GEOPHYSICS, VOL. 89, NO. 3 (MAY-JUNE 2024); P. 1–12, 12 FIGS., 2 TABLES.
10.1190/GEO2023-0465.1

# Robust multidimensional reconstruction via group sparsity with Radon operators

Ji Li[1] and Dawei Liu[2]

## ABSTRACT

Seismic data processing, specifically tasks like denoising and interpolation, often hinges on sparse solutions of linear systems. Group sparsity plays an essential role in this context by enhancing sparse inversion. It introduces more refined constraints, which preserve the inherent relationships within seismic data. To this end, we propose a robust orthogonal matching pursuit algorithm, combined with Radon operators in the frequency-slowness $(f\text{-}p)$ domain, to tackle the strong group sparsity problem. This approach is vital for interpolating seismic data and attenuating erratic noise simultaneously. Our algorithm takes advantage of group sparsity by selecting the dominant slowness group in each iteration and fitting Radon coefficients with a robust $\ell_1$-$\ell_1$ norm using the alternating direction method of multipliers (ADMM) solver. Its ability to resist erratic noise, along with its superior performance in applications such as simultaneous source deblending and reconstruction of noisy onshore data sets, underscores the importance of group sparsity. Synthetic and real comparative analyses further demonstrate that strong group sparsity inversion consistently outperforms corresponding traditional methods without the group sparsity constraint. These comparisons emphasize the necessity of integrating group sparsity in these applications, thereby indicating its indispensable role in optimizing seismic data processing.

## INTRODUCTION

Sparse representation has become a vital theoretical construct with crucial practical applications in seismic signal processing. The significance is underscored by its role in primaries estimation (Van Groenestijn and Verschuur, 2009; Lin and Herrmann, 2013), multiple attenuation (Sacchi and Ulrych, 1995; Herrmann et al., 2000), and data reconstruction (Sacchi et al., 1998; Trad, 2009; Herrmann, 2010), among others. Sparse representation asserts that the seismic signal is sparse in specific transform domains, such as Radon (Durrani and Bisset, 1984), Fourier (Gülünay, 2003; Xu et al., 2005), Wavelet (Jian et al., 2006), and Curvelet (Candès et al., 2006) transforms. Specifically, the number of nonzero coefficients is substantially fewer than the total number of coefficients. In contrast to direct applications of these mathematical transforms, the integration of sparsity constraints provides a more robust framework and enables a heightened ability to explore the inherent complexities of seismic data. Typically, the capability to get sparser solutions results in better outcomes when dealing with noisy, incomplete, or complex data. Therefore, researchers mainly work on developing more advanced sparse constraints and sparse solvers to enhance the sparse representation ability.

The implementation of an additional group structure — often referred to as group sparsity or group Lasso (Yuan and Lin, 2006; Bach, 2008; Nardi and Rinaldo, 2008) — has yielded promising results in improving the accuracy of sparse estimation in seismic data processing. Methods leveraging group sparsity have been innovatively applied in various contexts. For instance, Trad et al. (2002) use this concept to calculate the hyperbolic Radon transform efficiently. Naghizadeh (2012) and Li and Sacchi (2022) use group sparsity to classify coefficients in the frequency-wavenumber $(f\text{-}k)$ domain, enhancing seismic data denoising and interpolation. Similarly, Vera Rodriguez et al. (2012) apply group sparsity for microseismic data denoising, whereas Chen et al. (2019) fuse group sparsity with total variation for seismic signal denoising. Rodriguez et al. (2012) also use the idea of group sparsity for microseismic seismic moment tensor inversion, for which the six elements of the seismic moment tensor are treated as a group for sparse inversion. These methods demonstrate the ongoing evolution and potential of group sparsity, offering promising solutions for denoising and interpolation in seismic data processing. However,

Manuscript received by the Editor 10 August 2023; revised manuscript received 21 October 2023; published ahead of production 17 January 2024.
[1]University of Calgary, Department of Earth, Energy and Environment, Calgary, Alberta, Canada, E-mail: li.ji1@ucalgary.ca.
[2]Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an, China and University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: 409791715@qq.com (corresponding author)

their applications to prestack data, which are contaminated by high-amplitude erratic noise (statistical outliers) and missing traces, are less understood. The performance of these methods in simultaneously removing erratic noise and performing interpolation remains an area for further exploration.

The development of a robust sparse solver capable of addressing inversion problems in data contaminated by erratic noise has been a focal point of research for many scholars. Guitton and Symes (2003), for example, replace the conventional $\ell_2$-norm with the Huber norm for the residual term, enhancing the handling of seismic data with outliers. Similarly, Trickett et al. (2012) deploy a rank reduction filter to mitigate erratic noise. A more recent effort by Li and Sacchi (2021) has resulted in a sparse and robust Radon transform, estimated via matching pursuit (MP), for solving simultaneous source separation problems. A common thread in most of these methods is the utilization of robust M-estimators to replace the $\ell_2$-norm of the residual term, thereby increasing the robustness of the cost function to erratic noise (Maronna, 1976). This paper further explores this by integrating robust inversion with group sparsity, aiming to enhance the effectiveness of sparse estimation.

We propose a novel sparse inversion method with group sparsity for interpolating seismic data and attenuating erratic noise simultaneously. Our contributions to this study can be delineated as follows: initially, we create a robust algorithm that merges robust inversion with group sparsity, focusing on the linear Radon transform in the frequency-slowness ($f$-$p$) domain. Using the orthogonal matching pursuit (OMP) (Tropp and Gilbert, 2007), we iteratively build estimated results. This process involves dividing the Radon coefficients $\mathbf{m}$ into distinct groups $\{\mathbf{m}_{p_1}, \mathbf{m}_{p_2}, \ldots, \mathbf{m}_{p_n}\}$, categorized by different slowness $p$ in the $f$-$p$ domain. In each iteration, the algorithm selects the group $p_i$ that maximizes $\|\mathbf{m}_{p_i}\|_2^2$. Subsequently, we deploy an $\ell_1$-$\ell_1$ alternating direction method of multipliers (ADMM) solver (Wen et al., 2016) to fit the Radon coefficients within all currently selected groups. By limiting the number of best-correlated groups, the OMP enhances the sparsity of the groups, whereas the ADMM solver explores the sparsity within those selected groups. As a result, the proposed method actively encourages sparsity at group and individual coefficient levels, effectively resolving the inversion problem with a robust sparse constraint. To evaluate the resilience and efficacy of this novel reconstruction method, we present 2D and 3D synthetic examples and apply the technique to real-data scenarios. Synthetic and real-data applications reinforce the effectiveness of the proposed robust reconstruction method.

## THEORY

A seismic signal can be represented as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ denotes the signal, $\mathbf{x}$ is the coefficients, and $\mathbf{A}$ is a synthesis operator or matrix that maps the coefficients into the seismic signal. Similarly, $\mathbf{e}$ represents additive noise to the signal. In this situation, the sparse approximation problem can be expressed as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \le \delta, \tag{2}$$

where $\|.\|_0$ is the $\ell_0$-norm, which measures the number of nonzero coefficients, and $\|.\|_2$ symbolizes the $\ell_2$-norm (equal to the square root of the inner product of a vector with itself) used to fit the residuals. This problem is a combinatorial, nondeterministic polynomial time (NP)-hard problem for which finding an exact solution is prohibitively expensive. In general, two groups of methods can be used to solve the problem 2. One is the convex relaxation, replacing the $\ell_0$-norm with the $\ell_1$-norm (Chen et al., 2001), which transfers the problem into a convex optimization problem

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{3}$$

where $\lambda$ is a trade-off parameter. This problem also is known as the lasso problem (Tibshirani, 1996) and can be solved by many methods like Fast Iterative Shrinkage-Thresholding Algorithm (Beck and Teboulle, 2009), Iterative Reweighted Least Squares (Scales and Gersztenkorn, 1988), and ADMM (Boyd et al., 2011). The other group of approaches is the greedy method, like MP (Mallat and Zhang, 1993) and OMP (Pati et al., 1993; Tropp and Gilbert, 2007).

In many scenarios, incorporating a group structure constraint can yield a better sparsity estimation (Majumdar and Ward, 2009; Elhamifar and Vidal, 2011). This differs from the conventional sparsity assumption, for which sparsity is assessed by counting the number of nonzero coefficients. Assume the coefficients in $\mathbf{x}$ are partitioned into $N$ nonoverlapped groups $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots \mathbf{x}_N$. Within the framework of a general group sparsity constraint, elements are organized into these distinct groups, with either all or none of the elements in a group being zero, as follows:

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \beta \sum_{i=1}^{N} \|\mathbf{x}_i\|_2, \tag{4}$$

where $\beta$ is a trade-off parameter and $\mathbf{x}_i$ is the $i$th group of $\mathbf{x}$.

A further refinement, called strong group sparsity, can provide enhanced performance (Simon et al., 2013; Vincent and Hansen, 2014). This advanced form does not merely constrain sparsity at a group level but also imposes sparsity within the individual coefficients of the groups. In a strong group sparsity problem, not only must a few of the groups contain nonzero coefficients but the coefficients within those selected groups must themselves be sparse. This represents a more nuanced and adaptable method for evaluating sparsity:

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \beta \sum_{i=1}^{N} \|\mathbf{x}_i\|_2. \tag{5}$$

When seismic data are marred by erratic noise, which can be treated as statistical outliers, the use of an M-estimator becomes invaluable. An M-estimator is a statistical technique used for estimating the parameters of a statistical model, particularly in the realm of regression analysis and robust statistics. These estimators are intentionally crafted to exhibit resilience against outliers and departures from the assumption of normality in the data. Among the M-estimators commonly applied in robust denoising methods, the Huber norm, the $\ell_1$-norm, and the Tukey norm stand out as the most frequently used choices. In the proposed method, we use the $\ell_1$-norm to replace the $\ell_2$-norm. This substitution leads to a modification in equation 5 as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_1 + \lambda\|\mathbf{x}\|_1 + \beta \sum_{i=1}^{N} \|\mathbf{x}_i\|_2. \quad (6)$$

To solve this equation, we use the OMP with the $\ell_1$-$\ell_1$ ADMM. To adapt this method to the complex task of processing seismic data, we harness the power of the linear Radon transform, enabling us to synthesize the seismic signal in a more precise and controlled manner.

## Linear Radon transform

The linear Radon transform is time-invariant, which means it can be calculated either in the time or frequency domain. The Radon transform in the frequency domain can be represented as the following equation:

$$\hat{M}(p, f) = \sum_x D(x, f) e^{i\omega x p}, \quad (7)$$

$$D(x, f) = \sum_p M(p, f) e^{-i\omega x p}, \quad (8)$$

where $\omega = 2\pi f$. These equations also can be written in the matrix form

$$\hat{\mathbf{M}} = \mathcal{L}\mathbf{D}, \quad (9)$$

$$\mathbf{D} = \mathcal{L}^*\mathbf{M}, \quad (10)$$

where $\mathcal{L}$ is the adjoint Radon operator. Correspondingly, $\mathcal{L}^*$ is the forward Radon operator, which transfers the Radon coefficients in the $f$-$p$ domain to the $f$-$x$ domain. It also can be extended to a 3D form

$$\hat{M}(p_x, p_y, f) = \sum_x \sum_y D(x, y, f) e^{i\omega(p_x x + p_y y)}, \quad (11)$$

$$D(x, y, f) = \sum_{p_x} \sum_{p_x} M(p_x, p_y, f) e^{-i\omega(p_x x + p_y y)}, \quad (12)$$

which, in matrix form, can be written as

$$\hat{\mathbf{M}} = \mathcal{L}_{p_x} \mathbf{D} \mathcal{L}_{p_y}, \quad (13)$$

$$\mathbf{D} = \mathcal{L}^*_{p_x} \mathbf{M} \mathcal{L}^*_{p_y}. \quad (14)$$

In this section, we will demonstrate the application of group theory to the Radon transform. Figure 1 showcases a Radon panel in the $f$-$p$ domain. We have partitioned the Radon coefficient panel into $N$ equally sized groups, spanning from $p_{\min}$ to $p_{\max}$. Each group encompasses coefficients across all frequencies. The $\ell_2$-norm of each group is computed as follows:

$$\mathbf{E}(p_i) = \sum_{f_{\min}}^{f_{\max}} \|\hat{\mathbf{M}}(p_i, f)\|_2^2. \quad (15)$$

Consequently, the groups exhibiting the highest summation of $\ell_2$-norm in equation 15 will be considered the most correlated groups.

This method shares similarities with the approach proposed by Naghizadeh (2012), in which a similar technique is used to identify dominant dips in the $f$-$k$ domain. However, we find that working within the $f$-$p$ domain is more intuitive and convenient. This is because the groups in the $f$-$p$ domain manifest as vertical lines, eliminating the need for path calculations. Furthermore, it eliminates the aliasing issue present in the $f$-$k$ domain due to the Fourier transform. Because a sparse solver accommodates the chosen coefficients, the nonorthogonality of the Radon transform, in contrast to the Fourier transform, becomes inconsequential.

To demonstrate the efficacy of the group sparse method with the linear Radon transform, we analyze a synthetic 2D example featuring three linear events. Figure 2a shows the fully sampled clean data, and Figure 2b illustrates clean data with 75% of the traces missing. To emulate a more realistic scenario and increase the complexity, Figure 2c portrays fully sampled data contaminated with strong erratic noise, including sparse noise and three bad traces, whereas Figure 2d represents data affected by erratic noise and missing traces. Then, we perform the linear Radon transform to these data, respectively. Figure 2e–2h reveal the corresponding Radon coefficients in the $f$-$p$ domain. Because the synthetic linear events have varying velocities, the Radon coefficients are situated around three specific $p$ values in the $f$-$p$ domain. In addition, it is

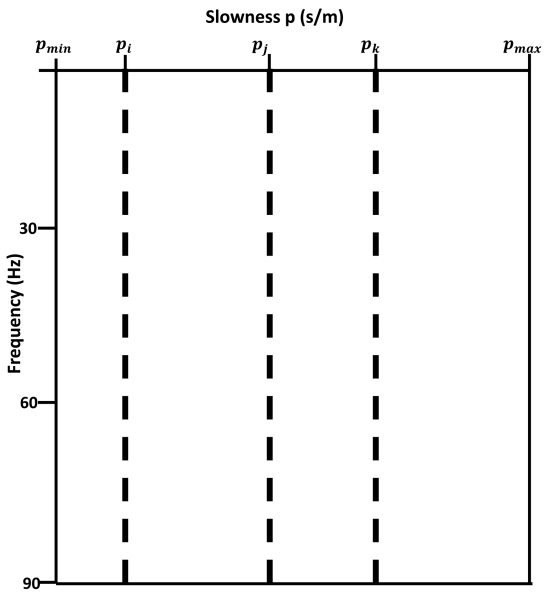


Figure 1. A cartoon shows how to find the best-correlated groups. The dashed line indicates the location of the coefficients for each group.

clear that the missing traces and erratic noise generated the random noise in the $f$-$p$ domain. Consequently, the dominant slowness in the $f$-$p$ domain is obscured when faced with strong erratic noise and many missing traces. However, Figure 2i–2l provides a clear visualization of the $\ell_2$-norm for each slowness $p$ across all frequencies. Even with the substantial erratic noise and many missing traces, we can still discern the three dominant slowness values effortlessly, as shown in Figure 2i–2l. These findings underscore the potential of the linear Radon transform in identifying key characteristics despite significant data disturbances.

Traditionally, the Radon transform is predominantly applied in the temporal domain ($\tau$-$p$), as opposed to the frequency-slowness ($f$-$p$) domain. In Figure 3, we compare Radon panels for a seismic gather similar to the one depicted in Figure 2d in $\tau$-$p$ and $f$-$p$ domains. In addition, we present corresponding $\ell_2$-norm maps in Figure 3c and 3d.

In the $\tau$-$p$ domain, the energy of the adjoint Radon transform is dispersed due to a smearing effect, resulting in less pronounced focus when compared to the $f$-$p$ domain. This reduced amplitude contrast between the dominant $p$ values and the noise complicates group selection. Consequently, our method benefits from the robustness of the $f$-$p$ domain in the group selection step, where the $\ell_2$-norm map is more informative and reliable.

By using the linear Radon operators in the $f$-$p$ domain, the problem for simultaneous interpolation and erratic noise removal becomes to minimize the following cost function:

$$\operatorname{argmin}\|\mathcal{L}^*\mathbf{m}-\mathbf{d}\|_1+\lambda\|\mathbf{m}\|_1+\beta\sum_{i=1}^{N}\|\mathbf{m}_i\|_2, \tag{16}$$

where $\mathbf{m}$ is the Radon coefficient in the $f$-$p$ domain. The data fitting term $\|\mathcal{L}^*\mathbf{m}-\mathbf{d}\|_1$ enhances the robustness of the cost function to erratic noise. Meanwhile, $\beta\sum_{i=1}^{N}\|\mathbf{m}_i\|_2$ is used to promote group sparsity, where $N$ designates the constraint number on group sparsity. The addition of $\lambda\|\mathbf{m}\|_1$ serves to augment the sparsity within these groups. To solve this problem, we use one commonly used greedy method, known as the OMP (Pati et al., 1993; Tropp and Gilbert, 2007).

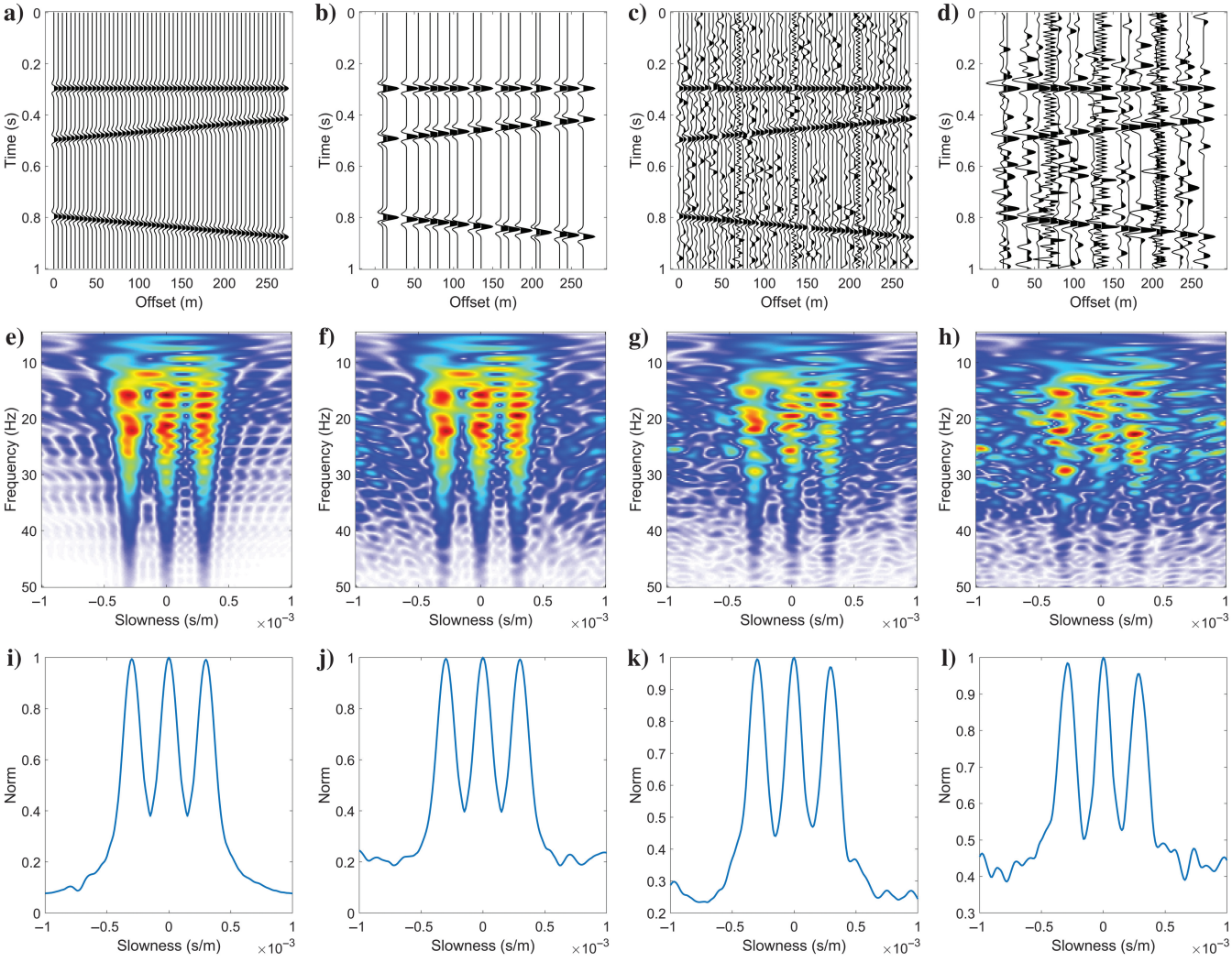


Figure 2. (a) Two-dimensional synthetic data. (b) Two-dimensional synthetic data with 75% of traces removed randomly. (c) Two-dimensional synthetic data with erratic noise. (d) Two-dimensional synthetic with erratic noise and missed traces. (e–h) The corresponding Radon coefficients of (a–d) in the $f$-$p$ domain. (i–l) The corresponding norm map of slowness $p$.

## Orthogonal matching pursuit

The MP (Mallat and Zhang, 1993) and OMP (Pati et al., 1993; Tropp and Gilbert, 2007) stand out as the two most frequently used greedy pursuit algorithms. Unlike MP, which updates one coefficient in each iteration, OMP is more intricate, considering all of the currently selected coefficients at each step. Therefore, to solve a general sparse linear problem like that in equation 2, OMP minimizes the following cost function in each iteration:

$$\hat{\mathbf{x}}^{[k]}_{T^{[k]}} = \underset{\tilde{\mathbf{x}}_{T^{[k]}}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}_{T^{[k]}} \tilde{\mathbf{x}}_{T^{[k]}}\|_2^2, \tag{17}$$

where $T^{[k]}$ refers to the set of indices of all coefficients selected up to iteration $k$. Consequently, the computational demands of OMP are greater than those of MP. Nevertheless, OMP often yields more refined results. A comprehensive outline of this process for a general sparse linear problem is provided in Algorithm 1. In this algorithm, OMP helps us select the best-correlated coefficient, and the minimization function equation 17 is used to fit these coefficients with data. Therefore, by combining the OMP with the minimization function of equation 17, we can find the minimum number of coefficients to fit the data and achieve a similar result as solving equation 3.

## Computing the group sparsity solution with OMP and Radon operators

Modifications to the OMP algorithm are required to address the minimization problem described in equation 16. Instead of selecting a single coefficient at each step, the approach used involves choosing an entire group of coefficients during the basis function selection phase. Specifically, coefficients are divided into various groups based on their slowness $p$. In our case, we evenly divide $p$ from $p_{\min}$ to $p_{\max}$ into N groups. Then, the group with the maximum $\ell_2$-norm in the frequency domain is selected. As illustrated in Figure 2, this selection method, which relies on the energy summation with $\ell_2$-norm, helps mitigate many effects from random noise in the $f$-$p$ domain, enhancing the robustness of the coefficient selection process. Subsequently, all coefficients within the currently chosen groups $T^k$ are used by minimizing the following cost function:

$$\hat{\mathbf{m}}^{k}_{T^k} = \underset{\tilde{\mathbf{m}}_{T^k}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{L}^*_n(\tilde{\mathbf{m}}_{T^k})\|_1 + \lambda \|\tilde{\mathbf{m}}_{T^k}\|_1. \tag{18}$$

This $\ell_1$-$\ell_1$ minimization problem can be effectively solved by the $\ell_1$-$\ell_1$ ADMM solver (Yang and Zhang, 2011; Wen et al., 2016). For complete details of this method, the robust group sparsity OMP is fully outlined in Algorithm 2. To interpolate the missing traces simultaneously, we add a sampling operator $\mathcal{S}$ to the cost function

$$\hat{\mathbf{m}}^{k}_{T^k} = \underset{\tilde{\mathbf{m}}_{T^k}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{S}\mathcal{L}^*_n(\tilde{\mathbf{m}}_{T^k})\|_1 + \lambda \|\tilde{\mathbf{m}}_{T^k}\|_1. \tag{19}$$

We use a meticulous selection process during each iteration to identify the most highly correlated group by using the $\ell_2$-norm map. Subsequently, we construct a model subspace denoted as

**Algorithm 1. OMP.**

Input: $\mathbf{y}$, $\mathbf{A}$, and $k$
Output: $\mathbf{r}^{[k]}$, $\hat{\mathbf{x}}^{[k]}$
Initialization: $\mathbf{r}^{[0]} = \mathbf{y}$, $\hat{\mathbf{x}}^{[k]} = 0$, and $T^{[0]} = \varnothing$
**for** $k = 1, 2, \ldots, K$ **do**
$l = \underset{j=1,2,\ldots,M}{\operatorname{argmax}} |\langle \mathbf{A}, \mathbf{r}^{[k-1]} \rangle|$
$T^{[k]} = T^{[k-1]} \cup \{l\}$
$\hat{\mathbf{x}}^{[i]}_{T^{[i]}} = \underset{\tilde{\mathbf{x}}_{T^{[i]}}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}_{T^{[i]}} \tilde{\mathbf{x}}_{T^{[i]}}\|_2^2$
$\mathbf{r}^{[k]} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{[k]}_{T^{[k]}}$
**end for**

**Algorithm 2. SG-OMP.**

Input: $\mathbf{d}$, $\mathcal{L}$, and $k$
Output: $\hat{\mathbf{m}}^{[k]}$
Initialization: $\mathbf{r}^{[0]} = \mathbf{d}$, $\hat{\mathbf{m}}^{[k]} = 0$, and $T^{[0]} = \varnothing$
**for** $k = 1, 2, \ldots, K$ **do**
$\mathbf{M} = \mathcal{L}\mathbf{r}$
Pick dominant slowness $p_k$
$T^k = T^{k-1} + p_k$
$\hat{\mathbf{m}}^{k}_{T^k} = \underset{\tilde{\mathbf{m}}_{T^k}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{S}\mathcal{L}^*_n(\tilde{\mathbf{m}}_{T^k})\|_1 + \lambda \|\tilde{\mathbf{m}}_{T^k}\|_1$
$\mathbf{r}^k = \mathbf{d} - \mathcal{S}\mathcal{L}^* \hat{\mathbf{m}}^{k}_{T^k}$
**end for**

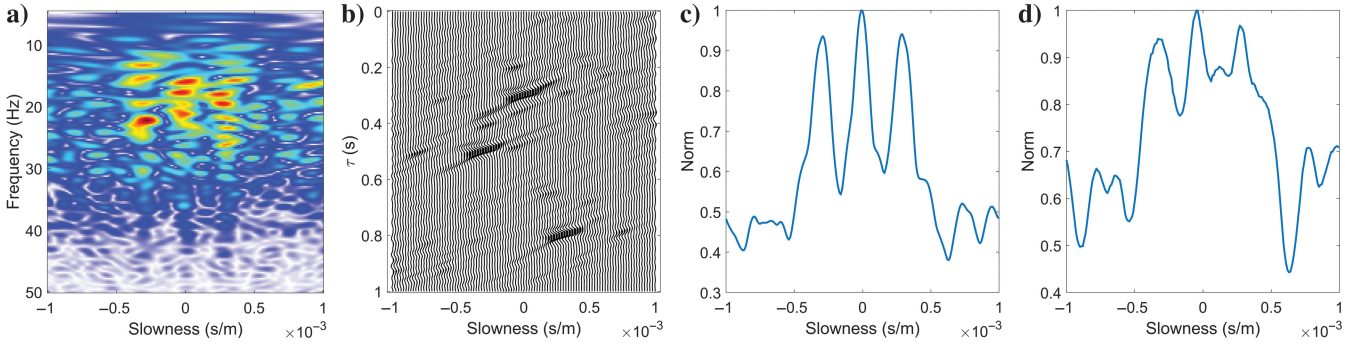


Figure 3. Comparing the $\ell_2$-norm maps for Radon coefficients in the $\tau$-$p$ and $f$-$p$ domain.

$\tilde{\mathbf{m}}_{T^k}$, encompassing coefficients from all presently selected groups up to the $k$th iteration. We then use an ADMM solver to minimize the cost function defined in equation 19, resulting in the solution $\hat{\mathbf{m}}^k_{T^k}$ at the $k$th iteration. Importantly, the Radon operator is exclusively applied within this selected subspace, $\tilde{\mathbf{m}}_{T^k}$, as opposed to the entire Radon space. This optimization strategy significantly reduces the computational burden, enhancing efficiency.

Using the OMP, we can fit the data with the minimum groups selected. The $\ell_1$-norm-regularized minimization problem inside the OMP helps us achieve sparsity inside the selected groups. Therefore, it produces the final result, not only with the minimum number of groups but also with sparsity inside the groups. If we use a minimization function without the $\ell_1$ regularized term, then this algorithm can be used to solve a general group sparse problem instead of strong sparse problems.

## EXAMPLES

In this section, we will assess the proposed algorithm's performance using a combination of synthetic and real-data examples. To gauge the quality of the denoised results, we will use the signal-to-noise ratio (S/N) calculation, which is defined as:

$$\mathrm{S/N_{out}} = 10 \log \frac{\|\mathbf{d}_c\|_2^2}{\|\mathbf{d}_c - \mathbf{d}_r\|_2^2}. \tag{20}$$

In this equation, $\mathbf{d}_c$ represents the original clean data, whereas $\mathbf{d}_r$ signifies the denoised output.

### 2D synthetic examples

We begin by assessing the robust interpolation performance of the proposed algorithm, using the same synthetic data shown in Figure 2. This test involves a comparison between the new method and three other cases: a traditional $\ell_2$-$\ell_1$ sparse inversion; an $\ell_2$-$\ell_1$ sparse inversion with a sparse group constraint; and an $\ell_1$-$\ell_1$ sparse inversion without a sparse group constraint. Figure 4 illustrates the reconstructed results using these different methods. As anticipated, the $\ell_2$-$\ell_1$ inversion in Figure 4a yields the worst outcome because it lacks robustness against erratic noise. Figure 4c demonstrates the benefits of the $\ell_2$-$\ell_1$ method with group sparsity, which significantly enhances the nonrobust $\ell_2$-$\ell_1$ inversion results. Meanwhile, the $\ell_1$-$\ell_1$ inversion results in Figure 4e offer an improvement compared with Figure 4a, though still inadequate in the reduction of strong erratic noise. Ultimately, Figure 4g presents the denoised result achieved through $\ell_1$-$\ell_1$ inversion coupled with group sparsity, standing out as the best among all tested methods.

Figure 5 shows the estimated Radon coefficients in the $f$-$p$ domain, whereas Figure 6 depicts the corresponding Radon coefficients in the $\tau$-$p$ domain. Through the application of the proposed algorithm, noise and smear effects within the Radon transform in the $\tau$-$p$ domain are successfully removed. This outcome highlights the superior robust interpolation results of the proposed robust group sparsity algorithm, particularly in managing erratic noise.

### 3D synthetic examples

Our algorithm is adaptable to 3D scenarios using the 3D Radon transform. Figure 7a illustrates a basic 3D cube featuring three linear events. Figure 7b presents the synthetic data contaminated with

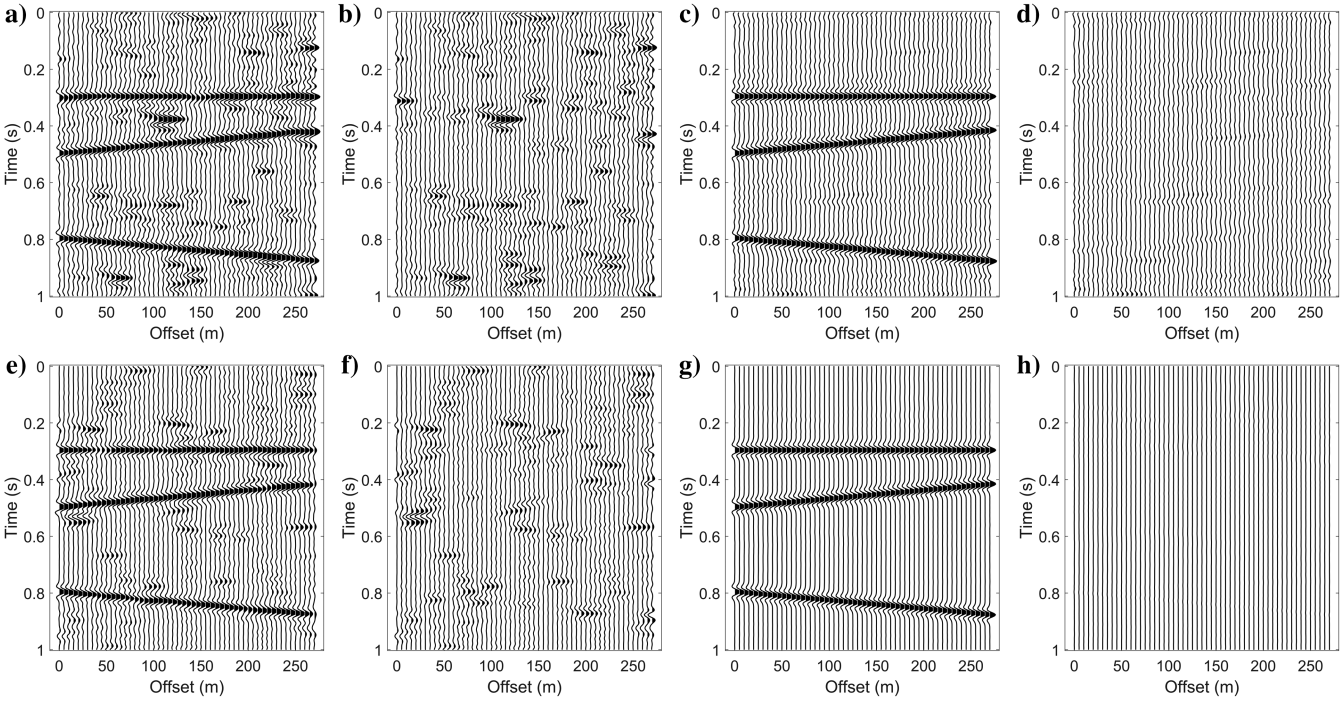


Figure 4. (a) Interpolated result of $\ell_2$-$\ell_1$ sparse inversion (S/N = 2.7 dB), (c) interpolated result of $\ell_2$-$\ell_1$ sparse inversion with group sparsity (S/N = 9.8 dB), (e) interpolated result of $\ell_1$-$\ell_1$ sparse inversion (S/N = 5.5 dB), and (g) interpolated result of $\ell_1$-$\ell_1$ sparse inversion with group sparsity (S/N = 26.4 dB). (b, d, f, and h) errors between corresponding interpolated result and Figure 2d.

erratic noise, with 90% of the traces missing. For a better comparison, Figure 7c shows a frequency slice of 3D Radon coefficients from the clean data, and Figure 7d reveals a corresponding slice from the noise-contaminated data. Whereas the peaks corresponding to the three original events can be identified in the clean data, discerning them in Figure 7d is challenging due to the missing traces and erratic noise. Figure 7e and 7f demonstrate the $\ell_2$-norm of each slowness group, enabling the correct peaks to be identified once more. Finally, the proposed robust reconstruction method, applied to Figure 7b in conjunction with the 3D Radon transform, is depicted in Figure 8. From it, it is clear that our algorithm effectively recovers the desired three linear events, demonstrating the effectiveness.

## Real-data example 1: Marine data with swell noise

We first apply the proposed method to eliminate a specific type of erratic noise in marine data. Figure 9 shows a shot gathered from a marine data set, consisting of 310 traces with a total sample time of 6.142 s and a sample interval of 0.002 s. This shot gather is notable for the presence of high-amplitude swell noise, considered a typical form of erratic noise (Bekara and van der Baan, 2010). When working with shot gathers containing curved events, our method can be applied to small windows, treating the events as linear within those confines. In this example, we partition the shot gather into small windows of every 40 traces, allowing for a 10-trace overlap between each window. We also use the stagewise weak OMP (SWOMP) (Blumensath and Davies, 2009) to accelerate the reconstruction process because this real-data example contains more dips than the previous synthetic counterpart. Unlike OMP, which selects only one group in each iteration, SWOMP identifies all groups within a factor of $\alpha$ of the largest group $\mathbf{m}_j$ as follows:

$$T^k = T^{k-1} \cup \left\{ i : \sum_{i=1}^{N} \|\mathbf{m}_i\|_2 \geq \alpha * \max \sum_{i=1}^{N} \|\mathbf{m}_j\|_2 \right\}. \quad (21)$$

Figure 10 showcases the denoised shot gathered using different methods. It is clear that our approach excels in noise elimination and maintaining the continuity of the useful data.

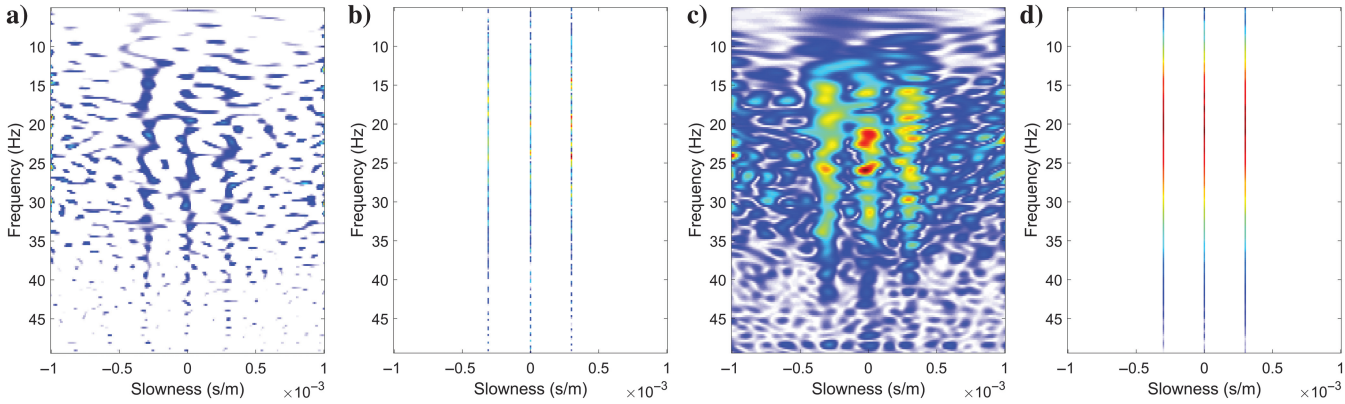


Figure 5. (a) Estimated $f$-$p$ panel by $\ell_2$-$\ell_1$ sparse inversion, (b) estimated $f$-$p$ panel by $\ell_2$-$\ell_1$ sparse inversion with group sparsity, (c) estimated $f$-$p$ panel by $\ell_1$-$\ell_1$ sparse inversion, and (d) estimated $f$-$p$ panel by $\ell_1$-$\ell_1$ sparse inversion with group sparsity.

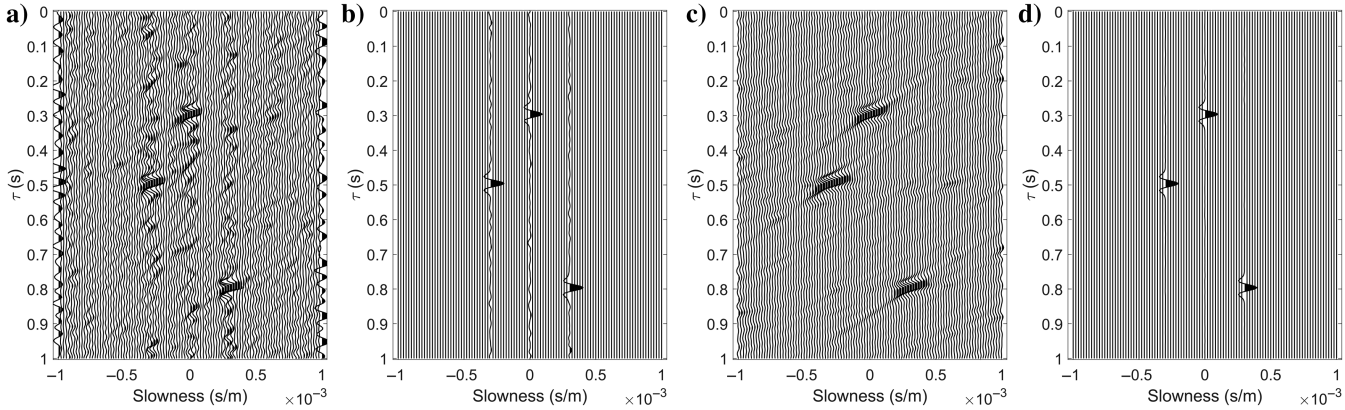


Figure 6. Corresponding $\tau$-$p$ panels. (a) $\ell_2$-$\ell_1$ sparse inversion, (b) $\ell_2$-$\ell_1$ sparse inversion with group sparsity, (c) $\ell_1$-$\ell_1$ sparse inversion, and (d) $\ell_1$-$\ell_1$ sparse inversion with group sparsity.

## Real-data example 2: Marine data deblending

In our final experiment, we test the proposed algorithm against a distinct noise type: the blending noise that arises during simultaneous source seismic data collection. This acquisition method, wherein seismic sources are fired at random with overlap, is used to conserve time and amplify source density (Beasley, 2008; Berkhout et al., 2008). Such random firing leads to what is known as blending noise.

Although blending noise appears coherent in the common-shot gather, its interferences manifest randomly in the common-receiver, offset, and midpoint gathers. This means that deblending can essentially be framed as a robust denoising challenge within these gathers. Our test uses 2D prestack marine shot records from an SEG

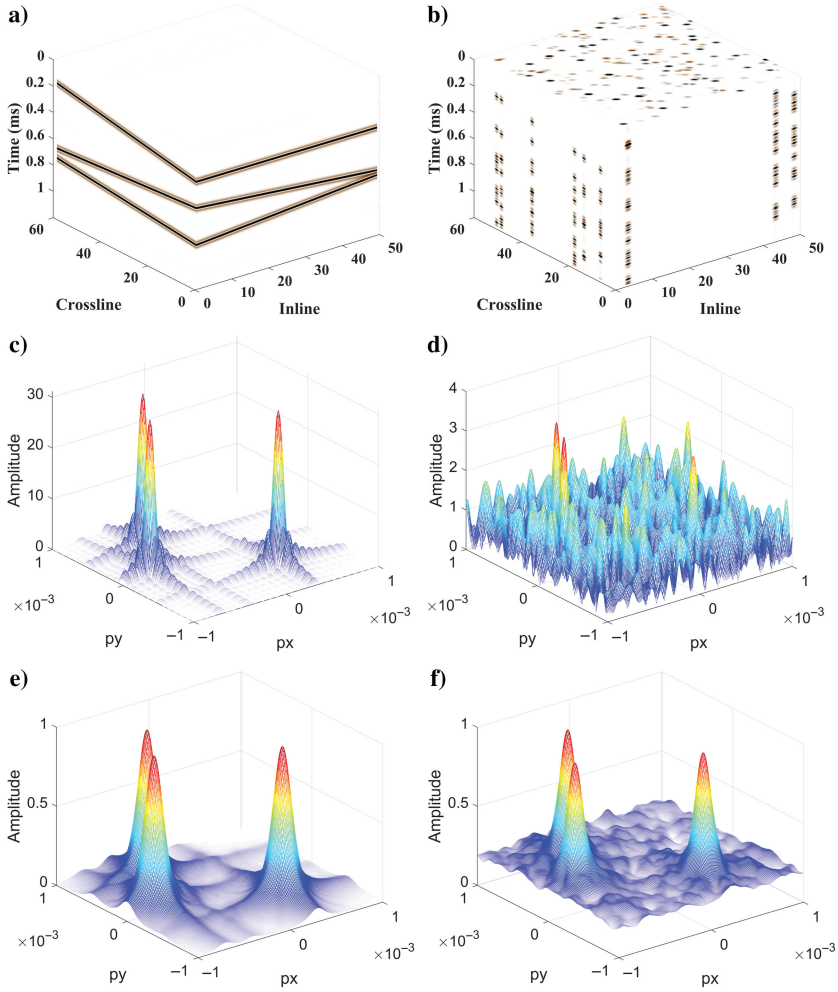


Figure 7. (a) Three-dimensional synthetic data, (b) 2D synthetic data with erratic noise and 90% of traces missed, (c) 3D Radon coefficients of (a) at one frequency slice, (d) 3D Radon coefficients of (b) at one frequency slice, (e) 2D norm map of slowness $p$ for clean data, and (f) 2D norm map of slowness $p$ for decimated noise data.

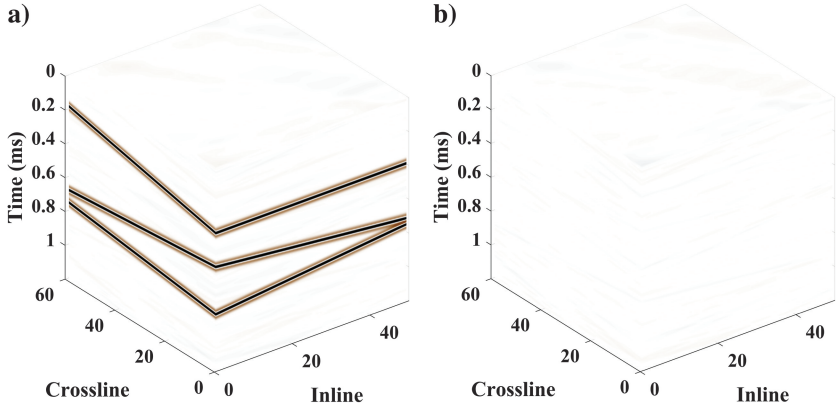


Figure 8. (a) Interpolated result (S/N = 27.3 dB) and (b) errors between (a) and Figure 7a.

open-source data set, comprising 1001 shot gathers and 120 traces each. By numerically blending three neighboring shots into one gather and arbitrarily eliminating half of the traces, we are able to simulate the deblending challenge. For this example, we apply the robust method with and without groups. Figure 11 shows the data in one of the common-offset gathers. To handle this blending noise and reconstruct the missing traces, we apply our method to 100 trace windows with 40 overlaps at the edges. This process is repeated across all common-offset gathers, leading to the final deblending result on the common-shot gathers. Figure 12 exhibits the final deblending and reconstruction result on one such common-shot gather. A close examination of Figure 12e reveals that our algorithm effectively eliminates the deblending noise and recovers useful signals. Comparing the results in Figure 12c and 12e, the result of the method with groups is better than the result without groups. The minimal differences (Figure 12f) between our results and the original data affirm a satisfactory removal of deblending noise.

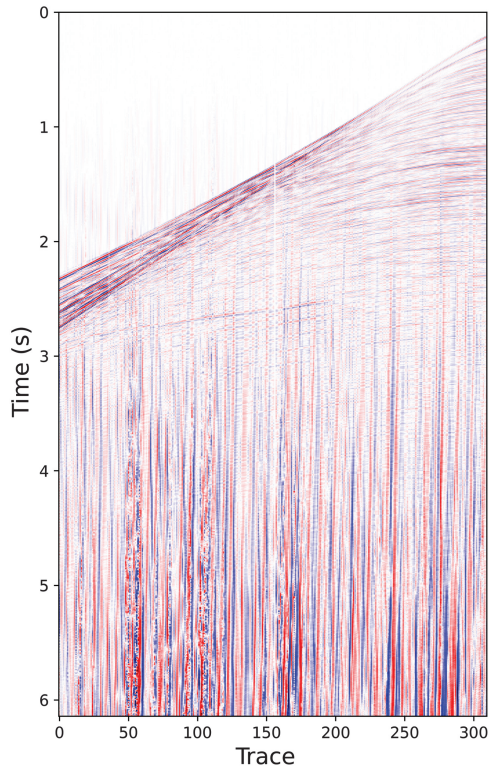


Figure 9. A 2D real data shot gather with strong swell noise.

## DISCUSSION

The paper introduces an innovative approach for implementing the robust Radon transform, focusing exclusively on the linear Radon transform. Given that the method operates within the frequency domain, it is not suitable for the hyperbolic Radon transform. When dealing with substantial real-data examples, it is advantageous to work with smaller windows to optimize computational efficiency. The choice of window size depends on the nature of the data; it should be such that the real data appear linear within these windows. However, excessively small window sizes are to be avoided because they can result in an imbalance between the energy of erratic noise and coherent signals, leading to increased computational costs.

Our proposed method is applied to address the simultaneous source separation problem. We acknowledge that various other methods exist for tackling this issue, such as those based on the robust Radon transform (Ibrahim and Sacchi, 2014, 2015) or those operating in different transform domains like the Fourier domain (Abma et al., 2010), the Seislet domain (Chen et al., 2014), and the Curvelet domain (Kontakis and Verschuur, 2017). Our method generated comparable results to other deblending methods. However, it is important to note that the primary objective of our paper is not to compare the deblending performance of our algorithm against other methods. Instead, our primary goal is to present a

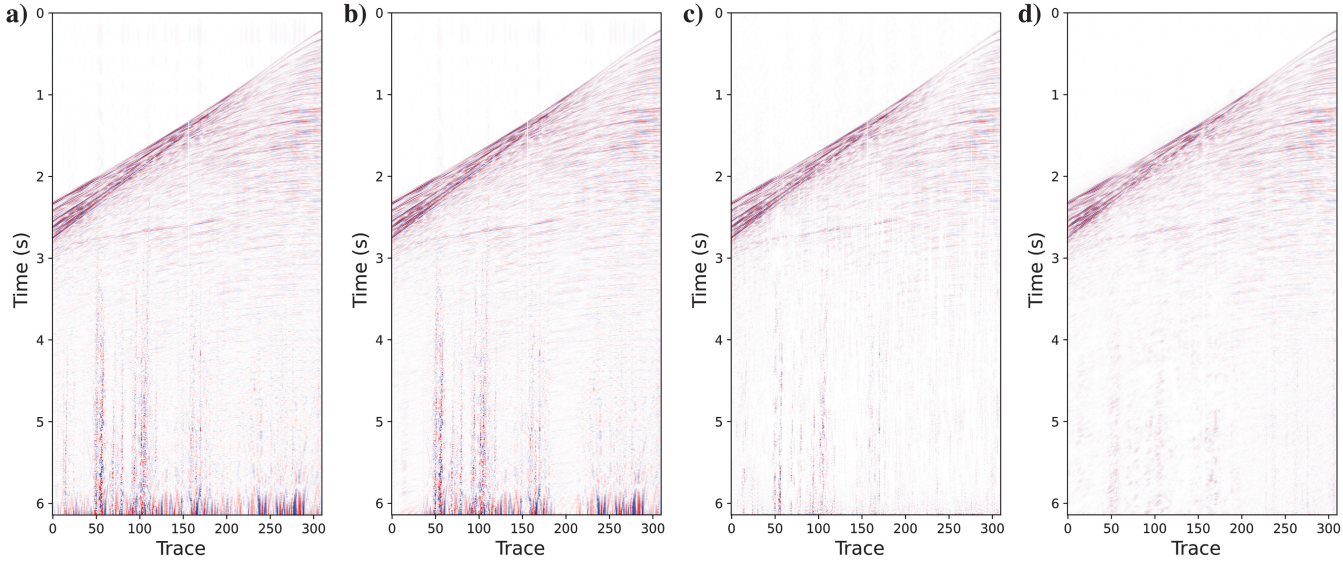


Figure 10. (a) Denoised result by $\ell_2$-$\ell_1$ sparse inversion, (b) denoised result by $\ell_2$-$\ell_1$ sparse inversion with group sparsity, (c) denoised result by $\ell_1$-$\ell_1$ robust sparse inversion, and (d) denoised result by $\ell_1$-$\ell_1$ sparse inversion with group sparsity.

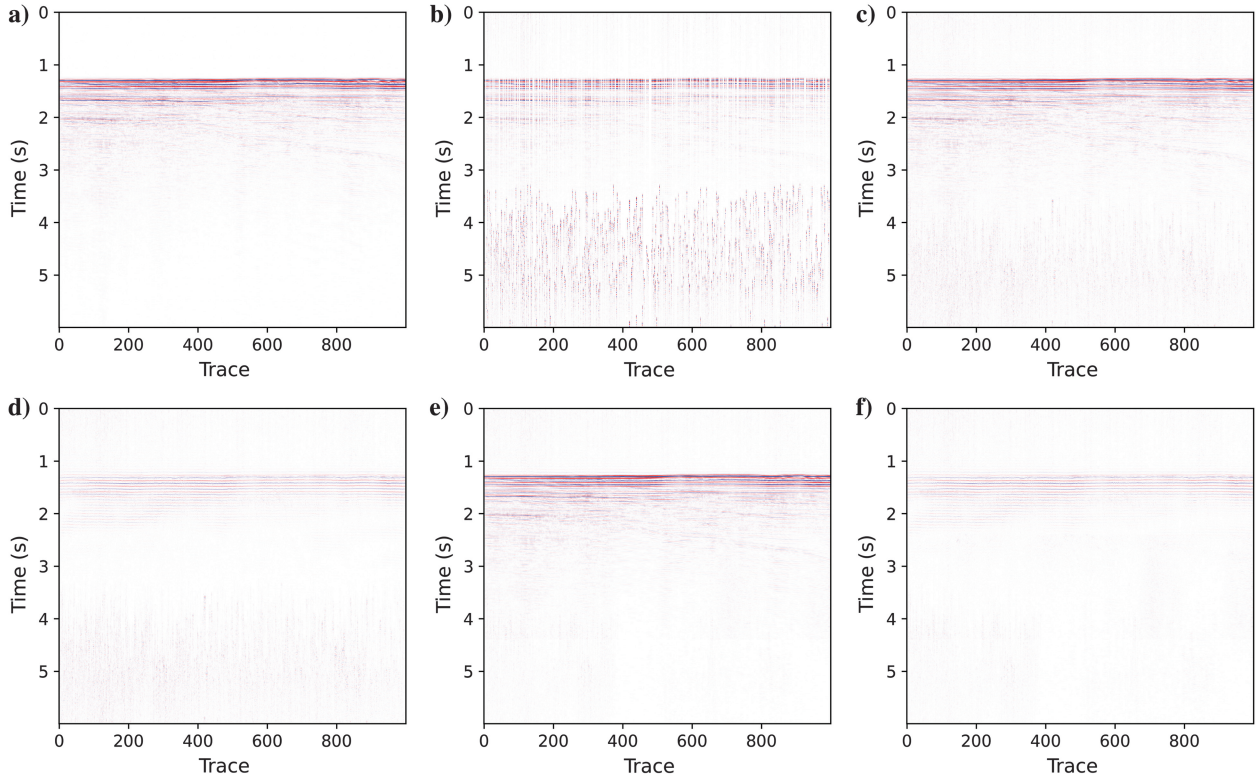


Figure 11. One common-offset gather. (a) Original data, (b) data with blended noise and 50% of random missed traces, (c) reconstructed result by $\ell_1$-$\ell_1$ method without using groups (S/N = 9.9 dB), (d) errors between the reconstructed result (c) and the original data, (e) reconstructed result by the proposed method (S/N = 12.2 dB), and (f) errors between the reconstructed result (e) and the original data.

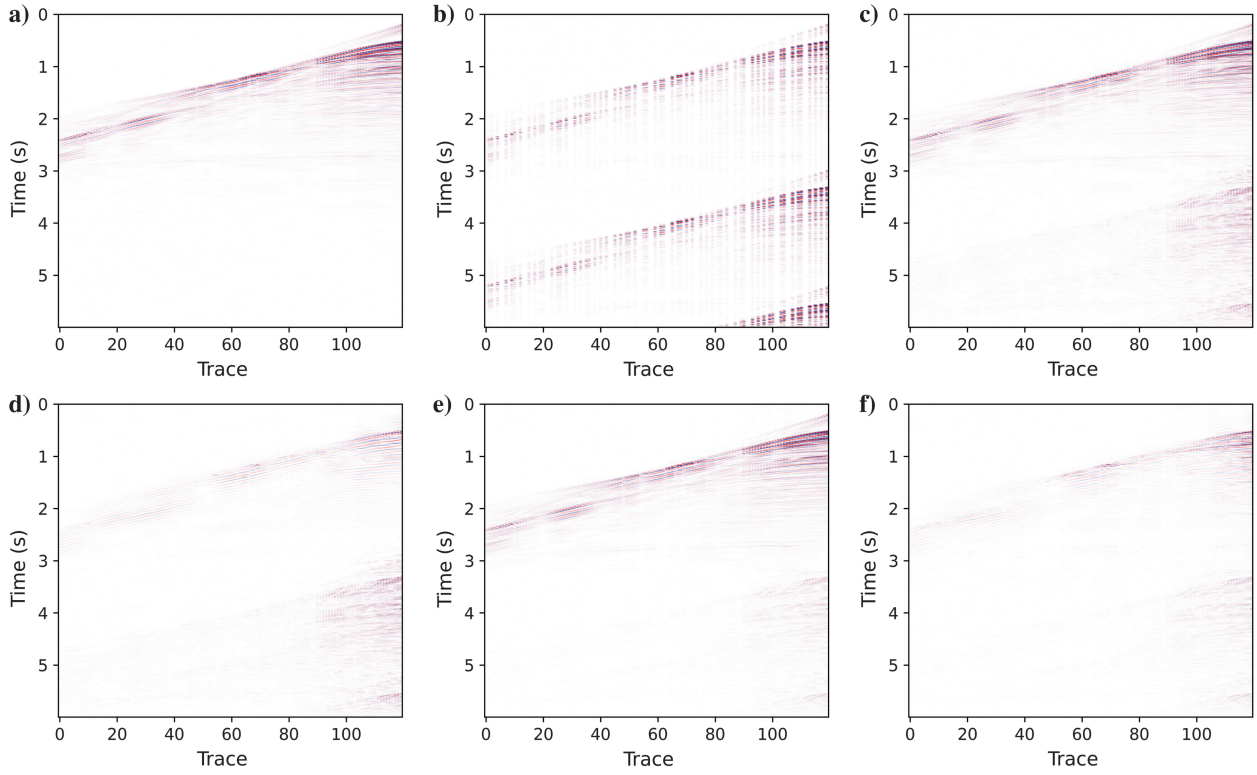


Figure 12. One common-shot gather. (a) Original data, (b) blended shot gather with three shots in one shot gather and 50% of random missed traces, (c) reconstructed result by $\ell_1$-$\ell_1$ method without using groups (S/N = 9.2 dB), (d) errors between the reconstructed result (c) and the original data, (e) reconstructed result by the proposed method (S/N = 11.3 dB), and (f) errors between the reconstructed result (e) and the original data.

robust OMP algorithm for solving the strong group sparsity problem, which can be used to compute a robust Radon denoiser for mitigating erratic seismic noise. Unlike most robust inversion methods that attempt to optimize all coefficients in the transform domains to best match the data, our approach focuses on optimizing coefficients within selected groups, with the Radon operator exclusively operating within the chosen subspace. Furthermore, when combined with the SWOMP, which selects multiple groups in each iteration, our method offers significant cost savings. In our case, for real-data examples, we use only a total of five iterations inside the OMP to process each window.

In general, solving the strong group sparsity problem entails minimizing the two regularized terms as outlined in equation 5, making it more complex than dealing with a cost function containing just one regularized term. The complexity increases further when the $\ell_2$ term is replaced with a robust M-estimator to address the robust strong group sparse problem provided in equation 6. Our approach leverages OMP, allowing us to minimize the cost function with a single regularized term in each iteration, and the algorithm terminates when all required groups are selected rather than computing all coefficients as in other sparse inversion methods. Therefore, our method is simple, faster, and more straightforward for robustly solving the strong sparse problem.

## CONCLUSION

We proposed a robust group sparse inversion algorithm that can provide a better sparse estimation and is robust to erratic noise simultaneously. The essence of the proposed method is based on the OMP. We divide the Radon coefficients in the $f$-$p$ domain into different slowness groups. During each iteration, the algorithm selected the group with the maximum norm. Subsequently, the Radon coefficients located within all currently selected groups directly fit the seismic data in the $t$-$x$ domain. The modified $\ell_1$-$\ell_1$ ADMM solver solves the coefficient optimization problem, which makes the algorithm robust to erratic noise. Our tests on synthetic and real-data examples prove the effectiveness and robustness of the proposed method.

## ACKNOWLEDGMENTS

The first author expresses gratitude to the CREWES industrial sponsors, Natural Science and Engineering Research Council of Canada (NSERC), through the grant CRDPJ 543578-19. The contribution of the second author is partially supported by the National Key R&D Program of China under grant 2021YFA0716904 and partially supported by the National Natural Science Foundation of China under grant 42374135.

## DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

## REFERENCES

Abma, R., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts, cp-161, doi: 10.3997/2214-4609.201400611.

Bach, F. R., 2008, Consistency of the group lasso and multiple kernel learning: Journal of Machine Learning Research, **9**, 1179–1225.

Beasley, C. J., 2008, A new look at marine simultaneous sources: The Leading Edge, **27**, 914–917, doi: 10.1190/1.2954033.

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM Journal on Imaging Sciences, **2**, 183–202, doi: 10.1137/080716542.

Bekara, M., and M. van der Baan, 2010, High-amplitude noise detection by the expectation-maximization algorithm with application to swell-noise attenuation: Geophysics, **75**, no. 3, V39–V49, doi: 10.1190/1.3428749.

Berkhout, A. G., G. Blacquière, and E. Verschuur, 2008, From simultaneous shooting to blended acquisition: 78th Annual International Meeting, SEG, Expanded Abstracts, 2831–2838, doi: 10.1190/1.3063933.

Blumensath, T., and M. E. Davies, 2009, Stagewise weak gradient pursuits: IEEE Transactions on Signal Processing, **57**, 4333–4346, doi: 10.1109/TSP.2009.2025088.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends in Machine Learning, **3**, 1–122, doi: 10.1561/2200000016.

Candès, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling & Simulation, **5**, 861–899, doi: 10.1137/05064182X.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129–159, doi: 10.1137/S003614450037906X.

Chen, Y., S. Fomel, and J. Hu, 2014, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, **79**, no. 5, V179–V189, doi: 10.1190/geo2013-0449.1.

Chen, Y., Z. Peng, M. Li, F. Yu, and F. Lin, 2019, Seismic signal denoising using total generalized variation with overlapping group sparsity in the accelerated ADMM framework: Journal of Geophysics and Engineering, **16**, 30–51, doi: 10.1093/jge/gxy003.

Durrani, T., and D. Bisset, 1984, The Radon transform and its properties: Geophysics, **49**, 1180–1187, doi: 10.1190/1.1441747.

Elhamifar, E., and R. Vidal, 2011, Robust classification using structured sparse representation: Proceedings/CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1873–1879.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319, doi: 10.1190/1.1598124.

Gülünay, N., 2003, Seismic trace interpolation in the Fourier transform domain: Geophysics, **68**, 355–369, doi: 10.1190/1.1543221.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, no. 6, WB173–WB187, doi: 10.1190/1.3506147.

Herrmann, P., T. Mojesky, M. Magesan, and P. Hugonnet, 2000, De-aliased, high-resolution Radon transforms: 70th Annual International Meeting, SEG, Expanded Abstracts, 1953–1956, doi: 10.1190/1.1815818.

Ibrahim, A., and M. D. Sacchi, 2014, Simultaneous source separation using a robust Radon transform: Geophysics, **79**, no. 1, V1–V11, doi: 10.1190/geo2013-0168.1.

Ibrahim, A., and M. D. Sacchi, 2015, Fast simultaneous seismic source separation using Stolt migration and demigration operators: Geophysics, **80**, no. 6, WD27–WD36, doi: 10.1190/geo2015-0044.1.

Jian, M., G. Jinghuai, and C. Wenchao, 2006, On the denoising method of prestack seismic data in wavelet domain: 76th Annual International Meeting, SEG, Expanded Abstracts, 2852–2856, doi: 10.1190/1.2370118.

Kontakis, A., and D. Verschuur, 2017, Using a hybrid focal: Curvelet transform for deblending: 87th Annual International Meeting, SEG, Expanded Abstracts, 4903–4908, doi: 10.1190/segam2017-17671812.1.

Li, J., and M. D. Sacchi, 2021, An lp-space matching pursuit algorithm and its application to robust seismic data denoising via time-domain Radon transforms: Geophysics, **86**, no. 3, V171–V183, doi: 10.1190/geo2020-0136.1.

Li, J., and M. D. Sacchi, 2022, Robust reconstruction via orthogonal matching pursuit with Fourier operators: Second International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 2947–2951, doi: 10.1190/image2022-3751045.1.

Lin, T. T., and F. J. Herrmann, 2013, Robust estimation of primaries by sparse inversion via one-norm minimization: Geophysics, **78**, no. 3, R133–R150, doi: 10.1190/geo2012-0097.1.

Majumdar, A., and R. K. Ward, 2009, Classification via group sparsity promoting regularization: IEEE International Conference on Acoustics, Speech and Signal Processing, 861–864.

Mallat, S. G., and Z. Zhang, 1993, Matching pursuits with time-frequency dictionaries: IEEE Transactions on Signal Processing, **41**, 3397–3415, doi: 10.1109/78.258082.

Maronna, R. A., 1976, Robust m-estimators of multivariate location and scatter: The Annals of Statistics, **4**, 51–67, doi: 10.1214/aos/1176343347.

Naghizadeh, M., 2012, Seismic data interpolation and denoising in the frequencywavenumber domain: Geophysics, **77**, no. 2, V71–V80, doi: 10.1190/geo2011-0172.1.

Nardi, Y., and A. Rinaldo, 2008, On the asymptotic properties of the group lasso estimator for linear models: Electronic Journal of Statistics, **2**, 605–633, doi: 10.1214/08-EJS200.

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

Rodriguez, I. V., M. Sacchi, and Y. J. Gu, 2012, Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory: Geophysical Journal International, **188**, 1188–1202, doi: 10.1111/j.1365-246X.2011.05323.x.

Sacchi, M. D., and T. J. Ulrych, 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, 1169–1177, doi: 10.1190/1.1443845.

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, doi: 10.1109/78.651165.

Scales, J. A., and A. Gersztenkorn, 1988, Robust methods in inverse theory: Inverse Problems, **4**, 1071, doi: 10.1088/0266-5611/4/4/010.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani, 2013, A sparse-group lasso: Journal of Computational and Graphical Statistics, **22**, 231–245, doi: 10.1080/10618600.2012.681250.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society: Series B (Methodological), **58**, 267–288.

Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, no. 6, V123–V132, doi: 10.1190/1.3245216.

Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with highresolution time-variant Radon transforms: Geophysics, **67**, 644–656, doi: 10.1190/1.1468626.

Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank-reduction filtering for erratic noise: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0129.1.

Tropp, J. A., and A. C. Gilbert, 2007, Signal recovery from random measurements via orthogonal matching pursuit: IEEE Transactions on Information Theory, **53**, 4655–4666, doi: 10.1109/TIT.2007.909108.

Van Groenestijn, G., and D. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, no. 3, A23–A28, doi: 10.1190/1.3111115.

Vera Rodriguez, I., D. Bonar, and M. Sacchi, 2012, Microseismic data denoising using a 3C group sparsity constrained time-frequency transform: Geophysics, **77**, no. 2, V21–V29, doi: 10.1190/geo2011-0260.1.

Vincent, M., and N. R. Hansen, 2014, Sparse group lasso and high dimensional multinomial classification: Computational Statistics & Data Analysis, **71**, 771–786, doi: 10.1016/j.csda.2013.06.004.

Wen, F., P. Liu, Y. Liu, R. C. Qiu, and W. Yu, 2016, Robust sparse recovery in impulsive noise via lp-l1 optimization: IEEE Transactions on Signal Processing, **65**, 105–118, doi: 10.1109/TSP.2016.2598316.

Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87–V95, doi: 10.1190/1.1993713.

Yang, J., and Y. Zhang, 2011, Alternating direction algorithms for l1-problems in compressive sensing: SIAM Journal on Scientific Computing, **33**, 250–278, doi: 10.1137/090777761.

Yuan, M., and Y. Lin, 2006, Model selection and estimation in regression with grouped variables: Journal of the Royal Statistical Society: Series B (Statistical Methodology), **68**, 49–67, doi: 10.1111/j.1467-9868.2005.00532.x.

Biographies and photographs of the authors are not available.