862.