GEOPHYSICS®

Noise attenuation in a low dimensional manifold

Journal:	*Geophysics*
Manuscript ID	GEO-2016-0509
Manuscript Type:	Technical Paper
Date Submitted by the Author:	04-Oct-2016
Complete List of Authors:	Yu, Shiwei Ma, Jianwei Osher, Stanley; Univeristy of California, Los Angeles, Shi, Zuoqiang; Univeristy of California, Los Angeles
Keywords:	signal processing, noise, filtering, sparse, optimization
Area of Expertise:	Signal Processing

Noise attenuation in a low dimensional manifold

Siwei Yu[1,2], *Jianwei Ma*[1], *Stanley Osher*[2], *Zuoqiang Shi*[2,3]

[1] *Center of Geophysics and Department of Mathematics, Harbin Institute of Technology, Harbin,* 150001, *China*

[2] *Department of Mathematics, University of California, Los Angeles, CA,* 90095, *USA*

[3] *Department of Mathematics, Tsinghua University, Beijing,* 100084, *China*

ABSTRACT

We show seismic data can be described in a low dimensional manifold, and then propose using a low dimensional manifold model (LDMM) method for extremely strong noise attenuation. The LDMM supposes the dimension of the patch manifold of seismic data should be low. In other words, the degree of freedom of the patches from seismic data should be low. Under the linear events assumption on a patch, the patch can be parameterized by the intercept and slope of the event, if the seismic wavelet is identical everywhere. The denoising problem is formed as an optimization problem including a fidelity term and an LDMM regularization term. We test LDMM on synthetic seismic data with different noise levels. LDMM achieves better denoised results in comparison with the Fourier, curvelet and nonlocal mean filtering methods, especially at strong noise or low signal-to-noise ratio situation. We also test the LDMM on field records, showing that LDMM is a potential method for handling relatively strong noise and preserving weak features.

INTRODUCTION

Background

Noise is unavoidable in seismic data acquisition. Strong background random noise is mainly caused by the surface conditions in an area. Strong noise makes processing procedures difficult, such as amplitude versus offset, amplitude versus azimuth, migration and inversion.

A more challenging scenario is microseismic monitoring, where very low amplitude data are typically acquired with strong coherent and random noise, and display a low signal-to-noise ratio (SNR) (Zhu et al., 2015). Accuracy and reliability of the location and other event attributes derived from microseismic traces is influenced by this strong noise content (Rodriguez et al., 2012).

Review on seismic noise attenuation methods

Typical time-domain methods include prediction (Abma and Claerbout, 1995) and stacking (Liu et al., 2009). These methods perform well in low noise distortion situations, however fail in low SNR situations (Harris and White, 1997). Various denoising methods have been proposed for low SNR data in the transformed domain.

Methods based on the Fourier transform are widely used in seismic noise attenuation, such as $f-x$ deconvolution (Canales, 1984). However, in the presence of strong noise, these methods actually struggle to denoise most of the frequencies. This shortcoming of the Canales $f-x$ denoising method is effectively overcome by the Naghizadeh's $f-k$ method (Naghizadeh, 2012) using all frequencies to find the dominant dips in the data.

Wavelets and directional wavelets are utilized in seismic noise attenuation, such as the curvelet transform (Candès and Donoho, 2004; Hennenfent and Herrmann, 2006), which uses a window segmentation in the frequency domain to reveal multiscale and multidirectional information. Seislet (Fomel and Liu, 2010) is a lifting wavelet based on prediction and updating. Seislet is slope based, which makes the seislet transform based on PWD

(PWD-seislet transform) occasionally fail in characterizing noisy signals with strong interference. The velocity-dependent (VD) seislet transform (Liu et al., 2014) has better resistance to strong random noise, which indicated the potential of VD seislets for random noise attenuation under a 1D earth assumption. The physical wavelet frame (Zhang and Ulrych, 2003) takes into account the characteristics of seismic data both in space and time, it can successfully remove strong noise from the data.

Predefined transforms may wipe out the details in seismic data at low SNR, since they are so structured (model driven). Dictionary learning methods were proposed for constructing a set of adaptive basis from the data itself (data driven), aiming to better preserve details. K-SVD (Aharon et al., 2006), sparse K-SVD (Rubinstein et al., 2010) and data driven tight frame (DDTF) (Cai et al., 2014) methods have been used in seismic noise attenuation (Beckouche and Ma, 2014; Liang et al., 2014; Yu et al., 2015, 2016; Zhu et al., 2015).

Rank reduction methods, such as singular spectrum methods (Sacchi, 2009), Cadzow filtering (Trickett et al., 2008), or multichannel singular spectrum analysis (MSSA) (Oropeza and Sacchi, 2011; Chiu, 2013), and low-rank constraint (or nuclear-norm minimization) methods (Yang et al., 2013; Ma, 2013; Kumar et al., 2015) are also used. For instance, the MSSA first arranges each frequency slice of the $f-x$ spectrum of seismic data into a Hankel matrix, then forces the Hankel matrix to be low rank to attenuate the noise. However, many numerical experiments suggest that the random noise cannot be completely removed using the MSSA algorithm. Huang et al. (2016) proposed a damped MSSA algorithm to handle strong noise situation. In the extremely strong noise situation, if we stress a very low rank, details will be wiped out by the global linear events.

The trade-off between regularization and detail preservation must be carefully considered, especially for strong noise situations. Removing strong noise requires strong regularization. However two problems may occur. First, the noise may be fit as a basis, e.g., in the sparsity promoting methods. Second, the local weak features may be wiped out. Patch and nonlocal based method seems to be an alternative for extremely high noise level situations.

An alternative way of denoising: patch based method

The nonlocal means (NLM) algorithm(Buades et al., 2005) was originally designed for image processing. NLM is a random noise filter which utilizes the structure similarities (some kind of redundancy) between the small windows (patches) of the image itself. Bonar and Sacchi (2012) first used the NLM in seismic data denoising, which seems to handle curvature and faults without losing the resolution of these features. NLM is achieved with the following equations:

$$\hat{f}(i) = \sum_j w(i,j)v(j), \tag{1}$$

$$w(i,j) = \frac{1}{Z(i)} exp\left(\frac{-D^2(i,j)}{h^2}\right), \tag{2}$$

where $Z(i)$ is the normalizing factor, $D(i,j)$ is the similarity between the pixels i and j (between the neighborhood around the pixels). Reader may refer to (Bonar and Sacchi, 2012) for details.

The structure similarities means that pixels of the patches are not randomly distributed, but restricted in less degrees of freedom (DOF). In the terms of mathematic, the patches are points on a low dimensional manifold.

Osher et al. (2016) proposed a new regularization method for image processing based on the patch manifold. The authors explained that patch manifolds of smooth images, cartoon images and texture images are all close to low dimensional manifolds. The so-called low dimensional manifold model (LDMM) achieves remarkable denoising results for low SNR inputs. In our work, we first show that the patch manifold of seismic data is also close to LDMM and then use the LDMM for seismic noise attenuation with extremely low SNR.

The remain sections are arranged as follows. The second section reviews the concept of manifold, patch manifold and shows that the patch manifold of seismic data is of low dimension. The third section gives the LDMM theory and algorithm. The fourth section illustrates numerical results on synthetic and field data. The last section concludes this paper.

METHOD

Manifold and patch manifold

A manifold is a collection of points in R^n, locating on a subspace of R^n. The subspace does not have to be linear. For instance, a sphere is a 2D manifold embedded in the 3D space. The points on the sphere are parameterized by longitude and latitude. They can not be obtained by linear combination of any other two points on the sphere. A manifold is locally identical to Euclidean space everywhere. A manifold is a powerful tool for describing natural images (Lee et al., 2003) and nonlinear dimensionality reduction (Tenenbaum et al., 2000).

A patch $p_x(f)$ of width τ extracted from data $f \in L^2([0,1]^2)$ around $x \in [0,1]^2$ is (Peyré, 2009)

$$\forall t \in [-\tau/2, \tau/2]^2,\ p_x(f)(t) = f(x+t). \tag{3}$$

where, the point x is the center of the patch $p_x(f)$.

Suppose $\Theta \subset L^2([0,1]^2)$ is a data set one is interested in. The patch manifold associated to Θ is

$$\mathcal{M} = \{p_x(g) | x \in [0,1]^2, g \in \Theta\}. \tag{4}$$

Figure 1 shows an illustration of a patch manifold from the sigmoid model. The seismic data is shown on the left. Small patches (size: $r \times r$) are extracted from the data and three of them are shown in the middle. The entire set of the patches consists a patch manifold in embedded R^{r^2} shown on the right. The patches are points on the manifold.

Low dimensional manifold model

The most important feature of the patch manifold is that it is close to a low-dimensional manifold for many natural images. Osher et al. (2016) gives three examples, as we summarized as follows.

If f is a C^2 function which corresponds to a smooth image, $p_x(f)$ for a given x can be approximated by a plane in R^d ($d = m \times n$, where m and n are the size of the patch) with three DOF. This fact implies that $M(f)$ is close to a 3D manifold.

If f is a piecewise constant function which corresponds to a cartoon image, then, any patch, $p_x(f)$ is well approximated by a straight edge patch. The edge has two DOF. $M(f)$ is also close to a 2D manifold.

If f is an oscillatory function corresponding to a texture image. $f(x) = a(x)\cos\theta(x)$, where $a(x)$ and $\theta(x)$ are both smooth functions. Then $M(f)$ is approximately a 6-dimensional manifold.

In this paper, we care about seismic data. In seismic data f, suppose the parameters of the seismic wavelet are constants everywhere, and each patch $p_x(f)$ consists of at most n events with two DOF for each. Then $M(f)$ is $2n$-dimension. If the amplitude and frequency of the wavelet not constrained, then $M(f)$ is $4n$-dimension. If there exist a fault, the data can be regarded as two texture image plus a straight edge.

We give examples to illustrate the manifold dimension of the cartoon image and seismic image. Left of Figure 2 shows the patches from a cartoon image, with lines indicating the boundary of the different regions. The lines are parameterized by the intercept and the angle, shown in the middle bottom figure. So the patch manifold of a cartoon image is 2D. The middle top figure shows a seismic Ricker wavelet. We convolve the patches on the left with the Ricker wavelet only on the vertical directional to obtain the patches on the right, which corresponding to patches from seismic data. This indicates the patch manifold of seismic image is also 2D with the linear events assumption on patch scale.

Low dimensional manifold model for denoising

Suppose the data acquisition model is

$$y = \Phi f + n \tag{5}$$

where Φ is the observation system, n is Gaussian noise, f is the original data and y is the degraded data. $\Phi = I$ for the application of denoising. We want to recover the original data such that the dimension of its patch manifold is as small as possible.

Figure 3 illustrates the dimension of the patch manifold from the data in Figure 1. The $x-$axis stands for the manifold dimension we use to approximate the patch manifold. The $y-$axis measures the residual between the approximation and the original patch manifold. Different marks stand for different noise level. This figure is generated with Isomap (Tenenbaum et al., 2000). For the noiseless situation, the residual changes litter after the dimension is larger than 7, which means that the patch manifold is approximately 7-D. Considering the fault in the data, we suppose each patch contains at most 2 events. Also we suppose the seismic wavelet changes across different patches. From the theoretical analysis in the previous section, the patch manifold should be 8-D, which is approximately consistent with the result in Figure 3. As the noise level increases, the residual also increases, since the noised patch manifold can not be approximated by low dimensional patch manifold.

The low dimensional manifold assumption formally gives the following optimization problem (Osher et al., 2016):

$$\min_f \dim(M(f)) + \lambda \|y - \Phi f\|_2^2, \tag{6}$$

where $\dim(\cdot)$ is the dimension of the manifold, λ is a balancing parameter. The optimization problem in equation 6 is very nonlinear and non-convex. An alternative direction method is utilized to solve it. First, fix the manifold M and update the data f. Then, the data is fixed and update the manifold, which is done by directly applying the patch operator on the data f. The derivation of the algorithm for the first step is achieved by replacing the dimension term with a coordinate function, then turning into an Laplace-Beltrami equation and finally using a point integral method to approximately solve it. We recommend interested readers to the appendix for the detailed algorithm.

NUMERICAL RESULTS

In this section, we test LDMM on synthetic and field seismic records. As a comparison, we present results with global filtering methods based on the Fourier transform, local filtering method based on curvelet transform, and patch filtering method based on NLM. For the parameters needed, we run experiments with different parameters and select those that achieve the highest SNR values. All tests were carried out on a personal laptop, with Matlab on Windows 10 operation system, 8G RAM and Intel Core i-7 CPU. The SNR value that is used to judge the quality of the noise attenuation is defined as follows

$$\text{SNR} = 10 \log_{10} \left(\frac{\|X^*\|_F^2}{\|X^* - X\|_F^2} \right)$$

where X^* and X denote the original data and its recovery, respectively. $\|\cdot\|_F$ is the Frobenious norm.

Figure 4a shows a synthetic seismic record, with 200 time samples and 256 space samples. Figure 4b shows the SNR obtained with different methods at different noisy levels (from SNR = -9.32 to SNR = 8.68). It is obvious that LDMM achieves much higher SNR, especially at low SNR situation. At lower noise situations, LDMM obtains similar SNR to the curvelet and NLM methods. Figure 4b shows that LDMM has a potential advantage of handling seismic data with extremely low SNR. Figure 5 shows some noise attenuation results corresponding to Figure 4b. The first to the fifth rows correspond to noisy data, with recovery results from the Fourier, curvelet, NLM and LDMM methods. Different columns correspond to different input noise levels. When the input noise level is low, all methods achieve pleasant results. However, at higher noise levels, the limitations of each method are obvious, especially at extremely high noisy situations. The Fourier method can hardly recover the sine function like events. The curvelet method fits part of the strong noise as small curvelet element, distributed all across the data. The artifacts from the NLM method degrade the result a lot. The LDMM achieves the best result both visually and with the highest SNR. However, due to the intense noise, the fault is also smoothed by LDMM as other methods.

The input SNRs and output SNRs for Figure 5 are summarized in Table 1. The last column of Table 1 also shows the elapsed time for different methods. LDMM is much more time consuming than the other three methods, which is a sacrifice for higher SNR. It is important to improve the efficiency of LDMM to handle high dimensional seismic data.

Input	8.68	-3.59	-8.49	Time (s)
Fourier	13.62	4.10	1.14	0.01
Curvelet	16.62	7.86	4.00	0.26
NLM	**16.85**	6.85	4.64	4.67
LDMM	16.78	**10.22**	**6.76**	30

Table 1: SNR and elapsed time

Figure 6a shows a seismic record with more complex structures. This record consists of 256 temporal samples and 256 spatial samples. We add strong noise (SNR = -7.35) to test LDMM. The events are barely notable in Figure 6b. Two circles stress where the results are easily comparable. The Fourier method (SNR = 2.73) in Figure 6c leaves notable noise in the recovery result and can hardly recovery the details. The curvelet method (SNR = 5.38) recovers features in the left circle well. However there are lots of artifacts caused by incorrect fitting of the noise with curvelets, as shown in Figure 6d. Also the details in the middle circle are not preserved. NLM (SNR = 5.53) can not recover the details at relatively low amplitude, as shown in Figure 6e. LDMM (SNR = 5.56) in Figure 6f achieves a little higher SNR compared to the curvelet and NLM methods. However, LDMM produces much better results visually. The details in the two circles are well preserved and the events seem very continuous. In order to compare the abilities of amplitude preservation by different methods, we give the recovery residual (i.e., recovery trace - clean trace) for a single trace in Figure 7. Figure 7a shows the original (solid blue) and noisy trace (dahsed green) from the data in Figure 6a and 6b (distance = $1000m$). Figure 7b shows the recovery residual of the Fourier, curvelet, NLM and LDMM methods from top to bottom. In the green rectangles, it is clear that the residual of LDMM is the smallest among the four methods, showing the

good behavior of LDMM in amplitude preservation.

Figure 8a shows a field record with strong distortion. The goal is to remove the distortion and preserve the structure of the events as much as possible. At the bottom of the two circles, the events shows jumping arches like features. The Fourier and curvelet methods (Figure 8b and 8c) smooth the arches too much. The patch based methods: NLM and LDMM (Figure 8d and 8e) preserve the arches much better. However NLM wipes out the energy of events with relatively low amplitude. LDMM eliminates the distortion in the record at the same time preserving the detailed structures. Figure 9 shows the comparisons of recovery for a single trace at distance $= 1000m$. LDMM achieves the closest amplitude to the original data, as shown in the red circles. It is understandable that all methods attenuate the amplitude of the trace to some extent, since there is noise in the original data.

DISCUSSION

As we mentioned, the trade-off between regularization and details preservation must be carefully considered, especially for strong noise situation. We now discuss the trade-off in different methods.

The $f-x$ deconvolution method is achieved with auto regression for each slice of the $f-x$ spectrum. The trade-off is controlled by the length of the regression operator (Canales, 1984). If we have a long regression operator, the details may be also degraded along with the noise.

The $f-k$ domain or other sparse transform based methods usually obtain the recovery data from an optimization problem, including a fidelity term and a sparsity term. The trade-off is controlled by the weights between these two terms (Trad et al., 2002; Naghizadeh and Innanen, 2011). When we have higher weights for the sparsity term, the noise may be seen as a basis and the weak features may be wiped out.

Dictionary learning methods are split into two steps. In the training step, a parameter

controls the sparsity of the data expanding on the learned dictionary. With more weight on the sparsity term, the trained filter is more structured, leading to more regularization in the sparsity promoting step. The sparsity promoting step is the same as the fixed basis method. However, as we mentioned, dictionary learning is difficult to learn valid dictionary in strong noise situations.

The low rank methods first arrange each frequency slice of seismic data into a Hankel matrix, then force the Hankel matrix to be low rank to attenuate the noise. In theory, the rank is equal to the number of events under a linear event assumption. The selection of the desired rank plays a role similar to the length of the prediction error filter in classical $f-x$ deconvolution (Oropeza and Sacchi, 2011). In strong noise situations, if we stress a very low rank, details will be wiped out by the global linear events.

For window strategies, details can not be preserved if the window is too big. If the size of the window is too small, the regularization will not work (Trad, 2014).

NLM filters one patch with the weighted average of the similar patches. The trade-off is controlled by the number of the patches utilized. When more patches are utilized, more noise is filtered. However, more patches with not so similar structures will destroy the structure of the desired patch.

LDMM first forms an optimization problem, including a fidelity term and a LDMM regularization term. The trade-off is controlled by the weights between these two terms. For strong noise situation, we should enlarge the weight of the LDMM regularization term. However, there exist a lower limitation of the dimension=$2n$ (n is the maximum number of events in the patches), as we discussed. So no over regularization exists, which is an advantage for preserving the details. Actually there is no obviously trade-off between removing the noise and preserving the details. The weighting parameter just controls how much noise is removed, since large weight almost does not smooth the original details. This can be better understood from Figure 3. Suppose we have obtained an approximately clean data from the noisy data by LDMM, where we may have reached dimension = 7 of the most

bottom line. If we continue enlarging the weight for the low dimensional regularization term, the dimension may decrease. However, the residual gap between dimension 7 and 6 is much larger than that after dimension > 7. So it is difficult to reduce the dimension any more, unless we impose a very large weight for the low dimensional regularization term. What's more, it is much more difficult to reach even lower dimension.

CONCLUSION

We introduce a novel conception, low dimensional patch manifold, for seismic data processing. LDMM assumes that the patches of a seismic records consists of a set which is a subset of a low dimensional patch manifold. LDMM utilizes nonlocal information to wipe out the noise in the record. We use LDMM for noise attenuation at extremely low SNR situation. The numerical results show that LDMM eliminates the strong distortion in the record while at the same time preserving the detailed structures. LDMM over performs the Fourier, curvelet and NLM methods, especially at low SNR situations. LDMM has many potential applications in seismic data processing, such as integrating in the POCS frame for interpolation, microseismic monitoring where very low amplitude data are typically acquired with strong coherent and random noise. Improving the efficiency of LDMM for industry applications is also a future direction.

ACKNOWLEDGEMENT

This work is supported by NSFC (grant number: NSFC 91330108, 41374121, 61327013), and the Fundamental Research Funds for the Central Universities (grant number: HIT.PIRS.A201501).

Appendices

We need to solve optimization problem:

$$\min_f \dim(\mathcal{M}(f)) + \lambda \|y - \Phi f\|_2^2. \tag{A-1}$$

The dimension of a manifold is calculated with (Osher et al., 2016):

$$\dim(\mathcal{M}(f)) = \sum_{j=1}^{d} \|\nabla_{\mathcal{M}} \alpha_j(x)\|^2_{L^2(\mathcal{M})}, \tag{A-2}$$

where $\alpha_i(x) = x_i$ is the coordinate function. Then optimization problem can be rewritten as:

$$\min_{f,\mathcal{M}} \sum_{j=1}^{d} \|\nabla_{\mathcal{M}} \alpha_j(x)\|^2_{L^2(\mathcal{M})} + \lambda \|y - \Phi f\|_2^2, \text{ s.t. } P(f) \subset \mathcal{M}, \tag{A-3}$$

where $P(f)$ is a discretized patch set extracted from seismic data f. An iterative method is used to solved the above problem approximately. First, the manifold is fixed, the data and the coordinate functions are computed,

$$(f^{n+1}, \alpha_1^{n+1}, \cdots, \alpha_d^{n+1}) = \text{argmin} \sum_{j=1}^{d} \|\nabla_{\mathcal{M}} \alpha_j(x)\|^2_{L^2(\mathcal{M})} + \lambda \|y - \Phi f\|_2^2, \text{ s.t. } \alpha_i(p_x(f^n)) = p_x^i(f), \tag{A-4}$$

where $p_x^i(f)$ is the ith element of patch $p_x(f)$. Then the manifold is updated using the new data and coordinate functions,

$$\mathcal{M}^{n+1} = \left\{ \alpha_1^{n+1}(x), \cdots, \alpha_d^{n+1}(x) : x \in \mathcal{M}^n \right\}, \tag{A-5}$$

then these two steps are repeated until convergence.

Bregman iteration is used to solve the constrained optimization problem A-4. Primary update for α,

$$(\alpha_1^{n+1,k+1}, \cdots, \alpha_d^{n+1,k+1}) = \text{argmin} \sum_{j=1}^{d} \|\nabla_{\mathcal{M}} \alpha_j(x)\|^2_{L^2(\mathcal{M})} + \mu \|\alpha(P(f^n)) - P(f) + d^k\|_F^2. \tag{A-6}$$

Primary update for f,

$$(f^{n+1,k+1}) = \text{argmin}\, \lambda \|y - \Phi f\|_2^2 + \mu \|\alpha(P(f^n)) - P(f) + d^k\|_F^2. \tag{A-7}$$

Dual update,

$$d^{k+1} = d^k + \alpha^{n+1,k+1}(P(f^n) - P(f^{n+1,k+1})). \tag{A-8}$$

In A-6, $\alpha_i^{n+1,k+1}, i = 1, \cdots, d$ can be solved separately,

$$\alpha_1^{n+1,k+1} = \text{argmin} \, \|\nabla_{\mathcal{M}} \alpha_j(x)\|^2_{L^2(\mathcal{M})} + \mu \|\alpha(P(f^n)) - P(f) + d^k\|_F^2, \tag{A-9}$$

which can be solve from a standard form,

$$\underset{u \in H^1(\mathcal{M})}{\text{argmin}} \ \|\nabla_{\mathcal{M}} u\|^2_{L^2(\mathcal{M})} + \mu \sum_{y \in \Omega} \|u(y) - v(y)\|_F^2. \tag{A-10}$$

It can be turned into an Laplace-Beltrami equation,

$$-\nabla_{\mathcal{M}} u + \mu \sum_{y \in \Omega} \delta(x - y)(u(y) - v(y)) = 0, x \in \mathcal{M} \tag{A-11}$$

$$\frac{\partial u}{\partial n} = 0, x \in \partial \mathcal{M}. \tag{A-12}$$

A point integral method is used to approximate equation A-11,

$$\int_{\mathcal{M}} (u(x) - u(y)) R_t(x, y) dy + \mu t \sum_{y \in \Omega} \bar{R}_t(x, y)(u(y) - v(y)) = 0 \tag{A-13}$$

where $R_t(x, y)$ and $\bar{R}_t(x, y)$ are distance functions and are usually set as follows

$$R_t(x, y) = \bar{R}_t(x, y) = C_t \exp\left(\frac{|x - y|^2}{4t}\right) \tag{A-14}$$

where C_t and t are parameters of the Gaussian function.

Finally the equation A-13 is discretized over the point set $P(f^n)$,

$$\frac{|\mathcal{M}|}{N} \sum_{j=1}^{N} R_t(x_i, x_j)(u_i - u_j) + \mu t \sum_{j=1}^{N} \bar{R}_t(x_i, x_j)(u_i - v_j) = 0 \tag{A-15}$$

where $v_j = v(x_j)$ and $|\mathcal{M}|$ is the volume of the manifold $\mathcal{M}$. Equation A-15 is rewritten in the matrix form

$$(L + \bar{\mu}\bar{W})u = \bar{\mu}\bar{W}v \tag{A-16}$$

where $v = (v_1, \cdots, v_N)$ and $\bar{\mu} = \frac{\mu t N}{\mathcal{M}}$, $L = D - W$, $W = (\omega_{ij})$, $\bar{W} = (\bar{\omega}_{ij})$. The weights are,

$$\omega_{ij} = R_t(x_i, x_j), \bar{\omega}_{ij} = \bar{R}_t(x_i, x_j), x_i, x_j \in P(f^n). \tag{A-17}$$

We summarize the final algorithm in Algorithm 1.

Algorithm 1 LDMM Algorithm

Input: Initial guess of the data $f^0, d^0 = 0$.

1: **repeat**

2: Compute the weight matrix $W = (\omega_{ij})$ from $P(f^n)$, where $i, j = 1, ..., N$ and $N = |P(f^n)|$ is the total number of points in $P(f^n)$,

$$\omega_{ij} = R_t(x_i, x_j),\ \bar{\omega}_{ij} = \bar{R}_t(x_i, x_j),\ x_i, x_j \in P(f^n),\ i, j = 1, ..., N.$$

And assemble the matrices L, W and $\bar{W}$ as follows:

$$L = D - W, W = (\omega_{i,j}), \bar{W} = (\bar{\omega}_{i,j}), i, j = 1, ..., N.$$

3: Solve the following linear systems

$$(L + \bar{\mu}\bar{W})U = \bar{\mu}\bar{W}V$$

where $V = P(f^n) - d^n$.

4: Update f by solving a least square problem.

$$f^{n+1} = \underset{f \in R^{m \times n}}{\operatorname{argmin}} \lambda \|y - \Phi f\|_{l^2}^2 + \bar{\mu}\|U - P(f) + d^n\|_F^2.$$

5: Update d^n,

$$d^{n+1} = d^n + U - P(f^{n+1})$$

6: **until** convergence

Output: Denoised data f.

REFERENCES

Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by tx and fx techniques: Geophysics, **60**, 1887–1896.

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.

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

Bonar, D., and M. Sacchi, 2012, Denoising seismic data using the nonlocal means algorithm: Geophysics, **77**, A5–A8.

Buades, A., B. Coll, and J.-M. Morel, 2005, A review of image denoising algorithms, with a new one: Multiscale Modeling & Simulation, **4**, 490–530.

Cai, J.-F., H. Ji, Z. Shen, and G.-B. Ye, 2014, Data-driven tight frame construction and image denoising: Applied and Computational Harmonic Analysis, **37**, 89–105.

Canales, L. L., 1984, Random noise reduction: the 54th Annual International Meeting, Society of Exploration Geophysicists.

Candès, E. J., and D. L. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities: Communications on Pure and Applied Mathematics, **57**, 219–266.

Chiu, S. K., 2013, Coherent and random noise attenuation via multichannel singular spectrum analysis in the randomized domain: Geophysical Prospecting, **61**, 1–9.

Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.

Harris, P., and R. White, 1997, Improving the performance of f–x prediction filtering at low signal-to-noise ratios: Geophysical Prospecting, **45**, 269–302.

Hennenfent, G., and F. J. Herrmann, 2006, Seismic denoising with nonuniformly sampled curvelets: Computing in Science & Engineering, **8**, 16–25.

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

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

Lee, A. B., K. S. Pedersen, and D. Mumford, 2003, The nonlinear statistics of high-contrast patches in natural images: International Journal of Computer Vision, **54**, 83–103.

Liang, J., J. Ma, and X. Zhang, 2014, Seismic data restoration via data-driven tight frame: Geophysics, **79**, V65–V74.

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

Liu, Y., N. Liu, and C. Liu, 2014, Adaptive prediction filtering in txy domain for random noise attenuation using regularized nonstationary autoregression: Geophysics, **80**, V13–V21.

Ma, J., 2013, Three-dimensional irregular seismic data reconstruction via low-rank matrix completion: Geophysics, **78**, V181–V192.

Naghizadeh, M., 2012, Seismic data interpolation and denoising in the frequency-wavenumber domain: Geophysics, **77**, V71–V80.

Naghizadeh, M., and K. A. Innanen, 2011, Seismic data interpolation using a fast generalized fourier transform: Geophysics, **76**, V1–V10.

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

Osher, S., Z. Shi, and W. Zhu, 2016, Low dimensional manifold model for image processing: Technical report, UCLA, Tech. Rep. CAM report 16-04.

Peyré, G., 2009, Manifold models for signals and images: Computer Vision and Image Understanding, **113**, 249–260.

Rodriguez, I., D. Bonar, and M. Sacchi, 2012, Microseismic data denoising using a 3c group sparsity constrained time-frequency transform: Geophysics, **77**, V21–V29.

Rubinstein, R., M. Zibulevsky, and M. Elad, 2010, Double sparsity: Learning sparse dictionaries for sparse signal approximation: IEEE Transactions on Signal Processing, **58**,

1553–1564.

Sacchi, M., 2009, Fx singular spectrum analysis: CSPG CSEG CWLS Convention, 392–395.

Tenenbaum, J. B., V. De Silva, and J. C. Langford, 2000, A global geometric framework for nonlinear dimensionality reduction: Science, **290**, 2319–2323.

Trad, D., 2014, Five-dimensional interpolation: New directions and challenges: CSEG Recorder, March, 40–46.

Trad, D., T. Ulrych, and M. Sacchi, 2002, Accurate interpolation with high-resolution time-variant Radon transforms: Geophysics, **67**, 644–656.

Trickett, S., et al., 2008, F-xy cadzow noise suppression: Presented at the 2008 SEG Annual Meeting, Society of Exploration Geophysicists.

Yang, Y., J. Ma, and S. Osher, 2013, Seismic data reconstruction via matrix completion: Inverse Problems and Imaging, **7**, 1379–1392.

Yu, S., J. Ma, and S. Osher, 2016, Monte carlo data-driven tight frame for seismic data recovery: Geophysics, **81**, V327–V340.

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

Zhang, R., and T. J. Ulrych, 2003, Physical wavelet frame denoising: Geophysics, **68**, 225–231.

Zhu, L., E. Liu, and J. H. McClellan, 2015, Seismic data denoising through multiscale and sparsity-promoting dictionary learning: Geophysics, **80**, WD45–WD57.

LIST OF FIGURES

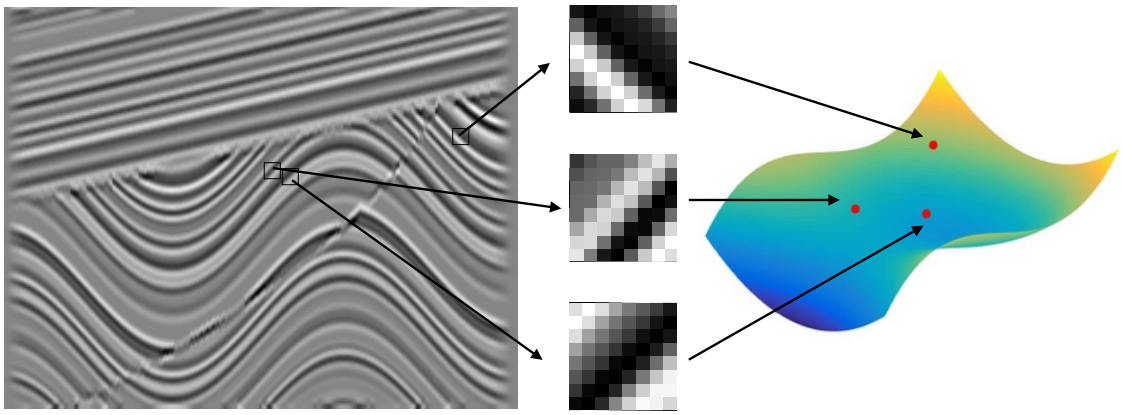


Figure 1: Illustration of the construction of a patch manifold from seismic patches. Left: the sigmoid seismic data. Middle: three patches from the seismic data. Right: a patch manifold composed by all the patches from the seismic data.

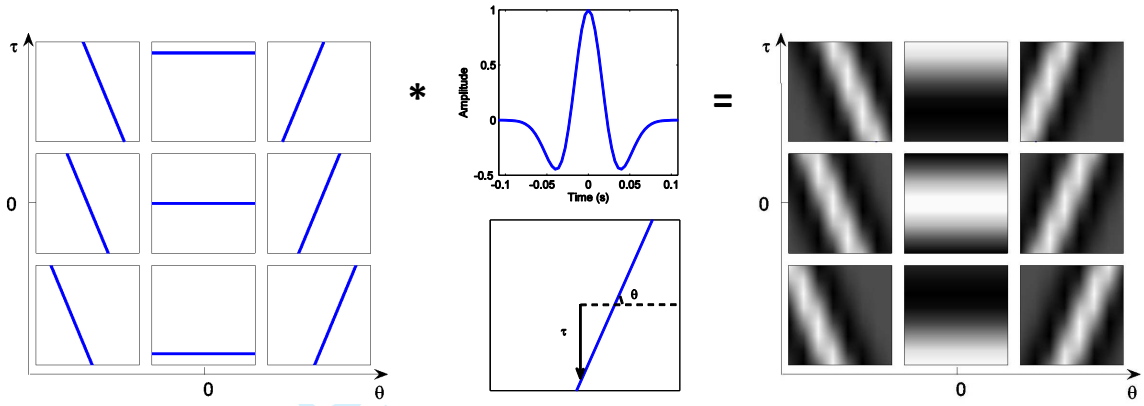


Figure 2: Illustration of the DOF of the seismic patches.

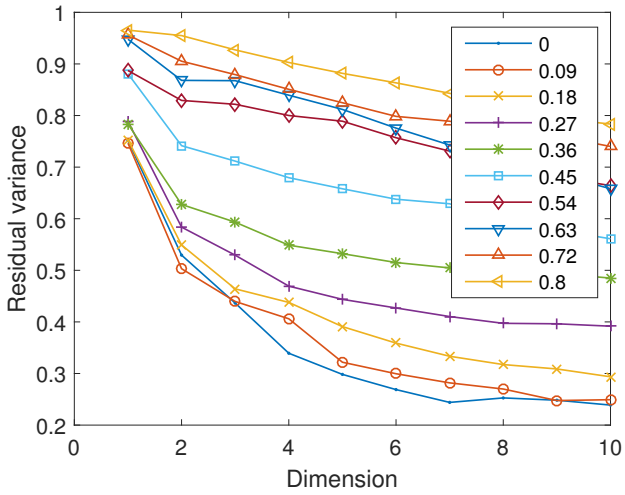
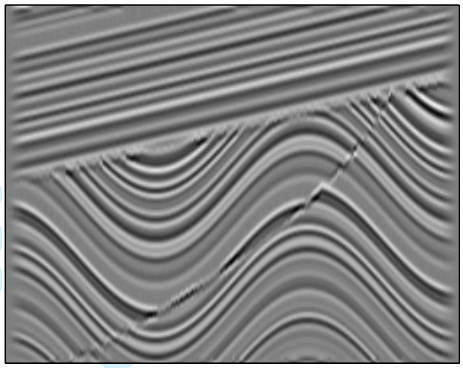
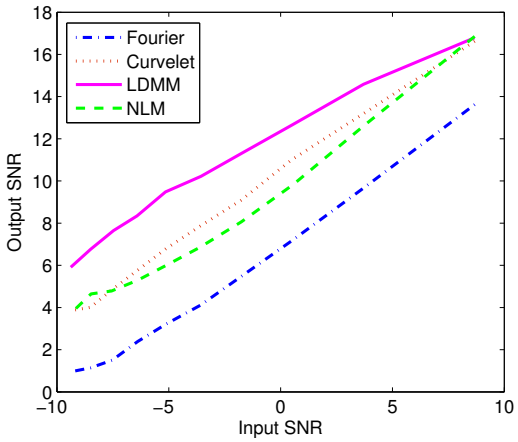


Figure 3: The residual variance of Isomap versus dimensionality on different noise level.



a



b

Figure 4: (a) A synthetic seismic record. (b) SNR obtained with different methods at different noisy levels.

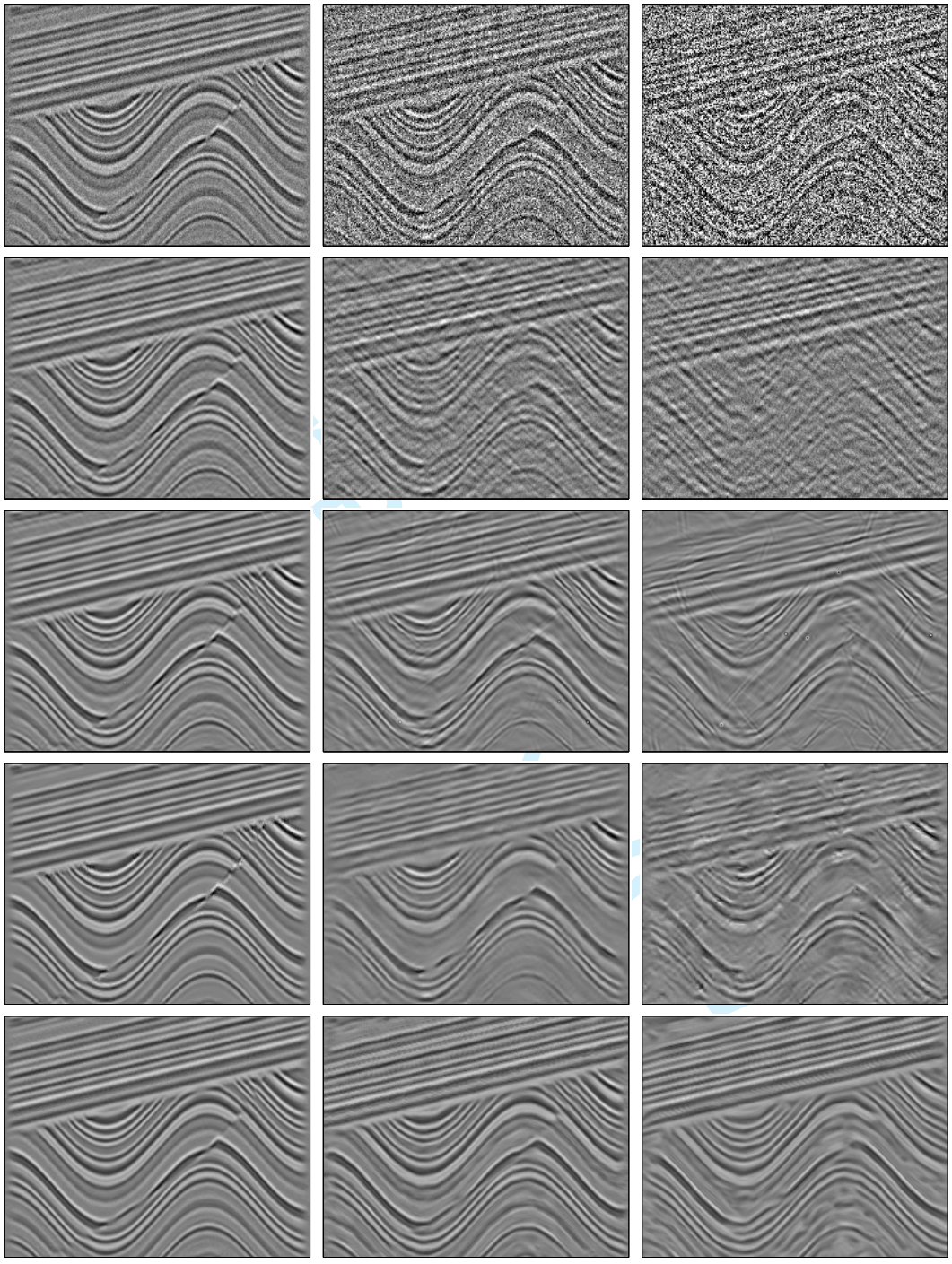


Figure 5: Denoise on a synthetic record. 1st-5th rows: noisy data, recovery results from the Fourier, curvelet, NLM, LDMM methods. 1st-3rd columns, different noise levels.

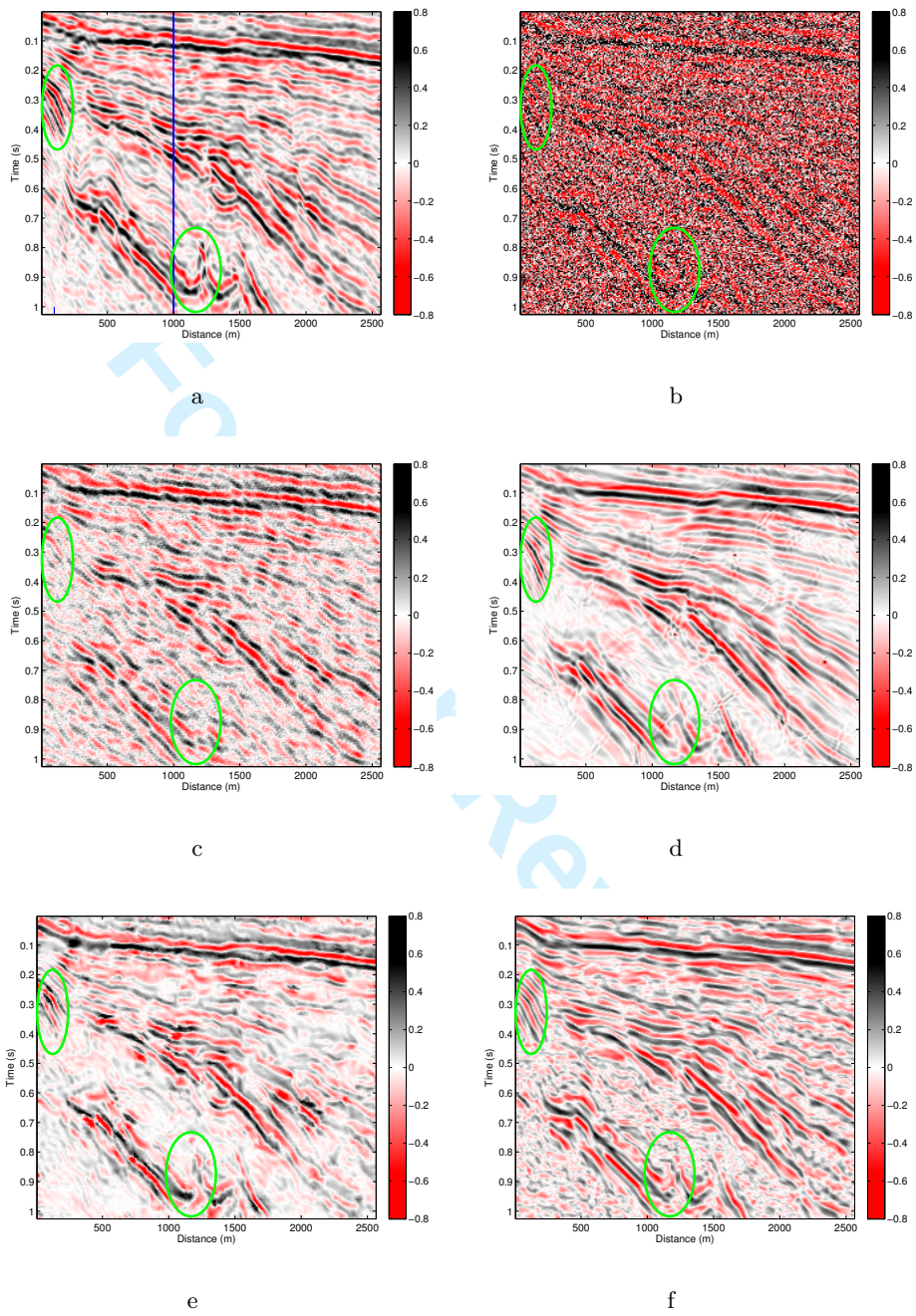
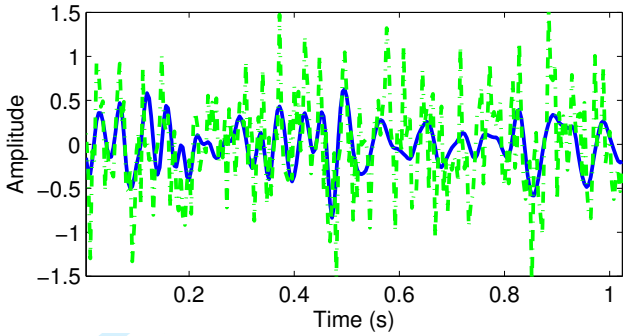
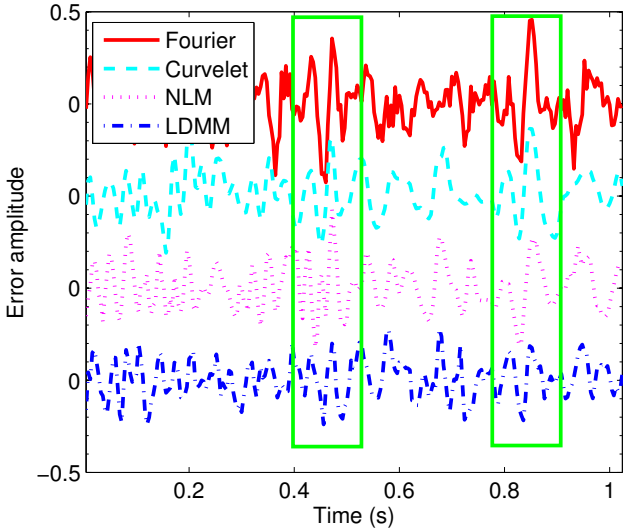


Figure 6: Noise attenuation on a seismic record with complex structures. (a) Clean data. (b) Noisy data (SNR = -7.35). Recovery results with (c) the Fourier method (SNR = 2.73), (d) the curvelet method (SNR = 5.38), (e) NLM (SNR = 5.53) and (f) LDMM (SNR = 5.56).



a



b

Figure 7: Comparisons of residual for a single trace (distance = $1000m$). (a) The original (solid blue) and noisy trace (dahsed green). (b) From top to bottom: recovery residual of the Fourier, curvelet, NLM and LDMM methods.

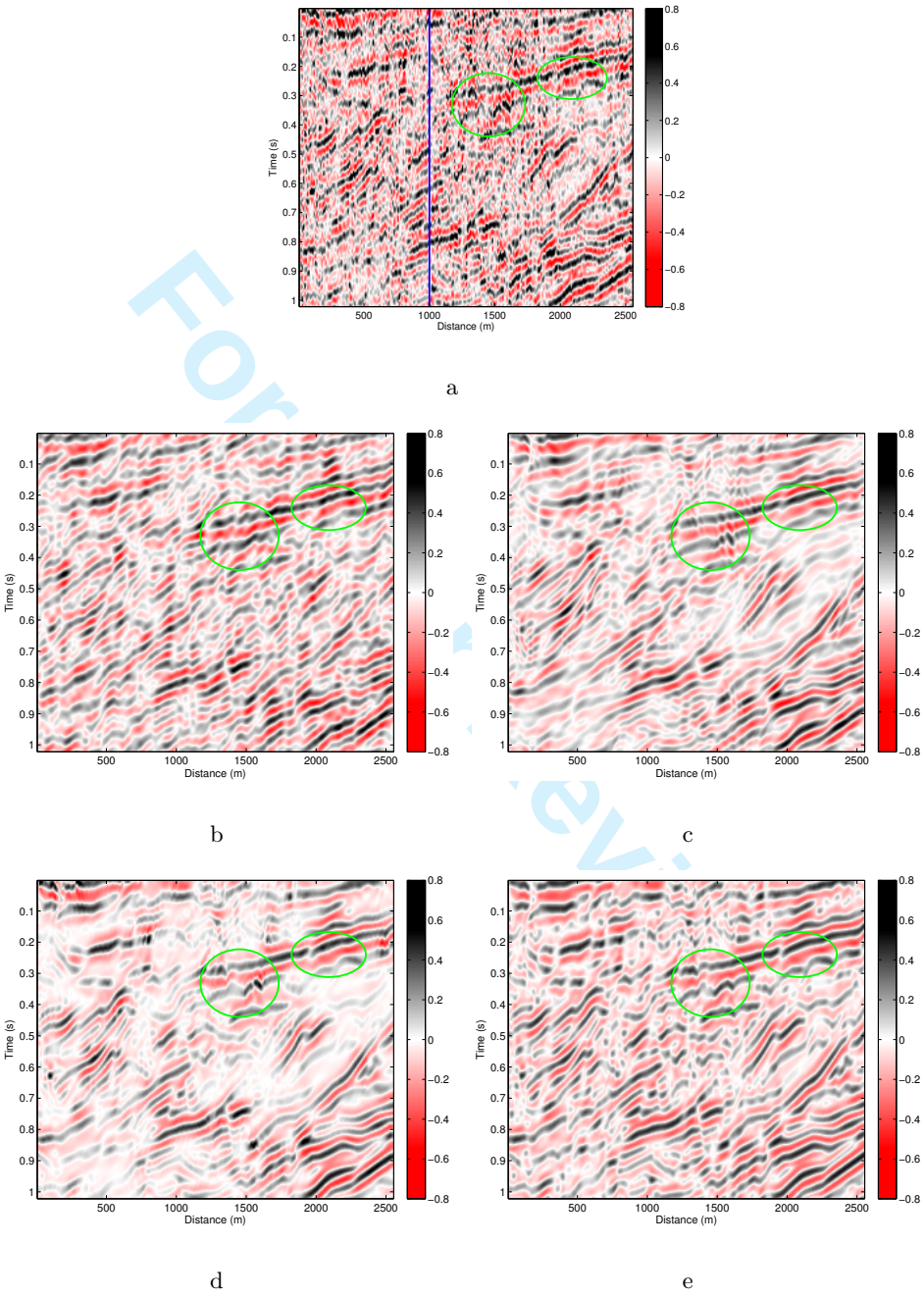


Figure 8: Noise attenuation on a field record. (a) A field record. (b)-(e) Recovery results from the Fourier, curvelet, NLM and LDMM methods.

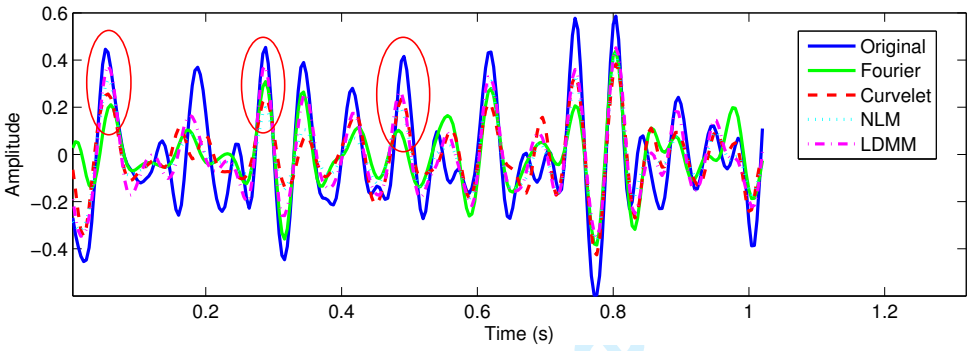


Figure 9: Comparisons of recovery for a single trace (distance = $1000m$).