

Research Article

Sparse Regularization Based on Orthogonal Tensor Dictionary Learning for Inverse Problems

Diriba Gemechu 问

Department of Mathematics, Mettu University, Mettu 318, Ethiopia

Correspondence should be addressed to Diriba Gemechu; dirrog1@gmail.com

Received 3 June 2022; Revised 23 February 2023; Accepted 16 December 2023; Published 15 February 2024

Academic Editor: Francisco Rossomando

Copyright © 2024 Diriba Gemechu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In seismic data processing, data recovery including reconstruction of the missing trace and removal of noise from the recorded data are the key steps in improving the signal-to-noise ratio (SNR). The reconstruction of seismic data and removal of noise becomes a sparse optimization problem that can be solved by using sparse regularization. Sparse regularization is a key tool in the solution of inverse problems. They are used to introduce prior knowledge and make the approximation of ill-posed inverses feasible. It deals with ill-posedness by replacing an ill-posed inverse problem with a well-posed problem that has a solution close to the true solution. In the last 2 decades, interest has shifted from linear toward nonlinear regularization methods even for linear inverse problems. In inverse problems, regularizations serve as stabilizing the solution of ill-posed inverse problems and give a solution that adequately fits measurements without producing unjustifiably complex artifacts. In this paper, we present a novel sparse regularization based on a tensor-based dictionary method for inverse problems (seismic data interpolation and denoising). This regularization avoids the vectorization step for sparse representation of seismic data during the reconstruction process. The key step in sparsifying signals is the choice of sparsity-promoting dictionary learning. The learning-based approach can adaptively sparsify datasets but has high computational complexity and involves no prior-constraint pattern information for the dataset. Many existing dictionary learning methods would be computationally infeasible for the high dimensional seismic data processing. These methods also destroy the essential information as well as it reduces the discriminability and expressibility of the signal, since they deal with vectorization. In this paper, the orthogonal tensor dictionary learning that learns a dictionary from the input data by employing orthogonality and separability is proposed as sparse regularization for the inverse problems. The performance of the proposed method was validated in seismic data interpolation and denoising individually as well as simultaneously. Numerical examples of synthetic and real seismic datasets demonstrate the validity of the proposed method. The SNR of the recovered data confirms that the proposed method is the most effective method than K-singular value decomposition and orthogonal dictionary learning methods.

1. Introduction

In exploration geophysics, seismic data processing is an essential task in processing various properties of the earth's subsurface. Due to the physical and budget constraint, only a subset of seismic data is acquired for this process. The acquired seismic data may be subjected to various noise contamination, and some important significant traces may also distort. To increase the resolution of the seismic records, random noise attenuation and interpolation are the two critical steps in seismic data processing. Improving the quality of seismic data by removing random noises and reconstructing the missing irregular seismic traces play a vital role in providing high-resolution processing, oil, and gas exploration, multiple suppression, and wave-equation migration [1]. Seismic event detection, migration, and inversion demand a high quality of seismic data [2]. Different methods have been proposed for seismic noise attenuation and interpolation [3–5]. Most methods in seismic noise suppression are applied in the transform domain in which the signal and noise are distinguishable [6]. Even though reconstructing the missing seismic traces (interpolation) and attenuating seismic noise (denoising) methods are different in their output, they are nearly identical operations in the process of obtaining clear seismic data at the final [7]. Based on the prediction method in the time–spatial t–x domain [8, 9], the frequency–spatial f–x domain [10, 11] based on the predictability of linear events without the prior knowledge of lateral coherence of the events has been proposed. In the frequency–wavenumber, f–x domain [7, 12] has been proposed by utilizing a convolutional prediction filter computed from the low-frequency parts to predict the high-frequency components. These methods are only applied to the seismic data with linear events and are applied for only regularly sampled seismic data.

Based on the rank-reduction Cadzow method such as truncated singular value decomposition (SVD)-based matrix rank reduction of constant frequency slices for trace interpolation [13], multichannel singular spectrum analysis [3], adaptive rank-reduction based on the energy entropy [14] in solving rank reduction were proposed. Low-rank matrix completion [15, 16], tensor higher order SVD [17], and nuclearnorm minimization-based matrix completion [18] have been proposed in seismic data restoration as an extension of Cadzow method.

Based on the wave-equation, seismic data restoration utilizes the inherent constraint of seismic data from wave equation to reconstruct seismic data [1]. Different methods have been proposed based on the wave-equation method [19]. This method attains good performance in the restoration of seismic data. However, it depends on only the known velocity model.

The fourth method is based on dictionary transform, promoting sparsity in seismic data processing. In seismic data processing, interpolation and denoising are the two inverse problems in which we can employ sparsity constraints [20]. The sparse dictionary learning (DL) algorithm has been also proposed to attenuate random noise in a transform-based denoising framework [21]. Sparse representation represents input data as the linear combination of basis elements such that most of the representation coefficients are zero [22]. For a given signal $x \in \mathbb{R}^n$, the sparsest representation is to find a sparse vector $\alpha \in \mathbb{R}^m$ such that $y = D\alpha$, where *D* is the dictionary and each of its columns is an atom. The idea of sparsity in seismic data processing has been used in two different ways: fixed (analytic) basis transform and learning (adaptive) dictionaries.

Fixed basis transforms have been applied in seismic data processing by using a known set of basis functions to estimate representation coefficients of the input data. Wavelet multiresolution analysis [23], physical wavelet transforms [24], Radon transforms [25, 26], Fourier transform [27, 28], curvelet transforms [29, 30], seislet transforms [31, 32], and shearlet transform [33] have been proposed. In the nonsubsampled contourlet transform domain, threshold shrink method [34] for denoising seismic data has been also proposed. In analytic DL approaches, the analytic construction is employed, and the mathematical model of the signal is formulated to represent the model. These methods are simple and have fast computation in creating data features for different applications in seismic data processing.

Adaptive/learned DL uses the learned basis functions from the input data to estimate the missing data and restore

the degraded data by a sparsity-promoting model. Based on adaptive/learned DL, different methods have been proposed in image processing. The adaptive DL algorithm has also been proposed to attenuate automatic coherent noise [35]. This approach can learn the features of both signals and coherent noise and leave obvious morphological differences in the dictionary atoms. Principal component analysis (PCA) [36], generalized PCA [37], method of optimal directions (MOD) [38, 39], and K-singular value decomposition (K-SVD) [40] are adaptive/learned dictionary proposed in image science. PCA is widely used in statistics for multivariate analysis. It reduces the dimensionality of the data by preserving the variability of data variables. K-SVD learns an overcomplete dictionary from the noisy image patches and uses the learned dictionary to sparsely represent model and image denoising [41]. K-SVD has been proposed for signal processing applications [42] and seismic data denoising [38]. However, the input data are considered as a vector in K-SVD, and then at each iteration step, SVD is used to transform the matrix to update each column of the dictionary [20]. It leads to the redundancy of dictionaries in which most of the atoms are similar or contribute little role in the representation of input data. K-SVD-based sparse representation has also been used in seismic data denoising and compression [43]. In this case, K-SVD is used to sparsely represent the 4D seismic dataset to find a representation with as few coefficients as possible while preserving the main characteristics of the data. In the learning dictionary for sparsity-based seismic data restoration, high redundancy dictionary leads to a poor sparse approximation of data. This approach is computationally infeasible when dealing with seismic data and requires high computational cost [44]. Data-driven tight frame (DDTF) [45], which learns dictionary with a prescribed block Toeplitz structure and satisfies the perfect reconstruction property for denoising, has been proposed to reduce the computational complexity of K-SVD. The high performance of DDTF in interpolation and denoising of high dimensional seismic data was analyzed [46, 47].

The hybrid method, which combines the advantage of fixed basis and learned-based DL, has been applied in seismic data processing. A double sparsity method which is based on seislets and DDTF [48] was proposed. However, most of these methods deal with vectors leading to loss of seismic data structure and poor sparse representation. To overcome this problem, tensor-based DL methods [49, 50] have been proposed for various sparsity-based restoration. In the tensor DL method, the input data are treated as tensors instead of vectors and the dictionary is learned by tensor decomposition [44]. The DDTF of Kronecker type (KronTF) [20], tensor decomposition [17], and a Kronecker-based dictionary in dynamic computed tomography [51] have been proposed to avoid the vectorization of input data and applying a learned dictionary based on tensor decomposition.

Many existing DL methods would be computationally infeasible when dealing with high-dimensional seismic data. Moreover, the vectorization method which likely destroys the input data's important information and reduces the discriminability and expressibility of the obtained representation are applied to those methods. In this paper, we have proposed the

orthogonal tensor DL that learns a dictionary from the input data by employing orthogonality and separability. We use this method for seismic data denoising and interpolation. The separability of the dictionary makes the proposed method highly scalable, and the orthogonality among dictionary atoms leads to a very efficient sparse coding computation. The scalability and computational efficiency make the proposed method suitable for processing seismic data in the tensor domain [44]. The rest part of this paper is organized as follows. In Section 2, we discussed the basic idea of orthogonal DL and its construction which greatly simplify the computation of both dictionary updating and sparse coding for the sparse representation of input data. In Section 3, orthogonal tensor-based DL and its property are verified. The orthogonal tensor DL is applied to seismic data denoising and interpolation problems. The numerical results are also presented in Section 4. Finally, the conclusion of this paper is addressed in Section 5.

2. Orthogonal Dictionary Learning

Sparse models in representing input data are an active research area in natural image processing. In a sparse model, the local input data patches are sparsely approximated by the linear combination of atoms in which the collection of these atoms forms a dictionary. The main challenge in sparse representation is the construction of a dictionary which is computationally feasible in seismic data representation. The orthogonal dictionary is the one which can greatly simplify the computation of both dictionary updating and sparse coding for the sparse representation of input data. In orthogonal DL, we aim at finding the orthogonal dictionary $\mathbf{D} = [\mathbf{A}, \mathbf{D}] \in \mathbb{R}^{n \times n}$ whose columns are dictionary atoms such that $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_{m}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{D} = 0$ to sparsely approximate the given data $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$ $y_3, ..., y_m$ $\in \mathbb{R}^{n \times m}$, where *n* and *m* are the size of orthogonal matrices. $\mathbf{A} \in \mathbb{R}^{n \times n-m}$ contains input orthogonal atoms and $\mathbf{D} \in \mathbb{R}^{n \times m}$ is the set of atoms to be learned from the input data. The corresponding minimization problem to learn orthogonal dictionary **D** is formulated as follows [52]:

$$\underset{\mathbf{D}\in\mathbb{R}^{n\times m},\mathbf{V}\in\mathbb{R}^{n\times m}}{\operatorname{argmin}} \|\mathbf{Y}-[\mathbf{A},\mathbf{D}]\mathbf{V}\|_{\mathrm{F}}^{2}+\lambda^{2} \|\mathbf{V}\|_{0}, \tag{1}$$

such that $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_m$ and $\mathbf{A}^{\mathrm{T}}\mathbf{D} = \mathbf{0}$, where $\|.\|_{\mathrm{F}}^2$ is the Frobenius norm which is defined as $\|\mathbf{V}\|_{\mathrm{F}}^2 = (\sum_{i,j} |\mathbf{V}_{i,j}|^2)^{\frac{1}{2}}$, where $\mathbf{V}_{i,j}$

are the vectors or entries of the matrix V and $||V||_0$ denotes the number of nonzero coefficients of V. $\lambda > 0$ is the parameter that balances the trade-off between the approximation term and the sparsity term. The solution of Equation (1) is disposed to increase the elements of $\hat{\mathbf{D}}$ such that V to become as small as possible. To solve Equation (1), we can take the alternating iterative scheme by decomposing the minimization problem into two steps. In this case, the K-SVD algorithm can solve Equation (1) by employing an iterative strategy that alternates between the two steps such that Equation (1) is reduced into only one unknown by fixing one as known. Step 1 (sparse coding): For a given orthogonal dictionary $\widehat{\mathbf{D}}$, we need to find the sparse code $\mathbf{V}^{(k)}$ by solving the following minimization problem:

$$\mathbf{V}^{(k)} = \underset{\mathbf{V} \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \|\mathbf{Y} - [\mathbf{A}, \mathbf{D}^{k}]\mathbf{V}\|_{\mathrm{F}}^{2} + \lambda^{2} \|\mathbf{V}\|_{0}.$$
(2)

To solve the minimization problem in step 1, we formulate the following remark which is equivalent to Equation (2).

Remark 1. Assume that $\widehat{\mathbf{D}}^{\mathrm{T}}\widehat{\mathbf{D}} = \mathbf{I}_n$, then $\operatorname{argmin}_{\mathbf{V}} ||\mathbf{Y} - \widehat{\mathbf{D}}\mathbf{V}||_F^2 + \lambda^2 ||\mathbf{V}||_0$ has a unique solution which is given by $\mathbf{V}^* = \mathrm{T}_{\lambda}(\widehat{\mathbf{D}}^{\mathrm{T}}\mathbf{Y})$.

Based on Remark 1 and for $\widehat{\mathbf{D}}^{\mathrm{T}}\widehat{\mathbf{D}} = \mathrm{I}_n$, Equation (2) has a unique solution which is given by $\mathbf{V}^{(k)} = \mathrm{T}_{\lambda}(\widehat{\mathbf{D}}^{\mathrm{T}}\mathbf{Y})$, where T_{λ} is the hard threshold which is defined as follows:

$$T_{\lambda}(x) = \begin{cases} x, |x| > \lambda \\ 0, |x| \le \lambda \end{cases}.$$
(3)

Step 2 (dictionary updating): For a given sparse code V, we want to update the dictionary \widehat{D} by solving the minimization problem:

$$\mathbf{D}^{(k+1)} = \underset{\mathbf{D} \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \| \mathbf{Y} - [\mathbf{A}, \mathbf{D}] \mathbf{V}^{(k)} \|_{\mathrm{F}}^{2}, \tag{4}$$

such that $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_m$ and $\mathbf{A}^{\mathrm{T}}\mathbf{D} = 0$. The minimization problem in Equation (4) can be solved by employing the following remark which is equivalent to step 2.

Remark 2. The optimization problem in step 2 is equivalent to the following equation:

$$\underset{\mathbf{V}}{\operatorname{argmin}} \|\mathbf{Y} - (\mathbf{A}\mathbf{V}_{\mathbf{A}} + \mathbf{D}\mathbf{V}_{\mathbf{D}})\|_{\mathrm{F}}^{2}, \tag{5}$$

such that $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_m$ and $\mathbf{A}^{\mathrm{T}}\mathbf{D} = 0$. This minimization problem has a unique solution which is given by the following equation:

$$\mathbf{D}^* = \mathbf{R}\mathbf{H}^{\mathrm{T}},\tag{6}$$

where \mathbf{R} and \mathbf{H} are the orthogonal matrices defined by the following SVD:

$$(\mathbf{I}_n - \boldsymbol{\Phi}_{\mathbf{A}})\mathbf{Y}\mathbf{V}_{\mathbf{D}}^{\mathrm{T}} = \mathbf{R}\boldsymbol{\Sigma}\mathbf{H}^{\mathrm{T}},\tag{7}$$

where Σ is the singular value matrix of \mathbf{RH}^T , and Φ_A defined from \mathbb{R}^n to the space spanned by the columns of the dictionary \mathbf{A} is an orthogonal projection operator given by: $\Phi_A l = \mathbf{A}(\mathbf{A}^T l)$, for all $l \in \mathbb{R}^n$. Finally, we adopt the following orthogonal DL algorithm. Input: Training dataset **Y**, Input orthogonal atoms **A**. Output: Learned dictionary **D**. 1 Initialize the dictionary $D^{(0)}$ 2 For k = 0, 1, 2, ..., K3 Compute the sparse coding by using hard threshold $V_{\mathbf{D}}^{k} = T_{\lambda}((\mathbf{D}^{k})^{T}\mathbf{V}),$ 4 Run the SVD for the following matrix $(\mathbf{I}_{n} - \boldsymbol{\Phi}_{\mathbf{A}})\mathbf{Y}(\mathbf{V}_{\mathbf{D}}^{k})^{T} = \mathbf{R}\Sigma\mathbf{H}^{T}.$ 5 Update the dictionary $\mathbf{D}^{k+1} = \mathbf{R}\mathbf{H}^{T}, \mathbf{D} : = \mathbf{D}^{k+1}.$ 6 end for



3. Orthogonal Tensor Dictionary Learning

In this section, we extend the orthogonal DL that we have discussed in Section 2 by introducing tensor and learn a tensor dictionary with separability and orthogonality. In the conventional DL approach, the extracted image patches are first transferred into vectors to form training data and then the K-SVD method is used to learn a vector-based dictionary [48]. In this strategy, the inherent spatial constraints of the input data structures can be lost or distorted when the dimension of input data is very high. To overcome this problem, the tensor-based method has been investigated for DL to achieve better performance in signal/image processing [53-55]. Tensor is a multidimensional array [56]. An *n*th order tensor is denoted as $\mathbf{X} \in \mathbb{R}^{i_1 \times i_2 \times i_3 \dots \times i_n}$ whose (i_1, \dots, i_n) $i_2, ..., i_n$) elements is $X_{i_1, i_2, ..., i_n}$, $1 \le i_k \le I_k$, k = 1, 2, 3, ..., n. Let $\mathbf{S}_{M} = {\mathbf{D} \in \mathbb{R}^{M \times M}}$ be the set of orthogonal square matrices. Then, we aim to learn the set of orthogonal dictionaries $\{D_i \in S_M, i = 1, 2\}$ to be learned from the input data simultaneously such that $D_i^T D_i = I$, for i = 1,2 and $D_2^T D_1 = 0$ in the tensor domain. In this paper, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ is considered as a tensor of two dimensions which can be written as follows:

$$\mathbf{Y} = \mathbf{V} \mathbf{x}_1 \mathbf{D}_1 \mathbf{x}_2 \mathbf{D}_2, \tag{8}$$

where \mathbf{D}_1 and \mathbf{D}_2 are orthogonal matrices from the set S_M with the sparse coding tensor \mathbf{V} and $\mathbf{V} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2$ is the product of sparse coding tensor \mathbf{V} with a matrix \mathbf{D}_n , n = 1, **2**, \times_i , i = 1, 2 are the tensor mode product operators. We formulate the following minimization problem to learn these dictionaries:

$$\underset{\mathbf{D}_{1},\mathbf{D}_{2},\mathbf{V}\in\mathbb{R}^{M_{1}\times M_{2}\times N}}{\operatorname{argmin}} \|\mathbf{Y}-\mathbf{V}\times_{1}\mathbf{D}_{1}\times_{2}\mathbf{D}_{2}\|_{F}^{2}+\lambda^{2}\|\mathbf{V}\|_{0}, \qquad (9)$$

where $\|.\|_0$ denotes the nonzero elements in the tensor domain, and $\|.\|_F^2$ is the Frobenius norm of the tensor which is defined as the square roots of the sum of the absolute squares of its elements. We decompose the problem in Equation (9) into two subproblems and solve them separately.

Step 1: For a given dictionaries D_1 and D_2 , we need to find the sparse tensor $V^{(k)}$ by solving the minimization problem:

$$\mathbf{V}^{(k)} = \underset{\mathbf{V}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{V} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2\|_{\mathrm{F}}^2 + \lambda^2 \|\mathbf{V}\|_0.$$
(10)

To solve for the sparse coding **V**, we consider the vectorized version of the tensor representation in Equation (8) in terms of Kronecker dictionaries which is formulated as follows:

$$\mathbf{y} = \operatorname{vec}(\mathbf{Y}) = (\mathbf{D}_2 \otimes \mathbf{D}_1)\operatorname{vec}(\mathbf{V}), \tag{11}$$

$$\mathbf{y} = (\mathbf{D}_2 \otimes \mathbf{D}_1)\mathbf{v},\tag{12}$$

where \otimes represents the Kronecker product. As we discussed in Section 2, we can compute the corresponding sparse coding $\mathbf{V} \in \mathbb{R}^{M_1 \times M_2 \times N}$ by applying a component-wise hard thresholding. Then, the solution for the sparse coding *V* is given by the following equation:

$$\mathbf{v}^{(k)} = [\mathbf{T}_{\lambda}(\mathbf{y} \otimes \mathbf{D}_{1}^{\mathrm{T}} \otimes \mathbf{D}_{2}^{\mathrm{T}})], \tag{13}$$

where T_{λ} is an operator that keeps the coefficients larger than λ and setting the other coefficients to be zero.

Step 2: For a given sparse coding $V^{(k)}$, we need to update the dictionaries D_1 and D_2 . The tensor representation Y in Equation (8) can be unfolded in order to update the two dictionaries D_1 and D_2 as follows:

$$\begin{aligned} \mathbf{Y}_{(1)} &= \mathbf{D}_1 \mathbf{V}_{(1)} \mathbf{D}_2^{\mathrm{T}} \\ \mathbf{Y}_{(2)} &= \mathbf{D}_2 \mathbf{V}_{(2)} \mathbf{D}_1^{\mathrm{T}} \end{aligned}$$
 (14)

Then, D_1 and D_2 can be updated by solving the following minimization problem:

The dictionary \mathbf{D}_n (n = 1, 2) can be updated by using an alternating least squares method while fixing the sparse coding $V^{(k)}$. To solve for the above two dictionaries \mathbf{D}_1 and \mathbf{D}_2 , we formulate the theorem which is related to the

minimization problems in Equations (14) and ((15). The proof of this theorem is given in the appendix section.

3.1. Application to Seismic Data Interpolation and Denoising. To evaluate the effect of the proposed method on seismic data processing in terms of quality and computational efficiency, we applied for interpolation (reconstruction of missing traces) and denoising (attenuation of random noise). Let **u** be the complete data for recovery, **f** be the observed data, **R** be the trace sampling matrix which contains only 1 or 0, and ϵ denotes the amount of noise to be added on the input data. Let these data be related by the following model:

$$\mathbf{f} = \mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}. \tag{16}$$

When $\epsilon = 0$, the model in Equation (16) is the interpolation problem that can reconstruct the missing traces and if **R** is identity matrix I_n and $\epsilon \neq 0$, it corresponds to the denoising problem. After the orthogonal tensor dictionary learned from the input seismic data simultaneously, we used it to propose sparse regularization which promotes sparsity in the seismic data restoration problem. Based on the learned orthogonal tensor dictionary model in Section 3, we formulate the seismic data restoration problem by employing a new sparse regularization. Let B_i be an operator which can extract features of seismic data **u**, then the following minimization problem is formulated to reconstruct the missing traces and denoising seismic data:

$$\operatorname{rgmin}_{\mathbf{u},\mathbf{D},\mathbf{V}^{i}} \quad \sum_{i} \frac{\lambda_{1}}{2} \left(\left\| \mathbf{V}^{(i)} \times_{1} \mathbf{D}_{1} \times_{2} \mathbf{D}_{2} - \mathbf{B}_{i} \mathbf{u} \right\|_{2}^{2} + \left\| \mathbf{V}^{(i)} \right\|_{0} \right) + \frac{\lambda_{2}}{2} \| \mathbf{R} \mathbf{u} - \mathbf{f} \|_{2}^{2}$$

$$(17)$$

with the parameters λ_1 and λ_2 need to be determined. The first term in Equation (17) is the regularization term that corresponds to tensor-based orthogonal DL and the second term is the fidelity data in the projection domain with the sampling matrix **R** and the degraded data **f**. Basically, we can write $V^{(i)} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2$ as $\mathbf{D}_1 V^{(i)} \mathbf{D}_2^T$. Now we can represent $\mathbf{D}_1 V^{(i)} \mathbf{D}_2^T$ by $\mathbf{D}v_i$ and then Equation (17) becomes:

$$\underset{\mathbf{u},\mathbf{D},\mathbf{v}_{i}}{\operatorname{argmin}} \quad \sum_{i} \frac{\lambda_{1}}{2} (\|\mathbf{D}\mathbf{v}_{i}-\mathbf{B}_{i}\mathbf{u}\|_{2}^{2}+\|\mathbf{v}_{i}\|_{0}) + \frac{\lambda_{2}}{2} \|\mathbf{R}\mathbf{u}-\mathbf{f}\|_{2}^{2}.$$
(18)

We employed the alternating direction method of multipliers (ADMM) to solve Equation (18) based on the augmented Lagrangian and the separable paraboloid surrogate method. ADMM method has been widely applied to solve different minimization problems in seismic data processing [57]. The surrogate method makes the cost function separable so that all traces of seismic data can be updated simultaneously. Let $c = B_i \mathbf{u} - \mathbf{D} \mathbf{v}_i$ and h be the two auxiliary variables used to transform the nonconvex optimization problem into a convex optimization problem, then we can rewrite Equation (18) as follows:

$$\left\{\mathbf{D}^{k+1}, \mathbf{v}_{i}^{k+1}, \mathbf{u}^{k+1}, c_{i}^{k+1}\right\} = \underset{\mathbf{u}, \mathbf{D}, \mathbf{v}_{i}, c}{\operatorname{argmin}} \sum_{i} \left(\frac{\lambda_{1}}{2} \left(\|c_{i}\|_{2}^{2} + \|\mathbf{v}_{i}\|_{0}\right) + \frac{\beta}{2} \left\|\mathbf{D}\mathbf{v}_{i} + c_{i} - \mathbf{B}_{i}\mathbf{u} - \frac{h_{i}}{\beta}\right\|_{2}^{2}\right) + \frac{\lambda_{2}}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_{2}^{2}.$$
(19)

$$h^{k+1} = h^k + \beta \left(-\mathbf{D}^{k+1} \mathbf{v}_i^{k+1} - c_i^{k+1} + \mathbf{B}_i \mathbf{u}^{k+1} \right).$$
(20)

We split the minimization problem in Equation (19) into three subproblems, where λ_1 , λ_2 , and β are the regularization parameters. Subproblem 1: updating DL and sparse coding:

$$\left\{ \mathbf{D}^{k+1}, \mathbf{v}_{i}^{k+1} \right\} = \underset{\mathbf{D}, \mathbf{v}_{i}}{\operatorname{argmin}} \quad \sum_{i} \|\mathbf{D}\mathbf{v}_{i} - \mathbf{B}_{i}\mathbf{u}\|_{2}^{2} + \frac{\lambda_{1}}{2} \|\mathbf{v}_{i}\|_{0}.$$
(21)

Subproblem 2: c-subproblem. The optimization problem for the auxiliary variable *c* is given by the following equation:

$$c_{i}^{k+1} = \underset{c_{i}}{\operatorname{argmin}} \sum_{i} \left(\frac{\lambda_{1}}{2} \|c_{i}\|_{2}^{2} + \frac{\beta}{2} \left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i} - \mathbf{B}_{i} \mathbf{u}^{k} - \frac{\mathbf{h}_{i}}{\beta} \right\|_{2}^{2} \right).$$
(22)

Subproblem 3: **u**-subproblem. The corresponding minimization problem for \mathbf{u} is as follows:

$$\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} \quad \frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{\mathbf{h}_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_{2}^{2}.$$
(23)

We have discussed the solution for the one that corresponds to updating DL and sparse coding for subproblem 1 in Section 2. To solve the second subproblem, we apply the optimality condition for Equation (22):

$$\frac{\partial}{\partial c_i} \left[\sum_i \left(\frac{\lambda_1}{2} (c_i)^2 + \frac{\beta}{2} \left(\mathbf{D}^k \mathbf{v}_i^k + c_i - \mathbf{B}_i \mathbf{u}^k - \frac{h_i}{\beta} \right)^2 \right) \right] = 0,$$
(24)
$$\frac{\lambda_1}{2} \left(2c_i^{k+1} \right) + \frac{\beta}{2} \left(2\mathbf{D}^{k+1} \mathbf{v}_i^{k+1} + 2c_i^{k+1} - 2\mathbf{B}_i \mathbf{u}^k - 2\frac{h_i}{\beta} \right) = 0.$$
(25)

Then,

$$c_i^{k+1} = \frac{\beta}{\lambda_1 + \beta} \left(\mathbf{B}_i \mathbf{u}^k - \mathbf{D}^{k+1} \mathbf{v}_i^{k+1} + \frac{h_i}{\beta} \right).$$
(26)

Input: Training data set **Y**, regularization parameters, patch size. Output: Learning dictionaries $\mathbf{D}_1, \mathbf{D}_2$ 1 Initialize $D_1^{(0)}, D_2^{(0)}$, and $\mathbf{u}^{(0)}, \mathbf{c}^{(0)}, \mathbf{h}^{(0)}$. 2 For k = 0, 1, 2, ..., N3 Update the sparse coding using hard thresholding, $\mathbf{v}^{(k)} = [\mathbf{T}_{\lambda}(\mathbf{y} \otimes \mathbf{D}_1^T \otimes \mathbf{D}_2^T)],$ 4 For n = 1, 2, Update the dictionaries \mathbf{D}_n by running the SVD in Equations (A.18) and (A.20). 5 End for 6 Update, h using Equation (20) 7 Update, c using Equation (26) 8 Solve for **u** using Equation (33) 9 end for



After the dictionary **D** updated in the iteration process, the seismic data **u** should be updated with v_i and **D**. To do so, the separable paraboloid surrogate method [58] is used to solve for **u** as follows:

$$\mathbf{u}^{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} \quad \frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{h_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_{2}^{2}.$$
(27)

By using separable paraboloid surrogate method, \mathbf{u}^{k+1} is given by the following equation:

$$\mathbf{u}^{k+1} = \mathbf{u}^{k} - \frac{\frac{\partial}{\partial \mathbf{u}} \left(\frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{h_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \left\| \mathbf{R} \mathbf{u} - \mathbf{f} \right\|_{2}^{2} \right)}{\left\| \frac{\partial^{2}}{\partial u^{2}} \left(\frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{h_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \left\| \mathbf{R} \mathbf{u} - \mathbf{f} \right\|_{2}^{2} \right) \right\|,}$$
(28)

where

$$\frac{\partial}{\partial u} \left[\frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{h_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \left\| \mathbf{R} \mathbf{u} - \mathbf{f} \right\|_{2}^{2} \right]$$

$$= \frac{\partial}{\partial u} \left[\frac{\beta}{2} \sum_{i} \left(\mathbf{D}^{k} \mathbf{v}_{i}^{k} + 2\mathbf{D}^{k} \mathbf{v}_{i}^{k} c_{i}^{k} - 2\mathbf{D}^{k} \mathbf{v}_{i}^{k} \mathbf{B}_{i}^{\mathrm{T}} \mathbf{u} - 2\mathbf{D}^{k} \mathbf{v}_{i}^{k} \left(\frac{h_{i}^{k}}{\beta} \right) - 2c_{i}^{k} \mathbf{B}_{i}^{\mathrm{T}} \mathbf{u} - 2c_{i}^{k} \left(\frac{h_{i}^{k}}{\beta} \right) + \mathbf{B}_{i}^{\mathrm{T}} \mathbf{B}_{i} \mathbf{u}^{2} + 2\mathbf{B}_{i}^{\mathrm{T}} \mathbf{u} \left(\frac{h_{i}^{k}}{\beta} \right) + \left(\frac{h_{i}^{k}}{\beta} \right)^{2} \right) + \frac{\lambda_{2}}{2} \left(\mathbf{R}^{\mathrm{T}} \mathbf{R} \mathbf{u}^{2} - 2\mathbf{R}^{\mathrm{T}} \mathbf{u} \mathbf{f} + \mathbf{f}^{2} \right) \right].$$
(29)

Then,

$$\frac{\partial}{\partial u} \left[\frac{\beta}{2} \sum_{i} \left(\left\| \mathbf{D}^{k} \mathbf{v}_{i}^{k} + c_{i}^{k} - \mathbf{B}_{i} \mathbf{u} - \frac{h_{i}}{\beta} \right\|_{2}^{2} \right) + \frac{\lambda_{2}}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_{2}^{2} \right]$$
$$= \beta \sum_{i} \mathbf{B}_{i}^{\mathrm{T}} \left(\mathbf{B}_{i} \mathbf{u}^{k} - \mathbf{D}^{k+1} \mathbf{v}_{i}^{k+1} - c_{i}^{k+1} + \left(\frac{\mathbf{h}_{i}^{k}}{\beta}\right) \right) + \lambda_{2} \mathbf{R}^{\mathrm{T}} \left(\mathbf{R}\mathbf{u}^{k} - \mathbf{f} \right).$$
(30)

The denominator of Equation (28) is the curvature of the paraboloid which can be obtained as follows:

$$\frac{\partial^2}{\partial u^2} \left(\sum_i \left(\left\| \mathbf{D}^k \mathbf{v}_i^k + c_i^k - \mathbf{B}_i \mathbf{u} - \frac{h_i}{\beta} \right\|_2^2 \right) + \frac{\lambda_2}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_2^2 \right) \\
= \frac{\partial}{\partial u} \left(\beta \sum_i \mathbf{B}_i^T \left(\mathbf{B}_i \mathbf{u}^k - \mathbf{D}^{k+1} \mathbf{v}_i^{k+1} - c_i^{k+1} + \left(\frac{h_i^k}{\beta} \right) \right) \\
+ \lambda_2 \mathbf{R}^T \left(\mathbf{R}\mathbf{u}^k - \mathbf{f} \right) \right) \\
= \beta \sum_i \mathbf{B}_i^T \mathbf{B}_i + \lambda_2 \mathbf{R}^T \mathbf{R}.$$
(31)



FIGURE 1: Interpolated results: (a) original seismic data; (b) seismic data with 50% with missing traces; (c) interpolation by K-SVD method; (d) interpolation by orthogonal DL method; and (e) interpolation result by the proposed method.

Mathematical Problems in Engineering



FIGURE 2: (a) Original seismic data and (b) seismic data with 50% missing trace. Interpolation results using by (c) K-SVD method, (d) orthogonal DL method, and (e) the proposed method.

Mathematical Problems in Engineering



FIGURE 3: Difference sections between (a) Figures 2(a) and 2(c), (b) Figures 2(a) and 2(d), and (c) Figures 2(a) and 2(e).

TABLE 1: SNR values of three different methods applied to synthetic data in Figure 2 with different sampling ratio.

Sampling ratio	0.3	0.4	0.5	0.6	0.7	0.8
K-SVD method	23.9923	25.6434	27.8589	29.3336	30.6696	31,1826
Orthogonal DL method	31.7795	32.9216	33.4244	36.3080	38.4702	41.3320
Proposed method	34.6241	36.3963	38.5860	39.9641	42.9120	43.5903

Hence,

Therefore,

$$\frac{\partial^2}{\partial u^2} \left(\sum_i \left(\left\| \mathbf{D}^k \mathbf{v}_i^k + c_i^k - \mathbf{B}_i \mathbf{u} - \frac{h_i}{\beta} \right\|_2^2 \right) + \frac{\lambda_2}{2} \|\mathbf{R}\mathbf{u} - \mathbf{f}\|_2^2 \right)$$
$$= \beta \sum_i \mathbf{B}_i^{\mathrm{T}} \mathbf{B}_i + \lambda_2 \mathbf{R}^{\mathrm{T}} \mathbf{R}.$$
(32)

$$\mathbf{u}^{k+1} = \mathbf{u}^{k} - \frac{\beta \sum_{i} \mathbf{B}_{i}^{T} \left(\mathbf{B}_{i} \mathbf{u}^{k} - \mathbf{D}^{k+1} \mathbf{v}_{i}^{k+1} - c_{i}^{k+1} + \left(\frac{h_{i}^{k}}{\beta}\right) \right) + \lambda_{2} \mathbf{R}^{T} \left(\mathbf{R} \mathbf{u}^{k} - \mathbf{f} \right)}{\beta \sum_{i} \mathbf{B}_{i}^{T} \mathbf{B}_{i} + \lambda_{2} \mathbf{R}^{T} \mathbf{R}}.$$
(33)



FIGURE 4: Single trace comparison of the reconstructed synthetic data: (a) trace from K-SVD result, (b) trace from orthogonal DL, and (c) trace from the proposed method.

Finally, we formulate the following algorithm for the proposed method.

4. Numerical Experiments

The numerical examples and the graphs used in this paper were simulated by MATLAB and the machine we used to run the code is DESKTOP-HKHFG0P: Processor: Intel(R) Core(TM) i7-7500U CPU @ 2.70 GHz 2.90 GHz and RAM: 8.00 GB.

We test the validity of the proposed method to noise free and noisy synthetic and real seismic datasets. The proposed method tested on both synthetic and real datasets by the different percentages of missing traces based on Jittered undersampled strategy which can help to obtain random properties as well as control the maximum gap between adjacent traces to meet the requirements of the compressive sensing theory. Throughout the numerical experiments, the patch size was 8×8 and the sparsity level was 4. The algorithm in Nazzal et al.'s [59] study is used to obtain the sparsity level. The regularization parameters are tuned based on the convergence condition. Accordingly, the choice of the parameters $\lambda_2 = 1$ and $\lambda_1 > 0$ is trivial. We empirically search for the optimal value of parameter λ_1 from the smallest set {0.0005, 0.0001, 0.005, 0.01} and search for the parameter β from the range of [0.001, 0.2]. To evaluate the quality of restored seismic data quantitatively,



FIGURE 5: Single trace comparison of the reconstructed synthetic data in Figure 3 with the original trace from Figure 3(a). Trace from (a) K-SVD result, (b) orthogonal DL result, and (c) the proposed method.

we introduced the signal-to-noise ratio (SNR) which is defined as follows:

$$SNR = 10\log_{10}\left(\frac{\|\mathbf{u}\|_{2}^{2}}{\|\mathbf{u} - \mathbf{v}\|_{2}^{2}}\right), \tag{34}$$

where **u** denotes the clean seismic data and **v** denotes the reconstructed seismic data. We simulate seismic data using synthetic datasets containing four layers with linear and curved events and synthetic datasets with six layers containing parabolic events. The 50% of missing traces is used based on the random sampling property as shown in Figures 1

and 2. We apply the proposed method to 2D synthetic data reconstruction with linear and curved events as shown in Figures 1 and 2, respectively. The result of the proposed method is compared with the results by K-SVD and orthogonal DL. In Figure 1, both K-SVD and orthogonal DL provide promising results in the reconstruction of missing traces when the seismic data are with the linear events. However, the reconstruction results by the K-SVD and orthogonal DL are not satisfactory when the seismic data are with the nonlinear events because of the elimination of some important features of the reconstructed data around the boundary, as shown in Figures 2(c) and 2(d). The effectiveness of the proposed method in the reconstruction of



FIGURE 6: (a) Original seismic data with strong noise. (b) Real seismic data with strong noise and 50% missing trace. Simultaneous denoising and interpolation of seismic data by using (c) K-SVD method, (d) orthogonal DL algorithm, and (e) the proposed method.



FIGURE 7: Difference sections between (a) Figures 6(b) and (c), (b) Figures 6(b) and (d), and (c) Figures 6(b) and (e).

missing seismic traces is presented in Figure 1(e). Furthermore, the residual which is defined as the difference between the interpolated and original seismic data is presented in Figure 3 to show the performance of K-SVD, orthogonal DL, and proposed methods in the reconstruction of missing traces. For the results by K-SVD and orthogonal DL, there are some important features of seismic signals left in the residual sections, as observed in Figures 3(a) and 3(b), respectively. There is no important information about seismic signals left in the residual sections as indicated in Figure 3(c). The minor residual indicates that the important features of the seismic signal are well preserved, and better interpolated seismic data can also be achieved. For further comparison, the SNR of the proposed method, K-SVD, and orthogonal DL are given in Table 1. The proposed method clearly shows much better signal preservation and SNR enhancement while reconstructing the missing seismic traces. For additional comparison and to see the detail of the interpolated seismic data by the three methods, we plot a single seismic trace from the interpolated data in Figures 4 and 5.



FIGURE 8: (a) Original seismic data. (b) Seismic data with missing traces. Reconstructed data using (c) K-SVD, (d) orthogonal DL, and (e) the proposed method.



FIGURE 9: Reconstructed single trace comparison by (a) K-SVD, (b) orthogonal DL, and (c) proposed method of the interpolated results in Figure 8.

In the next example, we focused on the real seismic data to verify the performance of the proposed method in simultaneous interpolation and denoising and compare it with the K-SVD and orthogonal DL methods. In this case, we randomly remove 50% seismic traces and apply the same method with synthetic data. The real data with 50% missing trace and covered by strong real noise in which detecting the useful seismic signals are hard, and hard to see some events of seismic data are presented in Figure 6(a). The interpolated and denoised results are shown in Figure 6(c)–6(e). The proposed method shows promising performance in removing strong random noise and interpolating the missing seismic traces. Because of the special information contained in the tensor DL, the proposed method achieves better results in the interpolation and denoising of seismic data in terms of visual quality and preservation of seismic features. As shown in Figures 6(c) and 6(d), some important parts of the interpolated and denoised data are deteriorated. The noise sections from the original data (with strong noise) and the recovered data are presented in Figure 7. As presented in Figures 7(a) and 7(b), there are some useful seismic signals left in the noise sections. As we observe from Figure 7(c), there are no parts of seismic signals left in the noise sections.

The effectiveness of the proposed method is also verified by interpolating aliasing seismic data containing different features. The original seismic data, observed data with 50% missing traces, and the interpolated results are presented in Figure 8. The reconstructed results by K-SVD and orthogonal DL are unsatisfactory because of the insertion of artifacts on some parts of the interpolated results as shown in



FIGURE 10: (a) Original seismic data. (b) Noisy data with SNR = 5.39. The result of random noise attenuation of synthetic seismic data by (c) K-SVD (SNR = 22.27 dB), (d) orthogonal DL (SNR = 23.51 dB), and (e) the proposed method (SNR = 24.36 dB).



FIGURE 11: The noise sections from the denoised results by (a) K-SVD, (b) orthogonal DL, and (c) the proposed method corresponding to the one in Figure 10.

Figures 8(c) and 8(d), respectively. The performance of the proposed method in the reconstruction of missing seismic traces is presented in Figure 8(e). The single trace is extracted from the reconstructed data by K-SVD, orthogonal DL, and the proposed method as presented in Figure 9 to compare the details of the obtained results.

For the next example, we discussed the validity of the proposed method for removing random noise and compared it with K-SVD and orthogonal DL. Both synthetic and real seismic datasets are considered for random noise attenuation. Both synthetic and real seismic datasets are contaminated by band-limited Gaussian noise which is spatially uncorrelated random noise. The synthetic seismic data with random noise are presented in Figure 10. The denoised results by K-SVD and orthogonal DL are not satisfactory because of the elimination of some primary features of seismic data around the boundary and they introduce some artifacts as presented in Figures 10(c) and 10(d). Since K-SVD uses the redundancy of the seismic features over the datasets to attenuate random noise, it is computationally more expensive than orthogonal DL and proposed methods. Moreover, the denoising performance of K-SVD, orthogonal DL, and proposed methods are assessed by using the residuals obtained by subtraction of the denoised results from the noisy seismic data. In Figures 11(a) and 11(b), K-SVD and orthogonal DL residuals show some removed important seismic signals. In Figure 11(c), we observe that almost there were no important seismic signals left in the residual part. For further comparison, a single seismic trace extracted from the denoised results in Figure 10 is displayed in Figure 12, and the comparison of SNR is presented in Table 2 with the different noise levels.

Finally, we test the denoising of 2D real data with bandlimited Gaussian noise. The denoised result by the proposed method is shown in Figure 13(e), and the result is clear and the primary features of the data are well preserved. The other two methods cause some damage to the useful seismic signals and some part of the seismic data left in the residual parts as



FIGURE 12: Single seismic trace comparison of the denoised results in Figure 10 and the original trace in Figure 10(a). Seismic trace from (a) original data, (b) noisy data, (c) K-SVD result, (d) orthogonal DL, and (e) the proposed method.

	TABLE 2: SNR values of three different methods applied to real seismic dataset in Figure 10.								
	Input SNR	5.39	6.65	8.1	9.83	11.9			
Real data	K-SVD method	22.27	23.13	24.04	26.77	28.76			
	Orthogonal DL method	23.51	25.18	26.86	27.74	28.76			
	Proposed method	24.36	26.17	28.40	30.47	32.09			



FIGURE 13: Continued.



FIGURE 13: (a) Real seismic data. (b) Noisy real seismic data (SNR = 2.48). Denoising results by (c) K-SVD (SNR = 16.58), (d) orthogonal DL algorithm (SNR = 17.55), and (e) the proposed method (SNR = 19.13).



FIGURE 14: The noise sections from the results by (a) K-SVD method, (b) orthogonal DL algorithm, and (c) the proposed method corresponding to the result in Figure 13.



FIGURE 15: Trace comparison between original seismic data in Figure 13(a) and denoised results of Figure 13: (a) original trace, (b) noisy trace, trace from the denoised result by (c) K-SVD, (d) orthogonal DL algorithm, and (e) the proposed method.

shown in Figures 14(a) and 14(b). In Table 2, the SNR values of the denoised seismic data for different noise levels are presented. Furthermore, a single seismic trace comparison is displayed in Figure 15 to demonstrate the performance and validity of the proposed method.

5. Conclusion

In this paper, we exploited sparse regularization which promotes sparsity for both seismic data interpolation and denoising. We have proposed a novel method to interpolate and attenuate random noise of seismic data based on orthogonal tensor DL. Due to the separability, orthogonality, and imposed tensor decomposition, the proposed method is computationally efficient and fast for learning the dictionary. The effectiveness of the proposed method is compared with both K-SVD and orthogonal DL methods in denoising and interpolation of seismic data. The numerical experiments demonstrate that the proposed method shows promising results in the denoising and interpolation of seismic data. Our approach uses the dictionary which can extract the local features and adapt to seismic data, with no introduced artifacts on the interpolated and denoised results. Compared with K-SVD and orthogonal DL methods, the proposed method is computationally effective. The experimental results and the performance of the methods in seismic data interpolation and denoising show the effective performance of the proposed method on synthetic and real seismic datasets.

Appendix

Theorem A.1. Let $\{\mathbf{D}_i: \mathbf{D}_i \in S_{\mathbf{M}_i}, i = 1, 2, 3, ..., N\}$ be the set of orthogonal matrices and for the given $\mathbf{Y}, \mathbf{V} \in \mathbb{R}^{\mathbf{M}_1 \times \mathbf{M}_2 \times \mathbf{M}_3 ... \times \mathbf{M}_N \times \mathbf{R}}$, the minimization problem:

$$\underset{\mathbf{E}\in S_{M_{i}}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{V} \times_{1} \mathbf{D}_{1} \times_{2} \mathbf{D}_{2} \dots \times_{i-1} \mathbf{D}_{i-1} \times_{i} \mathbf{E} \times_{i+1} \mathbf{D}_{i+1} \dots \times_{N} \mathbf{D}_{N}\|_{\mathrm{F}}^{2},$$
(A.1)

has a unique solution which is given by the following equation:

$$\mathbf{E} = \mathbf{W}\mathbf{X}^{\mathrm{T}},\tag{A.2}$$

with orthogonal matrices W and X such that

$$\mathbf{W} \mathbf{\Sigma} \mathbf{X}^{\mathrm{T}} = [\mathbf{Y} \mathbf{x}_{N} \mathbf{D}_{N}^{\mathrm{T}} \mathbf{x}_{N-1} \mathbf{D}_{N-1}^{\mathrm{T}} \dots \mathbf{x}_{i+1} \mathbf{D}_{i+1}^{\mathrm{T}}]_{i} [\mathbf{V} \mathbf{x}_{1} \mathbf{D}_{1} \mathbf{x}_{21} \mathbf{D}_{2} \dots \mathbf{x}_{i-1} \mathbf{D}_{i-1}]_{i}^{\mathrm{T}},$$
(A.3)

where $\mathbf{W} \Sigma \mathbf{X}^{\mathrm{T}}$ is the SVD of **E**.

Proof. From Theorem 2.1 in Chrétien and Wei's [60] study, every tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \dots \times n_R}$ can be written as follows:

$$\mathbf{X} = \mathbf{V}(\mathbf{X}) \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \dots \times_R \mathbf{D}_R, \qquad (A.4)$$

where each $\mathbf{D}_{R} \in \mathbb{R}^{n_{r} \times n_{r}}$ is an orthogonal matrix and $\mathbf{V}(\mathbf{X}) \in \mathbb{R}^{n_{1} \times n_{2} \times n_{3} \dots \times n_{r}}$ is tensor of the same size with **X**. Let \otimes be the standard Kronecker product for matrices, then:

$$\mathbf{X}_{(r)} = \mathbf{D}_{r} \mathbf{V}_{(r)}(\mathbf{X}) (\mathbf{D}_{r+1} \otimes \mathbf{D}_{r+2} \otimes \dots \otimes \mathbf{D}_{R} \otimes \mathbf{D}_{1} \dots \otimes \mathbf{D}_{r-1})^{\mathrm{T}},$$
(A.5)

where $\mathbf{V}_{(r)}(\mathbf{X})$ is the mode r matricization of $\mathbf{V}(\mathbf{X})$. For convenience we can use U, V, and A to represent $\mathbf{X}_{(r)}$, $\mathbf{D}_r \mathbf{V}_{(r)}(\mathbf{X})$, and $(\mathbf{D}_{r+1} \otimes \mathbf{D}_{r+2} \otimes ... \otimes \mathbf{D}_R \otimes \mathbf{D}_1 ... \otimes \mathbf{D}_{r-1})$, respectively. Then, by using the *r* mode unfolding and length preservation property of orthogonal transform, we can reformulate the minimization problem of this theorem as follows:

$$\underset{\mathbf{E}\in S_{M_i}}{\operatorname{argmin}} \|\mathbf{U} - \mathbf{V}\mathbf{A}^{\mathrm{T}}\|_{\mathrm{F}}^2, \tag{A.6}$$

such that $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$. We can write the objective function in Equation (A.6) as follows:

$$\|\mathbf{U} - \mathbf{V}\mathbf{E}^{\mathrm{T}}\|_{\mathrm{F}}^{2} = \mathrm{Tr}((\mathbf{U} - \mathbf{V}\mathbf{E}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{U} - \mathbf{V}\mathbf{E}^{\mathrm{T}})), \qquad (A.7)$$

$$= \operatorname{Tr}((\mathbf{U}^{\mathrm{T}} - \mathbf{E}\mathbf{V}^{\mathrm{T}})(\mathbf{U} - \mathbf{V}\mathbf{E}^{\mathrm{T}})), \qquad (A.8)$$

$$= Tr(\mathbf{U}^{\mathrm{T}}\mathbf{U} - \mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}} - \mathbf{E}\mathbf{V}^{\mathrm{T}}\mathbf{U} + \mathbf{E}\mathbf{V}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}}), \qquad (A.9)$$

$$= \operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{U}) + \operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{V}) - 2\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}}), \quad (A.10)$$

$$= \|\mathbf{U}\|_{\mathrm{F}}^{2} + \|\mathbf{V}\|_{\mathrm{F}}^{2} - 2\mathrm{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}}), \qquad (A.11)$$

$$\|\mathbf{U} - \mathbf{V}\mathbf{E}^{\mathrm{T}}\|_{\mathrm{F}}^{2} = \|\mathbf{U}\|_{\mathrm{F}}^{2} + \|\mathbf{V}\|_{\mathrm{F}}^{2} - 2\mathrm{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}}),$$
 (A.12)

such that $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$. Since the first two terms in Equation (A.12) are constant, the minimization problem in Equation (A.12) becomes:

$$\max_{\mathbf{E}} 2\mathrm{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{V}\mathbf{E}^{\mathrm{T}}), \qquad (A.13)$$

such that $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$. By considering the SVD of $\mathbf{U}^{\mathrm{T}}\mathbf{V} = \mathbf{P}\Sigma\mathbf{Q}^{\mathrm{T}}$, we can rewrite Equation (A.13) as follows:

$$\max_{\mathbf{E}} \operatorname{Tr}(\mathbf{P}\Sigma\mathbf{Q}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}), \mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}, \qquad (A.14)$$

$$= \max_{\mathbf{F}^*} \operatorname{Tr}(\mathbf{P}\Sigma(\mathbf{F}^*)^{\mathrm{T}}), \qquad (A.15)$$

where $F^* = EQ$.

$$= \max_{\mathbf{r}^*} \operatorname{Tr}((\mathbf{F}^*)^{\mathrm{T}} \mathbf{P} \Sigma), \qquad (A.16)$$

such that $(\mathbf{F}^*)^T = \mathbf{I}$, since \mathbf{Q} is orthogonal. Since Σ is diagonal matrix, then the maximization problem in Equation (A.16) is achieved when the diagonal of $(\mathbf{F}^*)^T \mathbf{P}$ is positive and

maximum. By Cauchy–Schwartz inequality, this is true when $\mathbf{F}^* = \mathbf{P}$ such that the diagonal elements are all 1. Therefore, $\mathbf{E} = \mathbf{P}\mathbf{Q}^{\mathrm{T}}$ is the explicit solution for the minimization problem in Equation (A.6).

Based on this theorem, we can solve for the dictionaries D_1 and D_2 in the same manner. Accordingly, for the minimization problem in Equation (14), the dictionary D_1 can be solved by fixing the dictionary D_2 and which is given by the following equation:

$$\mathbf{D}_1^{(k+1)} = \mathbf{W}_1 \mathbf{X}_1^{\mathrm{T}},\tag{A.17}$$

with the SVD

$$\mathbf{W}_{1} \Sigma \mathbf{X}_{1}^{\mathrm{T}} = \left(\left(\mathbf{D}_{2}^{\mathrm{k}} \right)^{T} \mathbf{Y} \right) \left(\mathbf{V} \right)_{(1)}^{\mathrm{T}}.$$
 (A.18)

Similarly, for a fixed \mathbf{D}_1 and \mathbf{D}_2 is given by the following equation:

$$\mathbf{D}_2^{(k+1)} = \mathbf{W}_2 \mathbf{X}_2^{\mathrm{T}},\tag{A.19}$$

with the SVD

$$\mathbf{W}_{2}\boldsymbol{\Sigma}\mathbf{X}_{2}^{\mathrm{T}} = \mathbf{Y}\left(\mathbf{V}_{(1)}\mathbf{D}_{1}^{(k)}\right)^{\mathrm{T}}.$$
 (A.20)

Data Availability

Every data used for this paper is free and can be accessed upon the request through the corresponding author.

Conflicts of Interest

The author declares that there is no conflict of interests about the publication of this paper.

Acknowledgments

This research was funded by Mattu University Research (Community Service, Technology Transfer, and Collaboration Office).

References

- Y. Chen, Y. Zhang, and L.-W. Mo, "Seismic data interpolation using nonlinear shaping regularization," *Journal of Seismic Exploration*, vol. 24, no. 5, pp. 327–342, 2015.
- [2] L. Zhu, E. Liu, and J. H. McClellan, "Joint seismic data denoising and interpolation with double-sparsity dictionary learning," *Journal of Geophysics and Engineering*, vol. 14, no. 4, pp. 802–810, 2017.
- [3] V. Oropeza and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011.

- [4] G. Tang and J. Ma, "Application of total-variation-based curvelet shrinkage for three-dimensional seismic data denoising," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 103–107, 2011.
- [5] B. Wang, X. Chen, J. Li, and J. Cao, "An improved weighted projection onto convex sets method for seismic data interpolation and denoising," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, pp. 228–235, 2016.
- [6] M. A. Nazari Siahsar, S. Gholtashi, V. Abolghasemi, and Y. Chen, "Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning," *Signal Processing*, vol. 141, pp. 309–321, 2017.
- [7] M. Naghizadeh, "Seismic data interpolation and denoising in the frequency–wavenumber domain," *Geophysics*, vol. 77, no. 2, pp. V71–V80, 2012.
- [8] S. Crawley, R. Clapp, and J. Claerbout, "Interpolation with smoothly nonstationary prediction-error filters," in 69th Annual International Meeting, pp. 1154–1157, SEG, Expanded Abstracts, 1999.
- [9] Y. Liu and S. Fomel, "Seismic data interpolation beyond aliasing using regularized nonstationary autoregression," *Geophysics*, vol. 76, no. 5, pp. V69–V77, 2011.
- [10] M. J. Porsani, "Seismic trace interpolation using half-step prediction filters," *Geophysics*, vol. 64, no. 5, pp. 1461–1467, 1999.
- [11] S. Spitz, "Seismic trace interpolation in the F-X domain," *Geophysics*, vol. 56, no. 6, pp. 785–794, 1991.
- [12] N. Gülünay, "Seismic trace interpolation in the Fourier transform domain," *Geophysics*, vol. 68, no. 1, pp. 355–369, 2003.
- [13] S. Trickett, L. Burroughs, A. Milton, L. Walton, and R. Dack, "Rank reduction- based trace interpolation," in SEG Technical Program Expanded Abstracts 2010, pp. 3829–3833, SEG Library, 2010.
- [14] Z. Lu, N. Xia, L. Sun et al., "An alternative adaptive method for seismic data denoising and interpolation," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6295902, 16 pages, 2020.
- [15] J. Ma, "Three-dimensional irregular seismic data reconstruction via low-rank matrix completion," *Geophysics*, vol. 78, no. 5, pp. V181–V192, 2013.
- [16] Y. Yang, J. Ma, and S. Osher, "Seismic data reconstruction via matrix completion," *Inverse Problems & Imaging*, vol. 7, no. 4, pp. 1379–1392, 2013.
- [17] N. Kreimer and M. D. Sacchi, "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation," *Geophysics*, vol. 77, no. 3, pp. V113–V122, 2012.
- [18] J. Wang, J. Ma, B. Han, and Y. Cheng, "Seismic data reconstruction via weighted nuclear-norm minimization," *Inverse Problems in Science and Engineering*, vol. 23, no. 2, pp. 277–291, 2015.
- [19] S. Fomel, "Application of plane-wave destruction filters," *Geophysics*, vol. 67, no. 6, pp. 1946–1960, 2002.
- [20] L. Liu, G. Plonka, and J. Ma, "Seismic data interpolation and denoising by learning a tensor tight frame," *Inverse Problems*, vol. 33, no. 10, Article ID 105011, 2017.
- [21] Z. Feng, "Seismic random noise attenuation using effective and efficient dictionary learning," *Journal of Applied Geophysics*, vol. 186, Article ID 104258, 2021.
- [22] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse representation: a novel approach," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1195–1198, 2013.
- [23] X. Luo, T. Bhakta, M. Jakobsen, and G. Nævdal, "An ensemble 4Dseismic history-matching framework with sparse representation

based on wavelet multiresolution analysis," SPE Journal, vol. 22, no. 3, 2017.

- [24] R. Zhang and T. J. Ulrych, "Physical wavelet frame denoising," *Geophysics*, vol. 68, no. 1, pp. 225–231, 2003.
- [25] M. M. N. Kabir and D. J. Verschuur, "Restoration of missing offsets by parabolic Radon transform," *Geophysical Prospecting*, vol. 43, no. 3, pp. 347–368, 1995.
- [26] J. Wang, M. Ng, and M. Perz, "Seismic data interpolation by greedy local Radon transform," *Geophysics*, vol. 75, no. 6, pp. WB225–WB234, 2010.
- [27] M. Naghizadeh and K. A. Innanen, "Seismic data interpolation using a fast-generalized Fourier transform," *Geophysics*, vol. 76, no. 1, pp. V1–V10, 2011.
- [28] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 31–38, 1998.
- [29] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, no. 1, pp. 233–248, 2008.
- [30] R. Shahidi, G. Tang, J. Ma, and F. J. Herrmann, "Application of randomized sampling schemes to curvelet-based sparsitypromoting seismic data recovery," *Geophysical Prospecting*, vol. 61, no. 5, pp. 973–997, 2013.
- [31] S. Fomel and Y. Liu, "Seislet transform and seislet frame," *Geophysics*, vol. 75, no. 3, pp. V25–V38, 2010.
- [32] S. Gan, S. Wang, Y. Chen, X. Chen, W. Huang, and H. Chen, "Compressive sensing for seismic data reconstruction via fast projection onto convex sets based on seislet transform," *Journal of Applied Geophysics*, vol. 130, pp. 194–208, 2016.
- [33] S. Hauser and J. Ma, "Seismic data reconstruction via shearletregularized directional inpainting preprint," 2012.
- [34] Y. Yang, Q. Ran, K. Chen et al., "Denoising seismic data via a threshold shrink method in the non-subsampled contourlet transform domain," *Mathematical Problems in Engineering*, vol. 2022, Article ID 1013623, 12 pages, 2022.
- [35] Y. Zhou, J. Yang, H. Wang, G. Huang, and Y. Chen, "Statisticsguided dictionary learning for automatic coherent noise suppression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.
- [36] I. Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science*, pp. 1094–1096, Springer, Berlin, 2011.
- [37] R. Vidal, Yi Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1945–1959, 2005.
- [38] S. Bechouche and J. Ma, "Simultaneous dictionary learning and denoising for seismic data," *Geophysics*, vol. 79, no. 3, pp. A27–A31, 2014.
- [39] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 2443–2446, 1999.
- [40] M. Aharon, M. Elad, and A. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [41] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736– 3745, 2006.
- [42] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral

images," SIAM Journal on Imaging Sciences, vol. 5, no. 1, pp. 33-56, 2012.

- [43] R. V. Soares, X. Luo, G. Evensen, and T. Bhakta, "4D seismic history matching: assessing the use of a dictionary learning based sparse representation method," *Journal of Petroleum Science and Engineering*, vol. 195, Article ID 107763, 2020.
- [44] Y. Quan, Y. Huang, and H. Ji, "Dynamic texture recognition via orthogonal tensor dictionary learning," in 2015 IEEE International Conference on Computer Vision (ICCV), pp. 73– 81, IEEE, Santiago, Chile, December 2015.
- [45] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.
- [46] J. Liang, J. Ma, and X. Zhang, "Seismic data restoration via data-driven tight frame," *Geophysics*, vol. 79, no. 3, pp. V65– V74, 2014.
- [47] S. Yu, J. Ma, X. Zhang, and M. D. Sacchi, "Denoising and interpolation of high-dimensional seismic data by learning tight frame," *Geophysics*, vol. 80, no. 5, pp. V119–V132, 2015.
- [48] Y. Chen, J. Ma, and S. Fomel, "Double sparsity dictionary for seismic noise attenuation," *Geophysics*, vol. 81, no. 2, pp. V103– V116, 2016.
- [49] R. Costantini, L. Sbaiz, and S. Susstrunk, "Higher order SVD analysis for dynamic texture synthesis," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 42–52, 2008.
- [50] T. Hazan, S. Polak, and A. Shashua, "Sparse image coding using a 3D non-negative tensor factorization," *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 1, pp. 50–57, 2005.
- [51] S. Tan, Y. Zhang, G. Wang et al., "Tensor-based dictionary learning for dynamic tomographic reconstruction," *Physics in Medicine & Biology*, vol. 60, no. 7, Article ID 2803, 2015.
- [52] C. Bao, J.-F. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," in 2013 IEEE International Conference on Computer Vision, pp. 3384–3391, IEEE, Sydney, NSW, Australia, December 2013.
- [53] C. F. Caiafa and A. Cichocki, "Multidimensional compressed sensing and their applications," WIREs Data Mining and Knowledge Discovery, vol. 3, no. 6, pp. 355–380, 2013.
- [54] G. Duan, H. Wang, Z. Liu, J. Deng, and Y.-W. Chen, "K-CPD: learning of overcomplete dictionaries for tensor sparse coding," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 493–496, IEEE, Tsukuba, Japan, November 2012.
- [55] S. Zubair and W. Wang, "Tensor dictionary learning with sparse TUCKER decomposition," in 2013 18th International Conference on Digital Signal Processing (DSP), pp. 1–6, IEEE, Fira, July 2013.
- [56] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Society for Industrial and Applied Mathematics*, vol. 51, no. 3, pp. 455–500, 2009.
- [57] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends[®] in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [58] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.
- [59] M. Nazzal, M. A. Aygül, and H. Arslan, "Estimating sparsity level for enabling compressive sensing of wireless channels and spectra in 5G and beyond," Arxiv, 2020.
- [60] S. Chrétien and T. Wei, "Von neumann's trace inequality for tensors," *Linear Algebra and Its Applications*, vol. 482, pp. 149–157, 2015.