Geophysical Journal International

Geophys. J. Int. (2017) **209,** 21–31
Advance Access publication 2017 January 11
GJI Marine geosciences and applied geophysics

doi: 10.1093/gji/ggw492

Fast dictionary learning for noise attenuation of multidimensional seismic data

Yangkang Chen*

National Center for Computational Sciences, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN 37831-6008, *USA.*
E-mail: chenyk2016@gmail.com

Accepted 2016 December 30. Received 2016 December 23; in original form 2016 November 7

SUMMARY

The K-SVD algorithm has been successfully utilized for ...ive...arning the sparse dictionary in 2-D seismic denoising. Because of the high co...put... cost o...ny singular value decompositions (SVDs) in the K-SVD algorithm... not ap... practical situations, especially in 3-D or 5-D problems. In this paper... nd the dicti... learning based denoising approach from 2-D to 3-D. To address t... tional efficiency problem in K-SVD, I propose a fast dictionary learning approac... based... sequential generalized K-means (SGK) algorithm for denoising multi... ional seis... ta. The SGK algorithm updates each dictionary atom by taking ar... metic average of several training signals instead of calculating an SVD as used in K-... algorith... I summarize the sparse dictionary learning algorithm using K-SVD, and intro... SGK al... hm together with its detailed mathematical implications. 3-D synthetic, 2-D a... -D f... data examples are used to demonstrate the performance of both K... and SGK... ms. It has been shown that SGK algorithm can significantly increase ... nal efficiency while only slightly degrading the denoising performance.

Key words: ...e pro... ; Inverse theory; Joint inversion; Time-series analysis; Computational ...gy; Se... noise.

1 INTRODUCTION

Seismic data are inevitably corru... y random n... field acquisition, with important cons... for oil and gas ...ploration. Thus, random noise attenuat... ays ...mental role in seismic data processing and inte...tation (Gu... 00; Qu *et al.* 2015; Zhuang *et al.* 2015; C... *et al.* 2016d; Li ... 2016a,b). Over the past few decades ...ge n... er of denoising methods for random noise have ...lev... ed. Pr...diction based methods utilize the predictable pr... useful ...gnals to construct prediction filters for ...cing ... ...jecting noise, for example, t-x predic... g (Abi... aerbout 1995), f-x deconvolution (C... s 198... he forward–backward prediction approach (Wang 19... ...g based approach (Liu *et al.* 2011), non-statio... edictive ...tering (Liu *et al.* 2012; Liu & Chen 2013). Mean a... an filters utilize the statistical difference between signal and ... to reject the Gaussian white noise or impulsive noise (Liu *et al.* 2009b; Liu 2013; Gan *et al.* 2016c). Decomposition based approaches decompose the noisy seismic data into different components and then select the principal components to represent the useful signals. Empirical mode decomposition and its variations (Huang *et al.* 1998), singular value decomposition based approaches (Bekara & van der Baan 2007; Chen & Ma 2014; Gan *et al.* (2015a), regularized non-stationary decomposition based approaches (Fomel 2013) are usually used to extract the useful components in multidimensional seismic data. Rank-reduction based approaches assume the seismic data to be low-rank after some data rearrangement steps, such methods include the Cadzow filtering (Trickett 2008), principal component analysis (Huang *et al.* 2016b), singular spectrum analysis (Oropeza & Sacchi 2011; Huang *et al.* 2017), damped singular spectrum analysis (Huang *et al.* 2016a; Zhang *et al.* 2016; Chen *et al.* 2016b,c).

The sparse transform based random noise attenuation is one of the most widely used approaches (Zhang *et al.* 2015; Chen 2016). Not only in seismic data processing, but also in all image processing fields, transformed domain thresholding approach has achieved very successful performance (Protter & Elad 2009; Cai *et al.* 2014). The denoising step can be implemented by simply applying a thresholding operator in the transformed sparse domain, followed by an inverse sparse transform. Sparse transform can be divided into two types: analytical transform, which has an exact basis, and learning-based dictionary, which iteratively updates the basis by learning (Chen *et al.* 2016a). I will use *transform* and *dictionary* to refer to these two types of sparse transform, respectively, in this paper.

A lot of transforms have been used in denoising seismic data. Gao *et al.* (2006) used the wavelet transform to denoise pre-stack seismic data. Wang *et al.* (2008) used the second-generation wavelet

* Previously at: Bureau of Economic Geology, John A. and Katherine G. Jackson School of Geosciences, The University of Texas at Austin, University Station, Austin, TX 78713-8924, USA.

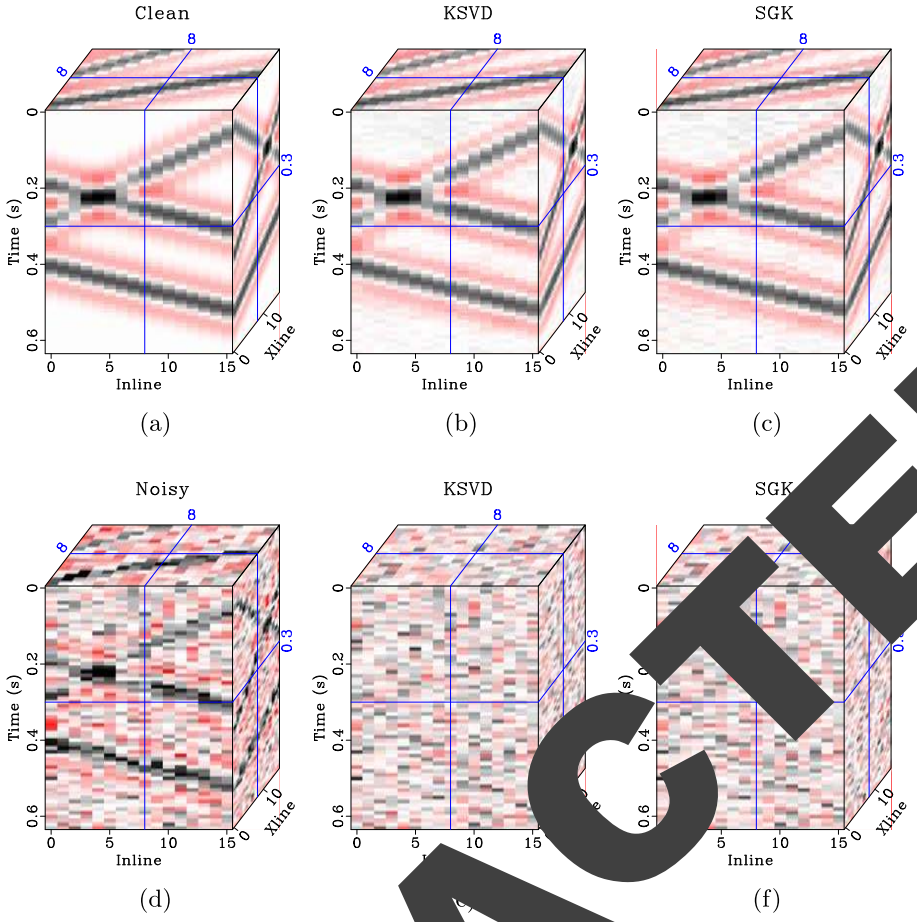


Figure 1. 3-D synthetic example. (a) Clean data. (b) Denoised data using [illegible]VD. [illegible]sed data using SGK. (d) Noisy data. (e) Noise using K-SVD. (f) Noise using SGK.

transform, which is based on the lifting scheme, t[illegible]oise[illegible]mic data with a percentile thresholding strategy. [illegible]fe[illegible] rmann (2006) and Neelamani *et al.* (200[illegible] app[illegible]curv[illegible] transform to attenuate both random and [illegible]rent no[illegible]eismic data. Zu *et al.* (2016) applied the c[illegible]t transform [illegible]rate simultaneous sources based on th[illegible] soft-thresho[illegible]ng algorithm. Fomel & Liu (2010) de[illegible]ied a[illegible] transform that is tailored specifically for seis[illegible]data, which [illegible]d seislet transform, for sparse represe[illegible]n based processi[illegible]of seismic data, including seismic den[illegible]g (C[illegible]2016; Wu *et al.* 2016), seismic deblending (Chen *et[illegible]et al.*[illegible]16b), and data restoration (Gan *et al.* 2015b, 201[illegible]*et al.*[illegible]). Chen & Fomel (2015a) used the ada[illegible]aratio[illegible]s of empirical mode decomposition [illegible]D; [illegible]g *et al.*[illegible]) for preparing the stable input for the [illegible]stati[illegible]D seislet transform and proposed a new EMD-s[illegible]noise seismic data with strong spatial heterogen[illegible]cently, Kong & Peng (2015) applied the shearlet transform to [illegible]random noise attenuation.

The learning[illegible]ed dictionaries are becoming more and more popular for seismic data processing in recent years since their superior performances in adaptively learning the basis that can sparsely represent the complicated seismic data (Sahoo & Makur 2013). Kaplan *et al.* (2009) used a data-driven sparse-coding algorithm to adaptively learn basis functions in order to sparsely represent seismic data and then perform denoising in the transformed domain. Based on a variational sparse-representation model, Beckouche & Ma (2014) proposed a denoising approach by adaptively learning dictionaries from noisy seismic data. Chen *et al.* (2016a) combined the learning based dictionaries and the fixed-basis transforms and proposed a double-sparsity dictionary to better handle the special features of seismic data, which can separate signals and noise more precisely.

K-SVD is one of the most effective dictionary learning algorithms (Aharon *et al.* 2006). However, the computational cost which requires thousands of singular value decomposition (SVD) hinders its wide application in seismic data processing, especially in practical 3-D or 5-D problems. In this paper, I propose to apply a fast dictionary learning algorithm, which is called sequential generalized K-means (SGK) algorithm (Sahoo & Makur 2013), to denoise multidimensional seismic data. Since sparse code is relatively new to the seismic community, I introduce the basic formulation of a sparse representation problem and mathematically analyse the principle of K-SVD algorithm and clarify its computational bottleneck. Then, I also introduce the SGK algorithm in detail and apply both K-SVD and SGK algorithms to denoise multidimensional seismic data. Three examples show that the SGK algorithm can significantly accelerate the dictionary learning process and cause no observably worse denoising performance.

2 METHOD

2.1 Problem formulation

Sparse representation via learning based dictionary consists of two main steps.

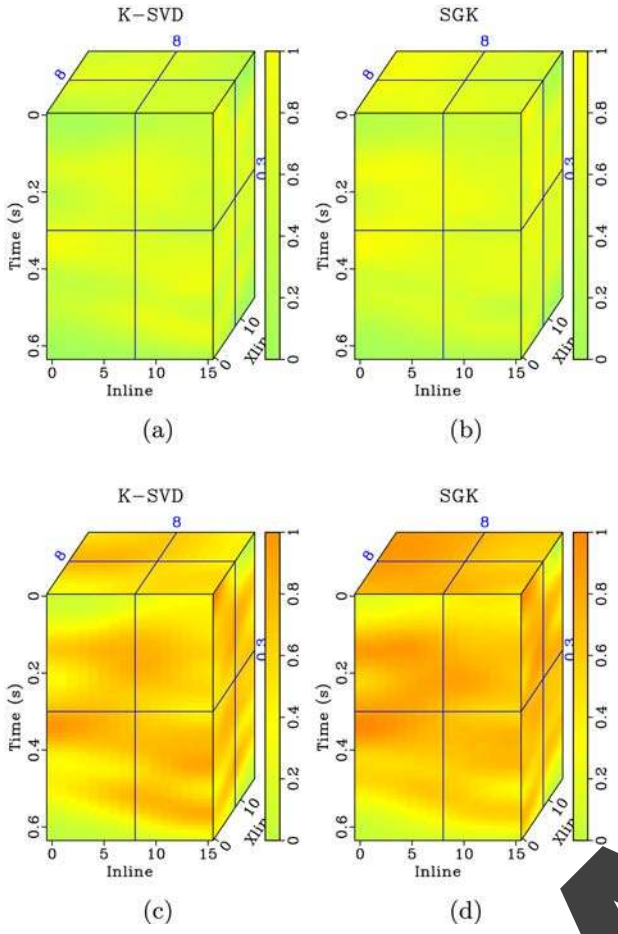


Figure 2. (a,b) Local similarity between denoised data and removed n… using K-SVD and SGK. (c,d) Amplified local similarity … between … noised data and removed noise using K-SVD and SG…

(i) *Sparse coding.* Given the obse… ata $\mathbf{d}$, … oding aims at solving the optimization probl…

$$\mathbf{m}^n = \arg\min_{\mathbf{m}} \| \mathbf{d} - \mathbf{F}^n\mathbf{m} \|_2^2 \; \| \; T, \tag{1}$$

where $\| \cdot \|_2$ and $\| \cdot \|$ … note the L_2 a… norms of an input vector, respectively … the n…mber of non-zero coefficients. $\mathbf{F}$ is the learned … nary … $\mathbf{m}$ is the sparse representation of $\mathbf{d}$.

(ii) *Dictionary* … For th… tained $\mathbf{m}^n$, update $\mathbf{F}^n$ such that

$$\mathbf{F}^{n+1} \; \| \mathbf{d} - \mathbf{Fm} \| \tag{2}$$

E… d (2) … iterated *Niter* times to learn the optimal diction… the sparest representation.

The mu… nsional seismic data is first reformulated into patch form $\mathbf{D}$. Each … olumn vector in $\mathbf{D}$ is extracted from the multidimensional seismic data matrix. An example is given in Yu *et al.* (2015) and Chen *et al.* (2016a). Eqs (1) and (2) then become

$$\forall_i \mathbf{m}_i^n = \arg\min_{\mathbf{m}_i} \| \mathbf{D} - \mathbf{F}^n\mathbf{M} \|_F^2, \text{ s.t.} \forall_i \| \mathbf{m}_i \|_0 \le T, \tag{3}$$

$$\mathbf{F}^{n+1} = \arg\min_{\mathbf{F}} \| \mathbf{D} - \mathbf{FM}^n \|_F^2, \tag{4}$$

where $\| \cdot \|_F$ denotes the Frobenius norm of an input matrix.

Problem (3) is an NP-hard problem, and directly finding the truly optimal $\mathbf{M}$ is impossible and is usually solved by an approximation pursuit method, such as the orthogonal matching pursuit (OMP) algorithm. To solve problem (4) for the adaptive dictionary $\mathbf{F}$, there are several different algorithms.

2.2 Dictionary learning by K-SVD

The K-SVD method (Aharon *et al.* 2006) is one approach that solves eq. (4) with good performance. The dictionaries in $\mathbf{F}$ are not obtained at a time. Instead, K columns in $\mathbf{F}$ are … ted one by one while fixing $\mathbf{M}$. In order to update the kth … can first write the objective function in eq. (4) as

$$\| \mathbf{D} - \mathbf{FM} \|_F^2 = \| \mathbf{D} - \sum_{j=1}^{K} \mathbf{f}_j \mathbf{m}_T^j \|$$

$$= \| \mathbf{D} - \sum_{j \neq} \mathbf{f}_j \mathbf{m}_T^j \|_F^2$$

$$= \| \mathbf{E}_k - \mathbf{f}_k \mathbf{m}_T^k \| \tag{5}$$

where $\mathbf{f}_j$ is t… nn vector in …, $\mathbf{m}_T^j$ is the jth row vector in $\mathbf{M}$.

Here … simply … a row vector. For simplicity, in eq… e following d… ations, I omit the superscript n shown i… (4). $\mathbf{E}_k$ is the fitting error using all column vectors other than kth diction… d their corresponding coefficients row vectors. … at in eq… $\mathbf{D}$ and $\mathbf{E}_k$ are of size $M \times N$, $\mathbf{F}$ is of size $M \times K$, a… … $\times N$.

Her… e length of each training signal, N is the number of …ining signals, and K is the number of atoms in the dictionary.

…ow obvious that the kth dictionary in $\mathbf{F}$ is updated by mi…izing the misfit between the rank-1 approximation of $\mathbf{f}_k\mathbf{m}_T^k$ and the $\mathbf{E}_k$ term. The rank-1 approximation is then solved using the SVD.

A problem in the direct use of SVD for rank-1 approximation of $\mathbf{E}_k$ is the loss of sparsity in $\mathbf{m}_T^k$. After SVD on $\mathbf{E}_k$, $\mathbf{m}_T^k$ is likely to be filled. In order to solve such problem, K-SVD restricts minimization of eq. (5) to a small set of training signals $\mathbf{D}_k = \{\mathbf{d}_i : \mathbf{m}_T^k(i) \neq 0\}$. To achieve this goal, one can define a transformation matrix $\mathbf{R}_k$ to shrink $\mathbf{E}_k$ and $\mathbf{m}_T^k$ by rejecting the zero columns in $\mathbf{E}_k$ and zero entries in $\mathbf{m}_T^k$. First one can define a set $r_k = \{i | 1 \le i \le N, \mathbf{m}_T^k(i) \neq 0\}$, which selects the entries in $\mathbf{m}_T^k$ that are non-zero. One then constructs $\mathbf{R}_k$ as a matrix of size $N \times N_r^k$ with ones on the $(r_k(i), i)$ entries and zeros otherwise.

Applying $\mathbf{R}_k$ to both $\mathbf{E}_k$ and $\mathbf{m}_T^k$, then the objective function in eq. (5) becomes

$$\| \mathbf{E}_k\mathbf{R}_k - \mathbf{f}_k\mathbf{m}_T^k\mathbf{R}_k \|_F^2 = \| \mathbf{E}_k^R - \mathbf{f}_k\mathbf{m}_R^k \|_F^2 \tag{6}$$

and can be minimized by directly using SVD:

$$\mathbf{E}_k^R = \mathbf{U}\Sigma\mathbf{V}^T. \tag{7}$$

$\mathbf{f}_k$ is then set as the first column in $\mathbf{U}$, the coefficient vector $\mathbf{m}_R^k$ as the first column of $\mathbf{V}$ multiplied by first diagonal entry σ_1. After K columns are all updated, one turns to solve eq. (3) for $\mathbf{M}$.

2.3 Fast dictionary learning by SGK

Although K-SVD can obtain very successful performance in a number of sparse representation based approaches, since there involves many SVD operations in the K-SVD algorithm, it is very

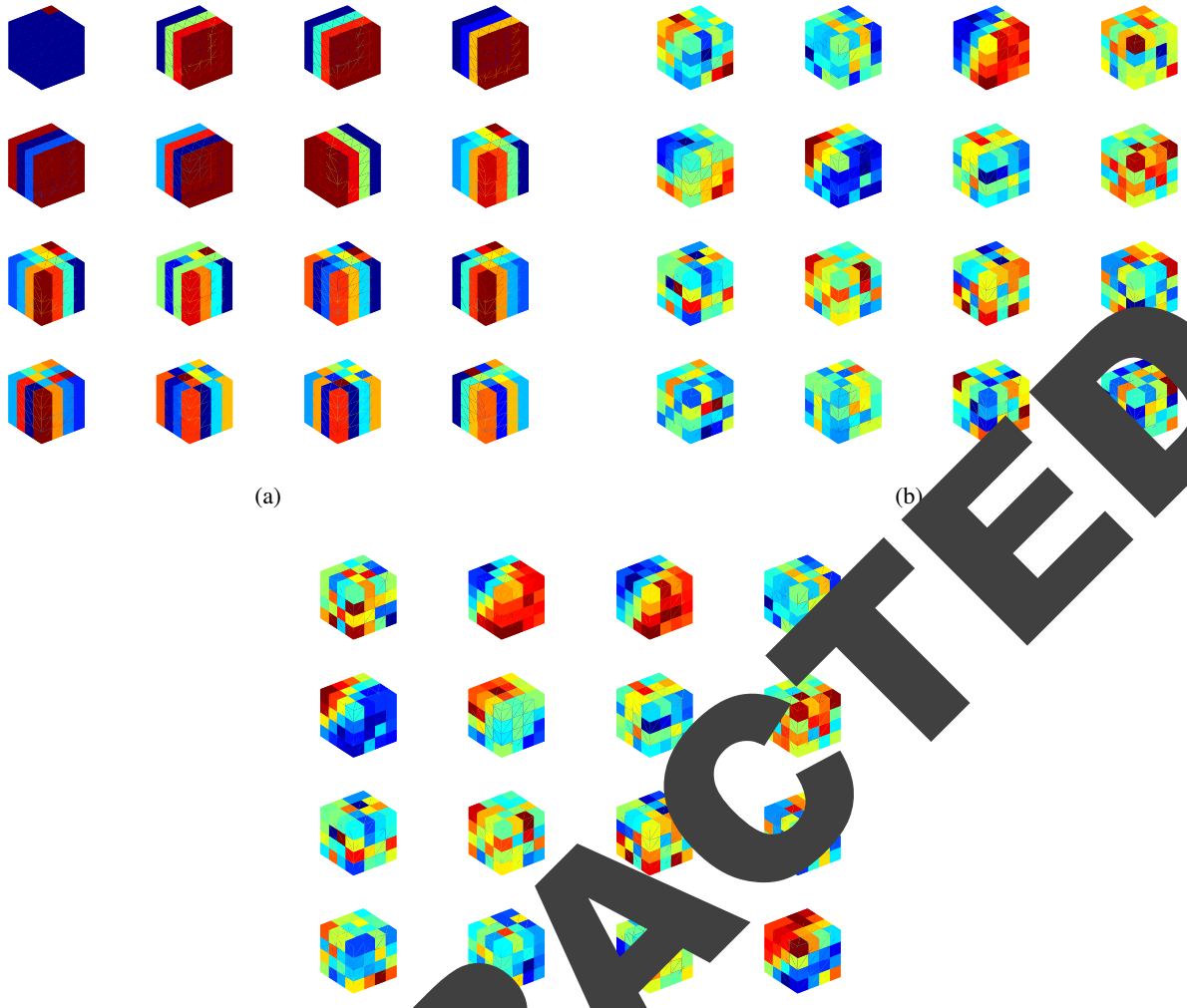


Figure 3. (a) Initial overcomplete DCT diction… (b) …diction… using K-SVD. (c) Learned dictionary using SGK. Only the first 16 atoms in the dictionary are displayed and each dictionary l… reshap… 3-D cube. It can see from (a) that the shapes of the atoms in the initial dictionary (discrete cosine transform) are rigid, which are not… al to represen… hly non-stationary seismic data. After dictionary learning, the updated dictionaries as shown in (b) and (c) contain atoms wit… d shapes. The… various atoms are more likely to capture the non-stationary features hidden in the seismic data.

Table 1. Co… son … noising performances with differ… S… etwee… K-SVD and SGK methods.

In…	…dB)	…dB)	SGK (dB)
		10.60	10.60
		9.61	9.32
		2.17	2.08
	33	1.27	1.11

Table 2. Comparison of computing time with different model sizes between K-SVD and SGK methods.

Model sizes	K-SVD (s)	SGK (s)
$64 \times 16 \times 16$	192.60	9.27
$128 \times 32 \times 32$	893.50	48.56
$128 \times 64 \times 64$	3059.3	201.34

computationally expensive. Especially when utilized in multidimensional seismic data processing (e.g. 3-D or 5-D processing), the computational cost is not tolerable. The SGK algorithm was proposed to increase the computational efficiency (Sahoo & Makur 2013). SGK tries to solve slightly different iterative optimization problem in sparse coding as eq. (3):

$$\forall_i \mathbf{m}_i^n = \arg\min_{\mathbf{m}_i} \| \mathbf{D} - \mathbf{F}^n \mathbf{M} \|_F^2, \text{s.t.} \forall_i \mathbf{m}_i = \mathbf{e}_t. \tag{8}$$

t indicates that $\mathbf{m}_i$ has all 0s except 1 in the tth position. The dictionary updating in SGK algorithm is also different. In SGK, eq. (6) also holds. Instead of using SVD to minimize the objective function, which is computationally expensive, SGK turns to use least-squares method to solve the minimization problem. Taking the derivative of $J = \| \mathbf{E}_k^R - \mathbf{f}_k \mathbf{m}_R^k \|_F^2$ with respect to $\mathbf{f}_k$ and setting the result to 0 gives the following equation:

$$\frac{\partial J}{\partial \mathbf{f}_k} = -2 \left(\mathbf{E}_k^R - \mathbf{f}_k \mathbf{m}_R^k \right) \left(\mathbf{m}_R^k \right)^T = 0 \tag{9}$$

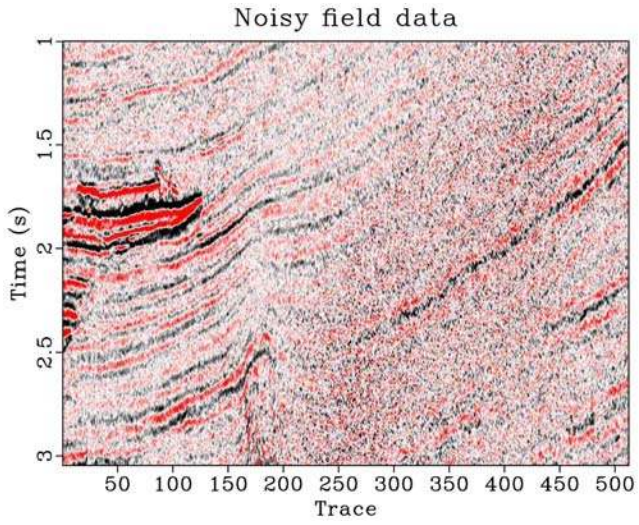


Figure 4. Noisy 2-D field data.

solving eq. (9) leads to

$$\mathbf{f}_k = \mathbf{E}_k^R \left(\mathbf{m}_R^k\right)^T \left(\mathbf{m}_R^k (\mathbf{m}_R^k)^T\right)^{-1}. \tag{10}$$

It can be derived further that

$$\begin{aligned}\mathbf{E}_k^R \left(\mathbf{m}_R^k\right)^T &= \left(\mathbf{D}_R - \sum_{j\neq k} \mathbf{f}_j \mathbf{m}_R^j\right) \left(\mathbf{m}_R^k\right)^T \\ &= \mathbf{D}_R \left(\mathbf{m}_R^k\right)^T + \sum_{j\neq k} \mathbf{f}_j \mathbf{m}_R^j \left(\mathbf{m}_R^k\right)^T.\end{aligned}$$

Here, $\mathbf{D}_R$ has the same meaning as $\mathbf{D}$ shown in eq. (5) except fo smaller size due to the selection set r_k that select s in $\mathbf{m}$ that are non-zero.

Since $\forall_i$, $\| \mathbf{m}_i \|_0 = 1$, as constrained in the

$$\forall_{j\neq k} \mathbf{m}_R^j \left(\mathbf{m}_R^k\right)^T = 0. \tag{12}$$

Since $\mathbf{m}_R^k$ is a smaller version of tor $\mathbf{m}_T^k$ and a ntries are all equal to 1, $\mathbf{D}_R(\mathbf{m}_R^k)^T$ is si tion over all the column vectors in $\mathbf{D}_R$. Considerin $\mathbf{D}_R =$ $\leq r_k\}$,

$$\mathbf{D}_R \left(\mathbf{m}_R^k\right)^T = \sum_{i\in r_k} \mathbf{d}_i \tag{13}$$

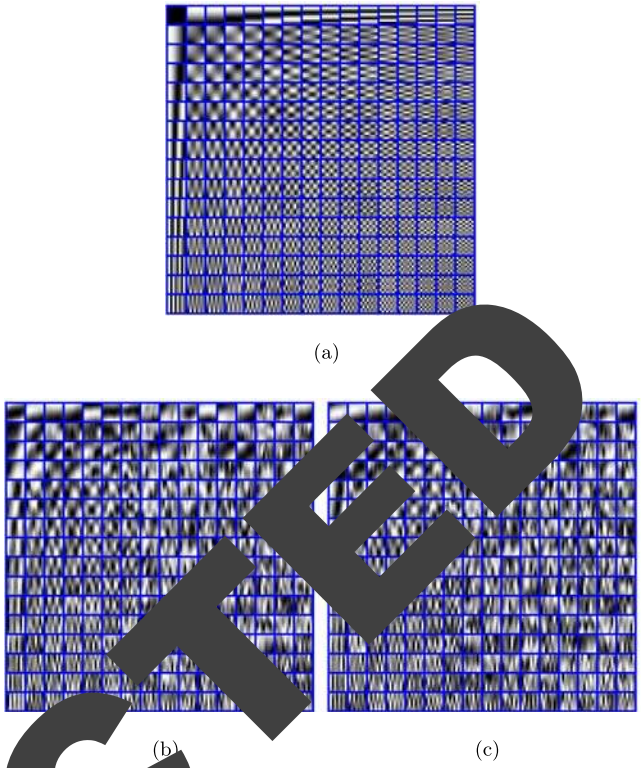
Following eq. (13 becor

$$\mathbf{E}_k^R \left(\mathbf{m}_R^k\right)^T \mathbf{d}_i. \tag{14}$$

It $\mathbf{m}_R^k \left(\mathbf{m}_R^k\right)^T = N_r^k$, where N_r^k denotes the numb ments in the set r_k, or the number of training signals associat he atom $\mathbf{f}_k$. The kth atom in $\mathbf{F}$ is

$$\mathbf{f}_k = \frac{\sum_{i\in r_k} \mathbf{d}_i}{N_r^k}. \tag{15}$$

Thus, in SGK, one can avoid the use of SVD. Instead the trained dictionary can be simply expressed as an average of several training signals. In this way, SGK can obtain significantly higher efficiency than K-SVD. In the next section, I will use several examples to show that the overall denoising performance does not degrade when one can obtain a much faster implementation.



e 5. (a) In vercomplete DCT dictionary. (b) Learned dictionar- i r K-SV Learned dictionaries using SGK. Each atom in the di reshaped into a 2-D matrix. As can be seen in either (b) or (re are some atoms in the middle part of the dictionary map taining linear patterns, indicating a better representation of the locally nts.

3 EXAMPLES

I will use three different examples to show the performance of SGK in denoising multidimensional seismic data. Please note that when using eqs (3) (or 8) and (4) for dictionary learning, the multidimensional seismic data is first mapped from the original form to a 2-D matrix according to some patching criteria. Some details about the patching method can be found in Yu *et al.* (2015) or Chen *et al.* (2016a). After iteratively solving eqs (3) (or 8) and (4) several times, the denoised data are expressed as

$$\hat{\mathbf{D}} = \mathbf{F}^{Niter} \mathbf{M}^{Niter}. \tag{16}$$

An inverse mapping is then applied to $\hat{\mathbf{D}}$ to output the finally denoised data.

For measuring the denoising performance of synthetic data examples, where one knows the clean data, I use the signal-to-noise ratio (SNR; Liu *et al.* 2009a; Huang *et al.* 2016a) measurement and the formula is expressed as follows:

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathbf{D}_{\mathrm{true}}\|_F^2}{\|\mathbf{D}_{\mathrm{true}} - \hat{\mathbf{D}}\|_F^2}, \tag{17}$$

where $\mathbf{D}_{\mathrm{true}}$ denotes the clean data and $\hat{\mathbf{D}}$ denotes the denoised data.

In addition to the commonly used SNR measurement, one can also use local similarity (Fomel 2007; Chen & Fomel 2015b) as a convenient tool to evaluate denoising performance. The abnormal area in the local similarity map with high similarity indicates the area that contains significant signal leakage in the removed noise.

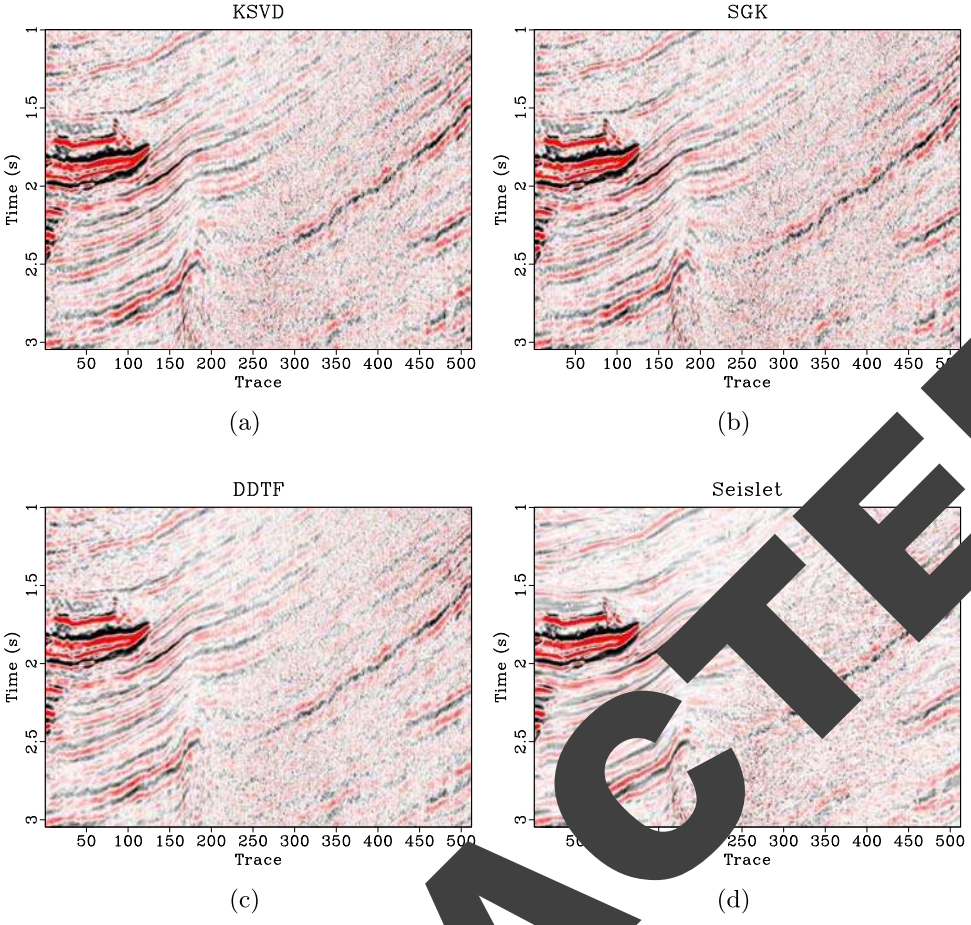


Figure 6. (a) Denoised data using K-SVD. (b) Denoised data using SGK. (c) Denoised data using DDTF. (d) Denoised data using seislet thresholding.

A local similarity map with values that are close to zero as the observed significant amount of removed noise provides valid support of a successful denoising performance. In addition, the local similarity measurement can be used in many cases since the input data of local similarity are just the denoised data and removed noise. It provides an alternative in the case of field data processing, where the clean data is unknown and the SNR based evaluation becomes unavailable. Besides, local similarity is a local measurement of denoising performance, whereas the SNR is a global measurement that does not give us to pick out the areas with poor denoising performances.

3.1 Synthetic

The first example is a 3-D synthetic example, as shown in Fig. 1. Fig. 1(a) is the clean data and Fig. 1(d) is the noisy data. Figs 1(b) and (c) show the denoised results using K-SVD and SGK, respectively. Figs 1(e) and (f) show the removed noise cubes of two approaches. It seems that the denoised results using both methods are very successful while the denoised result using SGK algorithm shows a little bit more residual noise, which is however negligible. The size of this example is $64 \times 16 \times 16$. I use a 3-D patch of size $4 \times 4 \times 4$ and the overlap between neighbour patches is 3 points in all time, inline, xline directions. In this example, the sample signals $\mathbf{D}$ is of size 64×10309. The K-SVD takes 192.60 s while SGK takes only 9.27 s. The SNRs of noisy data, K-SVD result and SGK result are 0.68, 9.61 and 9.32 dB respectively. While the SNRs using the K-SVD and SGK are very similar, the SGK method obtains about 20 times acceleration.

To further demonstrate the denoising performance and compare the two methods regarding the tiny differences, I plot the local similarity between denoised data and removed noise in Fig. 2. Figs 2(a) and (b) show the local similarity cubes without amplification that correspond to K-SVD and SGK methods, respectively. It can be seen that both methods cause negligible local similarity, or correlation, between denoised data and removed noise, confirming the extremely successful performance of the two methods. Figs 2(c) and (d) show the amplified local similarity cubes (similarity $\times$ 2) corresponding to K-SVD and SGK methods, respectively. It can be observed that there are some non-zero amplified similarity values around the events, indicating these tiny damages to the signal caused by both methods. It can also be observed that the amplified local similarity of SGK method is slightly higher than K-SVD method. Considering the slightly low SNR using SGK method, I conclude that SGK method might cause slightly worse performance than the K-SVD method while obtaining a huge improvement on computational efficiency. I also show some learned atoms of this example in Fig. 3. Figs 3(a)–(c) show the atoms from initial dictionary, K-SVD learned dictionary, and SGK learned dictionary, respectively. I set up the initial dictionary using the discrete cosine transform. It can be seen from Fig. 3(a) that the shapes of the atoms in the initial

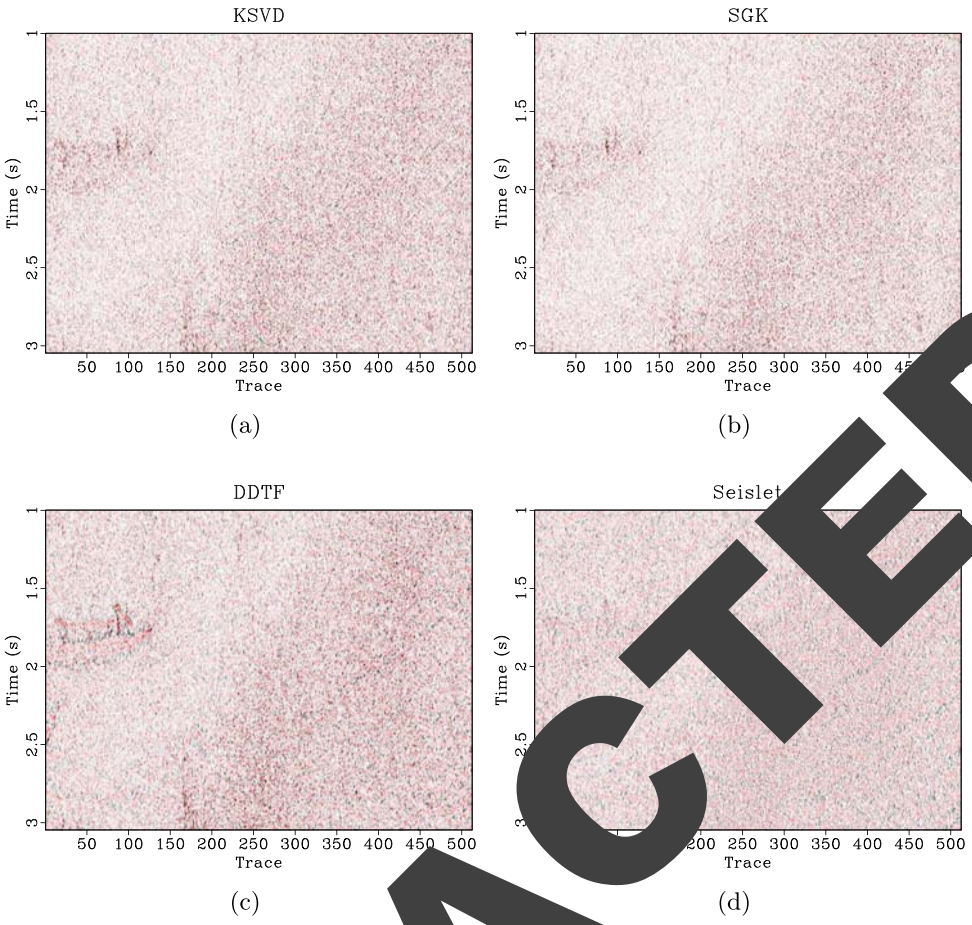


Figure 7. (a) Removed noise using K-SVD. (b) Removed noise using [illegible] (c) [illegible]oise using DDTF. (d) Removed noise using seislet thresholding.

dictionary are rigid, which are not opti[illegible]al t[illegible]t the [illegible]y non-stationary seismic data. After di[illegible]ary lea[illegible]he updated dictionaries as shown in Figs 3(b) [illegible]c) contain at[illegible]th much varied shapes. These various [illegible]ore likely to capture the non-stationary features in the [illegible]smic [illegible]d thus can potentially improve the sparse rep[illegible]tation of th[illegible]ved noisy seismic data. Please note tha[illegible] atom has been re[illegible]ped into a 3-D cube and only 16 atom[illegible] show[illegible]re.

In order to tes[illegible]g perf[illegible]nances of the two methods in different noise leve[illegible]culat[illegible]SNRs of two methods in four different [illegible]with i[illegible]oise levels (or decreasing input SNR[illegible]e [illegible]ising pe[illegible]nances of different input SNRs are co[illegible]ed in [illegible]. The SNR of input noisy data drops from 3.1[illegible]spondingly, the SNRs of the best results using [illegible]thods drop from above 10 to around 1 dB. Besides, the K-SVD [illegible] consistently obtains a slightly higher SNR than SGK metho[illegible]

In order to effectively compare the computational efficiency of two methods, I measure the computing time of two methods in different cases with increasing model sizes. The detailed comparison of computing time with different model sizes is shown in Table 2. The results show that the SGK method can consistently maintain a speedup of more than 15 times for different model sizes. In this paper, I compare the speed of SGK with the classic version of K-SVD algorithm (exact SVD calculation). Recently, there are a lot of new algorithms proposed to approximate SVD instead of exactly computing it, for example, Rubinstein *et al.* (2008), Foster *et al.* (2012) and Menon & Elkan (2011). These methods can hopefully improve the efficiency of K-SVD algorithm, but at the expense of slightly degrading the performance. Both K-SVD and SGK use the fast OMP algorithm for sparse coding in the whole dictionary learning process. In order to compare the computing time fairly, I repeat the same calculation three times for each model size and calculate the average time as the final measured computing time.

3.2 Field data example

I first use a 2-D field data example to compare the difference between K-SVD and SGK methods, as shown in Fig. 4. The data size of this example is 512 × 512. In this example, I choose a patch size of 8 × 8. The overlap between different patches is 7 points in both vertical and horizontal directions. Thus the atom size $M = 8 \times 8 = 64$, and the number of sample signals is thus $N = (512 - 7) \times (512 - 7) = 255\,025$. The size of $\mathbf{D}$ is $64 \times 255\,025$. For K-SVD, the dictionary updating process takes about 1590.2 s, while for SGK, the dictionary updating process takes only 87.23 s, which shows a great speedup. Fig. 5(a) shows the initial input dictionary. Figs 5(b) and (c) show the learned dictionaries using K-SVD

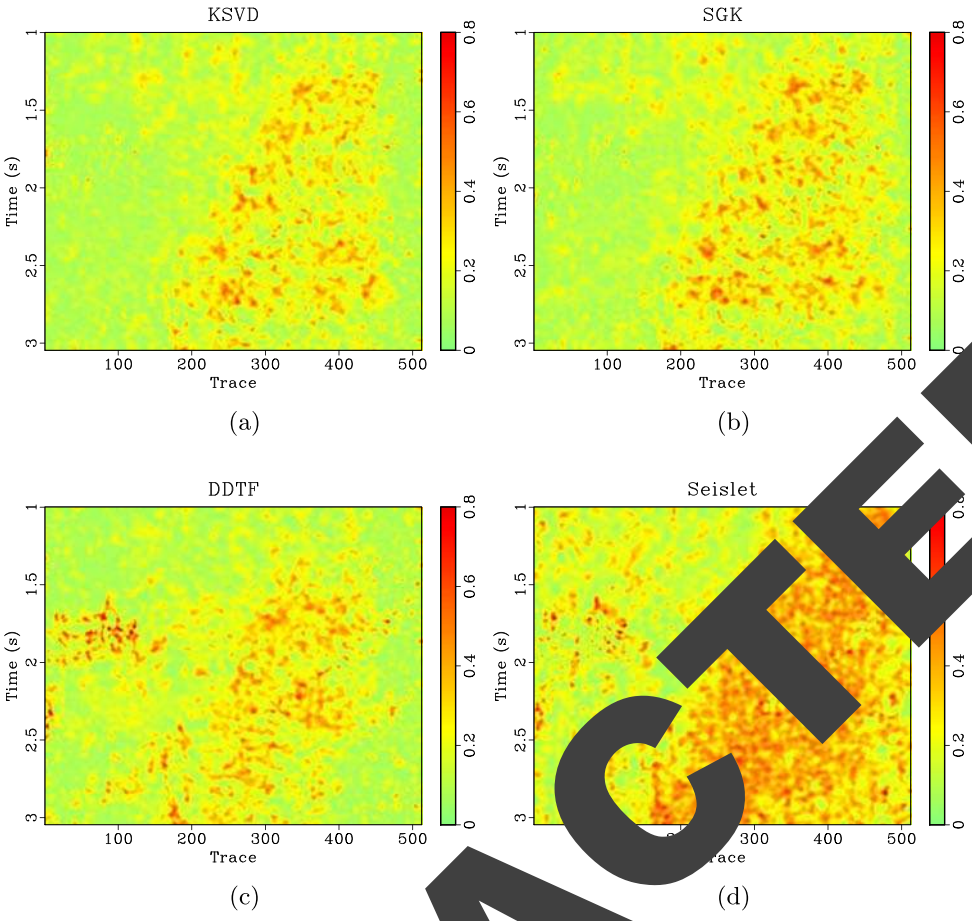


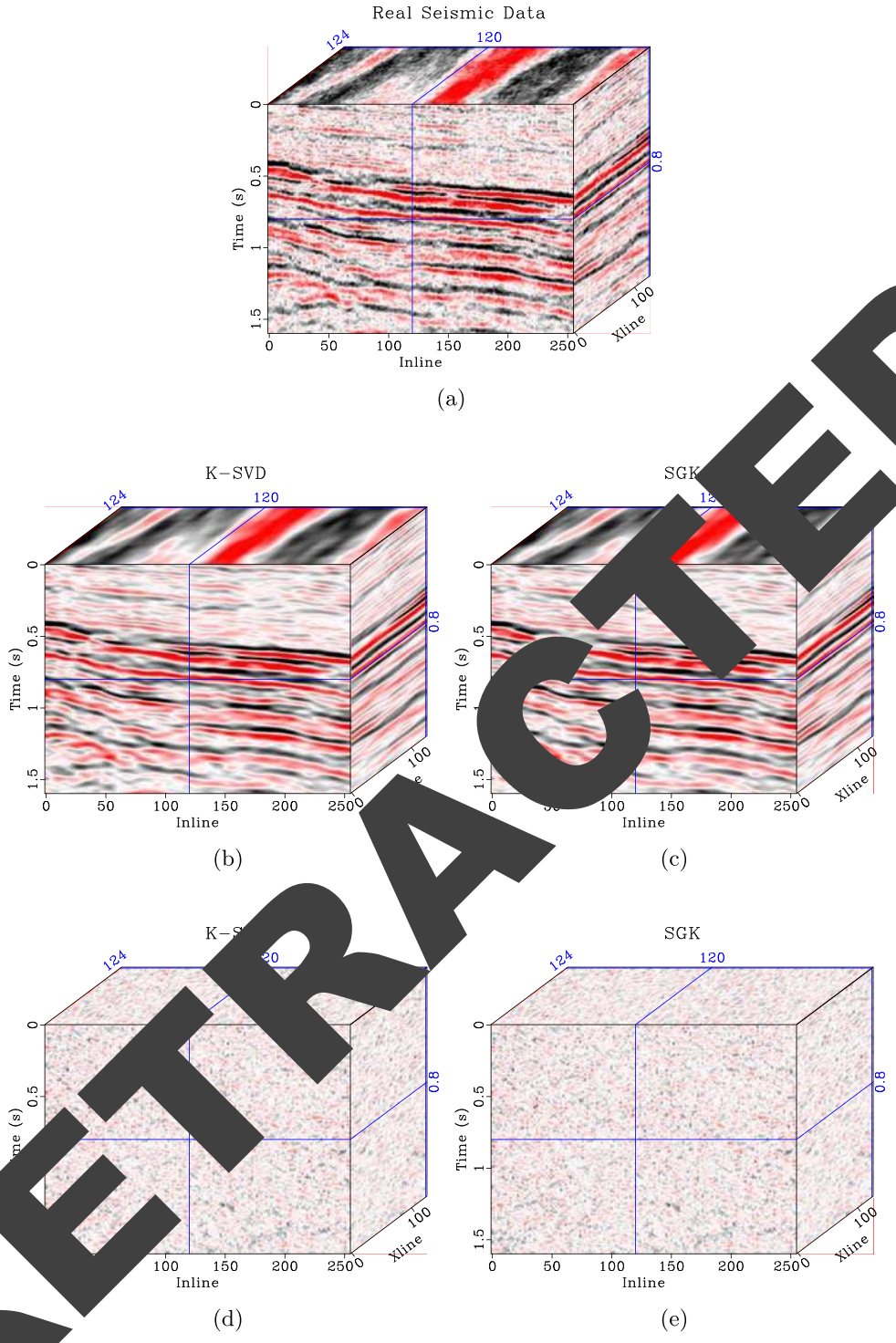
Figure 8. Local similarity between denoised data and removed noise using [illegible]-SV[illegible]K, (c) DDTF and (d) seislet thresholding.

and SGK, respectively. The two learned [illegible]tion[illegible]ow s[illegible] similarities but are not exactly the sa[illegible]s can [illegible] in either Fig. 5(b) or (c) that there are s[illegible]oms in the [illegible]part of the dictionary map containing [illegible]rns, indicating a better representation of the locally [illegible]ear e[illegible]n this example, I also compare the K-SVD a[illegible]GK methods [illegible]o other widely known methods, i.e. the [illegible]F method (Cai [illegible]. 2014) and the seislet transform met[illegible] (Fom[illegible] Liu 2010). The denoised results using four me[illegible]own in [illegible]g. 6. The corresponding noise sections are sho[illegible]. 7. [illegible]paring the results in both Figs 6 and [illegible] roug[illegible]e conclusions that K-SVD, SGK, an[illegible]hods a[illegible] to obtain successful denoised result[illegible] the [illegible] seislet transform is a bit over-smoothed, which [illegible]le low-frequency coherent energy in the noise [illegible] (Fig. 7d). A better evaluation of denoising performance can b[illegible]ed using the local similarity measurement and is shown in Fig[illegible] The local similarity confirms my observation in that the local similarity corresponding to seislet method is very high, which is followed by the DDTF method. The DDTF method obtains a successful performance in most part of the data but causes some damages to the highly curved signals around the 2s near the left boundary, as indicated from the local similarity map (Fig. 8c). The K-SVD and SGK methods obtain very close results but SGK results in a slightly higher local similarity in right part of the data.

I next use a 3-D field data example to demonstrate the performance, as shown in Fig. 9. Figs 9(a)–(c) show noisy data, K-SVD denoised data and SGK denoised data, respectively. Figs 9(d) and (e) show the noise sections of two approaches. It is clear that both approaches obtain approximate performance. It is computationally expensive to use K-SVD to learn the dictionary for this example. While it takes about half an hour to learn the dictionary using SGK algorithm, it takes more than half a day to learn the dictionary using the K-SVD algorithm. The local similarity cubes between denoised data cubes and removed noise cubes using two methods are shown in Fig. 10, which confirms the successful and comparable performance of both methods in that most part of the data is close to zero.

4 CONCLUSIONS

In this paper, I proposed a fast dictionary-learning based seismic denoising approach using the SGK algorithm. In the SGK algorithm, each atom in the dictionary is updated by an average of several sample signals while K-SVD uses computationally expensive SVD to update each atom. Thus, the SGK algorithm can be much faster than K-SVD algorithm for adaptively learning the dictionary. I applied both K-SVD and SGK to dictionary learning of seismic data for random noise attenuation. The results

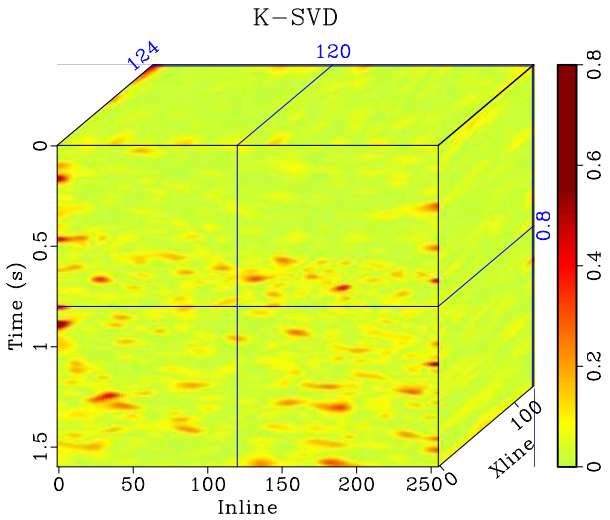


Figu Noisy [illegible]eld data. (b) & (c) Denoised data using K-SVD and SGK. (d) & (e) Noise cubes using K-SVD and SGK.

from three different examples show that SGK is much faster than K-SVD without sacrificing much denoising performance. I suggest substituting the K-SVD with SGK in any applications that require sparse coding. Future research direction may include applying the SGK based dictionary learning for multidimensional seismic data reconstruction.

ACKNOWLEDGEMENTS

I would like to thank editors Jorg Renner, Lapo Boschi and two anonymous reviewers for providing critical comments that improve the manuscript greatly. I also thank Shuwei Gan, Shaohuan Zu and Weilin Huang for helpful comments and suggestions on the topic



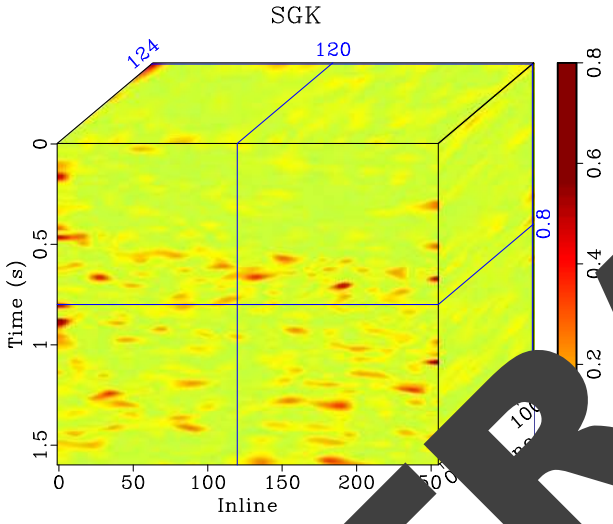


Figure 10. (a) Local similarity bet… dene… using K-SVD and the corresponding noise cube. (… local similari… en denoised data using SGK and the corresp… noise cube.

of seismic signal pro… d Jia… Ma, Sergey Fomel and Siwei Yu for in…ising d… ns … arse transform.

REF… N…

Abma, R. & … ut, J., 1995. Lateral prediction for noise attenuation by $t-x$ and f … iques, *Geophysics,* **60,** 1887–1896.

Aharon, M., Elad, … Bruckstein, A.M., 2006. The K-SVD: An algorithm for designing of evercomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.,* **54,** 4311–4322.

Beckouche, S. & Ma, J., 2014. Simultaneous dictionary learning and denoising for seismic data, *Geophysics,* **79,** A27–A31.

Bekara, M. & van der Baan, M., 2007. Local singular value decomposition for signal enhancement of seismic data, *Geophysics,* **72,** V59–V65.

Cai, J.-F., Ji, H., Shen, Z. & Ye, G.-B., 2014. Data-driven tight frame construction and image denoising, *Appl. Comput. Harmon. Anal.,* **37**(1), 89–105.

Canales, L., 1984. Random noise reduction, in 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

Chen, Y., 2016. Dip-separated structural filtering using seislet thresholding and adaptive empirical mode decomposition based dip filter, *Geophys. J. Int.,* **206**(1), 457–469.

Chen, Y. & Ma, J., 2014. Random noise attenuation by f-x empirical mode decomposition predictive filtering, *Geophysics,* **79,** V81–V91.

Chen, Y. & Fomel, S., 2015a. EMD-seislet transform, *85th Annual International Meeting*, SEG, Expanded Abstracts, 4775–4778.

Chen, Y. & Fomel, S., 2015b. Random noise attenuation using local signal-and-noise orthogonalization, *Geophysics,* **80,** WD1–WD9.

Chen, Y., Fomel, S. & Hu, J., 2014. Iterative deblending of simultaneous-source seismic data using seislet-domain shaping … n, *Geophysics,* **79**(5), V179–V189.

Chen, Y., Ma, J. & Fomel, S., 2016a. Double-spa… tionary … mic noise attenuation, *Geophysics,* **81,** V17–V30.

Chen, Y., Zhang, D., Huang, W. & Chen, W… 16b. A… our… tlab code package for improved rank-redu… D seismic … g and reconstruction, *Comput. Geosci.,* … –66.

Chen, Y., Zhang, D., Jin, Z., … X., … S., H…ang, W. & Gan, S., 2016c. Simultaneous …enoi… …recons…n of 5D seismic data via damped ra… …uctio… …d, *…ys. J. Int.,* **206**(3), 1695–1717.

Fomel, S., 2007. Loc… attributes, *Ge… sics,* **72,** A29–A33.

Fomel, S., 2013. S… mposition into spectral components using regularized nons…tionary … ession, *Geophysics,* **78,** O69–O76.

Fomel, S. & … X., 2010. Seis… orm and seislet frame, *Geophysics,* **75,** V…

Foste… Mahadevan, S. & Wang, R., 2012. A GPU-based approximate SV… gorithm, i… *llel Processing and Applied Mathematics, Book ch…* pp. 569–5… ds Wyrzykowski, R., Dongarra, J., Karczewski, K. &… …pringer.

Gan, S., … S., Qu, S. & Zhong, W., 2015a. Structure-oriented …ngular value decomposition for random noise attenuation of seismic …*eophys. Eng.,* **12,** 262–272.

G… …g, S., Chen, Y., Zhang, Y. & Jin, Z., 2015b. Dealiased seismic …ta interpolation using seislet transform with low-frequency constraint, *IEEE Geosci. Remote Sens. Lett.,* **12,** 2150–2154.

…n, S., Chen, Y., Wang, S., Chen, X., Huang, W. & Chen, H., 2016a. Compressive sensing for seismic data reconstruction using a fast projection onto convex sets algorithm based on the seislet transform, *J. Appl. Geophys.,* **130,** 194–208.

Gan, S., Wang, S., Chen, Y. & Chen, X., 2016b. Simultaneous-source separation using iterative seislet-frame thresholding, *IEEE Geosci. Remote Sens. Lett.,* **13**(2), 197–201.

Gan, S., Wang, S., Chen, Y., Chen, X. & Xiang, K., 2016c. Separation of simultaneous sources using a structural-oriented median filter in the flattened dimension, *Comput. Geosci.,* **86,** 46–54.

Gan, S., Wang, S., Chen, Y., Qu, S. & Zu, S., 2016d. Velocity analysis of simultaneous-source data using high-resolution semblance-coping with the strong noise, *Geophys. J. Int.,* **204,** 768–779.

Gao, J., Mao, J., Chen, W. & Zheng, Q., 2006. On the denoising method of prestack seismic data in wavelet domain, *Chin. J. Geophys.,* **49,** 1155–1163.

Gulunay, N., 2000. Noncausal spatial prediction filtering for random noise reduction on 3-D poststack data, *Geophysics,* **65,** 1641–1653.

Hennenfent, G. & Herrmann, F., 2006. Seismic denoising with nonunformly sampled curvelets, *Comput. Sci. Eng.,* **8,** 16–25.

Huang, N.E. *et al.*, 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. A,* **454,** 903–995.

Huang, W., Wang, R., Chen, Y., Li, H. & Gan, S., 2016a. Damped multichannel singular spectrum analysis for 3D random noise attenuation, *Geophysics,* **81,** V261–V270.

Huang, W., Wang, R., Zhou, Y., Chen, Y. & Yang, R., 2016b. Improved principal component analysis for 3D seismic data simultaneous reconstruction and denoising, in *86th Annual International Meeting,* SEG, Expanded Abstracts, 4777–4781.

Huang, W., Wang, R., Yuan, Y., Gan, S. & Chen, Y., 2017. Signal extraction using randomized-order multichannel singular spectrum analysis, *Geophysics,* **82**(2), V59–V74.

Kaplan, S.T., Sacchi, M.D. & Ulrych, T.J., 2009. Sparse coding for data-driven coherent and incoherent noise attenuation, in *79th Annual International Meeting,* SEG, Expanded Abstracts, 3327–3331.

Kong, D. & Peng, Z., 2015. Seismic random noise attenuation using shearlet and total generalized variation, *J. Geophys. Eng.,* **12,** 1024–1035.

Li, H., Wang, R., Cao, S., Chen, Y. & Huang, W., 2016a. A method for low-frequency noise suppression based on mathematical morphology in microseismic monitoring, *Geophysics,* **81**(3), V159–V167.

Li, H., Wang, R., Cao, S., Chen, Y., Tian, N. & Chen, X., 2016b. Weak signal detection using multiscale morphology in microseismic monitoring, *J. Appl. Geophys.,* **133,** 39–49.

Liu, G. & Chen, X., 2013. Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data, *J. Appl. Geophys.,* **93,** 60–66.

Liu, G., Fomel, S., Jin, L. & Chen, X., 2009a. Stacking seismic data using local correlation, *Geophysics,* **74,** V43–V48.

Liu, G., Chen, X., Du, J. & Song, J., 2011. Seismic noise attenuation using nonstationary polynomial fitting, *Appl. Geophys.,* **8,** 18–26.

Liu, G., Chen, X., Du, J. & Wu, K., 2012. Random noise attenuation using f-x regularized nonstationary autoregression, *Geophysics,* **77,** V61–V69.

Liu, W., Cao, S., Gan, S., Chen, Y., Zu, S. & Jin, Z., 2016. One-step slope estimation for dealiased seismic data reconstruction via iterative seislet thresholding, *IEEE Geosci. Remote Sens. Lett.,* **13**(10), 1462–1466.

Liu, Y., 2013. Noise reduction by vector median filtering, *Geophysics,* **78,** V79–V87.

Liu, Y., Liu, C. & Wang, D., 2009b. A 1D time-varying median filter for seismic random, spike-like noise elimination, *Geophysics,* **74,** V17–V24.

Menon, A.K. & Elkan, C., 2011. Fast algorithms for approximating the singular value decomposition, *J. ACM Trans. Knowl. Discovery [illegible] (TKDD),* **5,** 1–36.

Neelamani, R., Baumstein, A., Gillard, D., Hadidi, M. & Soroka, W., Coherent and random noise attenuation using the curvelet transf[illegible] *Leading Edge,* **27,** 240–248.

Oropeza, V. & Sacchi, M., 2011. Simultaneous seismic [illegible]ising a[illegible] reconstruction via multichannel singular spectrum [illegible]ysis, *[illegible]ophysics,* **76,** V25–V32.

Protter, M. & Elad, M., 2009. Image sequence [illegible] redundant representations, *IEEE Trans. [illegible] Pro[illegible]* 27–35.

Qu, S., Zhou, H., Chen, Y., Yu, S., Zha[illegible]Yuan, J., [illegible]& Qin, M., 2015. An effective method for red[illegible]rmonic distort[illegible]correlated vibroseis data, *J. Appl. Geophy[illegible]*28.

Rubinstein, R., Zibulevsky, M. & [illegible]ad, M[illegible]Efficient implementation of the K-SVD algorithm[illegible]g batch ortho[illegible]tching pursuit, *Tech. Rep*.

Sahoo, S.K. & Maku[illegible]2013. [illegible]onary training for sparse representation as generalizati[illegible]me[illegible]clusterin[illegible], *IEEE Signal Process. Lett.,* **20,** 587–590.

Trickett, S[illegible]. F-xy [illegible]p[illegible]uppression, in *CSPG CSEG CWLS Conv[illegible]*03–30[illegible]Society of Exploration Geophysicists (C[illegible]).

W[illegible]Liu[illegible]& Liu, G., 2008. Application of wavelet transform b[illegible]nd percentiles soft-threshold to elimination of seism[illegible]m noise, *Prog. Geophys. (in Chinese),* **23,** 1124–1130.

Wang, Y., [illegible]andom noise attenuation using forward-backward linear prediction[illegible]*sm. Explor.,* **8,** 133–142.

Wu, J., Wang, K., Chen, Y., Zhang, Y., Gan, S. & Zhou, C., 2016. Multiples attenuation using shaping regularization with seislet domain sparsity constraint, *J. Seism. Explor.,* **25**(1), 1–9.

Yu, S., Ma, J., Zhang, X. & Sacchi, M., 2015. Denoising and interpolation of high-dimensional seismic data by learning tight frame, *Geophysics,* **80,** V119–V132.

Zhang, C., Li, Y., Lin, H. & Yang, B., 2015. Signal preserving and seismic random noise attenuation by Hurst exponent based time–frequency peak filtering, *Geophys. J. Int.,* **203,** 901–909.

Zhang, D., Chen, Y., Huang, W. & Gan, S., 2016. Multi-step damped multichannel singular spectrum analysis for simultaneous reconstruction and denoising of 3D seismic data, *J. Geophys. Eng.,* **13,** 704–720.

Zhuang, G., Li, Y., Liu, Y., Lin, H., Ma, H. & Wu, N., 2015. Varying-window-length TFPF in high-resolution radon domain for seismic random noise attenuation, *IEEE Geosci. Remote Sens. Lett.,* **12,** 404–408.

Zu, S., Zhou, H., Chen, Y., Qu, S., Zou, X., Chen, H. & Liu, R., 2016. A periodically varying code for improving deblending of simultaneous sources in marine acquisition, *Geophysics,* **81,** V213–V225.

APPENDIX A: LOCAL SIMIL[illegible]Y

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the two signal vect[illegible]at are [illegible]d from a 2-D matrix or 3-D tensor. In the ca[illegible]valuatin[illegible]oising performance, $\mathbf{x}_1$ and $\mathbf{x}_2$ simply mea[illegible]signa[illegible]ise. T[illegible]implest way to measure the similarity [illegible]een two s[illegible]calculate the correlation coefficient,

$$c = \frac{\mathbf{x}_1^T \mathbf{x}_2}{\| \mathbf{x}_1 \|_2 \| \mathbf{x}_2 \|_2}, \tag{A1}$$

where c is the co[illegible]on coefficie[illegible]$_2$ denotes the dot product between $\mathbf{x}_1$ a[illegible] denotes th[illegible]$_2$ norm of the input vector. A locally c[illegible]ated [illegible]tion coefficient can be used to measure the local similarity bet[illegible]o signals,

$$c = \frac{\sum_{i_w=-N_w/2}^{N_w/2} x_1(i+i_w) x_2(i+i_w)}{\sqrt{\sum_{i_w=[illegible]}^{N_w} x_1(i+i_w)^2} \sqrt{\sum_{i_w=-N_w/2}^{N_w/2} x_2(i+i_w)^2}}, \tag{A2}$$

w[illegible](i) [illegible]$_2(i)$ denote the ith entries of vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, respe[illegible]denotes the index in a local window. $N_w + 1$ de[illegible]otes the length of each local window. The windowing is sometime[illegible]ome, since the measured similarity is largely dependent on [illegible]ndowing length and the measured local similarity might be discontinuous because of the separate calculations in windows. To avoid the negative performance caused by local windowing calculations, Fomel (2007) proposed an elegant way for calculating smooth local similarity via solving two inverse problems. The local similarity I use to evaluate denoising performance in this paper is defined as

$$\mathbf{s} = \sqrt{\mathbf{s}_1 \circ \mathbf{s}_2}, \tag{A3}$$

where $\mathbf{s}$ is the calculated local similarity, $\circ$ denotes Hadamard (or Schur) product, and $\mathbf{s}_1$ and $\mathbf{s}_2$ come from two least-squares inverse problem:

$$\mathbf{s}_1 = \arg\min_{\tilde{\mathbf{s}}_1} \|\mathbf{x}_1 - \mathbf{X}_2 \tilde{\mathbf{s}}_1\|_2^2, \tag{A4}$$

$$\mathbf{s}_2 = \arg\min_{\tilde{\mathbf{s}}_2} \|\mathbf{x}_2 - \mathbf{X}_1 \tilde{\mathbf{s}}_2\|_2^2, \tag{A5}$$

where $\mathbf{X}_1$ is a diagonal operator composed from the elements of $\mathbf{x}_1$: $\mathbf{X}_1 = \mathrm{diag}(\mathbf{x}_1)$ and $\mathbf{X}_2$ is a diagonal operator composed from the elements of $\mathbf{x}_2$: $\mathbf{X}_2 = \mathrm{diag}(\mathbf{x}_2)$. Eqs (A4) and (A5) are solved via shaping regularization

$$\mathbf{s}_1 = \left[\lambda_1^2 \mathbf{I} + \mathcal{T}\left(\mathbf{X}_2^T \mathbf{X}_2 - \lambda_1^2 \mathbf{I}\right)\right]^{-1} \mathcal{T} \mathbf{X}_2^T \mathbf{x}_1, \tag{A6}$$

$$\mathbf{s}_2 = \left[\lambda_2^2 \mathbf{I} + \mathcal{T}\left(\mathbf{X}_1^T \mathbf{X}_1 - \lambda_2^2 \mathbf{I}\right)\right]^{-1} \mathcal{T} \mathbf{X}_1^T \mathbf{x}_2, \tag{A7}$$

where $\mathcal{T}$ is a smoothing operator, and λ_1 and λ_2 are two parameters controlling the physical dimensionality and enabling fast convergence when inversion is implemented iteratively. These two parameters can be chosen as $\lambda_1 = \|\mathbf{X}_2^T \mathbf{X}_2\|_2$ and $\lambda_2 = \|\mathbf{X}_1^T \mathbf{X}_1\|_2$ (Fomel 2007).