Seismic sparse-spike deconvolution via Toeplitz-sparse matrix factorization

Lingling Wang[1], Qian Zhao[2], Jinghuai Gao[3], Zongben Xu[2], Michael Fehler[4], and Xiudi Jiang[5]

ABSTRACT

We have developed a new sparse-spike deconvolution (SSD) method based on Toeplitz-sparse matrix factorization (TSMF), a bilinear decomposition of a matrix into the product of a Toeplitz matrix and a sparse matrix, to address the problems of lateral continuity, effects of noise, and wavelet estimation error in SSD. Assuming the convolution model, a constant source wavelet, and the sparse reflectivity, a seismic profile can be considered as a matrix that is the product of a Toeplitz wavelet matrix and a sparse reflectivity matrix. Thus, we have developed an algorithm of TSMF to simultaneously deconvolve the seismic matrix into a wavelet matrix and a reflectivity matrix by alternatively solving two inversion subproblems related to the Toeplitz wavelet matrix and sparse reflectivity matrix, respectively. Because the seismic wavelet is usually compact and smooth, the fused Lasso was used to constrain the elements in the Toeplitz wavelet

matrix. Moreover, due to the limitations of computer memory, large seismic data sets were divided into blocks, and the average of the source wavelets deconvolved from these blocks via TSMF-based SSD was used as the final estimation of the source wavelet for all blocks to deconvolve the reflectivity; thus, the lateral continuity of the seismic data can be maintained. The advantages of the proposed deconvolution method include using multiple traces to reduce the effect of random noise, tolerance to errors in the initial wavelet estimation, and the ability to preserve the complex structure of the seismic data without using any lateral constraints. Our tests on the synthetic seismic data from the Marmousi2 model and a section of field seismic data demonstrate that the proposed method can effectively derive the wavelet and reflectivity simultaneously from band-limited data with appropriate lateral coherence, even when the seismic data are contaminated by noise and the initial wavelet estimation is inaccurate.

INTRODUCTION

Reflection seismology, a widely used method in geophysical exploration, is based on the general principle of sending seismic source wavelets (using an energy source, such as dynamite explosion or vibroseis) into the earth, recording the reflected waves at the earth's surface, and then using the recorded data for estimating the properties of the earth's subsurface. According to the convolutional model, a seismogram is the sum of the convolution of the seismic source wavelet with the subsurface reflectivity coefficients and added noise. Let $w(t)$ be the source wavelet, $r(t)$ denote the reflec-

tivity coefficients of the subsurface, and $n(t)$ be the noise; then, the seismogram trace $y(t)$ can be given by

$$y(t) = w(t) * r(t) + n(t), \qquad (1)$$

where * means convolution. From equation 1, we can see that the recorded seismic trace always bears the source wavelet, which smears adjacent events and reduces the resolution of the seismic image. Seismic deconvolution is an inverse problem for removing the source wavelet from the recorded seismic trace. In the ideal case, after deconvolution, the true seismic reflectivity is recovered. In

reality, because the source wavelet is always band-limited, the inverse problem is ill-posed, and requires regularization to achieve stable results. Because the bigger seismic reflectivity coefficients are the main contributors of seismic acoustic impedance, and they are usually sparse in time, sparsity is usually taken as a regularization constraint for reflectivity inversion. Using this constraint leads to a methodology called the sparse-spike deconvolution (SSD) method (Latimer et al., 2000; Nguyen, 2008).

SSD is a commonly used deconvolution method for acoustic-impedance inversion (Latimer et al., 2000). It assumes, as a prior, that the reflectivity is a sparse sequence of spikes. The main objective of SSD is to provide a significant increase in the bandwidth content from band-limited seismic observations, so that the result has high resolution and is suitable for acoustic-impedance inversion. There are many methods for performing SSD. Some use different search strategies to locate the spikes and rely on the optimization of different cost functions to satisfy a probabilistic model for the reflectivity (Kormylo and Mendel, 1978; Kaaresen and Taxt, 1998); some apply stochastic methods in a nonlinear optimization framework to find the locations and amplitudes of the least number of spikes (Velis, 2008); others proceed to optimize some norm that forces the solution to be sparse (Oldenburg et al., 1983; Debeye and van Riel, 1990; Sacchi et al., 1994; Wang, 2011). However, these methods face challenges when being applied to multidimensional data in a trace-by-trace basis: they may show bad lateral continuity and the image quality may be compromised in the presence of noise and wavelet estimation error. Although some scholars have proposed methods to reduce these problems (Wang et al., 2006; Nguyen, 2008; Lu, 2009; Kazemi and Sacchi, 2014; Guitton and Claerbout, 2015; Nose-Filho et al., 2016), the challenges remain for seismic data with complex structure.

We present a new seismic SSD method based on Toeplitz-sparsity matrix factorization (TSMF). Matrix factorization has been extensively studied in many research fields, such as numerical analysis (Golub and Van Loan, 2012), machine learning (Srebro and Jaakkola, 2003; Salakhutdinov and Mnih, 2008), computer vision (Okatani and Deguchi, 2007; Eriksson and van den Hengel, 2010), and signal processing (Hyvärinen and Oja, 2000; Cichocki et al., 2009). Various matrix models have been proposed that use different constraints. For example, by enforcing orthogonality constraints, we get the classical singular value decomposition (Golub and Van Loan, 2012); non-negative constraints lead to non-negative matrix factorization (Lee and Seung, 1999); and independent component analysis (Hyvärinen and Oja, 2000) can be formulated with statistical independence. These models have already been successfully applied to numerous problems, e.g., collaborative filtering (Lim and Teh, 2007; Porteous et al., 2010), distinguishing structure from motion (Tomasi and Kanade, 1992; Bregler et al., 2000), and blind source separation (Hyvärinen and Oja, 2000; Cichocki et al., 2009). For seismic data, the matrix form of a seismic trace modeled in equation 1 is

$$\mathbf{y} = \mathbf{A}\mathbf{r} + \mathbf{n}, \tag{2}$$

where $\mathbf{A}$ is a Toeplitz matrix with element $\mathbf{A}_{ij} = \mathbf{w}_{i-j+1}$, where $\mathbf{w}_{i-j+1}$ is the $i - j + 1$th sampling point of the wavelet $w(t)$. Here, $\mathbf{y}$, $\mathbf{r}$, and $\mathbf{n}$ are column vectors obtained by sampling $y(t)$, $r(t)$, and $n(t)$, respectively. For a 2D seismic profile, equation 2 becomes

$$\mathbf{Y} = \mathbf{A}\mathbf{R} + \mathbf{N}, \tag{3}$$

where the columns of $\mathbf{Y}$ are the successive traces of the seismic data, $\mathbf{R}$ and $\mathbf{N}$ are the corresponding reflectivity series and noise matrices, respectively. In practice, it is not easy to know the wavelet exactly; nevertheless, we could take advantage of the Toeplitz structure of $\mathbf{A}$ and the sparsity property of $\mathbf{R}$ to invert for them simultaneously. We note that a dictionary learning model (Aharon et al., 2006; Elad and Aharon, 2006) enforces sparsity constraints to $\mathbf{R}$ but does not constrain $\mathbf{A}$, whereas Kibangou and Favier (2007) consider $\mathbf{A}$ to have a Toeplitz structure but a nonsparse structure for $\mathbf{R}$, so their method cannot be directly applied to the seismic problem. We propose the TSMF method to combine the advantages of the above methods to invert for $\mathbf{A}$ and $\mathbf{R}$ simultaneously. This method is based on a matrix-factorization method that uses multiple traces to estimate the wavelet and reflectivity simultaneously; it thus obtains good lateral continuity and preserves structure well when applied to seismic data.

This paper is organized as follows: First, we present the formulation of the TSMF and derive an algorithm for achieving it. Next, we develop the TSMF-based SSD method. Finally, synthetic and field data examples are used to demonstrate the effectiveness of the proposed deconvolution method before drawing conclusions.

TOEPLITZ-SPARSITY MATRIX FACTORIZATION

In this section, we present our formulation of TSMF, and give a simple but effective algorithm to solve it.

Model formulation

Let $\mathbf{Y}$ be an $n \times m$ matrix factorizable as a product of an $n \times n$ Toeplitz matrix $\mathbf{A}$ with an $n \times m$ sparse matrix $\mathbf{R}$. The basic matrix factorization solves

$$J_0 = \min_{\mathbf{A},\mathbf{R}} \|\mathbf{Y} - \mathbf{A}\mathbf{R}\|_F^2. \tag{4}$$

Here, we enforce a Toeplitz structure to $\mathbf{A}$ and a sparse property to $\mathbf{R}$, respectively. Sparsity problems can be transformed into L_0-regularization problems. The general class of problems, for which some algorithms can provide an answer in polynomial time is called P. The computational complexity of the problem with L_0 regularization has been shown to be NP-hard (nondeterministic polynomial-time hard), which is a class of problems that are at least as hard as the hardest problems in NP. If $P \neq$ NP, the NP-hard problems cannot be solved in polynomial time, which means that the time required to solve the problem using any currently known algorithm increases very quickly as the size of the problem grows (Ge et al., 2011). Thus, many researchers suggest to relax the L_0 regularization and, instead, to consider the L_1 regularization. The L_1-regularization problem can be transformed into an equivalent convex quadratic optimization problem, and therefore, can be very efficiently solved (Xu et al., 2012). Hence, we use the L_1-norm (Tibshirani, 1996) to regularize the column vectors of $\mathbf{R}$, and add the Toeplitz constraint to $\mathbf{A}$, yielding the basic formulation of TSMF as follows:

$$\hat{J} = \min_{\mathbf{A},\mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{R}\|_F^2 + \lambda \sum_{j=1}^{m} \|\mathbf{r}_j\|_1, \tag{5}$$

$$\text{s.t. } \mathbf{A} \text{ is a Toeplitz matrix,}$$

where $\mathbf{r}_j \in \mathbb{R}^n$ is the jth column of $\mathbf{R}$ and $\|\cdot\|_1$ is the L_1-norm of a vector. The difficulty, however, is how to deal with the Toeplitz structure of $\mathbf{A}$. To tackle this, we first write $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & \cdots & \cdots & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & \vdots \\ \vdots & a_{-1} & a_0 & a_1 & \vdots \\ \vdots & & \ddots & a_{-1} & a_0 & a_1 \\ a_{-(n-1)} & \cdots & \cdots & a_{-1} & a_0 \end{pmatrix} = a_{-(n-1)} \begin{pmatrix} & & O \\ 1 & & \end{pmatrix}$$

$$+ \cdots + a_0 \begin{pmatrix} 1 & & O \\ & \ddots & \\ O & & 1 \end{pmatrix} + \cdots + a_{n-1} \begin{pmatrix} & & 1 \\ O & & \end{pmatrix} \tag{6}$$

based on its Toeplitz structure. Denoting $\mathbf{I}_k$ to be the matrix filled with zeros everywhere except along the kth-diagonal, which is filled with 1 s, the above equation can be written as

$$\mathbf{A} = a_{-(n-1)}\mathbf{I}_{-(n-1)} + \cdots + a_0\mathbf{I}_0 + \cdots + a_{n-1}\mathbf{I}_{n-1}$$

$$= \sum_{k=-n+1}^{n-1} a_k\mathbf{I}_k. \tag{7}$$

Substituting equation 7 into equation 5, we can then eliminate the Toeplitz constraint and simplify the original optimization problem to the following one:

$$\hat{J} = \min_{\boldsymbol{\alpha},\mathbf{R}} \frac{1}{2} \left\| \mathbf{Y} - \left(\sum a_k\mathbf{I}_k \right)\mathbf{R} \right\|_F^2 + \lambda \sum_{j=1}^{m} \|\mathbf{r}_j\|_1, \tag{8}$$

where $\boldsymbol{\alpha} = (a_{-(n-1)}, \ldots, a_0 \ldots, a_{n-1})^T$. Because in many applications, such as signal processing, $\mathbf{A}$ can have other properties, e.g., sparsity and smoothness, besides Toeplitz structure, it is necessary to incorporate a regularization term to $\boldsymbol{\alpha}$ to get the general form of TSMF

$$J = \min_{\boldsymbol{\alpha},\mathbf{R}} \frac{1}{2} \left\| \mathbf{Y} - \left(\sum a_k\mathbf{I}_k \right)\mathbf{R} \right\|_F^2 + \Psi(\boldsymbol{\alpha}) + \lambda \sum_{j=1}^{m} \|\mathbf{r}_j\|_1, \tag{9}$$

where $\Psi(\cdot)$ is some chosen regularization function that is assumed to be convex for convenience in this paper.

Algorithm

Equation 9 is generally difficult to solve because it is not convex for all variables. However, if either $\mathbf{A}$ (or equivalently $\boldsymbol{\alpha}$) or $\mathbf{R}$ is

fixed, the original problem can be reduced to much simpler convex subproblems. For this reason, it is natural to alternatively solve these subproblems to approach the solution of the original problem, which is in fact the block coordinate descent method (Tseng, 2001; Xu and Yin, 2013), and its convergence has been proved in Xu and Yin (2013). We describe the details as follows.

First, when $\mathbf{A}$ is fixed, the optimization becomes irrelevant to $\mathbf{A}$, and we get

$$J_\mathbf{R} = \min_{\mathbf{R}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{R}\|_F^2 + \lambda \sum_{j=1}^{m} \|\mathbf{r}_j\|_1. \tag{10}$$

It can be further decomposed into a series of subproblems based on the separability of $\mathbf{Y}$ and $\mathbf{R}$ as

$$J_\mathbf{r} = \min_{\mathbf{r}_j} \frac{1}{2} \|\mathbf{y}_j - \mathbf{A}\mathbf{r}_j\|_2^2 + \lambda \|\mathbf{r}_j\|_1 \tag{11}$$

for $j = 1, \ldots, m$, where $\mathbf{y}_j \in \mathbb{R}^n$ is the jth column of $\mathbf{Y}$. These are the standard L_1-norm regularization problems, which have been extensively studied and can be solved by several existing methods (e.g., Efron et al., 2004; Hale et al., 2007; Kim et al., 2007; Beck and Teboulle, 2009). Similarly, when $\mathbf{R}$ is fixed, we get

$$J_{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \frac{1}{2} \left\| \mathbf{Y} - \left(\sum a_k\mathbf{I}_k \right)\mathbf{R} \right\|_F^2 + \Psi(\boldsymbol{\alpha}). \tag{12}$$

By some simple algebra, it becomes

$$J_{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \frac{1}{2} \|\tilde{\mathbf{y}} - \tilde{\mathbf{R}}\boldsymbol{\alpha}\|_2^2 + \Psi(\boldsymbol{\alpha}), \tag{13}$$

where $\tilde{\mathbf{y}} = \text{vec}(\mathbf{Y})$ and $\tilde{\mathbf{R}} = (\text{vec}(\mathbf{I}_{-(n-1)}\mathbf{R}), \ldots, \text{vec}(\mathbf{I}_{n-1}\mathbf{R}))$. Equation 13 can also be solved without much difficulty if $\Psi(\boldsymbol{\alpha})$ is chosen to be a convex regularization function. By alternatively solving equations 11 and 13, it is expected that we can obtain a good solution to the original problem in equation 9.

Note that although any existing method can be used to solve equations 11 and 13, we highlight the proximal method (Combettes and Wajs, 2005; Bach et al., 2012) in this paper because it is easily implemented and its convergence is guaranteed. We now briefly review this method. Consider the following optimization problem:

$$\min_{x} f(x) + g(x), \tag{14}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a continuous convex function, and $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function with L-Lipschitz continuous gradient for $L > 0$, i.e., $\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x - y\|_2$ for every $x, y \in \mathbb{R}^n$. The basic proximal method obtains a solution using the following iteration:

$$x^{(t+1)} = \text{prox}_{\gamma g}(x^{(t)} - \gamma \nabla f(x^{(t)})), \tag{15}$$

where $\text{prox}_g(\cdot)$ is the proximity operator defined by

$$\text{prox}_g(x) = \arg_{y \in \mathbb{R}^n} \min g(y) + \frac{1}{2}\|x - y\|^2 \tag{16}$$

and γ is a proper parameter, such that $0 < \gamma < (1/L)$. Then, it can be shown (Combettes and Wajs, 2005) that $x^{(t)} \to x^*$ as $t \to \infty$, and x^* solves equation 15.

Applying above method to equations 11 and 13, we obtain the iterations

$$\mathbf{r}_j^{(t+1)} = \text{prox}_{\gamma_j \lambda \|\cdot\|_1}(\mathbf{r}_j^{(t)} - \gamma_j \mathbf{A}^T(\mathbf{A}\mathbf{r}_j^{(t)} - \mathbf{y}_j)) \quad (17)$$

and

$$\boldsymbol{\alpha}^{(t+1)} = \text{prox}_{\gamma_{\boldsymbol{\alpha}}\psi}(\boldsymbol{\alpha}^{(t)} - \gamma_{\boldsymbol{\alpha}}\tilde{\mathbf{R}}^T(\tilde{\mathbf{R}}\boldsymbol{\alpha}^{(t)} - \tilde{\mathbf{y}})) \quad (18)$$

for the subproblems. The proximity operator $\text{prox}_{\lambda\|\cdot\|_1}(\cdot)$ in equation 17 is in fact the well-known soft thresholding (Tibshirani, 1996) explicitly defined by

$$[\text{prox}_{\lambda\|\cdot\|_1}(x)]_i = \text{sign}(x_i)(|x|_i - \lambda)_+ \quad (19)$$

for $i = 1, \ldots, n$, with $(x)_+ = \max(x, 0)$. The proximity operator in equation 18 depends on the regularization function $\Psi(\cdot)$, which may not have an explicit form but can be efficiently computed in general (Bach et al., 2012).

The entire procedure is summarized in Algorithm 1.

The convergence of the sequence generated by Algorithm 1 is analyzed in Appendix A.

SPARSE-SPIKE DECONVOLUTION USING TSMF

As mentioned in the first section, following the convolution trace model, a section of 2D seismic data can be expressed by equation 3, where matrix $\mathbf{A}$ is the wavelet convolution matrix that has Toeplitz structure and matrix $\mathbf{R}$ is the reflectivity matrix that is sparse. Therefore, the previous proposed TSMF algorithm can be used to perform SSD for seismic data.

As we know, seismic wavelets are usually smooth and much shorter in time than the seismic trace. That is, if we assume that the length of the seismic wavelet is l sampling points, we usually have $l \ll n$. For equation 7, the points $a_{-(\lceil l/2 \rceil - 1)}, \ldots, a_{-(\lceil l/2 \rceil)}$ are

the values of the seismic wavelet, other points are zeros. That means $\boldsymbol{\alpha}$ is sparse to some extent in this case, and nonzero values are concentrated around the diagonal in matrix $\mathbf{A}$. Hence, in this paper, we choose the Fused Lasso (Tibshirani et al., 2005) penalty as the regularization function in equation 9, i.e.

$$\Psi(\boldsymbol{\alpha}) = \beta\|\boldsymbol{\alpha}\|_1 + \beta_1 \sum_k |a_{k+1} - a_k| + \beta_2 \|\boldsymbol{\alpha}\|_2, \quad (20)$$

where the first term forces $\boldsymbol{\alpha}$ to be sparse (i.e., forces $\boldsymbol{\alpha}$ to be compact), and the second and third terms constrain $\boldsymbol{\alpha}$ to be smooth. If we use only the second term to constrain $\boldsymbol{\alpha}$ to be smooth, the estimated wavelet may have successive identical values, so we add the third term to avoid this case. For simplicity, in this paper, we apply average smoothness to $\boldsymbol{\alpha}$ when we detect that there are successive identical values in the estimated waveform during iteration to replace the third term. The average smoothness also reduces the weight of the second term, when the S/N of the seismic data is high, β_1 can be set to 0. Then, substituting equation 20 into equation 13 gives the following Fused Lasso problem:

$$J_{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \frac{1}{2}\|\tilde{\mathbf{y}} - \tilde{\mathbf{R}}\boldsymbol{\alpha}\|_2^2 + \beta\|\boldsymbol{\alpha}\|_1 + \beta_1 \sum_k |a_{k+1} - a_k|, \quad (21)$$

which can be solved by any of the standard methods (e.g., Hoefling, 2010; Liu et al., 2010; Ye and Xie, 2011). Here, we use the function mexFistaFlat in SPAMS (Mairal, 2012) to help solve the objective function in equation 21.

In this paper, we use the fast iterative shrinkage-thresholding algorithm (FISTA) with backtracking method (Beck and Teboulle, 2009) to solve equation 11. Instead of λ, the sparsity K, assumed to be the upper bound of the number of nonzero reflectivity coefficients, is taken as the parameter of this FISTA method. We simply give the algorithm of FISTA with backtracking in Algorithm 2.

The principle of this algorithm can be found in Beck and Teboulle (2009) and Xu et al. (2012).

The sparsity parameter K can be estimated from the number of the local peaks of the seismic trace envelope. The local peaks of the seismic trace envelope contain information regarding the locations

---

Algorithm 1. Algorithm for TSMF.

Input: Observation matrix $\mathbf{Y} \in \mathbb{R}^{n \times m}$, sparsity parameter λ, parameters in $\Psi(\boldsymbol{\alpha})$, initialized $\mathbf{A}^{(0)} \in \mathbb{R}^{n \times n}$.

Output: Toeplitz matrix $\mathbf{A}$ and sparse matrix $\mathbf{R}$.

1:     $k = 1$

2:     **While** not converged **do**

3:        **for** $j = 1, \ldots, m$ **do**

4:          Solve $\mathbf{r}_j^{(k)} = \arg\min_{\mathbf{r}_j} \frac{1}{2}\|\mathbf{y}_j - \mathbf{A}^{(k-1)}\mathbf{r}_j\|_2^2 + \lambda\|\mathbf{r}_j\|_1$ via the proximal method (Algorithm 2 is used in this paper).

5:        **end for**

6:        $\mathbf{R}^{(k)} = (\mathbf{r}_1^{(k)}, \ldots, \mathbf{r}_m^{(k)})$, $\tilde{\mathbf{R}}^{(k)} = (\text{vec}(\mathbf{L}_{-(n-1)}\mathbf{R}^{(k)}), \ldots, \text{vec}(\mathbf{I}_{n-1}\mathbf{R}^{(k)}))$

7:        Solve $\boldsymbol{\alpha}^{(k)} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\tilde{\mathbf{y}} - \tilde{\mathbf{R}}^{(k)}\boldsymbol{\alpha}\|_2^2 + \Psi(\boldsymbol{\alpha})$ via the proximal method (function mexFistaFlat in SPAMS, a SPArse Modeling Software, is used in this paper), where $\tilde{\mathbf{y}} = \text{vec}(\mathbf{Y})$.

8:     Recover $\mathbf{A}^{(k)}$ from $\boldsymbol{\alpha}^{(k)}$ based on equation 7.

9:     $k = k + 1$.

10:     **end while**

11:     $\mathbf{A} = \mathbf{A}^{(k)}$, $\mathbf{R} = \mathbf{R}^{(k)}$

---

of the bigger seismic reflectivity coefficients. Specifically, there may be several bigger seismic reflectivity coefficients around each peak. For one trace, we denote the number of the local peaks of the trace envelope as K_0, and let the sparsity parameter be $K = cK_0$, where c is a constant. Then we could take the width of the relatively narrower waveform around a peak as a unit to estimate the average number of waveforms around each peak, and take a value a little larger than that average number as the value of c. For ease of understanding, Figure 1 gives an example for estimation of K. By convolving the reflectivity in Figure 1a with a 30° phase rotated Ricker wavelet with dominant frequency of 30 Hz, we get the synthetic seismic trace shown in Figure 1b and then we can obtain the trace envelope and the local peaks of the envelope. Comparing Figure 1a and 1b, we can easily see that the locations of the reflectivity coefficients and the locations of the local peaks of the envelope show a certain correspondence. In Figure 1b, the number of the local peaks K_0 is 16. Taking a region around the first local peak as a unit, we can see that there may be 1.5–2 reflectors around each peak on average. So the value of c can be 1.5–2, and the value of K can then be 24–32, which is reasonable compared with the true number of reflectors 22. Actually, for this noise-free seismic trace, when c increases to 3, the proposed method still works well, as will be shown later in the numerical examples. For the wavelet, the parameter β is related to
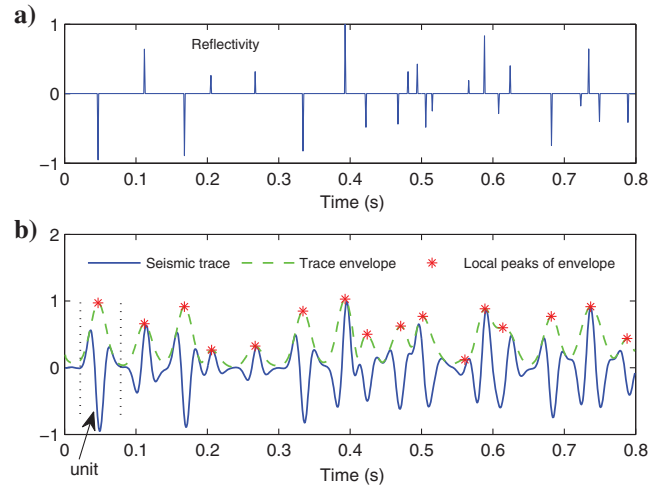

Figure 1. An example for estimation of K. (a) True reflectivity; (b) the synthetic seismic trace (solid) with its envelope (dashed) and local peaks of envelope (star). The number of the local peaks of the envelope K_0 is 16. Taking the waveform around the first local peak as a unit, we can see that there may be 1.5–2 reflectors around each peak on average. So the value of K can be 1.5–2 times of K_0, that is, 24–32.

Algorithm 2. Algorithm for FISTA with backtracking.

Input: Observed seismic trace $\mathbf{y} \in \mathbb{R}^{n\times1}$, sparsity K, initialized $\mathbf{A}_0 \in \mathbb{R}^{n\times n}$, $\mathbf{r}_0 \in \mathbb{R}^{n\times1}$, $t_0 = 1$, $\mathbf{z}_0 = \mathbf{r}_0$.
Output: Sparse reflectivity $\mathbf{r} \in \mathbb{R}^{n\times1}$.
1:   $k = 0$, $\mu_0 = \frac{1-\varepsilon}{\|\mathbf{A}_0\|^2}$ with any small $\varepsilon \in (0,1)$
2:   **While** not converged **do**
3:     $k = k + 1$, $\mu_k = \mu_0$,
4:     $\mathbf{e} = \mathbf{y} - \mathbf{A}_{k-1} * \mathbf{r}_{k-1}$, $\mathbf{u}_k = \mathbf{A}_{k-1}^T \mathbf{e}$, $\mathbf{b}_k = \mathbf{r}_{k-1} + \mu_k * \mathbf{u}_k$
5:     Put $\mathbf{b}_k$ in descending order, let $\tilde{\mathbf{b}}_k$ be the sorted $\mathbf{b}_k$ and $\mathbf{J}$ be the original index of the element in $\tilde{\mathbf{b}}_k$, that is $\tilde{\mathbf{b}}_k(\mathbf{J}) = \mathbf{b}_k$
6:     $\lambda\mu = \tilde{\mathbf{b}}_k(K+1)$
7:     Let $\mathbf{JK} = \mathbf{J}(1:K)$, $\mathbf{z}_k = \text{zeros}(n,1)$, $\mathbf{z}_k(\mathbf{JK}) = \mathcal{T}_{\lambda\mu}(\mathbf{b}^{(k)}) = (|\mathbf{b}^{(k)}(\mathbf{JK})| - 0.5 * \lambda\mu)\text{sign}(\mathbf{b}^{(k)}(\mathbf{JK}))$
8:     $t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$,
9:     $\mathbf{r}_k = \mathbf{z}_k + \left(\frac{t_{k-1}-1}{t_k}\right)(\mathbf{z}_k - \mathbf{z}_{k-1})$,
10:   **end while**
11:   $\mathbf{r} = \mathbf{r}^{(k)}$

Table 1. Influence and selection method for the parameters.

Parameter	Influence	How to choose
λ	Sparsity of reflectivity series	Not used (replaced by K)
K	Upper bound of the number of reflectors	Count local peaks of the seismic trace envelope (K_0), and estimate the average number of pulses around each peak (c), then we have $K = cK_0$ (usually, the value of c can be 1.5–3)
β	Compactness of wavelet	Related to trace length and wavelet dominant frequency
β_1	Smoothness of wavelet	Zero if high S/N
β_2	Smoothness of wavelet	Not used (replaced by average smoothness during iteration)

the dominant frequency of the wavelet and the length of the seismic trace, which will be discussed later. In practice, we can get a reference value for β from synthetic experiments with the same dominant frequency wavelet and trace length. The parameter β_1 can be given a small value when the seismic data are somewhat noisy. If well logs are available, we can use them for quality control to improve the choice of these parameters. For clarity, we list the influence and selection method for all the parameters appeared in this paper in Table 1.

APPLICATION

In the following, we will test the proposed TSMF-based SSD method using synthetic and field data.

Synthetic data examples

Most SSD methods require a known wavelet (Nguyen, 2008). Though several methods exist to estimate the seismic wavelet that yield reasonable results even for nonminimum phase sources, often the quality of the derived wavelets is data dependent. Clearly, ignoring wavelet inaccuracies would lead to partially successful deconvolution results (Velis, 2008). As a result, the knowledge of the wavelet is critical in carrying out the deconvolution.

In the first experiment, we test our TSMF-based SSD method on synthetic seismic data, where we assume that we have no advanced knowledge of the wavelet. The reflectivity profile is produced from a portion of the Marmousi2 model, as shown in Figure 2, which contains the section that Velis (2008) used for testing. The size of this section is 800×200 (number of time sampling points × number of traces). For simplicity, we assume constant density, and also assume traces have 800 points with 1 ms sampling interval. Thus, this data set consists of 200 traces with an offset interval of 12.5 m and a time window of 800 ms with $\Delta t = 1$ ms. The reflectivity section is shown in Figure 3b. By convoluting this reflectivity section with a 30° phase rotated Ricker wavelet with dominant frequency of 30 Hz, we get the noise-free synthetic seismic section shown in Figure 3a. After adding Gaussian noise with approximately 10% of the maximum trace value to the seismic profile in Figure 3a, we get the noisy synthetic profile in Figure 3d. As this section is relatively large compared with our computer memory, we divide it into seven sections along its lateral direction. Each of the first six sections has 30 seismic traces with dimension of

800×30. Consequently, the last section has 20 seismic traces, and its size is 800×20. We average the amplitude spectra of all traces and perform an inverse Fourier transform to obtain the initial estimation of the wavelet, and then form the initial estimation of matrix **A** following equations 6 and 7 for each section of data. Then we use the wavelet solution of the previous section as the initial model for the next section, and take the average of the source wavelets deconvolved from all the sections via TSMF-based SSD as the final estimation of the source wavelet for all sections to deconvolve the reflectivity using equation 11. Thus, the lateral continuity of the seismic data can be maintained.
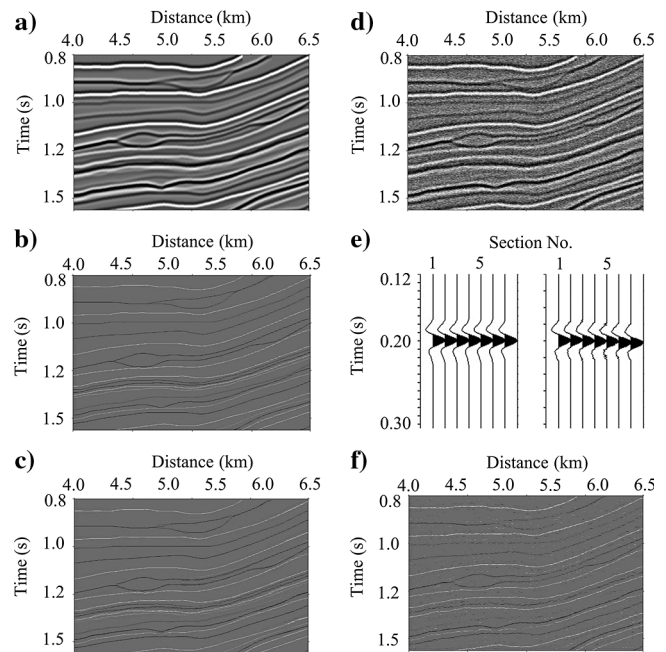

Figure 3. TSMF-based SSD on the Marmousi2 model. (a) The noise-free synthetic data, (b) the true reflectivity, (c) the estimated reflectivity from noise-free synthetic data, (d) the noisy synthetic data with Gaussian noise approximately 10% of the maximum trace value, (e) the estimated wavelet from (left) noise-free and (right) noisy synthetic data, and (f) the estimated reflectivity from noisy synthetic data.

Figure 2. (a) Marmousi2 P-wave velocity model and (b) an enlarged portion in the rectangular area that will be used to generate the reflectivity profile for the synthetic examples.
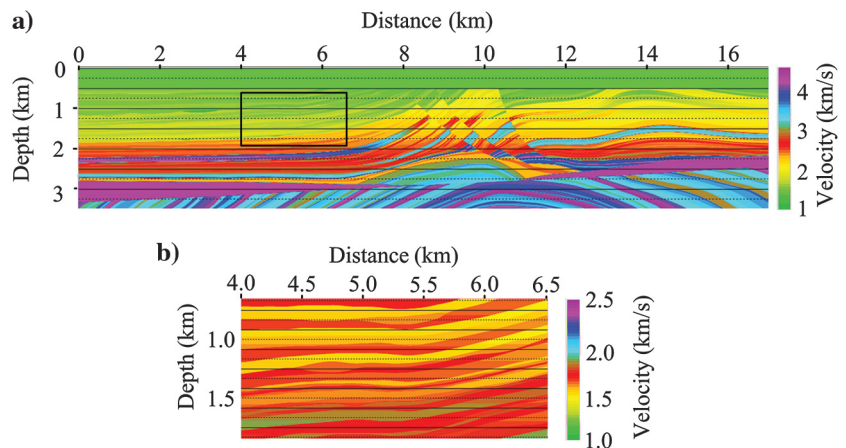
Figure 3c shows the TSMF-based SSD result of the noise-free input synthetic data shown in Figure 3a with $K = 40$, $\beta = 6$, and $\beta_1 = 0$. We can see that nearly all the reflectors are recovered and resolved accurately with appropriate lateral coherence. Figure 3f shows the result obtained with $K = 30$, $\beta = 6$, and $\beta_1 = 0.01$ when the input data are contaminated with Gaussian noise of approximately 10% of the maximum trace amplitude. As expected, the variability of the solution obtained with noisy data increases. Nevertheless, most reflectors are recovered well and the main structure is appropriately preserved. Here, the average number of the local peaks of the envelopes of the synthetic seismic traces is 20.6. Let $K_0 = 20$, so we have $K = 2K_0 = 40$ for noise-free data,
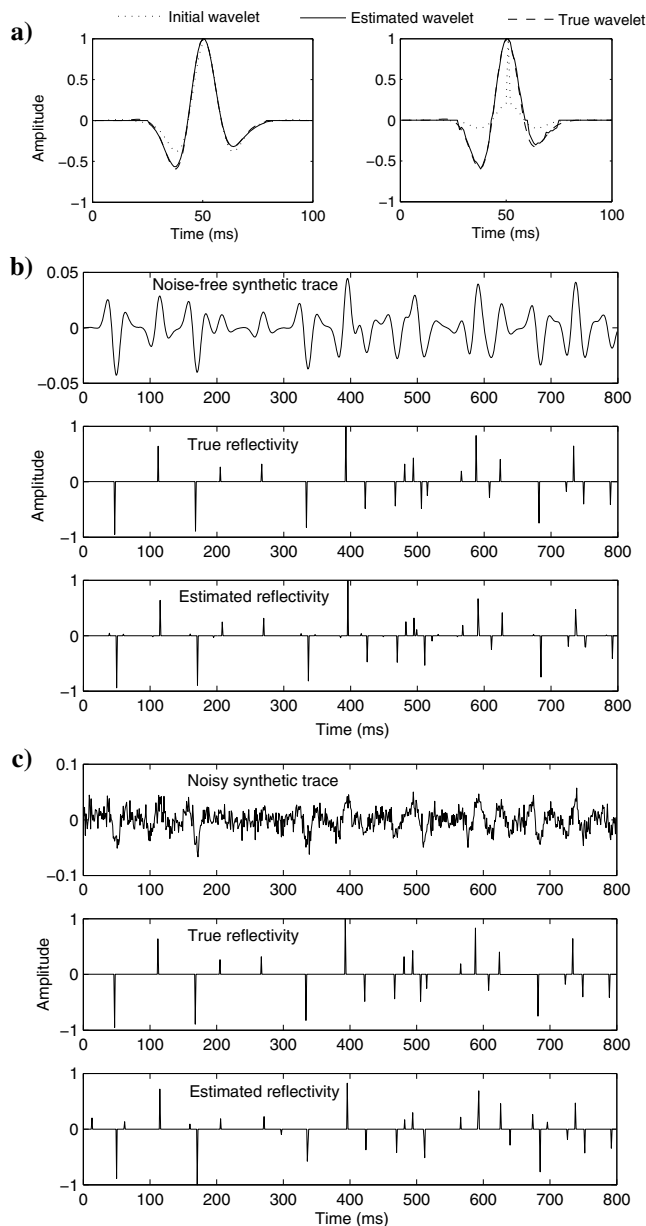

Figure 4. Wavelet and reflectivity comparison for the examples in Figure 3. (a) Wavelet comparison for (left) the noise-free and (right) noisy cases. The first trace of (b) Figure 3a–3c and (c) Figure 3d, 3b, and 3f.

and $K = 1.5K_0 = 30$ for noisy data. Taking synthetic experiments as reference, we choose $\beta = 6$, which will be discussed in detail later in the example of the sensitivity analysis for β. Figure 3e shows the estimated wavelets for (left) noise-free and (right) noisy data for each section. Figure 4 shows wavelet and one trace comparison for the noise-free and noisy cases in Figure 3. Figure 4a shows the wavelet comparison, where the estimated wavelets are the average of the wavelets in Figure 3e for (left) noise-free and (right) noisy data, respectively. We can see that, although the estimation of the initial wavelet is inaccurate, after applying our method, the wavelet is recovered very well. Figure 4b illustrates that almost all reflectors are resolved for the noise-free case. Figure 4c shows that the noise affects the accuracy of the deconvolution. Comparing the noise-free synthetic trace in Figure 4b and the noisy trace in Figure 4c, we can see that the synthetic trace is badly contaminated by the noise, nevertheless, by decreasing the value of K from 40 to 30, most reflectors are recovered from the noisy data. Compared with the stochastic SSD method of Velis (2008), the proposed method has a better tolerance to errors in the initial wavelet estimation. Velis's (2008) method uses constant-phase shift to calibrate the initial wavelet, which just calibrates the phase of the wavelet. Our method can automatically calibrate the whole wavelet, not just the phase of the wavelet, as shown in Figure 4a.

Figure 5 shows a sensitivity analysis of the result obtained with various values of the sparsity K for noise-free data. The true number of reflectors is 22. Figure 5b shows the estimated reflectivity when K equals to 22, 25, 30, 40, and 50 in turn, where $\beta = 6$ and $\beta_1 = 0$ in all cases. From this analysis, we can see that, as K increases, our method induces some small artificial reflectors. Nevertheless, it works well even when K is more than two times of the true number of reflectors.
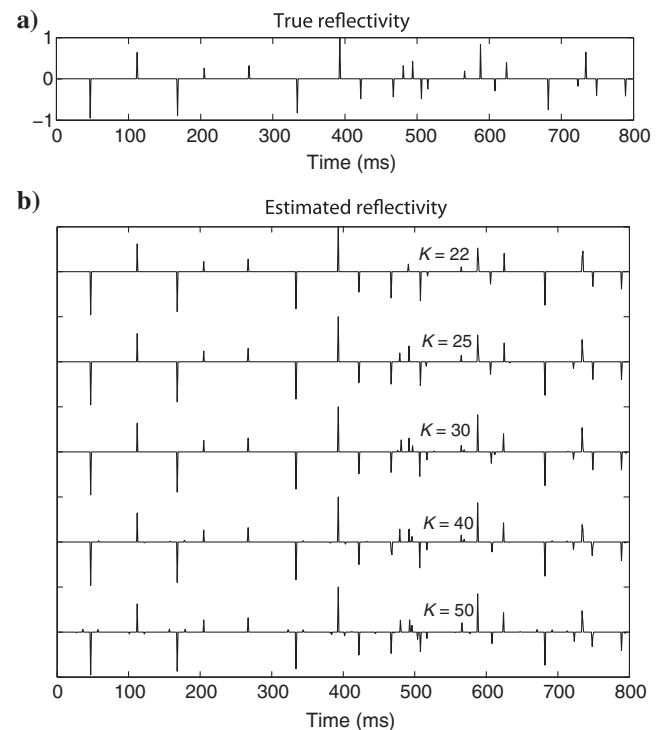

Figure 5. Sensitivity analysis for the choice of sparsity K. (a) True reflectivity; (b) estimated reflectivity when K is 22, 25, 30, 40, and 50 from top to bottom, $\beta = 6$ and $\beta_1 = 0$ in all cases.

Figure 6 gives a sensitivity analysis for β. Figure 6b shows the estimated wavelet and reflectivity when β equals to 0.01, 2, 6, 10, 20, 50, and 100 in turn, where $K = 30$ and $\beta_1 = 0$ in all cases. From this figure, we can see that, as β increases, the wavelet becomes more compact. When β equals to 2–20, the proposed method works well, which means β can be selected within a wide range. Because β controls the compactness of the wavelet, it is related to the dominant frequency of the wavelet and the length of the seismic trace, which is demonstrated in Figure 7. Compared with the above examples, the length of the seismic trace is reduced to 300 sampling points in Figure 7. Figure 7b shows the estimated wavelet and reflectivity when β equals to 1, 2.3, 5, and 10 in turn, where the wavelet is still a 30 Hz Ricker wavelet, $K = 10$ and $\beta_1 = 0$ in all cases. From this figure, we can see that, when β equals to 2.3 and 5, the proposed method works well, when β increases to 10, the result is beginning to degrade. The reason is that when the dominant frequency of the wavelet remains unchanged, whereas the length of the seismic trace is reduced, the sparsity of the wavelet of matrix A decreases, so β should be given a smaller value. Compared with the example in Figure 6, in this example, β should be selected as almost 3/8 (ratio of the trace length) of the β value in Figure 6. Figure 7c shows the estimated wavelet and reflectivity when β equals to 6, 15, 20, and 60 in turn, where the wavelet is changed to a 60 Hz Ricker wavelet, with $K = 10$ and $\beta_1 = 0$ in all cases. From this figure, we can see that, the selection value of β in this case is larger than in Figure 7b. The reason is that when the dominant frequency of the wavelet in-

creases, whereas the length of the seismic trace remains unchanged, the sparsity of the wavelet of matrix A increases, so β should be given a larger value. Comparisons between the true reflectivity and the estimated reflectivity show that β can be determined according to the length of the seismic trace and the dominant frequency of the wavelet by comparing with a reference data set. In practice, we can get a reference value for β from synthetic experiments.

Figure 8 shows examples of the tolerance of the proposed method to the phase error of the initial wavelet. We use the same amplitude spectra averaging method mentioned above to obtain the initial estimation of the wavelet. That means the initial wavelet is zero phase. Figure 8b and 8c shows the estimated wavelet and reflectivity when the true phase of the synthetic data changes from −90° to 90°, where $K = 30$, $\beta = 6$, and $\beta_1 = 0$ in all cases. We can see that in all cases,
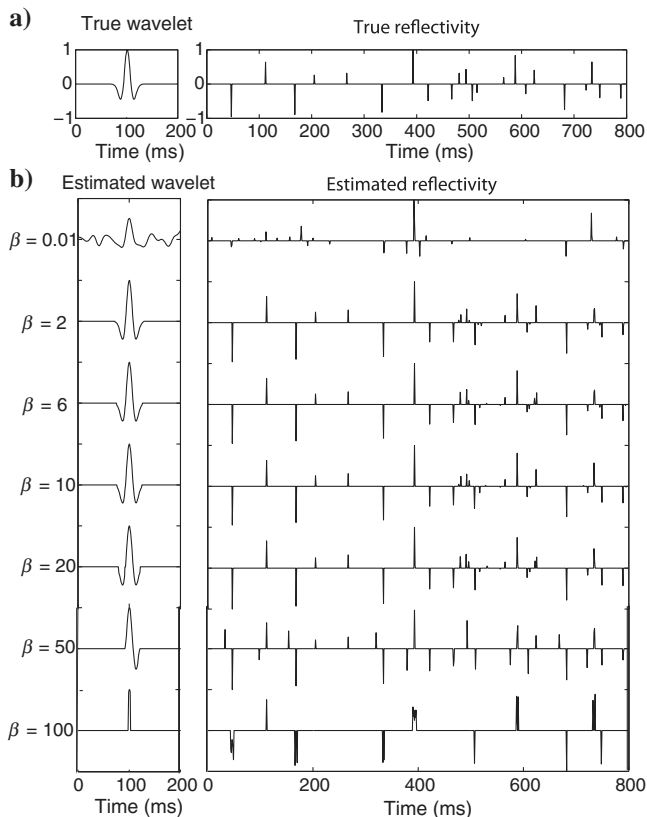

Figure 6. Sensitivity analysis for the choice of β. (a) True wavelet and true reflectivity; (b) estimated wavelet and estimated reflectivity when β is 0.01, 2, 6, 10, 20, 50, and 100 from top to bottom, $K = 30$ and $\beta_1 = 0$ in all cases.
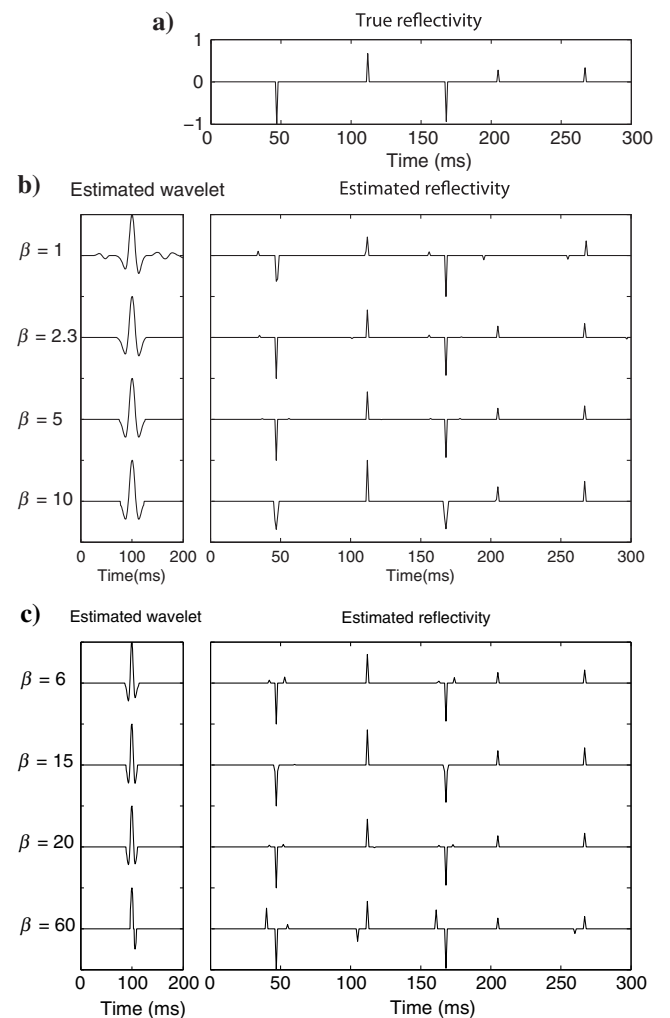

Figure 7. Analysis of the relation of β to the length of the seismic trace and the dominant frequency of the wavelet. Comparing with the above examples, the length of the seismic trace is reduced to 300 sampling points in this example. (a) True reflectivity; (b) estimated wavelet and estimated reflectivity when β is 1, 2.3, 5, and 10 from top to bottom, where the wavelet is still a 30 Hz Ricker wavelet; (c) estimated wavelet and estimated reflectivity when β is 6, 15, 20, and 60 from top to bottom, where the wavelet is changed to a 60 Hz Ricker wavelet. Here, $K = 10$ and $\beta_1 = 0$ in all cases.

our method can resolve the wavelet and reflectivity well, except for the case, where the phase equals 90°. In that case, the estimated wavelet and reflectivity all have their polarity reversed. Thus, in practice, we need to know the polarity of the seismic data first, and give the correct polarity to the initial zero-phase wavelet.

Field data example

Figure 9 illustrates a poststack field data example. Figure 9a shows a portion of a poststack field data profile provided by China National Offshore Oil Corporation (CNOOC). The size of this seismic profile is 300×801 (number of time sampling points × number of traces, $\Delta t = 2$ ms), which is relatively large. Thus, we divide this seismic profile into 27 sections along its lateral direction. Each of the first 26 sections has 30 seismic traces with size 300×30. Consequently, the last section has 21 seismic traces, and its size is 300×21. We process each seismic section separately using Algo-
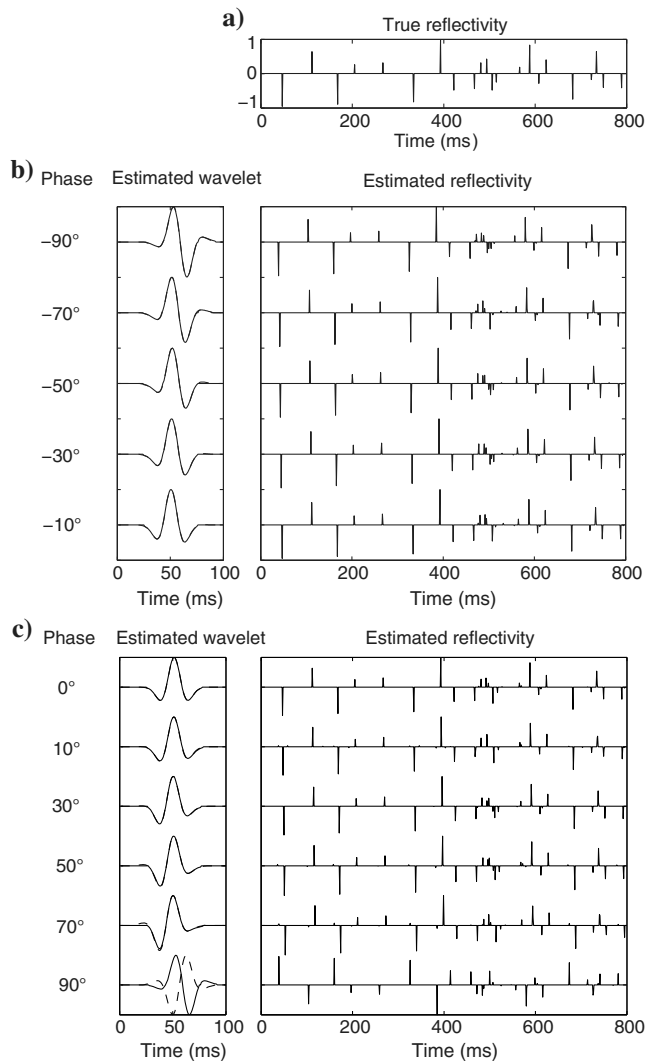

Figure 8. Analysis of the tolerance to phase error of the initial wavelet. (a) True reflectivity; (b and c) estimated wavelet (solid line), true wavelet (dash line), and estimated reflectivity, data phase changes in the range from −90° to 90°, $K = 30$, $\beta = 6$, and $\beta_1 = 0$ in all cases.

rithm 1. Here, we take the wavelet estimated from the first section as the initial wavelet of the second section, wavelet estimated from the second section as the initial wavelet of the third section, and so on. Moreover, we normalize the amplitude of the estimated wavelet during each iteration to keep the energy balanced between sections. Then, we take the average of the wavelets estimated from these sections as the final estimation of the wavelet to deconvolve all sections for the reflectivity following equation 11. Thus, we obtain the TSMF-based SSD result that can maintain the lateral continuity of the seismic data, as shown in Figure 9b. In this example because the mean number of the local peaks of the seismic traces' envelope is 27, and usually there are much more layers than observed in practice, so we choose 80 that is almost three times of the number of the peaks of the seismic trace envelope as the value for K. Because the dominant frequency of this seismic profile is near to 30 Hz, taking the example in Figure 7b as reference, we select $\beta = 2.3$. Because this section of seismic profile is contaminated by some noise, so we choose $\beta_1 = 0.2$ to reduce the effects of noise. Comparing Figure 9a and 9b, we can see that the TSMF-based SSD method cannot only effectively derive the reflectivity from seismic data, but it also achieves a good laterally continuous result and preserves the complex structure of the seismic data: see the complex structures, such as the fault indicated by arrows, the trap shown using an ellipse and the anticlines within the rectangles. A well log of impedance filtered by a low-pass filter with 300 Hz cutoff frequency is inserted in these two profiles. We can see that the estimated reflectivity trace near the position of the well matches the locations of the boundaries between layers of constant impedance very well.

The well-log data are not required by our inversion method, but they can be used for quality control purposes. Due to the resolution differences between the well logs and the seismic data, the crosscorrelation calculated between the reflectivity derived from the well logs filtered with a zero-phase band-pass filter, and the inverted trace reflectivity near the position of the well filtered with the same filter is used for quality control. Should this not match as expected, the requisite parameterization can be varied until a suitable match is obtained. In Figure 10a, the synthetic produced by convolving the reflectivity obtained from well logs with a 25 Hz Ricker wavelet is correlated with the trace near the well location from input seismic data in Figure 9a. In Figure 10b, the reflectivity obtained from well logs is filtered by a zero-phase band-pass filter with bandwidth 15–120 Hz, and correlated with the seismic trace obtained by passing the estimated reflectivity trace near the well location from Figure 9b through the same filter. Comparing the seismic profile (black) in Figure 10a and 10b, we can see that the vertical resolution of the seismic data is significantly improved by our method. The correlation coefficient between the synthetic and the seismic data increases from 0.3942 to 0.58 after the frequency enhancement provided by our method. This match is as expected, so we accepted this group of parameters and the result.

Figure 11 shows the estimated wavelet and comparisons between some seismic traces and their corresponding estimated reflectivity for the field data in Figure 9. Figure 11a is the estimated wavelet, which is the average of the 27 estimated wavelets from the seismic sections. Figure 11b–11d shows the 270th, 550th, and 700th traces and their estimated reflectivity. We can see that our method works well to estimate the reflectivity from the original data, and the vertical resolution of the seismic data has been significantly improved by the removal of the wavelet.

Figure 12a and 12b shows a comparison between the amplitude spectra of the 270th trace of the field data shown in Figure 9 before and after the application of the TSMF-based SSD. After being processed by the TSMF-based SSD, the shape of the low-frequency component of the amplitude spectrum of the trace is similar to that of the original data, whereas the high-frequency component is enhanced as expected, leading to a significant improvement on the vertical resolution. Figure 12c shows the normalized amplitude spectrum of the reflectivity derived from well logs near the 270th trace, we can see that the distribution of this reflectivity is somewhat like blue noise. Comparing Figure 12a, 12b, and 12c, we can see that the distribution of the amplitude spectrum of the processed trace is closer to that of the reflectivity derived from well logs, which further demonstrates the effectiveness of our method.

DISCUSSION

It should be pointed out that the proposed deconvolution method requires significant computer memory because of the matrix computation. In practice, the size of field seismic data is usually large, so we need to divide the seismic data into proper sections before processing. We can similarly divide 3D poststack seismic data into blocks, rearrange each block into a matrix, and then use our method to process the data. To keep lateral continuity, we take the average of the source wavelets deconvolved from these sections (blocks) via TSMF-based SSD as the final estimation of the source wavelet for all sections (blocks) to deconvolve the reflectivity. Moreover, our method can be developed for the case that the seismic wavelet varies in the lateral direction, in which case we can take the average of the wavelet solutions obtained for one section (block) and the sections (blocks) around it as the final wavelet solution for this section (block) to keep the lateral continuity. We can divide the seismic data according to the capacity of computer memory. If necessary, we can use parallel computation to reduce the computation time.

Another issue is that despite the convergence of the TSMF algorithm, it is not guaranteed to converge to the optimum solution. Subsequently, although the proposed method has a favorable error tolerance to the choice of the initial wavelet, giving a more accurate initial value to wavelet matrix **A** can result in a more reliable result. In addition, when the reflectors are closely spaced, the interference of the reflected wavelets may affect the estimation of wavelet and

Figure 9. Field data example. (a) A part of post-stack field data with complex structures, such as the fault indicated by an arrow, the trap in the ellipse, and the anticlines in the rectangle; (b) the SSD result of the field data. A well log of impedance is inserted in these two profiles.
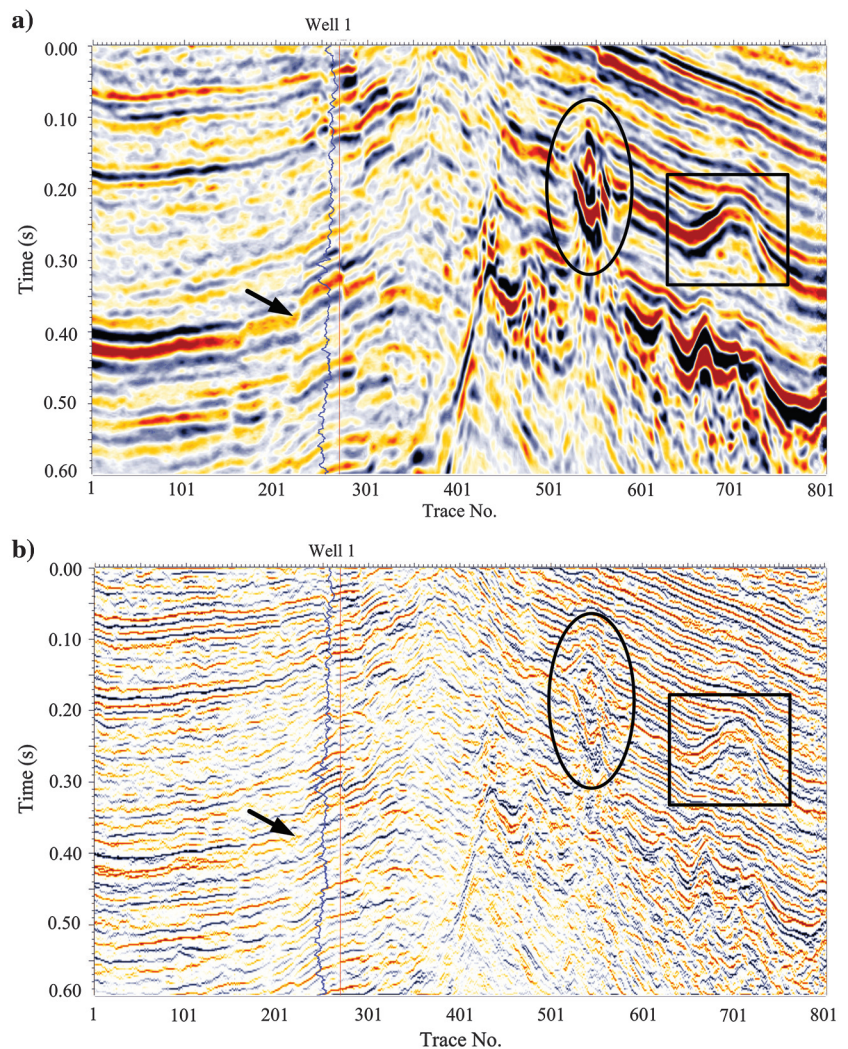
reflectivity. The proposed method can alleviate this problem in some cases, but it cannot thoroughly solve this problem.

CONCLUSION

We have proposed a novel SSD method based on TSMF. The TSMF algorithm computes the factors of a matrix bilinear decomposition, where the factors are Toeplitz and sparse matrices. When a section of 2D seismic data can be approximately described by the product of a wavelet convolution matrix that has Toeplitz structure and a reflectivity matrix with sparse structure, the TSMF algorithm is ideally suited for SSD. The TSMF-based SSD simultaneously deconvolves the seismic matrix into a wavelet matrix and a reflectivity matrix by alternatively solving two inversion subproblems related to Toeplitz wavelet matrix and sparse reflectivity matrix, respectively. As the seismic wavelet is usually compact and smooth,

Fused Lasso is used to constrain the elements in the Toeplitz wavelet matrix. After using average smoothing, the parameters in Fused Lasso can be determined according to the length of the seismic trace and the dominant frequency of the wavelet by comparing with a reference data set. An L_1-norm is used to constrain the reflectivity, where upper bound of the number of reflectors is used as the parameter, which can be easily determined by counting peaks of the seismic trace's envelope. Because the TSMF algorithm estimates the wavelet using multiple traces, it is more reliable in the presence of random noise. In practice, due to the limitations of computer memory, large seismic data sets are divided into blocks, and the average of the source wavelets deconvolved from these blocks via TSMF-based SSD is used as the final estimation of the source wavelet for all blocks to deconvolve the reflectivity, thus the lateral continuity of the seismic data can be maintained. Examples using

a) Density   Velocity   Impedance   A section of original field data in Figure 8a

Correlation coefficient: 0.3942

b) Density   Velocity   Impedance   A section of estimated reflectivity in Figure 8b filtered by
                                         15−120 Hz bandpass filter
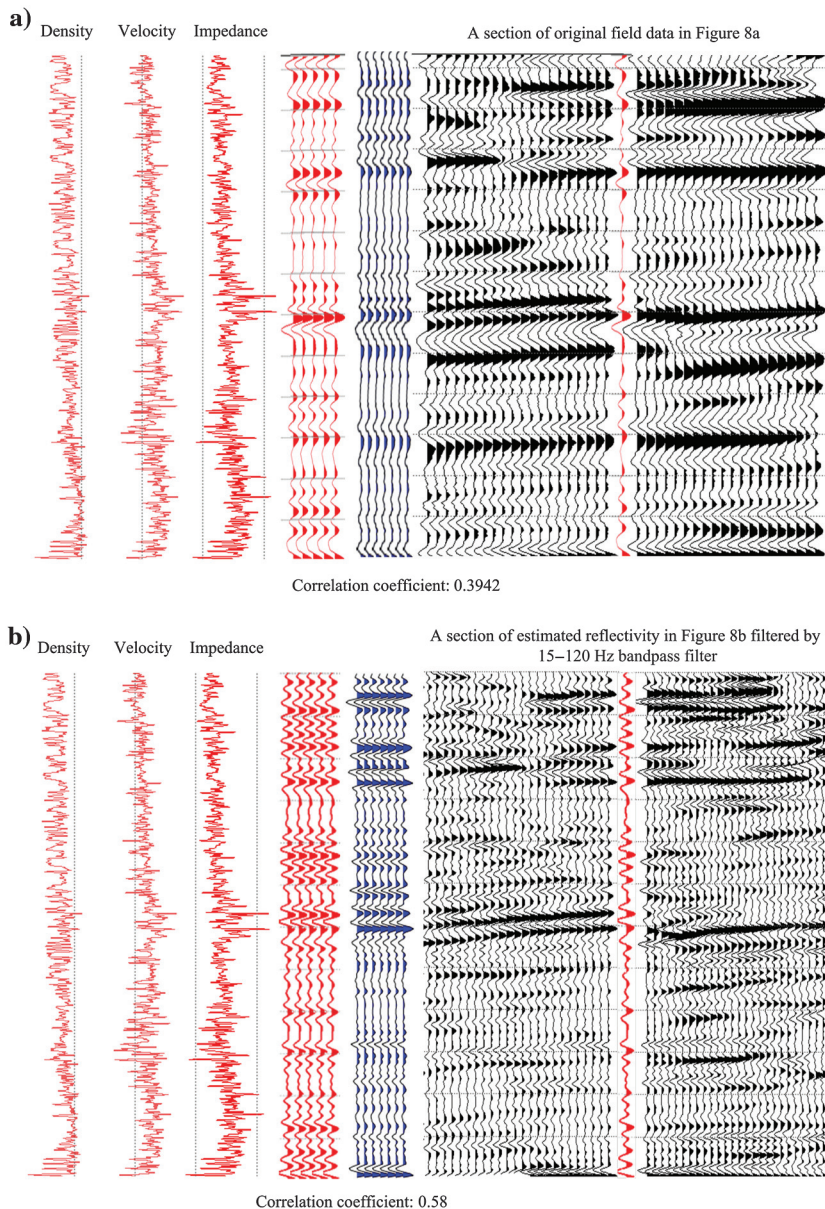
Correlation coefficient: 0.58

Figure 10. Correlation between the synthetic trace derived from well logs and the trace near the well location. (a) Correlation between a synthetic trace derived from well logs (red) and a trace from Figure 9a near the well location (blue). (b) Correlation between filtered reflectivity derived from well logs (red) and filtered reflectivity from Figure 9b near the well location (blue). Comparison between the seismic profile (black) in panels (a and b) shows that the vertical resolution of the seismic data is significantly improved by our method.
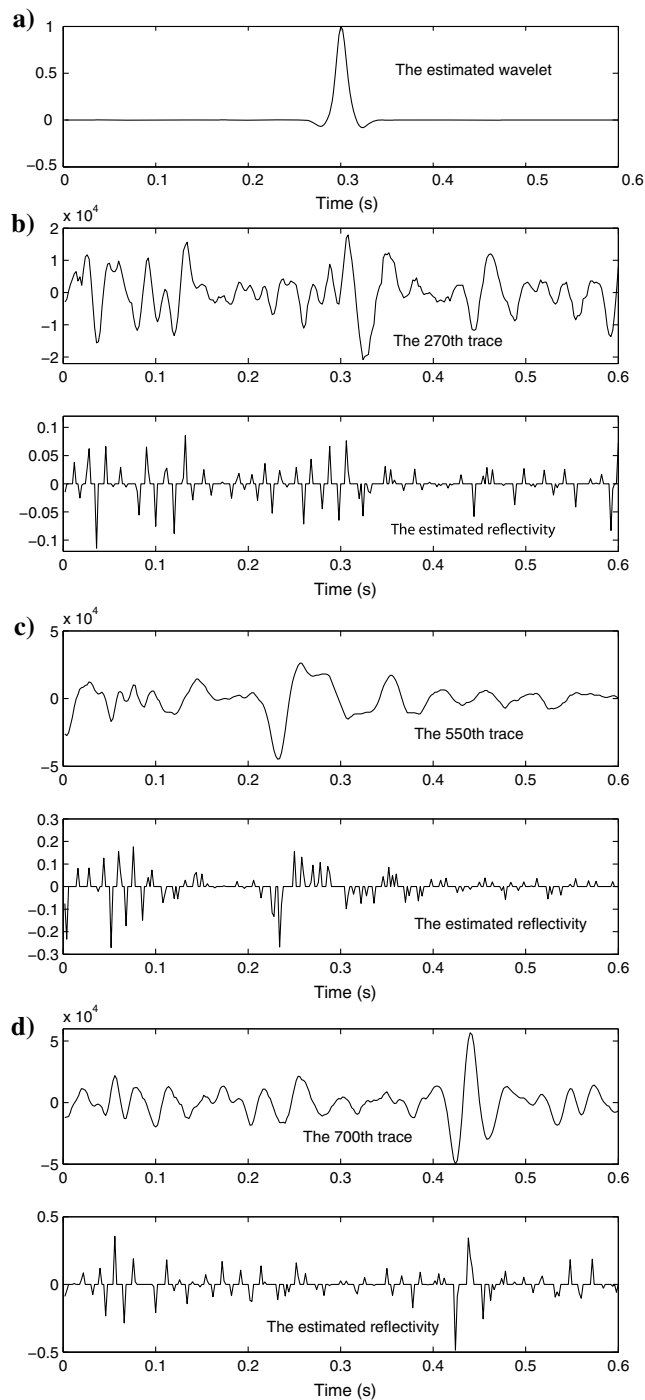
Figure 11. Estimated wavelet and traces comparison of the field data in Figure 9. (a) Estimated wavelet; (b) the 270th trace and its estimated reflectivity; (c) the 550th trace and its estimated reflectivity; and (d) the 700th trace and its estimated reflectivity.
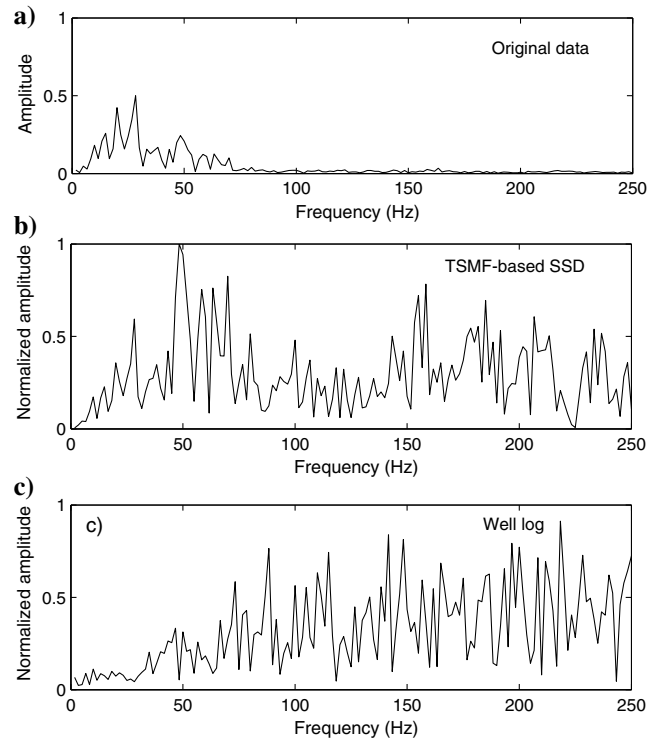

Figure 12. The amplitude spectra of the 270th trace of the field data in Figure 9 (a) before and (b) after the application of the TSMF-based SSD, and (c) the normalized amplitude spectrum of the reflectivity derived from well logs near the location of the 270th trace.

tolerance to errors in the initial wavelet estimation even when the phase error is up to −90° or near to 90°.

ACKNOWLEDGMENTS

APPENDIX A

CONVERGENCE ANALYSIS

It is obvious that the value of the objective function in equation 9 monotonically decreases during the iterations of Algorithm 1 because the subproblems of equations 11 and 13 are convex, to which the optimal solution can be guaranteed. Considering that the objective function clearly has a lower bound of zero, we can simply conclude that the algorithm converges in the sense of the objective function. In this appendix, we will explore the convergence of the sequence generated by Algorithm 1.

the synthetic seismic data from the Marmousi2 model and a section of field seismic data demonstrate that the proposed method can effectively derive wavelet and reflectivity simultaneously from band-limited data with appropriate lateral coherence, even when the seismic data are contaminated by noise, and it has a favorable

THEOREM 1:

The sequence $\{\mathbf{A}^{(k)}, \mathbf{R}^{(k)}\}$ $(k = 1,2, \ldots)$ generated by Algorithm 1 converges, provided Ψ is coercive, i.e., $\lim_{\|\alpha\| \to +\infty} \Psi(\alpha) = +\infty$.

To prove this theorem, we first introduce some operators. Denote $\mathbf{z} = \{\alpha, \mathbf{r}_1, \ldots, \mathbf{r}_m\} \in \mathbb{R}^{n(m+1)}$, we define operator $\mathcal{T}_\alpha : \mathbb{R}^{n(m+1)} \to \mathbb{R}^{n(m+1)}$ as

$$\mathcal{T}_\alpha \mathbf{z} = \{\text{prox}_{\gamma_\alpha \Psi}(\alpha - \gamma_\alpha \tilde{\mathbf{R}}^T(\tilde{\mathbf{R}}\alpha - \tilde{\mathbf{y}})), \mathbf{r}_1, \ldots, \mathbf{r}_m\} \quad \text{(A-1)}$$

and $\mathcal{T}_j : \mathbb{R}^{n(m+1)} \to \mathbb{R}^{n(m+1)}$ as

$$\mathcal{T}_j \mathbf{z} = \{\alpha, \mathbf{r}_1, \ldots, \mathbf{r}_{j-1}, \text{prox}_{\gamma_j \lambda \|.\|_1}(\mathbf{r}_j$$
$$- \gamma_j \mathbf{A}^T(\mathbf{A}\mathbf{r}_j - \mathbf{y}_j)), \mathbf{r}_{j+1}, \ldots, \mathbf{r}_m\} \quad \text{(A-2)}$$

for $j = 1, \cdots, m$. Denote the sets of fixed points of $\mathcal{T}_\alpha$ and $\mathcal{T}_j(j = 1, \ldots, m)$ as $\text{Fix}(\mathcal{T}_\alpha) = \{\mathbf{z} \in \mathbb{R}^{n(m+1)} : \mathcal{T}_\alpha \mathbf{z} = \mathbf{z}\}$ and $\text{Fix}(\mathcal{T}_j) = \{\mathbf{z} \in \mathbb{R}^{n(m+1)} : \mathcal{T}_j \mathbf{z} = \mathbf{z}\}(j = 1, \ldots, m)$, respectively. Then, we have the following lemmas that are useful for proving Theorem 1.

LEMMA 1:

The operators defined in equations A-1 and A-2 are nonexpansive, i.e., for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{n(m+1)}$, it holds that

$$\|\mathcal{T}_\alpha \mathbf{z}_1 - \mathcal{T}_\alpha \mathbf{z}_2\| \leq \|\mathbf{z}_1 - \mathbf{z}_2\|, \quad \text{(A-3)}$$

$$\|\mathcal{T}_j \mathbf{z}_1 - \mathcal{T}_j \mathbf{z}_2\| \leq \|\mathbf{z}_1 - \mathbf{z}_2\|, \quad \text{(A-4)}$$

for $j = 1, \cdots, m$.

Proof:

Following Lemma 2.4 of Combettes and Wajs (2005), the proximity operator prox_g defined in equation 16 is nonexpansive. Consequently, $\text{prox}_{\gamma_\alpha \Psi}$ and $\text{prox}_{\gamma_j \lambda \|.\|_1}$ $(j = 1, \cdots, m)$ are nonexpansive. Then, it is easy to check that $\mathcal{T}_\alpha$ and $\mathcal{T}_j(j = 1, \cdots, m)$ are nonexpansive by definition.

LEMMA 2:

The following statements are true:

1) $\text{Fix}(\mathcal{T}_j) \neq \varnothing$ for $j = 1, \cdots, m$.
2) $\text{Fix}(\mathcal{T}_\alpha) \neq \varnothing$, provided Ψ is coercive.
3) $F = \text{Fix}(\mathcal{T}_\alpha) \bigcap (\bigcap_{j=1}^m \text{Fix}(\mathcal{T}_j))$, provided Ψ is coercive.

Proof:

1 and 2 follow immediately from Proposition 3.1 in Combettes and Wajs (2005).

It is obvious that the objective function in equation 9 is coercive and has a lower bound of zero. Therefore, the optimal solution to equation 9 exists. Assume $\mathbf{z}^* = \{\alpha^*, \mathbf{R}^*\} = \{\alpha^*, \mathbf{r}_1^*, \cdots, \mathbf{r}_m^*\}$ is one of such optimal solutions, it is easy to check that $\mathcal{T}_\alpha \mathbf{z}^* = \mathbf{z}^*$ and $\mathcal{T}_j \mathbf{z}^* = \mathbf{z}^*$ $(j = 1, \cdots, m)$. In fact, taking $\mathcal{T}_\alpha$ as an example, if $\mathcal{T}_\alpha \mathbf{z}^* \neq \mathbf{z}^*$, then $\alpha^* \neq \text{prox}_{\gamma_\alpha \Psi}(\alpha^* - \gamma_\alpha \tilde{\mathbf{R}}^{*T}(\tilde{\mathbf{R}}^* \alpha^* - \tilde{\mathbf{y}}))$, which implies that there exists an $\tilde{\alpha} \neq \alpha^*$, such that $\tilde{\alpha}$ gives lower value of the objective function in equation 13 than α^*, with $\mathbf{R}$ fixed to $\mathbf{R}^*$. Consequently, $\tilde{\mathbf{z}} = \{\tilde{\alpha}, \mathbf{r}_1^*, \cdots, \mathbf{r}_m^*\}$ gives lower value of the objective function in equation 9 than $\mathbf{z}^*$, which is a contradiction to the fact that $\mathbf{z}^*$ is an optimal solution to equation 9.

Therefore, we have found $\mathbf{z}^* \in F$, which proves 3. Now, we can give the proof of Theorem 1.

Proof of Theorem 1:

It is easy to check that the iterations in Algorithm 1 can be described by the product of the operators from $\{\mathcal{T}_\alpha, \mathcal{T}_1, \cdots, \mathcal{T}_m\}$

$$\mathbf{z}^{(k+1)} = \mathcal{T}_\alpha \cdots \mathcal{T}_\alpha \mathcal{T}_m \cdots \mathcal{T}_m \cdots \mathcal{T}_1 \cdots \mathcal{T}_1 \mathbf{z}^{(k)},$$
$$\text{(A-5)}$$

where $\mathbf{z}^{(k)} = \{\alpha^{(k)}, \mathbf{r}_1^{(k)}, \cdots, \mathbf{r}_m^{(k)}\}$, which is equivalent to $\{\mathbf{A}^{(k)}, \mathbf{R}^{(k)}\}$, noting that $\mathbf{r}_j^{(k)}$ is the jth column of $\mathbf{R}^{(k)}$, and $\alpha^{(k)}$ represents the free parameters of $\mathbf{A}^{(k)}$.

Then, by Lemmas 1 and 2, it suffices show that $\mathbf{z}^{(k)}$ converges to a point $\mathbf{z}^* \in F$ by applying Corollary 1.1 in Bruck (1982).

Though the sequence generated by Algorithm 1 converges according to Theorem 1, we cannot guarantee it converges to a local minimum of equation 9. However, it is easy to verify that the limit, denoted as $\mathbf{z}^* = \{\alpha^*, \mathbf{r}_1^*, \cdots, \mathbf{r}_m^*\}$, is a coordinatewise minimum point of the objective function, i.e., denote

$$F(\mathbf{z}^*) = \left\| \mathbf{Y} - \left(\sum a_k \mathbf{I}_k \right) \mathbf{R} \right\|_F^2 + \Psi(\alpha) + \lambda \sum_{j=1}^m \|\mathbf{r}_j\|_1,$$
$$\text{(A-6)}$$

we have

$$F(\mathbf{z}^* + (0, \cdots, 0, d_j, 0, \cdots, 0)) \leq F(\mathbf{z}^*), \quad \text{(A-7)}$$

where $d_j(j = 1, \cdots, m)$ correspond to coordinate block $\alpha^*, \mathbf{r}_1^*, \cdots, \mathbf{r}_m^*$, respectively because $\mathbf{z}^*$ is a common fixed point of $\mathcal{T}_\alpha$ and $\mathcal{T}_j(j = 1, \ldots, m)$.

REFERENCES

Aharon, M., M. Elad, and A. Bruckstein, 2006, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation: IEEE Transactions on Signal Processing, **54**, 4311–4322, doi: 10.1109/TSP.2006.881199.

Bach, F. R., R. Jenatton, J. Mairal, and G. Obozinski, 2012, Optimization with sparsity-inducing penalties: Foundations and Trends in Machine Learning, **4**, 1–106, doi: 10.1561/2200000015.

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.

Bregler, C., A. Hertzmann, and H. Biermann, 2000, Recovering nonrigid 3D shape from image streams: IEEE Conference on Computer Vision and Pattern Recognition, **2**, 690–696, doi: 10.1109/CVPR.2000.854941.

Bruck, R. E., 1982, Random products of contractions in metric and Banach spaces: Journal of Mathematical Analysis and Applications, **88**, 319–332, doi: 10.1016/0022-247X(82)90195-0.

Cichocki, A., R. Zdunek, A. H. Phan, and S.-I. Amari, 2009, Non-negative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation: Wiley.

Combettes, P. L., and V. R. Wajs, 2005, Signal recovery by proximal forward-backward splitting: Multiscale Modeling and Simulation, **4**, 1168–1200, doi: 10.1137/050626090.

Debeye, H. W. J., and P. van Riel, 1990, L_p-norm deconvolution: Geophysical Prospecting, **38**, 381–403, doi: 10.1111/j.1365-2478.1990.tb01852.x.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004, Least angle regression: The Annals of Statistics, **32**, 407–499, doi: 10.1214/009053604000000067.

Elad, M., and M. Aharon, 2006, Image denoising via sparse and redundant representations over learned dictionaries: IEEE Transactions on Image Processing, **15**, 3736–3745, doi: 10.1109/TIP.2006.881969.

Eriksson, A., and A. van den Hengel, 2010, Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L_1 norm: IEEE Conference on Computer Vision and Pattern Recognition, 771–778.

Wang et al.

Ge, D. D., X. Jiang, and Y. Ye, 2011, A note on the complexity of L_p minimization: Mathematical Programming, Series B, **129**, 285–299, doi: 10.1007/s10107-011-0470-2.

Golub, G. H., and C. F. Van Loan, 2012, Matrix computations (3rd ed.): The Johns Hopkins University Press.

Guitton, A., and J. Claerbout, 2015, Nonminimum phase deconvolution in the log domain: A sparse inversion approach: Geophysics, **80**, no. 6, WD11–WD18, doi: 10.1190/geo2015-0016.1.

Hale, E. T., W. Yin, and Y. Zhang, 2007, A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing: Rice University, CAAM Technical Report 07–07.

Hoefling, H., 2010, A path algorithm for the fused lasso signal approximator: Journal of Computational and Graphical Statistics, **19**, 984–1006, doi: 10.1198/jcgs.2010.09208.

Hyvärinen, A., and E. Oja, 2000, Independent component analysis: Algorithms and applications: Neural Networks, **13**, 411–430, doi: 10.1016/S0893-6080(00)00026-5.

Kaaresen, K. F., and T. Taxt, 1998, Multichannel blind deconvolution of seismic signals: Geophysics, **63**, 2093–2107, doi: 10.1190/1.1444503.

Kazemi, N., and M. D. Sacchi, 2014, Sparse multichannel blind deconvolution: Geophysics, **79**, no. 5, V143–V152, doi: 10.1190/geo2013-0465.1.

Kibangou, A. Y., and G. Favier, 2007, Toeplitz–Vandermonde matrix factorization with application to parameter estimation of Wiener–Hammerstein systems: IEEE Signal Processing Letters, **14**, 141–144, doi: 10.1109/LSP.2006.882101.

Kim, S. J., K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, 2007, An interior-point method for large-scale l_1-regularized least squares: IEEE Journal of Selected Topics in Signal Processing, **1**, 606–617, doi: 10.1109/JSTSP.2007.910971.

Kormylo, J., and J. Mendel, 1978, On maximum-likelihood detection and estimation of reflection coefficients: 48th Annual International Meeting, SEG, Expanded Abstracts, 45–46.

Latimer, R. B., R. Davidson, and P. van Riel, 2000, An interpreter's guide to understanding and working with seismic-derived acoustic impedance data: The Leading Edge, **19**, 242–256, doi: 10.1190/1.1438580.

Lee, D. D., and H. S. Seung, 1999, Learning the parts of objects by non-negative matrix factorization: Nature, **401**, 788–791, doi: 10.1038/44565.

Lim, Y., and Y. Teh, 2007, Variational Bayesian approach to movie rating prediction: KDD Cup and Workshop, 15–21.

Liu, J., L. Yuan, and J. Ye, 2010, An efficient algorithm for a class of fused lasso problems: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 323–332.

Lu, W. K., 2009, Frequency recovery of band-limited seismic data based on sparse spike train deconvolution and lateral coherence constraint: CPS/SEG Beijing 2009 International Geophysical Conference and Exposition, 117, doi: 10.1190/1.3603642.

Mairal, J., 2012, SPAMS: A SPArse Modeling Software, v2.3, http://spams-devel.gforge.inria.fr/index.html, accessed 23 May 2012.

Nguyen, T. H., 2008, High resolution seismic reflectivity inversion: Ph.D. thesis, University of Houston.

Nose-Filho, K., A. K. Takahata, R. Lopes, and J. M. T. Romano, 2016, A fast algorithm for sparse multichannel blind deconvolution: Geophysics, **81**, no. 1, V7–V16, doi: 10.1190/geo2015-0069.1.

Okatani, T., and K. Deguchi, 2007, On the Wiberg algorithm for matrix factorization in the presence of missing components: International Journal of Computer Vision, **72**, 329–337, doi: 10.1007/s11263-006-9785-5.

Oldenburg, D. W., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms: Geophysics, **48**, 1318–1337, doi: 10.1190/1.1441413.

Porteous, I., A. U. Asuncion, and M. Welling, 2010, Bayesian matrix factorization with side information and Dirichlet process mixtures: Association for the Advancement of Artificial Intelligence.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency domain constraints: Geophysics, **59**, 938–945, doi: 10.1190/1.1443653.

Salakhutdinov, R., and A. Mnih, 2008, Probabilistic matrix factorization: Advances in Neural Information Processing Systems 20 (NIPS'07), 1257–1264.

Srebro, N., and T. Jaakkola, 2003, Weighted low-rank approximations: Proceedings of the 20th International Conference on Machine Learning, 720–727.

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

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight, 2005, Sparsity and smoothness via the fused lasso: Journal of the Royal Statistical Society: Series B (Statistical Methodology), **67**, 91–108, doi: 10.1111/j.1467-9868.2005.00490.x.

Tomasi, C., and T. Kanade, 1992, Shape and motion from image streams under orthography: A factorization method: International Journal of Computer Vision, **9**, 137–154, doi: 10.1007/BF00129684.

Tseng, P., 2001, Convergence of a block coordinate descent method for non-differentiable minimization: Journal of Optimization Theory and Applications, **109**, 475–494, doi: 10.1023/A:1017501703105.

Velis, D. R., 2008, Stochastic sparse-spike deconvolution: Geophysics, **73**, no. 1, R1–R9, doi: 10.1190/1.2790584.

Wang, J., X. Wang, and M. Perz, 2006, Structure preserving regularization for sparse deconvolution: 76th Annual International Meeting, SEG, Expanded Abstracts, 2072–2076.

Wang, Y. F., 2011, Seismic impedance inversion using l_1-norm regularization and gradient descent methods: Journal of Inverse and Ill-Posed Problems, **18**, 823–838, doi: 10.1515/jiip.2011.005.

Xu, Y., and W. Yin, 2013, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion: SIAM Journal on Imaging Sciences, **6**, 1758–1789, doi: 10.1137/120887795.

Xu, Z. B., X. Y. Chang, and F. M. Xu, 2012, $L_{1/2}$ regularization: A thresholding representation theory and a fast solver: IEEE Transactions on Neural Networks and Learning Systems, **23**, 1013–1027, doi: 10.1109/TNNLS.2012.2197412.

Ye, G. B., and X. Xie, 2011, Split Bregman method for large scale fused lasso: Computational Statistics and Data Analysis, **55**, 1552–1569, doi: 10.1016/j.csda.2010.10.021.