Inverse Problems in Science and Engineering
Vol. 19, No. 5, July 2011, 679–704

Taylor & Francis
Taylor & Francis Group

A review on restoration of seismic wavefields based on regularization and compressive sensing

Jingjie Cao[ab], Yanfei Wang[b]*, Jingtao Zhao[ab] and Changchun Yang[b]

[a]*Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing 100029, P.R. China;* [b]*Graduate University of Chinese Academy of Sciences, Beijing 100049, P.R. China*

(*Received 15 March 2011; final version received 16 March 2011*)

Restoration of seismic data as an ill-posed inverse problem means to recover the complete wavefields from sub-sampled data. Since seismic data are typically sparse in the curvelet domain, this problem can be solved based on the compressive sensing theory. Meanwhile three major problems are modelling, sampling and solving methods. We first construct l_0 and l_1 minimization models and then develop fast projected gradient methods to solve the restoration problem. For seismic data interpolation/restoration, the regular sub-sampled data will generate coherence aliasing in the frequency domain, while the random sub-sampling cannot control the largest sampling gap. Therefore, we consider a new sampling technique in this article which is based on the controlled piecewise random sub-sampling scheme. Numerical simulations are made and compared with the iterative soft thresholding method and the spectral gradient-projection method. It reveals that the proposed algorithms have the advantages of high precision, robustness and fast calculation.

Keywords: wavefield recovery; curvelet transform; compressive sensing; inverse problems; ill-posedness; projected gradient method

AMS Subject Classifications: 65J22; 86A22

1. Introduction

In exploration seismology, the process of acquisition records the continuous wave field which is generated by the source. In order to restore the seismic data correctly, the acquisition should satisfy the *Nyquist/Shannon* sampling theorem, i.e. the sampling frequency should at least be twice the maximum frequency of original signal. In seismic acquisition, because of the influence of obstacles at land surface, rivers, bad receivers, noise, acquisition aperture, restriction of topography and investment, the obtained data usually does not satisfy the sampling theorem. A direct effect of the limitations of acquisition is that the sub-sampled data will generate aliasing in the frequency domain; therefore, it may affect the subsequent processing such as filtering, de-noising, AVO (amplitude versus offset) analysis, multiple eliminating and migration imaging. In order to remove the influence of sub-sampled data, the seismic data restoration/interpolation

*Corresponding author. Email: yfwang_ucf@yahoo.com; yfwang@mail.iggcas.ac.cn

ISSN 1741–5977 print/ISSN 1741–5985 online

DOI: 10.1080/17415977.2011.576342
http://www.informaworld.com

technique is often used. On the other hand, seismic data interpolation can generate high-resolution data which will save the cost and improve the imaging results. In summary, seismic data restoration is an important research direction in exploration seismology [1,2].

Many seismic data restoration methods have been developed in the past few years. Mostafa and Sacchi [3], classify the methods for seismic wavefield recovery into two kinds: wave-equation-based methods and signal processing methods. According to our knowledge these methods can be classified into four classes. The first class is the prediction error filter methods. Seismic data can be obtained by finding a convolution filter that predicts the data in such a way that the error is white noise. The well-known f–x predictive method in [4,5] can restore the spatially aliased regularly sub-sampled data. The low-frequency data components are utilized to recover the high-frequency data. A modified projective prediction error filter method was proposed in [6]. However, these methods are applicable only if the original data is regularly sampled in space and will cost a lot in computation. Similar methods in the f–k domain [7] for band-limited signals use the Fourier transform to predict the complete wave field. The interpolation method that rely on the t–x domain predictive filters were introduced in [8–10], which first estimate the dip in a sliding time window and then interpolate along the dip direction in each time window. All the above-mentioned methods are based on the assumption of linear events; the restoration will deteriorate for cross events. The second kind of method is based on the wave equation. These methods utilize the physical properties of wave propagation to restore seismic data. An integral with a continuous operator is often used to obtain the complete wave field [11,12]. Some information of velocity distribution in the interior of earth is needed for these methods. The third class is the mathematical interpolation in the time domain. These methods include sinc interpolation after alias reduction via NMO (normal movement) correction [13], most coherent dip interpolation [14], interpolation using event attributes and power diversity slant–stack interpolation. The sinc interpolation methods are not appropriate for far offset data, while the most coherent dip interpolation is only suitable for data consisting of linear events. The last kind of method is based on the Fourier analysis. Some *a priori* knowledge such as the seismic signals being band-limited or sparseness of seismic data is needed. The seismic data are decomposed through a transform. These methods are robust for noisy data. Methods based on Fourier transform [15–19] can be applied to seismic signals with spatially band-limited property. The methods based on linear Radon transform [20] and parabolic Radon transform [21] can focus the energy of signals in the transformed domain. The curvelet transform was introduced in seismic restoration in [22,23], and was proved to be the most sparse one for seismic data.

The model of seismic acquisition can be written as

$$Ax = b, \tag{1}$$

where $A \in \mathbb{R}^{l \times m}$ is the sampling operator, $x \in \mathbb{R}^m$ is the reflectivity model, and $b \in \mathbb{R}^l$ denotes the sampled data. The restoration problem is to solve x from A and b, thus it is an inverse problem. A problem is called well-posed if the solution of (1) exists, is unique and continuous. If any one of these conditions is violated, the problem is ill-posed. Equation (1) can be solved by finding a least squares solution, i.e. solving the problem

$$\min \frac{1}{2} \|Ax - b\|_2^2, \tag{2}$$

where $\|\cdot\|_2$ is the Euclidean norm of a vector. It is equivalent to solving the normal equation

$$A^T Ax = A^T b, \tag{3}$$

where A^T is the transpose of A. However, the direct inversion of (3) produces an unstable solution. In order to obtain a stable solution, regularization is necessary.

A popular method for solving ill-posed problems is the *Tikhonov* regularization, which refers to solving a regularized functional

$$\min J_\alpha(x) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{\alpha}{2}\Omega(x), \tag{4}$$

where $\Omega(\cdot)$ is the so-called *Tikhonov* stabilizer and $\alpha > 0$ is the regularization parameter. We often choose $\Omega(x) = \|D^{1/2}x\|_2^2$, where D is a positive semi-definite and bounded operator. Then, the minimizer of $J_\alpha(x)$ is given by

$$x_\alpha = (A^T A + \alpha D)^{-1} A^T b. \tag{5}$$

For solving Equation (5), several methods such as LU decomposition, singular value decomposition and QR decomposition can be used [24]. However, any direct method should be avoided and iterative solvers are preferred for geophysical inversion problems. The commonly used iterative methods involve conjugate gradient-type methods, Newton-type methods and statistical methods based on Bayesian inference [24]. However, these methods may be time consuming and other methods which possess fast convergence and stability are desirable for large-scale seismic data restoration problems.

For seismic interpolation problem, let us denote by x the original seismic wave field, b the sampled data and A the sampling operator, the expression again can be written as (1). Our purpose is to restore x from the sampled data b. Since b is usually incomplete, A is an under-determined operator. This indicates that there are infinite solutions satisfying the seismic interpolation Equation (1). Hence, seismic data interpolation is an ill-posed inverse problem.

The compressive sensing in information theory could be used to recover the original data from insufficient sampling, which breaks the limitations of the *Nyquist/Shannon* sampling theorem. This theory has been applied in many application fields. Theoretically, if the seismic data is sparse under a transform, then we can restore the complete data from sub-sampled data according to the compressive sensing. Another potential application of the compressive sensing theory in seismology is the design of new sampling methods in seismic acquisition. As we all know, seismic acquisition expends a lot of time and money. The number of sources and receivers can be reduced greatly based on compressive sensing, which will save the acquisition cost. At the same time, the complete wave field can still be recovered from partial sampled data.

Finding the sparse solution of an under-determined system is also a crucial issue in seismic processing. Due to huge storage of the data volume and the specific sampling, methods for solving general problems cannot be used in seismic data restoration directly. So far, developed methods include the iterative soft thresholding method [25,26], the spectral projection-gradient method [27] and the iterative re-weighted least squares method [28]. However, these methods do not converge very fast which may lead to deficient usage.

It is well-known that the best model for finding a sparse solution of an under-determined system is the solution in l_q space with $q \to 0$. Therefore, to meet the above

requirements of applying compressive sensing to seismology, we consider the smoothing l_0 method and develop two fast and robust algorithms for solving the sparse restoration problem. As far as we know, the methods we consider have not been addressed in the literature for restoration of seismic wavefields.

The outline of this article is as follows. We give a short review of the concept of compressive sensing in Section 2.

In Section 3, we introduce the sparse transform. In particular, the popular curvelet transform is introduced. As a multi-scale, multidirectional, anisotropic and local transform, it has great advantages over the Fourier transform, Radon transform and the wavelet transform for seismic data processing. Therefore, it can be used in seismic data recovery.

Our recently developed sampling scheme [2] for seismic restoration is introduced in Section 4. Random sub-sampling cannot control the largest sampling gap which is unfavourable for curvelet-based wave field restoration [25], and the regularly sub-sampled data will generate aliasing in the frequency domain. Our sampling technique can both control the sampling gap and reduce the aliasing in the frequency domain. It is worthwhile to mention that a lot of sampling methods were proposed in different applications, but they cannot be introduced directly to solve seismic restoration because of the specialty of seismic acquisition.

In Section 5, we first give a brief review of the classical methods, and then we devote ourselves to build the l_q model with $q=0$ and $q=1$ and develop two fast and robust methods for solving the sparse restoration problem.

Numerical simulations are given in Section 6. In Section 6.1, we show that our controlled sampling method can greatly improve the restoration; its performance is comparable with the jittered sampling [22]. In Sections 6.2 and 6.3, we show that our proposed methods based on l_q model with $q=0$ and $q=1$ along with our sampling technique are fast and robust in sparse restoration of seismic wavefield. In Section 6.4, a field data example is presented. The performance of different methods for the practical data are compared. Some remarks on classical matching pursuit and orthogonal matching pursuit methods are addressed in Section 6.5. Finally in Section 7, we discuss the potential usage of the compressive sensing and other interesting study fields.

2. The basics of compressive sensing

The *Nyquist/Shannon* sampling theorem as the foundation of digital signal and image processing is widely used in many areas. But it takes a lot of time in computation and storage since the amount of sampled data is huge. Because of the restriction of sampling time and sampling numbers in applications, the sampled data usually disobey this theorem. Therefore, it is necessary to find new sampling methods to break the restriction.

If the sampling process is linear, then it can be written as

$$b = \Phi f, \tag{6}$$

where $f \in \mathbb{R}^N$ is the original signal, $b \in \mathbb{R}^M$ is the sampled data and $\Phi \in \mathbb{R}^{M\times N}$ is the sampling matrix. If the sampled data is incomplete, i.e. $M < N$, then Equation (6) is under-determined. This indicates that inversion of (6) is ill-posed.

The recently posed compressive sensing theory can restore the original data from the sub-sampled data. However, this theory is based on two basic conditions. First, the

original signal f must be sparse under some transform, i.e. f can be expressed as $f = \Psi s$, where s has a few non-zero entries. The Fourier transform is commonly used in signal processing, and signals are usually decomposed in the frequency domain. The discrete cosine transform and wavelet transform are two famous transforms for signal/image processing. The wavelet transform can also be used to restore a medical image using magnetic resonance imaging (MRI) [29]. For seismic data processing, the commonly used transforms include the Fourier transform [17,18], the linear Radon transform [20], the parabolic Radon transform [21] and the Gabor transform. If there exists a transform Ψ such that $s = \Psi^T f$ is sparse, then Equation (6) can be changed into

$$b = As, \tag{7}$$

where $A = \Phi\Psi$. Another condition for compressive sensing is that A must satisfy the restricted isometry property (RIP) [30,31]. If Φ is the Gaussian random matrix, the partial Fourier matrix, the uniform spherical matrix, the binary random matrix, the partial Hadamard matrix or the Toeplitz matrix, then A satisfies the RIP. If $K < C \cdot M/\log(N/M)$, the sparse solution can be solved, where C is a universal constant and K the number of non-zero elements of s [31].

If the above two conditions are satisfied, we can find s by solving Equation (7), which is actually an optimization problem

$$\min \|s\|_0, \quad \text{s.t. } As = b, \tag{8}$$

where $\|\cdot\|_0$ denotes the number of non-zero entries. As a combinatorial optimization, solving this problem essentially requires exhaustive searches over all subsets of columns of A [30,32].

3. Sparse transforms

3.1. *Transform classification*

Sparse transform is an important part of compressive sensing. If the coefficients in the transform domain are very sparse, then only small sampling numbers are enough. In the seismic processing, the most commonly used transforms are the Fourier transform, the linear Radon transform, the parabolic Radon transform and the curvelet transform. The linear Radon transform can focus the energy of linear events; the parabolic Radon transform can compress events with parabolic shapes. Although applications of wavelets have become increasingly popular in scientific and engineering fields, traditional wavelets perform well only at those points which possess singularities, and the geometric properties, e.g. waveforms of wave fronts in seismic data are not utilized. The ridgelet transform can deal with the linear events, but fail with curve events. Curvelet transform as a multi-scale ridgelet transform, and is multi-scale, multi-directional, anisotropic and local [33]. Curves are approximated by piecewise line segments in the curvelet domain; therefore, seismic signals can sparsely be compressed.

3.2. *The curvelet transform*

Similar to the wavelet and ridgelet transforms, curvelet transform can be represented by inner product of the curvelet functions φ and the signal $f(x)$

$$c(j,l,k) = \langle f, \varphi_{j,l,k} \rangle = \int_{R^2} f(x)\overline{\varphi_{j,l,k}(x)}\mathrm{d}x, \tag{9}$$

where $f \in L^2(\mathbb{R}^2)$, $\varphi_{j,l,k}$ denotes curvelet function and $\bar{\varphi}$ is the conjugate of φ, which is indexed by three parameters: a scale parameter 2^{-j}, $j \in \mathbb{N}_0$ ($\mathbb{N}_0$ is the positive integer set), a sequence of rotation angles $\theta_{j,l} = 2\pi l \cdot 2^{-\lfloor j/2 \rfloor}$, $0 \le l \le 2^{-\lfloor j/2 \rfloor} - 1$ and a position $x_k^{(j,l)} = R_{\theta_{j,l}}^{-1}(k_1 2^{-j}, k_2 2^{-\lfloor j/2 \rfloor})^T, (k_1, k_2) \in \mathbb{Z}^2$ ($\mathbb{Z}$ denotes the integer set), where

$$R_{\theta_{j,l}} = \begin{pmatrix} \cos\theta_{j,l} & \sin\theta_{j,l} \\ -\sin\theta_{j,l} & \cos\theta_{j,l} \end{pmatrix}$$

is the rotation matrix with angle $\theta_{j,l}$. Curvelets consist of directional entries $\varphi_{j,l,k}(x)$ in fine scales and isotropic entries $\varphi_{j_0,k}(x)$ in coarse scales. In fine scales, the curvelet function can be written as

$$\varphi_{j,l,k}(x) = \varphi_j\Big(R_{\theta_{j,l}}(x - x_k^{(j,l)})\Big), \tag{10}$$

while in coarse scales, the curvelet function can be denoted as $\varphi_{j_0,k}(x) = \varphi_{j_0}(x - 2^{-j_0}k)$, which indicates that the curvelets are isotropic in the coarse scale. The function $\varphi_j(x)$ is the waveform by means of its Fourier transform $\hat{\varphi}_j(\omega)$, which serves as a 'mother' curvelet in the sense that all curvelets at scale 2^{-j} are obtained by rotations and translations of φ_j. Numerical implementations of curvelet transform can be outlined in three steps: applying 2D FFT, product with frequency windows and applying 2D inverse FFT for each window. Through the curvelet transform, the original signals are decomposed into various scales and angles. Its discrete form can be written as $c = Sf$, where c is a vector denoting the discrete set of curvelet coefficients, f is the discrete form of the data and $S = W_F F_2$ is the curvelet transform matrix, where F_2 is the 2D Fourier transform matrix and W_F denotes the windowing operator followed by 2D inverse Fourier transform in each scale and in each direction. The computational cost of the forward and inverse curvelet transform is $\mathcal{O}(N^2 \log N)$ for an $N \times N$ data. We refer to [34,35] for details of the implementation of the curvelet transform by involving FFT and IFFT. The inverse curvelet transform can be written as $f = S^* c$, where S^* denotes the adjoint operator of S. Since seismic data are sparse under the curvelet transform, we use it as the sparse transform in this article.

4. Sampling

The sampling technique is important for compressive sensing. It is closely related to the quality of restoration, e.g. correct sampling in MRI and radar imaging [29] can improve the quality of restoration. In seismology, because the receivers can only receive the signals on the ground, traditional sampling methods cannot be employed in seismic exploration directly. There are mainly two sub-sampling methods: the regular sub-sampling and the random sub-sampling. In the former, the receivers are located at equal distance, but they may not satisfy the *Nyquist/Shannon* theorem. A direct consequence of the regular sub-sampling is the coherent aliasing which is not straightforward for transform-based restoration and is difficult to remove. In the random sub-sampling, the receivers are placed randomly on the survey line. The aliasing caused by random sub-sampling in the frequency domain can be seen as random noise, thus the restoration problem can be treated as a de-noising problem [25]. If the gap size of random sampling is large, the recovered data will lose some information. Therefore, a sampling method which can control the gap size and preserve the randomness is necessary for seismic restoration.

The jittered sub-sampling proposed in [22] can both control the gap size and preserve the randomness, so it is better than the random sampling. However, this method is not flexible. In view of these sampling methods, a piecewise random sampling method, developed by us recentley, will be utilized. This sampling technique can both control the gap size and keep the randomness and is flexible. For the random sampling, the probability of been sampled is the same for each position. If the *Nyquist* sampling number is N, and the actual chosen number is K, then the largest gap may be $N-K$. If the gap size is larger than the scale of the curvelet, then some information cannot be restored. Instead, if the N positions are first divided into M pieces, the length of each piece is N/M, and random sampling with the ratio of K/N is in each piece, then the largest gap size is $2(N/M)(1-K/N)$, therefore, the gap size is controlled. And at the same time, the randomness is also retained.

The piecewise sampling method can be described as follows: (i) divide the N positions into M pieces, where the length of each piece N/M should be less than a scale of the curvelet; (ii) randomly choose K/M positions at each piece, thus the total sampling number is K. If the number of pieces being divided is large enough, then the gap size is small. Thus, it will improve the quality of restoration.

It deserves attention that in a recent work, Naghizadeh and Saachi [23] succeed in reconstructing regularly decimated aliased seismic data based on choosing a special mask function. This method relies on choosing a user-defined threshold or a nearest neighbour operator. Although additional effort is added, it is a promising sampling method.

5. Methods of solution

5.1. *Classical methods*

Finding the sparse solution of under-determined problems has been studied in many areas, meanwhile the commonly used methods are based on the l_1 norm optimization. These methods find the sparse solution by solving

$$\min \|s\|_1, \quad \text{s.t. } As = y, \tag{11}$$

where $\|\cdot\|_1$ denotes the l_1 norm. This problem can be changed into linear programming, and then solved by interior point methods [1,36–38]. Suppose that the sampling process has additive noise n, we obtain $b = As + n$; therefore, the corresponding problem becomes

$$\min \|s\|_1, \quad \text{s.t. } \|As - y\|_2 < \epsilon, \tag{12}$$

where ϵ is a nonnegative real parameter which controls the noise level in the data. This problem can be solved as a second-order cone program. Problems (11) and (12) are closely related to the following two problems:

$$\min \|As - y\|_2^2 + \lambda \|s\|_1 \tag{13}$$

and

$$\min \|As - y\|_2^2, \quad \text{s.t. } \|s\|_1 < \delta, \tag{14}$$

where λ is the Lagrange multiplier. With appropriate parameter choices of ϵ, λ and δ, the solutions of (12), (13) and (14) coincide, and these problems are in some sense equivalent.

A lot of methods can be used to solve problems (11)–(14). The homotopy method was originally designed for solving noisy over-determined l_1-penalized least squares problems [39]. It was also used to solve under-determined problems in [40]. This method for the model problems (11)–(14) has been considered in [41,42]. The least angle regression method as a modification of the homotopy method, considered in [43], was investigated solving problem (14). When the solution is sufficiently sparse, these two methods are more rapid than general-purpose linear programming methods even for the high-dimensional cases. If the solution is not rigorously sparse, the homotopy method may fail to find the correct solution in certain scenarios.

Another popularly used method is the interior point method. In [1], an interior point method based on equality and inequality constraints are addressed. In [38], problems (11) and (13) are solved by first reformulating them as 'perturbed linear programs', then applying a standard primal-dual interior point method. The linear equations are solved by iterative methods such as LSQR or conjugate gradient method (CGM). Another interior point method given in [44] is based on changing (13) into a quadratic programming (QP) problem solved by preconditioned CGM. Problem (12) can also be solved by reconsidering it as a second-order cone programming and then applying a log-barrier method, e.g. methods developed in [36,45] are used to solve the land surface parameter retrieval problems.

Iterative soft thresholding methods were used to solve problem (13) [46]. But they are sensitive to the initial values and are not always stable. In addition, many iterations are required for convergence.

Recently, the spectral gradient-projection method was developed for solving problem (12) [27]. The method relies on root-finding of the parameter δ by solving the non-linear convex, monotone equation $\|As(\delta) - y\|_2 = \epsilon$. This method can handle both noisy and noiseless cases. However, it is clear that the root-finding method is the famous 'discrepancy principle' in regularization theory for ill-posed problems [47].

Matching pursuit (MP) and orthogonal matching pursuit (OMP) [48–50] are also used for sparse inversion. This is known as a kind of greedy approach. The vector b is approximated as a linear combination of a few columns of A, where the active set of columns to be used is built in a greedy fashion. At each iteration, a new column is added to the active set. OMP includes an extra orthogonal step which is known to perform better than the standard MP. The computational cost is small if the solution is very sparse. However, only small dimensional problems are suitable for these methods. If $Ax = b$, with x being sparse and the columns of A being sufficiently incoherent, then OMP finds the sparsest solution. It is also robust for small levels of noise [49].

The projected gradient method developed in [51] express (13) as a nonnegative constraint quadratic programming by separating s as $s = (s)_+ - (-s)_+$, where $(s)_+ = \max\{s, 0\}$. However, this method can only tackle real numbers, thus it is unsuitable for curvelet-based seismic data restoration. The iterative re-weighted least squares method developed in [52] changes the weights of entries of s at each iteration. A similar method was proposed in [53].

Methods based on non-convex objects are studied in [54,55]. In [54], the l_0 norm is replaced by a non-convex function. However, initial values must be carefully chosen to prevent the local optimal solution. Other methods such as iterative support detection method [56] and fix point method are also developed.

As an inverse problem with *a priori* knowledge the above problems fall into the general $l_p - l_q$ model [57]:

$$\min \|As - b\|_p^p + \lambda \|s\|_q^q, \quad p, q \geq 0. \tag{15}$$

Regularizing active set (projection) method can be applied.

These methods have extensive applications in various fields, such as image reconstruction, MRI [29], seismic images [22], model selection in regression [43] and texture/geometry separation.

5.2. *Two fast methods for seismic restoration*

Usually, the seismic restoration is a large-scale problem, thus fast solving methods are crucial for seismic processing. Because of the large dimension of seismic data, traditional methods such as interior methods [38], homotopy methods [40] and least angle regression methods [43] cannot be used to large-scale seismic restoration. The iterative soft thresholding (IST) method was introduced to recover the seismic data in [22]. A robust spectral gradient-projection (SPGL1) method was proposed in [27] to solve problem (14). In [58], the iterative re-weighted least squares was used for interpolation. Unfortunately, these methods have slow convergence rate for large-scale problems [59], therefore, it is urgent to find fast convergent methods.

We develop two fast methods for seismic restoration in this section. They are very fast compared with the above three methods. The CPU time takes only about 1/3 of the SPGL1 method, and about 1/4 of the IST method. Numerical experiments in Section 6 verify the quick convergence of the developed methods.

5.2.1. *A smooth l_0 method*

The original problem for l_0 minimization is the following equality constrained optimization problem:

$$\min \|x\|_0, \quad \text{s.t. } Ax = b. \tag{16}$$

However, direct solution of (16) is hard to obtain because it is time consuming. Obviously, solving the true l_0 norm optimization is superior to the l_1 norm optimization though both methods can yield sparse solutions. We consider approximation of l_0 minimization problem. Denote $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$ as a function of σ and t. This function satisfies the following properties: (a) $f_\sigma(t)$ is continuous and differentiable; (b) $f_\sigma(t)$ tends to the l_0 norm when σ tends to 0, i.e.

$$\lim_{\sigma \to 0} f_\sigma(t) = \begin{cases} 0, & t = 0, \\ 1, & t \neq 0. \end{cases} \tag{17}$$

Thus, we can construct a continuous function to approximate the l_0 quasi-norm, and then obtain the optimal solution. In this way, problem (16) is approximated by

$$\min J_\sigma(x) := \sum_{i=1}^{N} f_\sigma(x_i), \quad \text{s.t. } Ax = b. \tag{18}$$

The object function $J_\sigma(x)$ is differentiable and is closely related to the parameter σ: the smaller the value of σ, the closer the behaviour of $J_\sigma(x)$ to the l_0 quasi-norm. For small values of σ, $J_\sigma(x)$ is highly non-smooth and contains a lot of local minima, hence its minimization is not easy. On the other hand, for larger values of σ, $J_\sigma(x)$ is smoother and contains fewer local minima, and its minimization is easier. Practically, we use a decreasing sequence values of σ: for minimizing $J_\sigma(x)$ for each value of σ, the initial value of the

minimization algorithm is the minimum of $J_\sigma(x)$ from the previous value of σ. Then we apply a simple projected gradient method to solve Equation (18). Details of the procedure are given in Algorithm 1. For the convergence of a similar method for general signal processing problems, we refer to [54] for details. Below we present a similar algorithm to [54] based on the above l_0 norm approximation and use it for seismic wavefield restoration.

Algorithm 1 (Smooth l_0 method)

(1) Initialization:

  (1) Let $\hat{x}_0$ be the minimum l_2-norm solution of $Ax = b$, which can be obtained by applying the pseudo-inverse of A.
  (2) Choose the inner loop number L, outer loop number J and the step-length τ; set a decreasing sequence values of $\sigma : [\sigma_1, \ldots, \sigma_J]$.

(2) Iteration: for $j = 1, \ldots, J$

  (1) Let $\sigma = \sigma_j$.
  (2) Minimize the function $J_\sigma(x)$ on the feasible set $\mathcal{S} = \{x | Ax = b\}$ using L iterations of gradient descent method.

    ① Let $x = \hat{x}_{j-1}$;
    ② For $l = 1, \ldots, L$:

  (a) Let $g_\sigma = [\nabla J_\sigma(x_1), \nabla J_\sigma(x_2), \ldots, \nabla J_\sigma(x_n)]$.

  (b) (The gradient decent iteration): $x = x + \tau d$ (τ is the step-length, $d = \gamma(g_\sigma)$).
  (c) (Projection): Project x on the feasible set $\mathcal{S} = \{x | Ax = b\}$:

$$x = x - A^T(AA^T)^{-1}(Ax - b).$$

  (3) Set $\hat{x}_j = x$, $\sigma = \sigma/2$.

(3) Final solution is $\hat{x} = \hat{x}_J$.

We can also choose other functions to approximate the l_0 quasi-norm, e.g. the 'truncated hyperbolic' function:

$$f_\sigma(t) = \begin{cases} 0, & |t| \le \sigma, \\ 1 - (t/\sigma)^2, & |t| \ge \sigma \end{cases} \tag{19}$$

and

$$f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2). \tag{20}$$

Remark 1 In Step 2 of Algorithm 1, if the gradient descent step is based on steepest descent (SD) step, i.e. $\gamma(g_\sigma) = -g_\sigma$ and $\tau = \tau^{\mathrm{SD}}$, then the algorithm corresponds to projected steepest descent method.

In addition, in Algorithm 1, the inner loop number L need not be too large, and according to our experience, the step-length τ should be greater than 2 to ensure fast convergence. However, it is clear that this choice of the step-length is not optimal. Usually, we need to calculate an optimal step-length τ^* by line search for the one-dimensional problem $\tau^* = \mathrm{argmin}_\tau J_\sigma(x + \tau d)$. One may readily see that some fast gradient methods

based on non-monotone gradient descent step can be applied, e.g. the Barzilai–Borwein step used in seismic migration inversion [60], where the function $\gamma(g_\sigma)$ is updated by the former iterative information instead of the current iterative information.

For curvelet-based restoration, because of the orthogonality of the curvelet transform, the inverse of AA^T is the identify matrix, thus the projection onto $\mathcal{S} = \{x|Ax = b\}$ can be simply solved by $x = x - A^T(Ax - b)$.

5.2.2. *Approximations of the l_1 norm optimization*

As it is well-known that the objective function based on l_1 norm is non-differentiable at the original point, most of the solvers for l_1 norm optimization are based on the interior point methods to solve a linear programming problem. In this section, we consider using smooth functions to approximate the l_1 norm. We consider the function

$$f_\epsilon(t) = \sqrt{|t|^2 + \epsilon}, \tag{21}$$

which is continuous, convex and differentiable. If ϵ is small enough, $f_\epsilon(t)$ will approximate the l_1 norm sufficiently, thus it can be used to replace the l_1 norm [24]. Another similar function is

$$f_\theta(t) = \begin{cases} |t|, & |t| \geq 0.01, \\ \dfrac{1}{\theta}[\log(1 + \exp(-\theta t)) + \log(1 + \exp(\theta t))], & |t| < 0.01. \end{cases} \tag{22}$$

If θ is large enough, then f_θ will approximate the l_1 norm. This function has been used for solving non-linear and mixed complementarity problems and feature classification [61,62].

Since the object function based on $f_\epsilon(t)$ or $f_\theta(t)$ is an approximation of the l_1 norm, the errors are inevitable. However, the cost of computation will be much less than the IST method and SPGL1 method. At each iteration of the SPGL1 method, one needs to project the iteration point to the active set specified by the l_1 norm, which will increase the amount of computation [27]; while the gradient projection for sparse reconstruction (GPSR) method cannot deal with complex numbers [51]. We found that the smaller the values of ϵ, the better the approximation of $f_\epsilon(t)$ to the l_1 norm; the smaller the values of θ, the worse the approximation of $f_\theta(t)$ to the l_1 norm. Thus, we can solve the following problems to get the sparse solution:

$$\min F_\epsilon(x) = \sum_{i=1}^{N} f_\epsilon(x_i), \quad \text{s.t. } Ax = b \tag{23}$$

or

$$\min F_\theta(x) = \sum_{i=1}^{N} f_\theta(x_i), \quad \text{s.t. } Ax = b. \tag{24}$$

We use a projected gradient algorithm to solve these two problems. Details are outlined in Algorithm 2. In our algorithm, a projection to $\mathcal{S} = \{x|Ax = b\}$ is also needed. However, the projection is easy to be computed. Since A is orthogonal, the initial l_2 norm solution can be obtained by $x = A^T b$, and the projection can be easily calculated through

$x^{k+1} = x_{pre}^{k+1} - A^T(Ax_{pre}^{k+1} - b)$. However, it is different from the smooth l_0 algorithm because it does not need the outer iteration.

Algorithm 2 (Projected gradient method for approximation of the l_1 norm optimization)

(1) Initialization:

  (1) Choose the maximum loop number L, set $\epsilon = 1.0e-16$ (or $\theta = 1000$), and $k = 0$.
  (2) Let x^0 be the l_2 norm solution of $Ax = b$, which can be obtained by the pseudo-inverse of A.

(2) Iteration:

  (1) Let $g_\epsilon = \nabla F_\epsilon(x^k)$ (or $g_\theta = \nabla F_\theta(x^k)$).
  (2) Perform the gradient decent iteration: $x_{pre}^{k+1} = x^k + \tau d$ (τ is the step length, $d = \gamma(g_\epsilon)$ or $d = \gamma(g_\theta)$, x_{pre}^{k+1} denotes the predicted step).
  (3) Projection: Project x_{pre}^{k+1} on the feasible set $\mathcal{S} = \{x|Ax = b\}$:

$$x^{k+1} = x_{pre}^{k+1} - A^T(AA^T)^{-1}(Ax_{pre}^{k+1} - b).$$

  (4) Let $x^k = x^{k+1}$, $k = k+1$.

(3) Final solution is $x = x^L$.

In Step 2 of Algorithm 2, if the gradient descent step is based on steepest descent step, i.e. $\gamma(g_\epsilon) = -g_\epsilon$ or $\gamma(g_\theta) = -g_\theta$ and $\tau = \tau^{SD}$, then the algorithm corresponds to projected steepest descent method. More advanced choices of the function γ and the step-length τ are given in Remark 1.

6. Numerical examples

6.1. *Piecewise sampling vs. random sampling*

Considering a synthetic shot data of a layered earth model with six plane layers, modelled with a 15 m receiver interval, 2×10^{-3} s sampling interval and a source function given by a Ricker wavelet with a central-frequency of 15 Hz. The dataset contains 256 traces of seismic data with 256 time samples in each trace. The original shot gather and its frequency spectrum are shown in Figure 1(a) and (b). The resulting shot gather after regular sub-sampling and its corresponding frequency spectrum are shown in Figure 1(c) and (d). In this simulation, the sampling number is 1/3 of the *Nyquist* sampling number. Serious coherent aliasing can be seen in Figure 1(d). With the same number of regular sampling, the random sampling and its frequency spectrum are shown in Figure 2(a) and (b). It reveals that the aliasing in Figure 2(b) is just like random noise; the recovery results by SPGL1 and the corresponding frequency spectrum are given in Figure 2(c) and (d). To show the degree of restoring ability, we adopt the signal-to-noise ratio (SNR) defined by $SNR = 10 * \log(\frac{\|\text{data}_{\text{orig}}\|^2}{\|\text{data}_{\text{orig}} - \text{data}_{\text{rest}}\|^2})$. The SNR using the random sampling is about 7.1606, where $\text{data}_{\text{orig}}$ denotes the complete data and $\text{data}_{\text{rest}}$ denotes the recovered data. With the same number of regular sampling, the jittered sampling and its frequency spectrum are shown in Figure 3(a) and (b); the restoration and the restored frequency spectrum are given in Figure 3(c) and (d). The SNR using the jittered sampling is 9.3008. We also

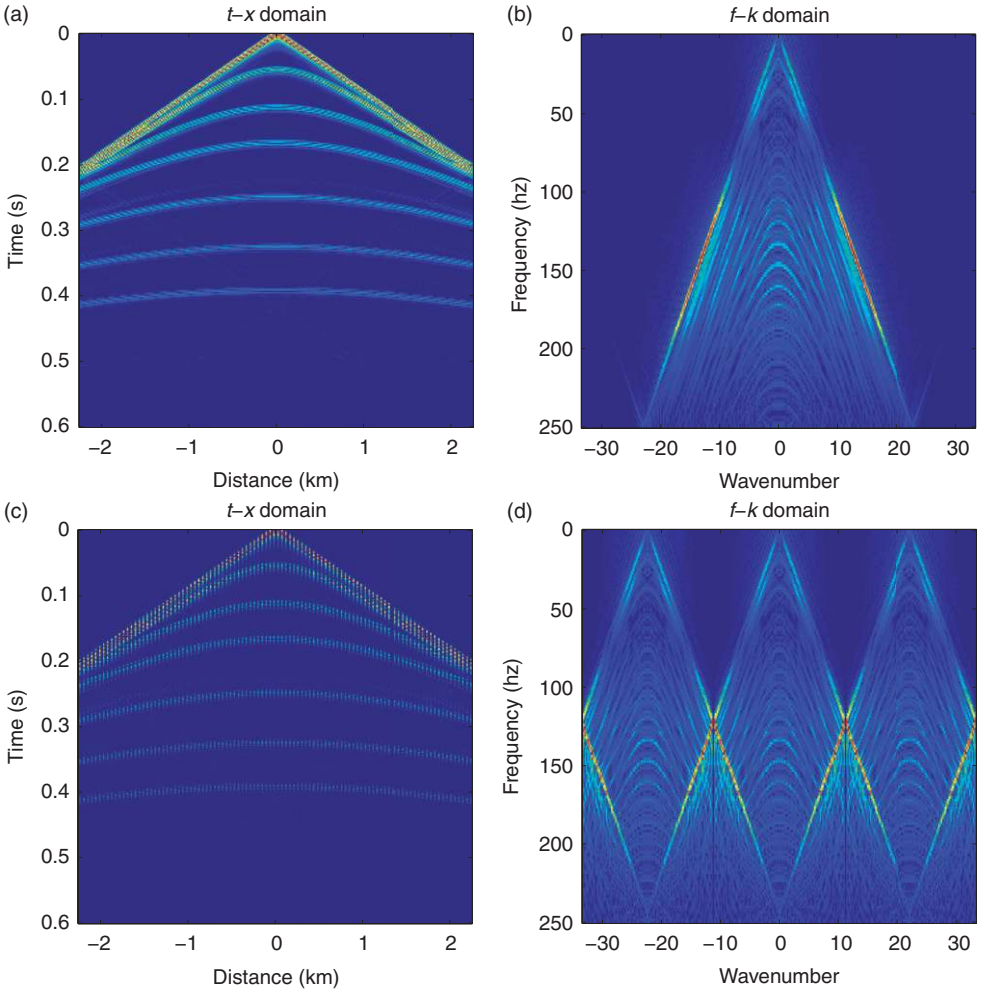

Figure 1. (a) The original data and (b) its frequency; (c) the regular sub-sampling and (d) the frequency of the sub-sampled data.

perform our piecewise sampling simulations. The piecewise sampling and its frequency spectrum can be seen in Figure 4(a) and (b); the restoration results and the restored frequency are shown in Figure 4(c) and (d). The SNR of restoration using the piecewise sampling is 9.8417. The numerical performance indicates that the piecewise sampling can improve the restoration greatly.

6.2. *Results of the smooth l_0 method*

We give examples to show the superiority of the smooth l_0 method to the IST and SPGL1 methods. The synthetic data consists of 300 traces with spacing 15 m. The temporal sample interval is 2×10^{-3} s and the sampling number is 300. The full data and the incomplete acquisition are depicted in Figure 5(a) and (b), respectively, where the sampling number in

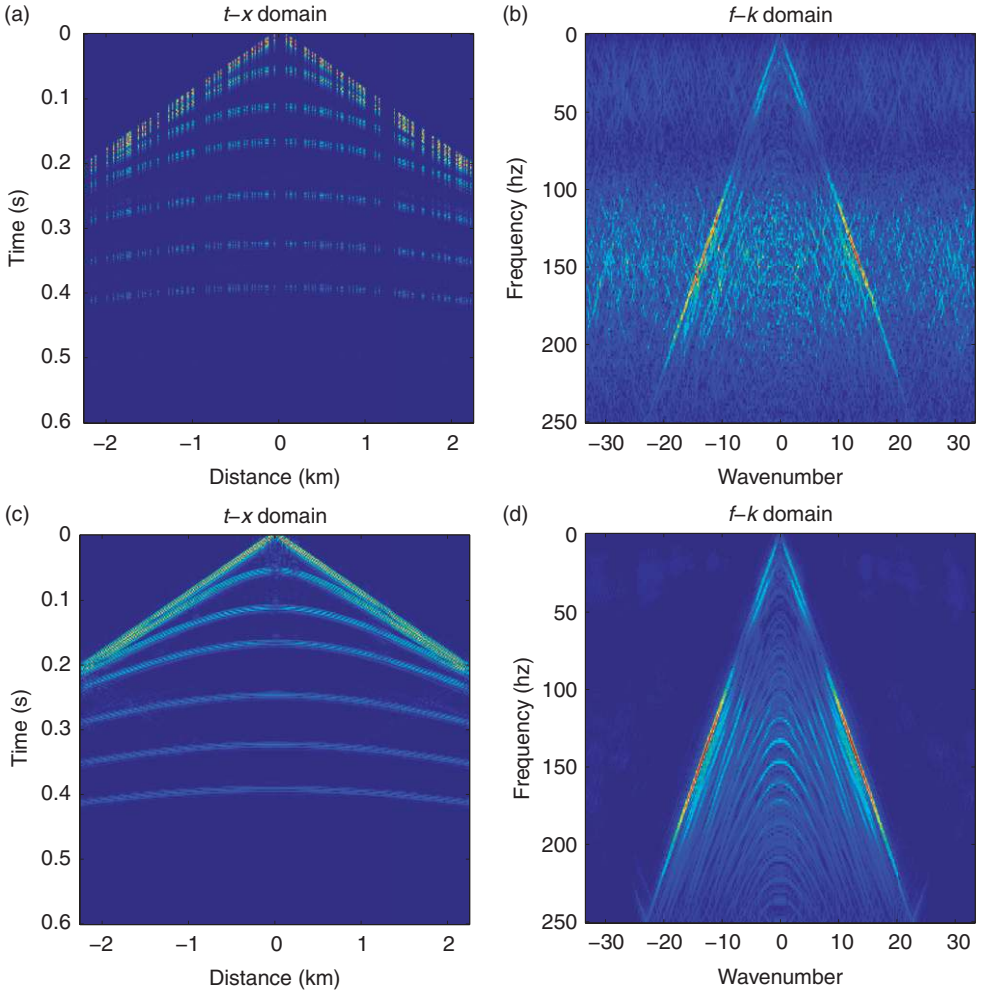

Figure 2. (a) The random sampling and (b) its frequency; (c) the restoration results and (d) the frequency of the restored data.

Figure 5(b) is half of the *Nyquist* sampling number. The restoration results by the smooth l_0 method and the difference between the restoration and the original data are shown in Figure 6, where we choose $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$. The restoration results by the smooth l_0 method using the approximate function $f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2)$ and difference between the restoration and the original data are given in Figure 7(b). Similar results of smooth l_0 method using the approximate function

$$f_\sigma(t) = \begin{cases} 0, & |t| \le \sigma, \\ 1 - (t/\sigma)^2, & |t| \ge \sigma \end{cases}$$

are shown in Figure 8. The restoration results by the IST method and the difference between the restoration and the original data are shown in Figure 9. While the results using the SPGL1 method are shown in Figure 10. Table 1 summarizes details of the SNR,

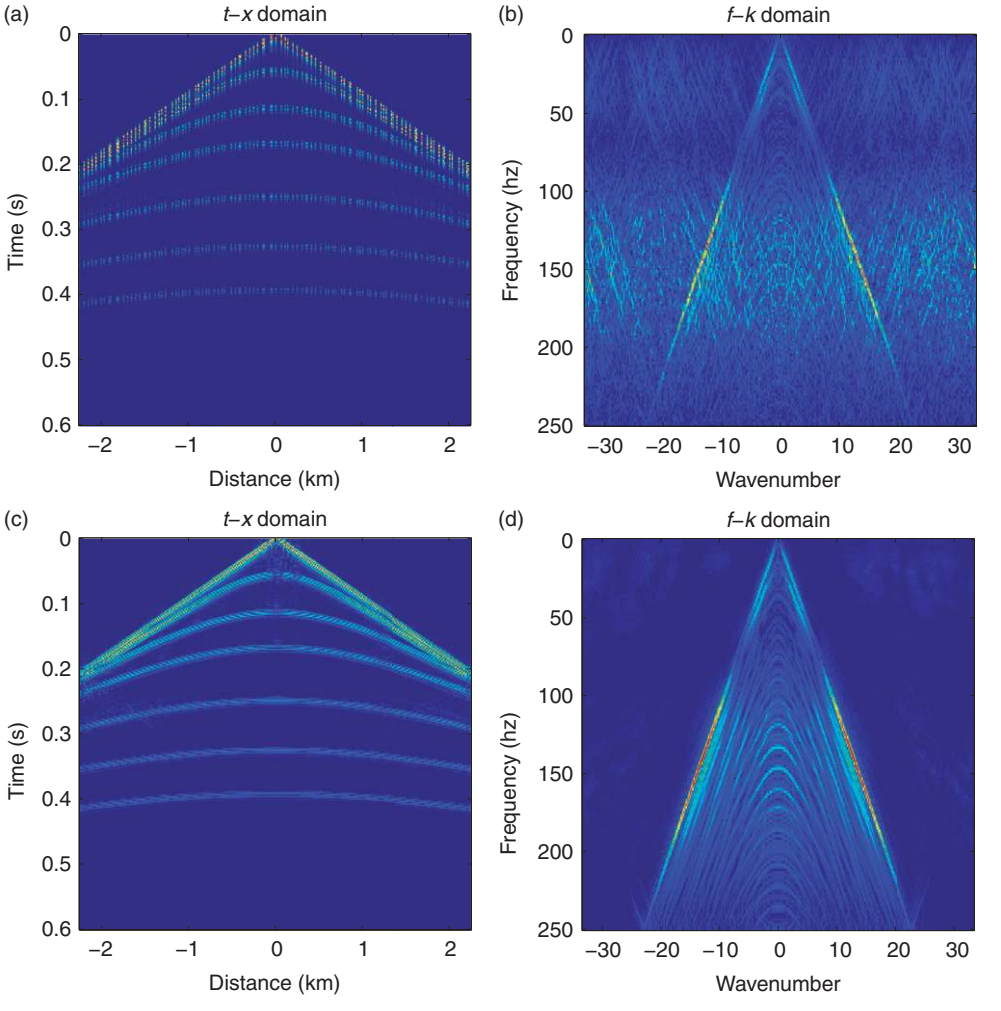

Figure 3. (a) The jittered sampling and (b) its frequency; (c) the restoration results and (d) the frequency of the restored data.

the relative error and the cost of computation of these methods. The results show that the smooth l_0 method is much faster than the IST and SPGL1 methods. It uses the CPU time about 1/4 of the IST method, and about 1/3 of the SPGL1 method to yield the similar results. Thus, we conclude that the smooth l_0 method is suitable for seismic data restoration.

6.3. *Results of the l_1 norm optimization using projected gradient methods*

We use the same full data and sampled data as in Figure 5 and perform Algorithm 2. The number of sampling is half of the *Nyquist* number. When the l_1 norm is replaced by $F_\epsilon(x)$, where $\epsilon = 1.0e - 20$, then after 30 iterations, the SNR of the restored results approaches 22.8120 and the CPU time is 229 s. The restoration results and the difference between the

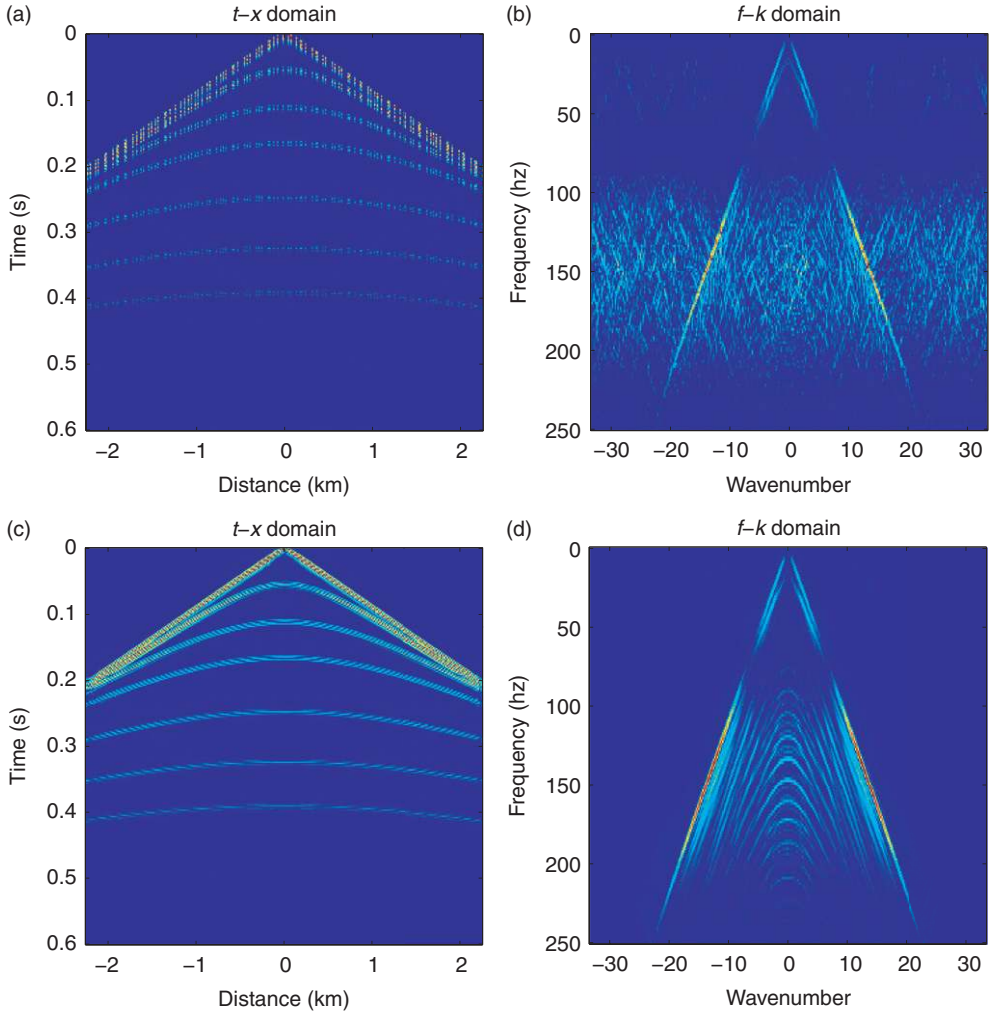

Figure 4. (a) The piecewise sampling and (b) its frequency; (c) the restoration results and (d) the frequency of the restored data.

restoration and the original data are shown in Figure 11. When $F_\theta(x)$ is used to replace the l_1 norm, where $\theta = 1000$, then after 30 iterations, the SNR approaches 23.2334 and the CPU time is 241 s. The restoration results and the difference between the restoration and the original data are shown in Figure 12. It is evident that both the convergence rate is fast and the solution is acceptable. Though the restoration has random noise, the key information can still be restored; therefore, this method can be applied to seismic data restoration. However, to better apply this method to seismic data processing, the next major work is to remove the random noise. Some tricks may be used: (i) normalization of the original data, which will lead to small coefficients in the curvelet domain; (ii) since curvelet coefficients are focused on the first half of the vector, the latter half part of vector can be truncated to boost the SNR.

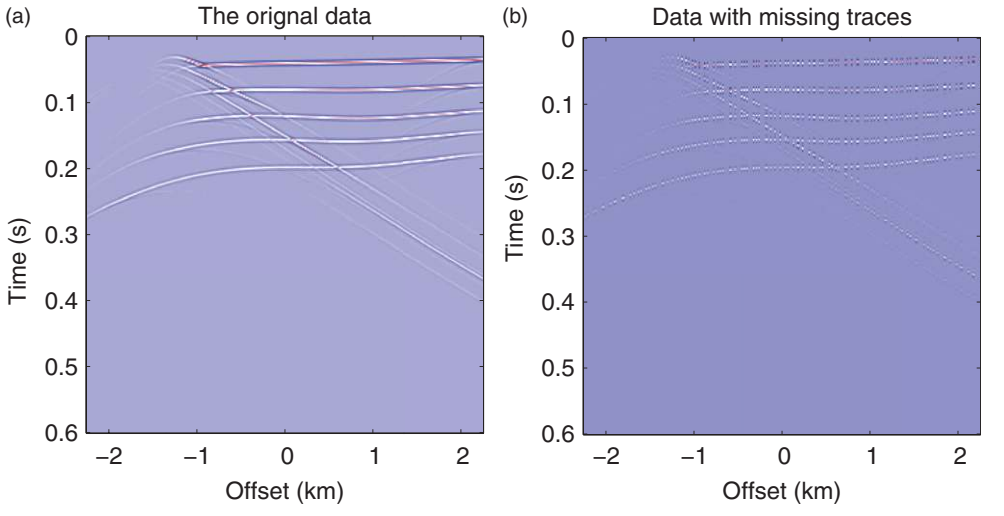

Figure 5. (a) The original data; (b) the sampled data with randomly removed half of the receiver positions.

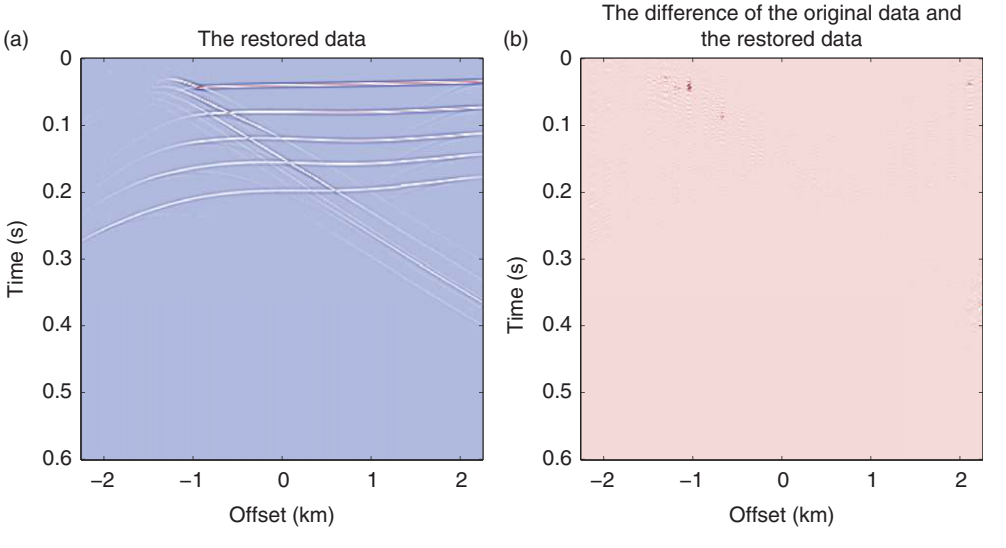

Figure 6. (a) Restoration results by the smooth l_0 method with $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$; (b) the difference between (a) and the original data.

6.4. *Field data applications*

We further examine the efficiency of the new methods with field data. A marine shot gather is provided in Figure 13(a) which consists of 160 traces with spacing 25 m and 800 time samples with interval 2×10^{-3} s. There are damaged traces in the gather. The sub-sampled gather is shown in Figure 13(b) with half of the original traces randomly deleted. This sub-sampled gather was used to restore the original gather with different methods. The restoration using the smooth l_0 method with $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$ is displayed in

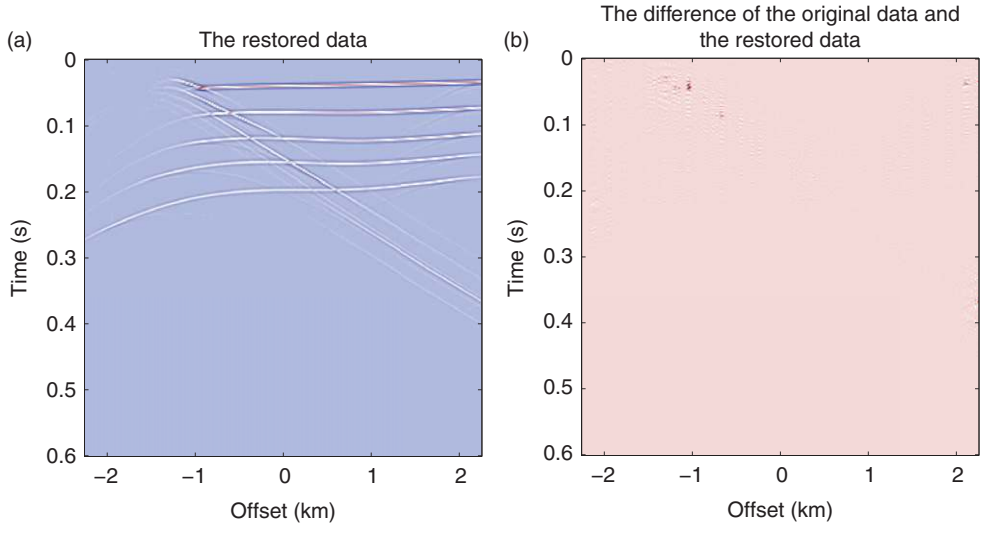

Figure 7. (a) Restoration results by the smooth l_0 method with $f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2)$; (b) the difference between (a) and the original data.

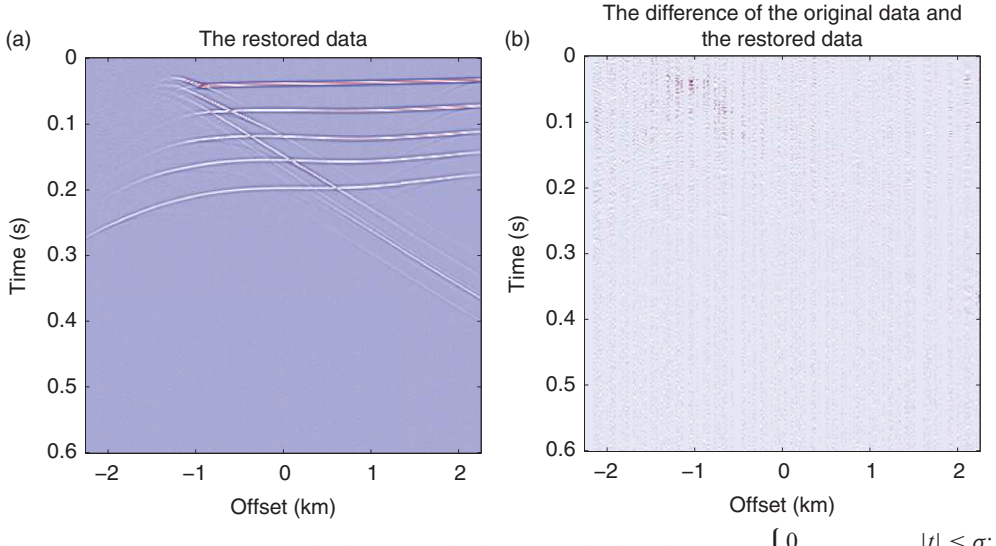

Figure 8. (a) Restoration results by smooth the l_0 method with $f_\sigma(t) = \begin{cases} 0, & |t| \le \sigma; \\ 1 - (t/\sigma)^2, & |t| \ge \sigma. \end{cases}$; (b) the difference between (a) and the original data.

Figure 14(a), while the restoration with $f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2)$ is given in Figure 14(b). Meanwhile, the restoration results using the SPGL1 and IST methods are shown in Figure 15. Comparison of the SNR, relative error and computational time of these methods are shown in Table 2. These results show that the smooth l_0 method with the projected gradient algorithm is much faster than the IST and SPGL1 methods. The CPU time is about 1/3 of the IST and SPGL1 methods. In addition the damaged trace in the original gather was restored as a good trace. Thus, the new methods in this article are

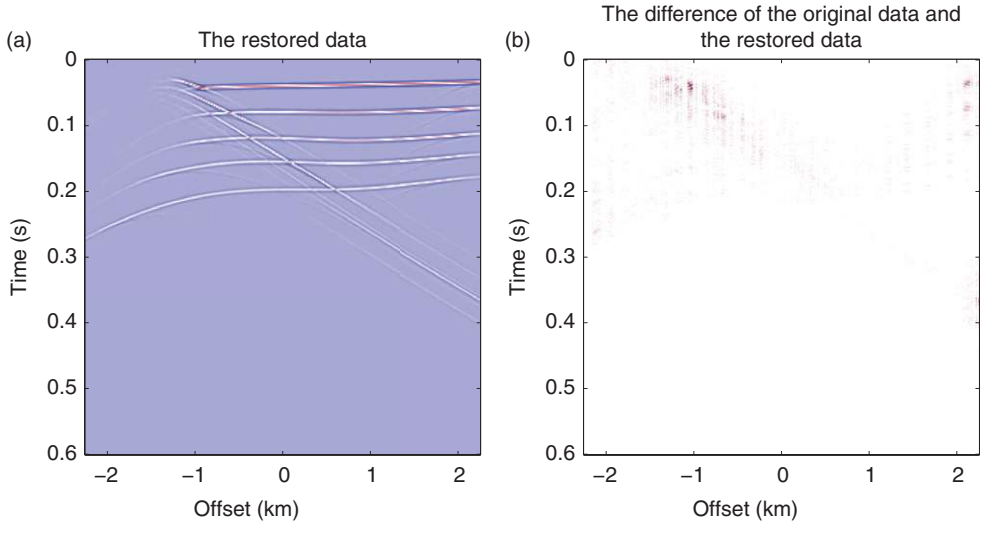

Figure 9. (a) Restoration results by the IST method; (b) the difference between (a) and the original data.

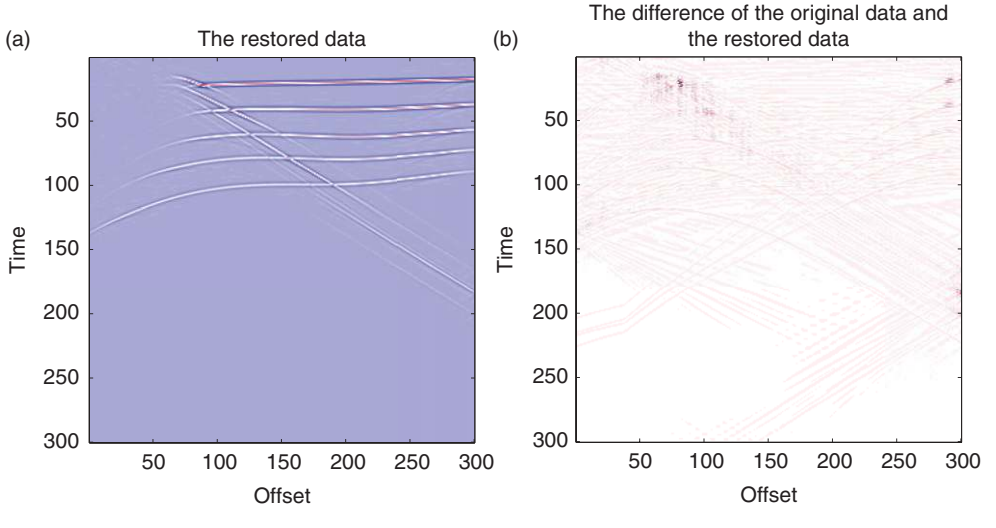

Figure 10. (a) Restoration results by the SPGL1 method; (b) the difference between (a) and the original data.

efficient and can reduce the amount of computational cost greatly for seismic data restoration.

6.5. *Further remarks*

It deserves pointing out that matching pursuit as a greedy algorithm has been widely used in time-frequency analysis. Signals in the time domain can be changed into another

Table 1. Comparison of the smooth l_0 method, IST method and the SPGL1 method.

Algorithm		SNR	Relative error	Time (s)
Smooth	l_0 with $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$	25.5282	0.0529	306
Smooth	l_0 with $f_\sigma(t) = 1 - \sigma^2/(t^2+\sigma^2)$	25.6524	0.0522	318
Smooth	l_0 with $f_\sigma(t) = \begin{cases} 0, & \lvert t\rvert \le \sigma; \\ 1-(t/\sigma)^2, & \lvert t\rvert \ge \sigma. \end{cases}$	23.6445	0.0657	323
IST		24.8597	0.0571	1326
SPGL1		25.1210	0.0555	1069

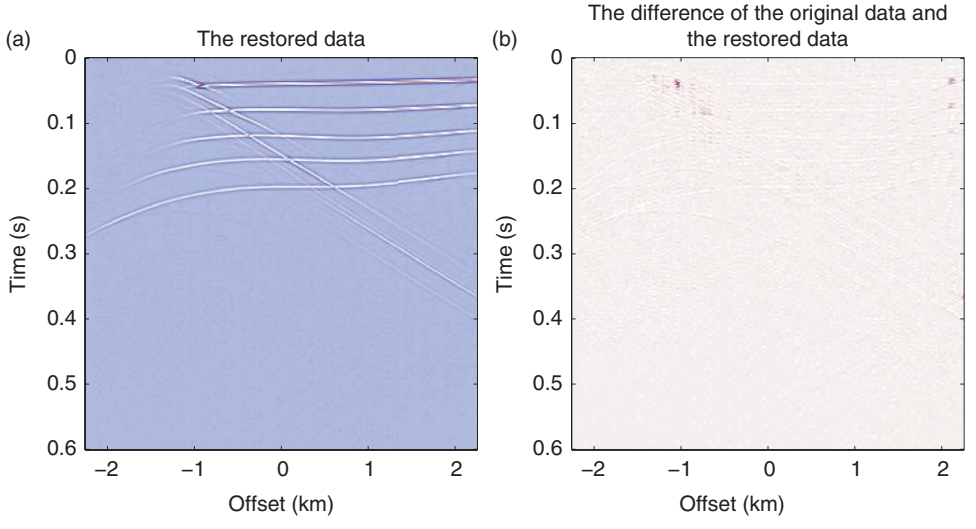

Figure 11. (a) Restoration results using $F_\epsilon(x)$ to replace the l_1 norm; (b) the difference between (a) and the original data.

domain under a suitable transform. These transforms are called a dictionary. Each dictionary is a collection of waveforms. The commonly used transforms are usually over-completed, i.e. the number of atoms is greater than the length of the signals. In discrete case, this process can be written as

$$Ax = b, \tag{25}$$

where A is a $m \times n$ transform matrix, $m < n$. Each column of A is called an atom, b is a signal in the time domain and x is the signal in the transformed domain. Since A is under-determined there are infinite numbers of x satisfying (25).

The matching pursuit consists of two steps in each iteration: choose a new atom and update the residual. Let a_i, $i = 1, \ldots, n$ denote the i-th column of A. The initial residual is $r_0 = b$, the initial solution is $x = 0$. At the k-th step, we choose the atom which has the largest correlation with the residual r_{k-1}

$$a_k = \text{argmax}_{\{a_i\}_{1\le i\le n}} |\langle r_{k-1}, a_i\rangle|, \tag{26}$$

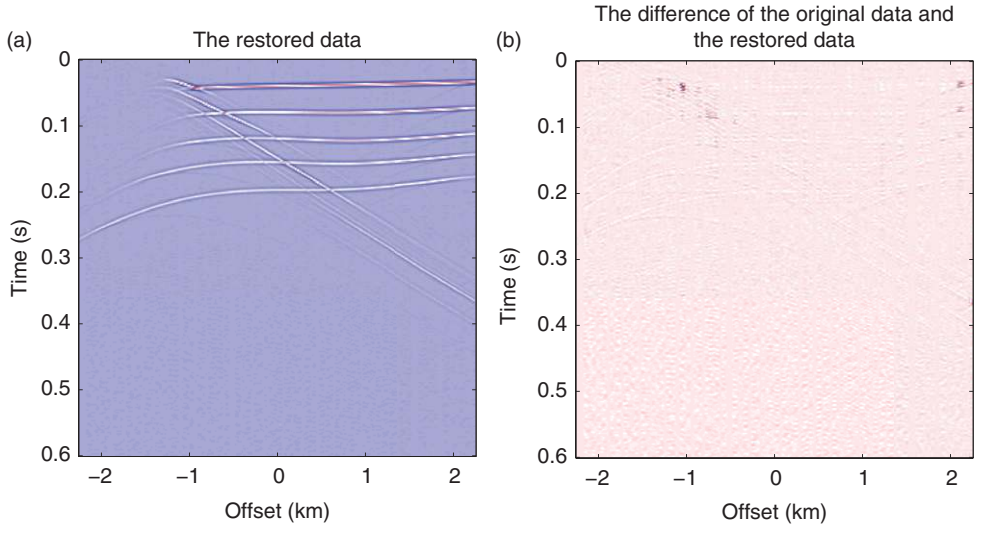

Figure 12. (a) Restoration results using $F_\theta(x)$ to replace the l_1 norm; (b) the difference between (a) and the original data.

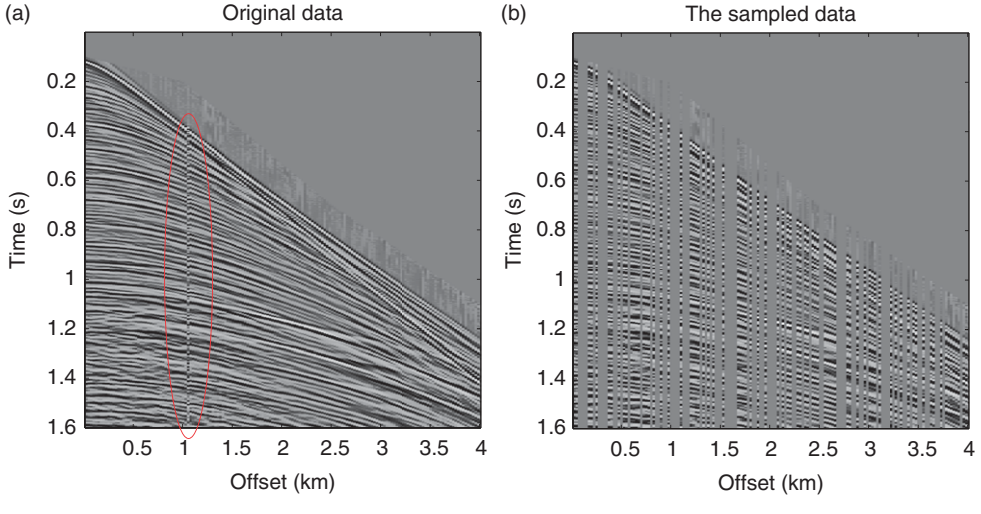

Figure 13. (a) The original marine shot gather; (b) the sub-sampled gather.

which will give the new residual $r_k = r_{k-1} - \langle r_{k-1}, a_k \rangle a_k$. The algorithm of matching pursuit is outlined in Algorithm 3.

Algorithm 3 (Matching pursuit algorithm)

(1) Initialization: set $r = b$, $x = 0$.
(2) Iteration:

- Choose: $y = A^T r$, $k = \max_i |y_i|$.
- Update: $c = \langle r, a_k \rangle$, $x[k] = x[k] + c$, $r = r - ca_k$.

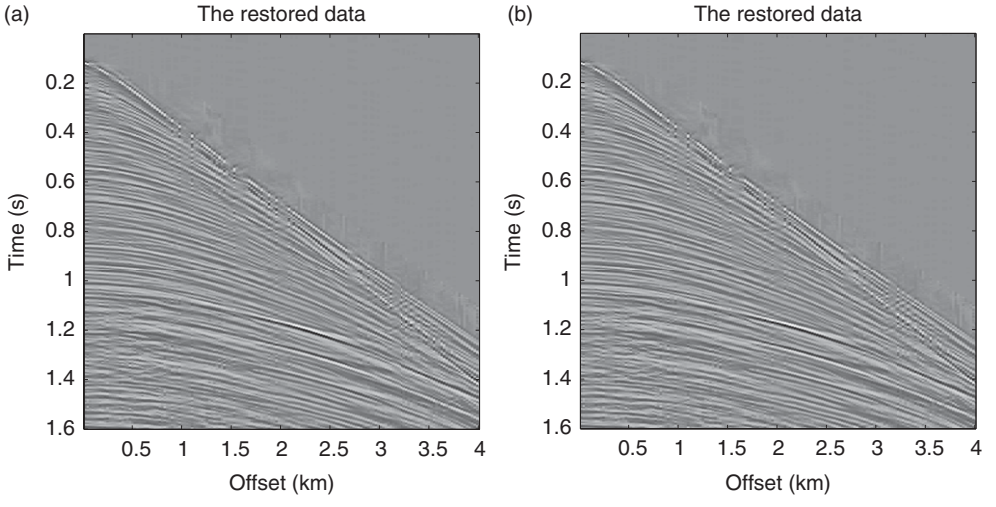

Figure 14. (a) Restoration results by the smooth l_0 method with $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$; (b) restoration results by the smooth l_0 method with $f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2)$.

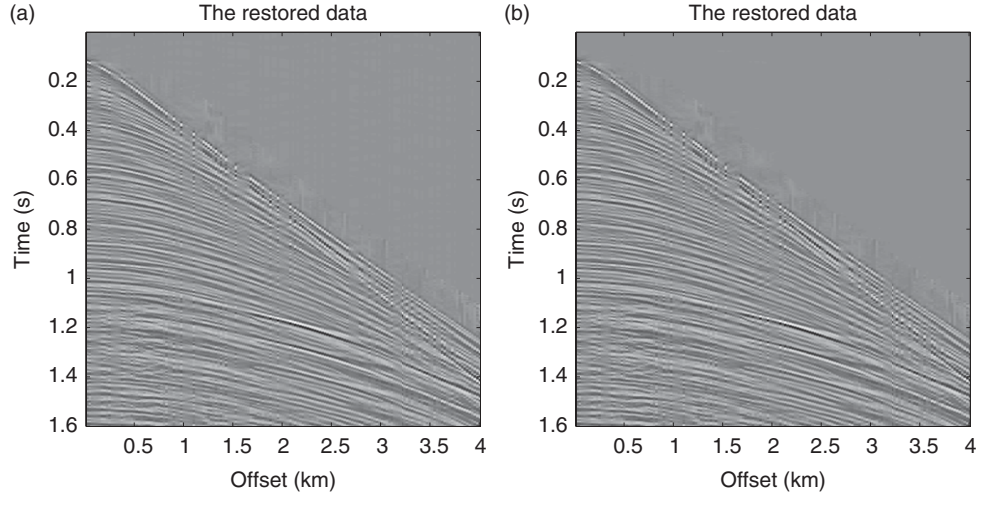

Figure 15. (a) Restoration results by the SPGL1 method; (b) restoration results by the IST method.

Table 2. Comparison of the smooth l_0 method, IST method and the SPGL1 method for field data.

Algorithm		SNR	Relative error	Time (s)
Smooth	l_0 with $f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2))$	6.1309	0.4937	83.985168
Smooth	l_0 with $f_\sigma(t) = 1 - \sigma^2/(t^2 + \sigma^2)$	6.1546	0.4923	91.954318
IST		6.1015	0.4954	269.382163
SPGL1		6.1426	0.4930	255.340876

(3) If the stopping criterion is satisfied, stop; otherwise, go to Step 2.

In Step 3, the stopping criterion is based on the norm of the residual r_k. If the value of $\|r_k\|$ is less than a preassigned small positive number, the iterations are terminated. The algorithm is very simple, but the convergence is slow and needs a lot of iterations. Because the coefficient is not optimized an atom will still be chosen in the following iterations even if it has been selected.

The orthogonal matching pursuit algorithm can overcome this drawback by projecting x onto the subspace spanned by the selected atoms. Since the orthogonal process can remove the energy of the selected atoms, the selected atoms will not be chosen anymore. The atom selection method is the same as in matching pursuit. If x has k non-zero elements, the iteration will need k steps, but the iteration number of matching pursuit is usually larger than m.

For the synthetic example in Section 6.2, the scale of dataset is 300×300. The number of coefficients in the curvelet domain is 649,161. For the matching pursuit method, it takes 361 s for 50 iterations. The number of the curvelet coefficients whose absolute values are larger than 0.01 is 92,901; and the number of the curvelet coefficients whose absolute values are larger than 0.001 is 265,444. This indicates that there are many more non-zero coefficients, which conflicts the sparsity assumption. Therefore, millions of seconds will be spent to find the right solution. For the orthogonal matching pursuit method, $A_k^T A_k$ will be ill-conditioned after 12 iterations, so the solution will be unstable, where A_k denotes the sub-matrix constituted by the selected atoms at the k-th iteration. It spends 1821 s for 30 iterations. Therefore, both methods are not quite efficient for curvelet-based restoration.

7. Conclusion

In this article we use the compressive sensing theory to solve the seismic interpolation and data restoration problem. The curvelet transform is used as the sparse transform, and a piecewise sampling was introduced to improve the quality of restoration. The piecewise sampling method can control the sampling gap and keep the randomness. Numerical results show that this sampling method is superior to the random sampling. In computation, we introduce two fast methods to solve the l_0 and l_1 minimization models. The methods we developed were proved to be much faster than the IST and SPGL1 methods.

We argue that the curvelet transform can be used for de-noising, multiple remove and migration; sampling method and sparse transform have great effect on the restoration while fast methods will improve the restoring efficiency which is quite important for seismic restoration. The three parts are major issues in compressive sensing and deserve further attention for seismic data processing.

Apart from the seismic data restoration, compressive sensing can also be used for high-resolution deconvolution. The high-resolution deconvolution problem was introduced in [63], which is also an under-determined problem, thus it can be solved based on the compressive sensing theory. We believe that the compressive sensing theory will serve wide applications in geophysics.

Acknowledgements

We are grateful to reviewers' and Daniel Lesnic's helpful comments and suggestions on this article. This work is supported by the National Natural Science Foundation of China under grant numbers

10871191, 40974075 and Knowledge Innovation Programs of Chinese Academy of Sciences KZCX2-YW-QN107.

References

[1] Y.F. Wang, C.C. Yang, and J.J. Cao, *On Tikhonov regularization and compressive sensing for seismic signal processing*, accepted (2011), to appear in Math. Mod. Meth. Appl. Sci.
[2] Y.F. Wang, J.J. Cao, and C.C. Yang, *Recovery of seismic wavefields based on compressive sensing by a l_1-norm constrained trust region method and the piecewise random sub-sampling*, Geophys. J. Int. (2011), submitted for publication.
[3] N. Mostafa and D.M. Sacchi, *Multistep autoregressive reconstruction of seismic records*, Geophysics 72 (2007), pp. V111–V118.
[4] S. Spitz, *Seismic trace interpolation in the F-X domain*, Geophysics 56 (1991), pp. 785–794.
[5] M.J. Porsani, *Seismic trace interpolation using half-step prediction filters*, Geophysics 64 (1999), pp. 1461–1467.
[6] R. Soubaras, *Spatial interpolation of aliased seismic data*, Soc. Expl. Geophys. (SEG) Expanded Abstract 16 (2004), pp. 1167–1170.
[7] N. Gulunay, *Seismic interpolation in the Fourier transform domain*, Geophysics 68 (2003), pp. 355–369.
[8] S. Fomel, *Application of plane-wave destruction filters*, Geophysics 67 (2002), pp. 1946–1960.
[9] J. Claerbout, *Earth Soundings Analysis: Processing Versus Inversion*, Blackwell Science, Boston, 1992.
[10] S. Crawley, *Seismic trace interpolation with non-stationary predictionary-error filters*, PhD thesis, Stanford University, CA, 2000.
[11] J. Ronen, *Wave-equation trace interpolation*, Geophysics 52 (1987), pp. 973–984.
[12] R.H. Stolt, *Seismic data mapping and reconstruction*, Geophysics 67 (2002), pp. 890–908.
[13] H. Jakubowica, *Seismic data acquisition*, US Patent 5 (1997), p. 648.
[14] K. Larner, B. Gibson, and D. Rothman, *Trace interpolation and the design of seismic surverys*, Geophysics 46 (1981), pp. 407–409.
[15] M.D. Sacchi, *Estimation of the discrete Fourier transform, a linear inversion approach*, Geophysics 61 (1996), pp. 1128–1136.
[16] M.D. Sacchi, T.J. Ulrych, and C.J. Walker, *Interpolation and extrapolation using a high-resolution discrete Fourier transform*, IEEE Trans. Signal Process. 46 (1998), pp. 31–38.
[17] A.J.W. Duijndam, M.A. Schonewille, and C.O.H. Hindriks, *Reconstruction of band-limited signals, irregularly sampled along one spatial direction*, Geophysics 64 (1999), pp. 524–538.
[18] S. Xu, Y. Zhang, D. Pham, and G. Lambare, *Antileakage Fourier transform for seismic data regularization*, Geophysics 70 (2005), pp. V87–V95.
[19] B. Liu, *Multi-dimensional reconstruction of seismic data*, PhD thesis, University of Alberta, Alberta, Canada, 2004.
[20] D.O. Trad, T.J. Ulrych, and M.D. Sacchi, *Accurate interpolation with high-resolution time-variant Radon transforms*, Geophysics 67 (2002), pp. 644–656.
[21] G. Darche, *Spatial interpolation using fast parabolic transform*, 60th Annual Internattional Society Explaing Geophysics, San Francisco, CA, Expanded Abstracts, 1990, pp. 1647–1650.
[22] G. Hennenfent and F.J. Herrmann, *Simply denoise: Wavefield reconstruction via jittered undersampling*, Geophysics 73 (2008), pp. V19–V28.
[23] M. Naghizadeh and M.D. Sacchi, *Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data*, Geophysics 75 (2010), pp. WB189–WB202.
[24] Y.F. Wang, *Computational Methods for Inverse Problems and Their Applications*, Higher Education Press, Beijing, 2007.
[25] F.J. Herrmann and G. Hennenfent, *Non-parametric seismic data recovery with curvelet frames*, Geophys. J. Int. 173 (2008), pp. 233C–248.

[26] K.K. Herrity, A.C. Gilbert, and J.A. Troop, *Sparse approximation via iterative thresholding*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Proceeding, Toulouse, France, pp. 624–627.
[27] V.B. Ewout and P.F. Michael, *Probing the pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput. 31 (2008), pp. 890–912.
[28] I.F. Gorodnitsky and B.D. Rao, *Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm*, IEEE Trans. Signal Process. 45 (1997), pp. 600–616.
[29] M. Lustig, D.L. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Reson. Med. 58 (2007), pp. 1182–1195.
[30] E. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory 51 (2005), pp. 4203–4215.
[31] E. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory 52 (2006), pp. 489–509.
[32] E. Candes, *Compressive sampling*, Proceedings of International Congress of Mathematicians, European Mathematical Society Publishing House, Madrid, Spain, 2006, pp. 33–52.
[33] E. Candes and D.L. Donoho, *Curvelets: A surprisingly effective non-adaptive representation for objects with edges*, in *Curves and Surfaces Fitting*, A. Cohen, C. Rabut, and L.L. Schumaker, eds., Vanderbilt University Press, Nashville, TN, 2000, pp. 105–120.
[34] E. Candes, L. Demanet, D. Donoho, and L.X. Ying, *Fast discrete curvelet transforms*, Multiscale Model. Simul. 5 (2006), pp. 861–899.
[35] E. Candes and D. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise singularities*, Commun. Pure Appl. Math. 57 (2004), pp. 219–266.
[36] Y.F. Wang, S.Q. Ma, H. Yang, J.D. Wang, and X.W. Li, *On the effective inversion by imposing a priori information for retrieval of land surface parameters*, Sci. China D 52 (2009), pp. 540–549.
[37] Y.F. Wang, S.F. Fan, and X. Feng, *Retrieval of the aerosol particle size distribution function by incorporating a priori information*, J. Aerosol Sci. 38 (2007), pp. 885–901.
[38] S. Chen, D. Donoho, and M. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. 20 (1998), pp. 33–61.
[39] M.R. Osborne, B. Presnell, and B.A. Turlach, *A new approach to variable selection in least squares problems*, IMA J. Numer. Anal. 20 (2000), pp. 389–403.
[40] D. Donoho and Y. Tsaig, *Fast solution of l_1-norm minimization problems when the solution may be sparse*, Technical Report, Stanford University, CA, 2006.
[41] D.M. Malioutov, M. Cetin, and A.S. Willsky, *Homotopy continuation for sparse signal representation*, Proc. IEEE Int. Conf. Acoustics Speech Signal Process. 5 (2005), pp. 733–736.
[42] B. Turlach, *On algorithms for solving least squares problems under an L_1 penalty or an L_1 constraint*, Proceedings of the American Statistical Association, Statistical Computing Section, Alexandria, VA, 2005, pp. 2572–2577.
[43] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Annal. Stat. 32 (2004), pp. 407–499.
[44] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinvesky, *A method for large-scale l_1-regularized least squares problems with applications in signal processing and statistics*, Technical Report, Stanford University, Palo Alto, CA, 2007.
[45] Y.F. Wang, X.W. Li, S.Q. Ma, Z. Nashed, and Y.N. Guan, *BRDF model inversion of multi-angular remote sensing: Ill-posedness and interior point solution method*, Proc. 9th Int. Symp. Phys. Measurements Signature Remote Sens. XXXVI (2005), pp. 328–330.
[46] M.A.T. Figueiredo and R.D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process. 12 (2003), pp. 906–916.
[47] Y.F. Wang and T.Y. Xiao, *Fast realization algorithms for determining regularization parameters in linear inverse problems*, Inverse Probl. 17 (2001), pp. 281–291.
[48] G. Davis, S. Mallat, and M. Avellaneda, *Greedy adaptive approximation*, J. Constructive Approx. 12 (1997), pp. 57–98.

[49] D. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory 52 (2006), pp. 6–18.
[50] J. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory 50 (2004), pp. 2231–2242.
[51] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, *Gradient projection for sparse reconstruction: application to compressed sensing and other inversed problems*, IEEE J. Selected Topics Signal Process. 1 (2007), pp. 586–597.
[52] B.D. Rao and K. Kreutz-Delgado, *An affine scaling methodology for best basis selection*, IEEE Trans. Signal Process. 47 (1999), pp. 187–200.
[53] E. Candes, M.B. Wakin, and S.P. Boyd, *Enhancing sparsity by reweighted l_1 minimization*, J. Fourier Anal. Appl. 14 (2008), pp. 877–905.
[54] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, *A fast approach for overcomplete sparse decomposition based on smoothed l^0 norm*, IEEE Trans. Signal Process. 57 (2009), pp. 289–301.
[55] G. Gasso, A. Rakotomamonjy, and S. Canu, *Recovering sparse signals with a certain family of non-convex penalties and DC programming*, IEEE Trans. Signal Process. 57 (2009), pp. 4686–4698.
[56] Y.L. Wang and W.T. Yin, *Sparse signal reconstruction via iterative support detection*, Technical Report, Rice University, TX, 2009.
[57] Y.F. Wang, J.J. Cao, Y.X. Yuan, C.C. Yang, and N.H Xiu, *Regularizing active set method for non-negatively constrained ill-posed multi-channel image restoration problem*, Appl. Optics 48 (2009), pp. 1389–1401.
[58] P.M. Zwartijes and M.D. Sacchi, *Fourier reconstruction of nonuniformly sampled, aliased seismic data*, Geophysics 52 (2007), pp. V21–V32.
[59] G. Hennenfent, E. Van den Berg, M.P. Friedlander, and F. Herrmann, *New insights into one-norm solvers from the Pareto curve*, Geophysics 73 (2008), pp. A23–A26.
[60] Y.F. Wang and C.C. Yang, *Accelerating migration deconvolution using a nonmonotone gradient method*, Geophysics 75 (2010), pp. S131–S137.
[61] C.H. Chen and O.L. Mangasarian, *A class of smoothing functions for non-linear and mixed complementarity problems*, Comput. Optim. Appl. 5 (1996), pp. 97–138.
[62] M. Schmidt, G. Fung, and R. Rosales, *Fast optimization methods for L1 regularization: A comparative study and two new approaches*, Proceedings of European Conference on Machine Learning, Springer-Verlag, Heidelberg, 2007, pp. 286–297.
[63] H.L. Taylor, S. Banks, and J. McCoy, *Deconvolution with the l_1 norm*, Geophysics 44 (1979), pp. 39–52.