Geophys. J. Int. (2010) **000**, 1–19

Recovery of Seismic Wavefields Based on Compressive Sensing by an l_1-Norm Constrained Trust Region Method and the Piecewise Random Sub-sampling

Yanfei Wang[†], Jingjie Cao[†‡], Changchun Yang[†]

[†] *Key Laboratory of Petroleum Resources Research,*

Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O.Box 9825, Beijing, 100029, P.R.China

E-mail: yfwang@mail.iggcas.ac.cn

[‡] *Graduate University of Chinese Academy of Sciences,*

Beijing, 100049, P.R.China

Received 2010 August 19; in original form 2010 August 4

SUMMARY

Due to the influence of variations in landform, geophysical data acquisition is usually sub-sampled. Reconstruction of the seismic wavefield from sub-sampled data is an ill-posed inverse problem. Compressive sensing can be used to recover the original geophysical data from the sub-sampled data. In this paper, we consider the wavefield reconstruction problem as a compressive sensing and propose a piecewise random sub-sampling scheme based on the wavelet transform. The proposed sampling scheme overcomes the disadvantages of uncontrolled random sampling. In computation, an l_1-norm constrained trust region method is developed to solve the compressive sensing problem. Numerical results demonstrate that the proposed sampling technique and the trust region approach are robust in solving the ill-posed compressive sensing problem and can greatly improve the quality of wavefield recovery.

Key words: Computational seismology – inverse theory – wavefield reconstruction – compressive sensing – regularization – sparse optimization

1 INTRODUCTION

In seismology, due to limitations of the observations, the observed data is incomplete, e.g., some traces are lost. In that situation, a key obstacle is how to invert the model using only incomplete, sub-sampled data (Herrmann & Hennenfent, 2008). Restoration of the original wavefield from incomplete observed data is an ill-posed problem in general. Recently, the recovery of seismic wavefield based on compressive sensing was developed (Herrmann & Hennenfent, 2008). Meanwhile, two main problems are how to establish a proper mathematical compressive sensing model and how to solve the minimization model.

For solving a compressive sensing problem, there are many methods available such as the (orthogonal) matching pursuit method (Chen et al., 1998; Tropp & Gilbert, 2007), interior point solution method (Wang et al., 2010), operator splitting method (Wang, 2011), preconditioning conjugate gradient method (Kim et al., 2007) and the gradient projection method (Figueiredo et al., 2007; Ewout & Michael, 2008). The gradient descent methods usually yield a local solution (Yuan, 1993). In geophysics, we are required to find a global optimization solution of the minimization model. Therefore proper globally convergent optimized algorithms are urgently needed. To obtain a global minimizer, we develop an l_1 constrained trust region method in this paper. The trust region method was proved to be a regularization method for ill-posed inverse problems (Wang & Yuan, 2005), and hence can be employed to solve ill-posed wavefield restoration problems based on compressive sensing. To solve for a Lagrangian parameter of the trust region subproblem, we propose a Newton's method which possesses quadratic convergence rate. Numerical experiments on signal processing and seismic wavefield restoration problem indicate the robustness and applicability of our algorithms.

2 COMPRESSIVE SENSING THEORY

Signal acquisition systems based on the Nyquist-Shannon sampling theorem require that the sampling rate needed to recover a signal without error is twice the bandwidth. This sampling theorem is hard to satisfy in practice. As an alternative, compressive sensing (CS) has recently received a lot of attention in the signal and image processing community. Instead of relying on the bandwidth of the signal, the CS uses the basic assumption: sparsity. The sparsity can lead to efficient estimations and compression of the signal via a linear transform, e.g., sine, cosine, wavelet and curvelet transforms (Herrmann et al., 2008). The method involves taking a relatively small number of non-traditional samples in the form of projections of the signal onto random basis elements or random vectors (Donoho, 2006; Candes et al., 2006; Candes & Wakin, 2008). Therefore, if the signal has a sparse representation on some basis, it is possible to reconstruct the signal using few linear measurements.

2.1 Sparse transform

For a signal x in N-dimensional space, we have M observation data $d_i = A_i x$, $i = 1, 2, \cdots, M$, where A_i for each i is a row vector, which represents the impulse response of the i-th sensor. The product of A_i with x yields the i-th component of data d. Denote $A = [A_1, A_2, \cdots, A_M]^T$, the observation data can be reformulated as $d = Ax$. The aim of the compressive sensing is to use limited observations d_i $(i = 1, 2, \cdots, M)$ with $M \ll N$ to restore the input signal x.

Suppose x is the original wavefield which can be spanned by a series of orthogonal bases $\Psi_i(t)$. These bases for all i constitute an orthogonal transform matrix Ψ such that

$$x(t) = (\Psi m)(t) = \sum_i m_i \Psi_i(t), \tag{1}$$

where $m_i = (x, \Psi_i)$. Using operator expression, $m = \Psi^* x$. The vector m is thought of as the sparse or compressive expression of the signal x. Letting $L = A\Psi$, the reconstruction problem of the sparse signal m reduces to solving a simple problem $d = Lm$. Note that if m_i is the weight or coefficient of linear combinations for the signal x, the reconstruction of the signal x in turn becomes to find the coefficient vector m.

There are many ways to choose an orthogonal transform matrix based on some orthogonal bases, e.g., sine curve, wavelet, curvelet and framelet, and so forth (Herrmann et al., 2008). As we are mainly concerned with new methods to solve the linear system $d = Lm$ in this paper, we choose a simple wavelet orthogonal bases to form the transform matrix Ψ.

2.2 Relations with Tikhonov regularization

The compressive sensing is closely related with Tikhonov's regularization for solving ill-posed problems (Wang, 2007). Let us begin with a compact problem

$$Lm = d, \tag{2}$$

where L is a compact operator (e.g., a finite rank measurement matrix) maps m from parameter space into observation space. One may readily see that the problem (2) can be regarded as a forward model to generate seismic data: $m(t)$ represents the model or reflectivity and L the scattering matrix that generates the data d. Problem (2) is usually ill-posed due to the fact that existence, uniqueness and stability of the solution may be violated. Conventional methods for solving such an ill-posed problem are Tikhonov's regularization (Tikhonov & Arsenin, 1977)

$$\min J^{\alpha}_{\mathrm{Tikh}}[m] = \frac{1}{2}\|Lm - d\|^2_{L_2} + \alpha\Omega[m], \tag{3}$$

where $\alpha > 0$ is the regularization parameter and $\Omega[m]$ is the stabilizer which provides some *a priori* constraints on the solution. Tikhonov's regularization method is usually used for solving non-sparse problems, and has also been used in solving sparse problems with proper choice of $\Omega[m]$.

2.3 Minimization in l_0 space

A natural model to satisfy the sparse solutions of the linear system $Lm = d$ is the equality constrained minimization model with l_0 norm:

$$\|m\|_{l_0} \longrightarrow \min, \text{ subject to } Lm = d, \tag{4}$$

where $\|\cdot\|_{l_0}$ is defined as: $\|x\|_{l_0} = \{\mathrm{num}(x \neq 0), \text{ for all } x \in \mathbb{R}^N\}$, where $\mathrm{num}(x \neq 0)$ denotes the cardinality of nonzero components of the vector x. Minimization of $\|x\|_{l_0}$ means the number of nonzero components of x to be minimal. It is well known that the minimization of $\|x\|_{l_0}$ is an NP-Hard problem, i.e., optimization algorithms solving the l_0 minimization problem cannot be finished in polynomial times. This indicates that this model is doomed to be infeasible in practice.

2.4 Minimization in l_1 space

Because of the numerical infeasibility of the l_0 minimization problem, we relax it to solve the approximation model based on l_1 norm:

$$\|m\|_{l_1} \longrightarrow \min, \text{ subject to } Lm = d. \tag{5}$$

The presence of the l_1 term encourages small components of m to become exactly zero, thus promoting sparse solutions. Introducing the Lagrangian multiplier λ, equation (5) is equivalent to the following unconstrained problem

$$\|Lm - d\|_{l_2}^2 + \lambda \|m\|_{l_1} \longrightarrow \min. \tag{6}$$

The minimization model based on l_1 norm approximates the minimization model based on l_0 norm quite well, while the sparsity is retained (Cao et al., 2011; Figueiredo et al., 2007; Ewout & Michael, 2008).

2.5 Minimization in l_p-l_q space

In (Wang et al., 2009), the authors proposed a general l_p-l_q model for solving multi-channel ill-posed image restoration problem,

$$J^\alpha[m] := \frac{1}{2}\|Lm - d\|_{l_p}^p + \frac{\alpha}{2}\|m\|_{l_q}^q \longrightarrow \min, \text{ for } p, q \geq 0, \tag{7}$$

which includes most of the regularization models thus far. Straightforward calculation yields the gradient and Hessian (the matrix of the second-order partial derivatives) of $J^\alpha[m]$ as

$$\mathrm{grad}_{J^\alpha}[m] = \tfrac{1}{2}pL^T \begin{bmatrix} |r_1|^{p-1}\mathrm{sign}(r_1) \\ |r_2|^{p-1}\mathrm{sign}(r_2) \\ \vdots \\ |r_m|^{p-1}\mathrm{sign}(r_m) \end{bmatrix} + \tfrac{1}{2}\alpha q L^T \begin{bmatrix} |m_1|^{q-1}\mathrm{sign}(m_1) \\ |m_2|^{q-1}\mathrm{sign}(m_2) \\ \vdots \\ |m_n|^{q-1}\mathrm{sign}(m_n) \end{bmatrix} \tag{8}$$

and

$$\begin{aligned} \mathrm{Hess}_{J^\alpha}[m] = \; & \tfrac{1}{2}p(p-1)L^T\mathrm{diag}(|r_1|^{p-2}, |r_2|^{p-2}, \cdots |r_m|^{p-2})L \\ & + \tfrac{1}{2}\alpha q(q-1)\mathrm{diag}(|m_1|^{q-2}, |m_2|^{q-2}, \cdots, |m_n|^{q-2}), \end{aligned} \tag{9}$$

respectively, where $r = (r_1, r_2, \cdots, r_M)^T = Lm - d$ is the residual, $\mathrm{sign}(\cdot)$ denotes a function which returns -1, 0, or $+1$ when the numeric expression value is negative, zero, or positive respectively, $\mathrm{diag}(v)$ is the diagonal matrix whose i-th diagonal entry is the same as the i-th component of the vector v. Evidently, when $p = 2$ and $q = 0$ or $q = 1$, the l_p-l_q model becomes the l_0 minimization model or the l_1 minimization model, respectively. We remark that the l_p-l_q regularization model does not require the convexity of the objective function, hence could be used to solve inverse problems with complex structure.

3 SOLVING THE CS MODEL

3.1 Classical Solution Methods

Several optimization algorithms have been developed to solve the l_1 minimization model (5), e.g., the basis pursuit denoising (BPDN) criterion (Chen et al., 1998; Tropp & Gilbert, 2007) and the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). Both BPDN and LASSO approaches can reduce to the regularizing problem (6). Many methods can be involved to solve (6), e.g., conjugate gradient methods with preconditioning techniques (Kim et al., 2007), gradient projection methods (Dai & Fletcher, 2005; Wang & Ma, 2007; Figueiredo et al., 2007; Ewout & Michael, 2008). The BPDN problem with $\|Lm - d\|_{l_2}^2 = \delta = 0$ (δ is the upper bound of the norm of the misfit) is equivalent to (5), a particular method called the interior point (IP) solution method can be employed (Wang et al., 2007; Wang et al., 2010). However the IP solutions may be physically meaningless for geophysical problems. In addition, (orthogonal) matching pursuit method, a popular method in engineering, can also be used for solving a sparse recovery problem (Chen et al., 1998; Tropp & Gilbert, 2007). The method greedily picks up a series of columns of the measurement matrix as atoms and applies the Gram-Schmidt orthogonalization upon chosen atoms for efficient computation of projections. However this method is non-related with optimization.

Recalling that the original problem (2) is ill-posed and has infinite solutions if the number of observations is insufficient, therefore theoretically, the above mentioned methods using only the gradient information usually give local solutions. In geophysics, we are eager to find a global optimized solution. Therefore it is desirable to find methods that give a global minimum.

3.2 Trust Region Method

3.2.1 Trust region scheme

Trust region methods have been widely used for solving nonlinear problems, and provide globally convergent solutions (Yuan, 1993). Using the notations in the optimization community, we consider the optimization problem

$$\min_{x \in X} J(x), \tag{10}$$

where $J(x)$ is the objective function about the variable x in its domain of definition X. Refer to our problem, the objective function is $J^\alpha[m]$. At each iteration a trial step is calculated by solving the subproblem

$$\min_{\xi\in X} \psi_k(\xi) := (\mathrm{grad}_k(J), \xi) + \frac{1}{2}(\mathrm{Hess}_k(J)\xi, \xi), \tag{11}$$

$$\text{subject to } \|\xi\| \le \Delta_k, \tag{12}$$

where $\mathrm{grad}_k(J)$ is the gradient of J at the k-th iterative point x_k,

$$\mathrm{grad}(J)(x) = \frac{d}{dx}J(x) = \nabla J(x), \tag{13}$$

$\mathrm{Hess}_k(J)$ is the Hessian of J at the k-th iterative point x_k,

$$\mathrm{Hess}(J)(x) = \nabla^2 J(x), \tag{14}$$

and Δ_k is the trust region radius. The trust region subproblem (11)-(12) is an approximation to the original optimization problem (10) with a trust region constraint which prevents the trial step being too large.

One may readily see that the minimal problem (10) can be solved by the Gauss-Newton method, i.e., solving the following problem at the k-th iteration

$$\mathrm{Hess}_k(J)\xi = -\mathrm{grad}(J)_k,$$

$$x_{k+1} = x_k + \xi.$$

However the method is unstable for ill-posed problems and converges locally (Wang, 2007).

At each iteration, a trust region algorithm generates a new point in the trust region, and has the procedure to determine the acceptance and rejection of the new point. At each iteration, the trial step ξ_k is normally calculated by solving the trust region subproblem (11)-(12). Generally, a trust region algorithm uses

$$r_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k} \tag{15}$$

to decide whether the trial step ξ_k is acceptable and how the next trust region radius is chosen, where

$$\mathrm{Pred}_k = \psi_k(0) - \psi_k(\xi_k) \tag{16}$$

is the predicted reduction in the approximate model, and

$$\mathrm{Ared}_k = J(x_k) - J(x_k + \xi_k) \tag{17}$$

is the actual reduction in the objective functional.

Now we recall the trust region algorithm for solving nonlinear ill-posed problems.

Algorithm 3.1. (Trust region algorithm for nonlinear ill-posed problems)

STEP 1 Choose parameters $0 < \tau_3 < \tau_4 < 1 < \tau_1$, $0 \le \tau_0 \le \tau_2 < 1$, $\tau_2 > 0$ and initial values x_0, $\Delta_0 > 0$; Set $k := 1$.

STEP 2 If **the stopping rule is satisfied** then STOP; Else, solve (11)-(12) to give ξ_k.

STEP 3 Compute r_k;

$$x_{k+1} = \begin{cases} x_k & \text{if} \quad r_k \le \tau_0, \\ x_k + \xi_k & \text{otherwise.} \end{cases} \tag{18}$$

Choose Δ_{k+1} that satisfies

$$\Delta_{k+1} \in \begin{cases} [\tau_3 \|\xi_k\|, \tau_4 \Delta_k] & \text{if} \quad r_k < \tau_2, \\ [\Delta_k, \tau_1 \Delta_k] & \text{otherwise.} \end{cases} \tag{19}$$

STEP 4 Evaluate $\mathrm{grad}_k(J)$ and $\mathrm{Hess}_k(J)$; $k := k+1$; GOTO STEP 2.

In Step 2, the stopping rule is based on the discrepancy principle, i.e., the iteration should be terminated at the first occurrence of the index k such that the energy of the residual of the observation to model is less than the preassigned tolerance. Global convergence and regularity properties of the trust region method for ill-posed inverse problems are discussed in (Wang & Yuan, 2005).

The constants τ_i $(i = 0, \cdots, 4)$ can be chosen by users. Typical values are $\tau_0 = 0,\ \tau_1 = 2,\ \tau_2 = \tau_3 = 0.25,\ \tau_4 = 0.5$. The parameter τ_0 is usually zero or a small positive constant. The advantage of using zero τ_0 is that a trial step is accepted whenever the objective function is reduced. When the objective function is not easy to compute, it seems that we should not throw away any "good" point that reduces the objective function (Yuan, 1993; Wang, 2007).

3.2.2 Solving the compressive sensing model in the form of a trust region subproblem

Let us go back to the l_p-l_q minimization model with $p = 2$ and $q = 1$. In this case, the model reads as

$$f(m) = \|Lm - d\|_{l_2}^2 + \alpha \|m\|_{l_1} \longrightarrow \min. \tag{20}$$

The regularization parameter α is set *a priori* value. It is evident that the above function f is nondifferentiable at $m = 0$. To make it easy to be calculated by computer, we approximate $\|m\|_{l_1}$ by $\sum_{i=1}^{l} \sqrt{(m_i, m_i) + \epsilon}$ $(\epsilon > 0)$ and l is the length of the vector m. For notational simplicity, we let $A = L^T L$, $\gamma(m^k) = \left(\frac{m_1^k}{\sqrt{(m_1^k)^T m_1^k + \epsilon}}, \frac{m_i^k}{\sqrt{(m_i^k)^T m_i^k + \epsilon}}, \cdots, \frac{m_n^k}{\sqrt{(m_n^k)^T m_n^k + \epsilon}} \right)^T$ and

$$\chi_p(m^k) = \mathrm{diag}\Big(\frac{\epsilon}{((m_1^k)^T m_1^k + \epsilon)^{p/2}}, \cdots, \frac{\epsilon}{((m_i^k)^T m_i^k + \epsilon)^{p/2}}, \cdots, \frac{\epsilon}{((m_n^k)^T m_n^k + \epsilon)^{p/2}} \Big),$$

where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with only nonzero components on the main diagonal line. Straightforward calculation shows the gradient of f at m^k

$$g_k := g(m^k) \approx L^T(Lm^k - d) + \alpha \gamma(m^k)$$

and the Hessian of f at m^k

$$H_k := H(m^k) \approx L^T L + \alpha \chi_3(m^k).$$

With above preparation, a new trust region subproblem for the compressing model can be formulated from (11)-(12) as

$$\min_{\xi\in X}\phi_k(\xi) := (g_k,\xi) + \frac{1}{2}(H_k\xi,\xi), \tag{21}$$

$$\text{subject to } \|\xi\|_{l_1} \le \Delta_k. \tag{22}$$

To solve the trust region subproblem (21)-(22), we introduce the Lagrangian multiplier λ and solve an unconstrained minimization problem

$$L(\lambda,\xi) = \phi_k(\xi) + \lambda(\Delta_k - \|\xi\|_{l_1}) \longrightarrow \min. \tag{23}$$

Straightforward calculation yields that the solution satisfies

$$\xi = \xi(\lambda) = -(H_k + \lambda\epsilon^{-1}\chi_1(\xi))^{-1}g_k. \tag{24}$$

From (24), we find that the trial step ξ can be obtained iteratively

$$\xi^{j+1}(\lambda) = -(H_k + \lambda\epsilon^{-1}\chi_1(\xi^j))^{-1}g_k. \tag{25}$$

And at the k-th step, the Lagrangian parameter λ can be solved via the nonlinear equation

$$\|\xi_k(\lambda)\|_{l_1} = \Delta_k. \tag{26}$$

Denoting $\Gamma(\lambda) = \frac{1}{\|\xi_k(\lambda)\|_{l_1}} - \frac{1}{\Delta_k}$, the Lagrangian parameter λ can be iteratively solved by Newton's method

$$\lambda_{l+1} = \lambda_l - \frac{\Gamma(\lambda_l)}{\Gamma'(\lambda_l)},\ l = 0,1,\cdots. \tag{27}$$

The derivative of $\Gamma(\lambda)$ can be evaluated as $\frac{d}{d\lambda}\left(\frac{1}{\rho(\lambda)}\right) = -\frac{\rho'(\lambda)}{\rho^2(\lambda)} = -\frac{\rho'(\lambda)}{\|\xi_k(\lambda)\|_{l_1}^2}$, where $\rho(\lambda) := \|\xi_k(\lambda)\|_{l_1}$. One may readily derive that at the k-th step

$$\rho'(\lambda) \approx \begin{pmatrix} \frac{\xi_1^k(\lambda)}{\sqrt{\xi_1^k(\lambda)^T\xi_1^k(\lambda)+\epsilon}} \\ \vdots \\ \frac{\xi_i^k(\lambda)}{\sqrt{\xi_i^k(\lambda)^T\xi_i^k(\lambda)+\epsilon}} \\ \vdots \\ \frac{\xi_n^k(\lambda)}{\sqrt{\xi_n^k(\lambda)^T\xi_n^k(\lambda)+\epsilon}} \end{pmatrix}^T * \frac{d}{d\lambda}\xi_k(\lambda) = \gamma(\xi_k)^T[\epsilon H_k + \lambda\chi_1(\xi_k)]^{-1}\chi_1(\xi_k)\xi_k(\lambda). \tag{28}$$

Hence the optimal Lagrangian parameter λ^* can be obtained from iteration formula (27). Once λ^* is reached, the optimal step ξ^* is obtained, and the trust region scheme in Algorithm 3.1 can be driven to another round of iteration.

Remark. Interesting properties can be obtained for the trust region method:

(1) The Lagrangian parameters $\{\lambda_k\}$ is uniformly bounded. This assertion can be obtained by noticing the fact that $\Delta_k \ge \omega\|g_k\|$ ($\omega > 0$), hence $\{\lambda_k\}$ is uniformly bounded. This property indicates that the Lagrangian parameter λ also plays a role of regularization.

(2) Unlike the smooth regularization, where the $\|\xi_k(\lambda)\|_{l_2}$ solved by the corresponding trust region method is monotonically decreasing

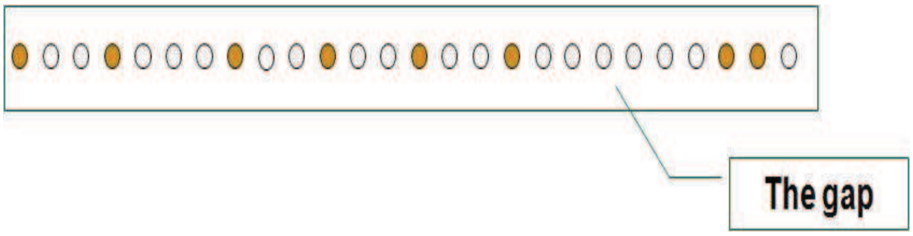

Figure 1. Random sampling: the solid circle (yellow one) represents receivers, the hollow circle (white one) means no receivers. Large gap occurs for some sampling points.

and $\|\xi_k(\lambda)\|_{l_2}$ is a decreasing function of λ (Wang & Yuan, 2005; Wang, 2007); for the sparse regularization model, the $\|\xi_k(\lambda)\|_{l_1}$ solved by the above trust region method is not necessarily decreasing, but $\|\xi_k(\lambda)\|_{l_1}$ can be bounded so long as noticing the expression (24).

(3) The trust region subproblems (21)-(22) can be also solved by some gradient-type methods. This may relax the computational cost of factorization of matrices.

4 SAMPLING

4.1 Random sampling

Regular incomplete sampling takes a number of observations in a measurement line with equidistance. This kind of sampling may not satisfy the Shannon/Nyquist sampling theorem. As the coherence noise in frequency-wavenumber domain occurs in this type of sampling, hence it is not suitable for orthogonal transform-based wavefield reconstruction. Random incomplete sampling refers to taking a number of independent observations in a measurement line with randomly allocated geophones. This sampling technique is better than the regular incomplete sampling, however a large sampling interval is not suitable for wavefield reconstruction, e.g., reconstruction using short-time Fourier transform and curvelet transform. This lack of control over the size of the gaps during random sampling may lead to an occasional failed recovery. Figure 1 illustrates the problem of the uncontrolled random sampling. Another sampling technique is the jittered undersampling (Hennenfent & Herrmann, 2008). The basic idea of jittered undersampling is to regularly decimate the interpolation grid and subsequently perturb the coarse-grid sample points on the fine grid. However the jittered undersampling takes only integer partition of the complete sampling, which may not satisfy the practical wavefield reconstruction.

4.2 A new sampling technique: piecewise random sub-sampling

Usually in field applications, because of the influence of ground geometry such as valleys and rivers, the sampling is difficult to allocate properly. Therefore the above sampling techniques would not be able to overcome such kind of difficulties completely. Considering their shortcomings, we propose a new sampling scheme: a piecewise random sub-sampling, see Figure 2. We first partition the measurement line into several subintervals; then perform random sampling on each subinterval. As the number of partitions is sufficient enough, the sampling scheme will control the size of the sampling gaps while keeping the randomicity of the sampling.

Figure 2. Piecewise random sub-sampling: : the solid circle (yellow one) represents receivers, the hollow circle (white one) means no receivers. Large gap are controlled for any sampling points.

5 NUMERICAL RESULTS

To verify the feasibility of our algorithm, we consider two numerical examples. We start our simulation from a simple one-dimensional sparse signal reconstruction. Then we consider an example of reconstruction of seismic shot gathers.

5.1 Random signal reconstruction

We consider a sparse signal $m \in \mathbb{R}^N$, which is measured (sensed) by a random measuring matrix $L \in \mathbb{R}^{M\times N}$ $(M < N)$. Then $d = Lm \in \mathbb{R}^M$ is the measurement vector. Every row of the matrix L can be seen as a measuring operator, whose inner product with m is a measurement. $M < N$ means the number of measurements is smaller than the length of the signal, thus the number of measurements is compressed. In our simulation, M is chosen as 140 and N equals 200. Our problem is to recover the original signal m from the measurement d. This is a severely ill-posed problem. Since the measurement is random, therefore, the data is randomly recorded. To show the randomness, we plot the measurement data in Figure 3. The original sparse signal is shown in Figure 4 with legend "o" lines. Using our trust region algorithm and the piecewise random sub-sampling, the restoration results ("+" lines) comparing with the original signal is shown in Figure 4. It is evident from the comparison that our algorithm is robust in reconstruction of sparse signals. This example shows that our method works for any random generated data (e.g., in Figure 3) using random measurement matrix. Therefore, it would be a reliable and stable method for potential practical problems.

We have shown in Section 3.2.2 that the Lagrangian parameter λ is uniformly bounded. In fact, we noticed from numerical tests that this parameter is also in a decreasing tendency with process of iterations, see Figure 5. Hence the Lagrangian parameter λ plays a role of regularization.

5.2 Reconstruction of shot gathers

Now we consider a 7 layers geologic velocity model, see Figure 6. With the spatial sampling interval of 15 meters and the time sampling interval of 0.002 second, the shot gathers are shown in Figure 7. The data with missing traces are shown in Figure 8. Using our trust region approach and the piecewise random sub-sampling, the recovered wavefield is shown in Figure 10. It is clear from the reconstruction that most of the details of the wavefield are preserved. To show the good performance of our method, we plot the frequency information of the sub-sampled data the restored data in Figures 9 and 11, respectively. It is clear that the aliasing (just like noise) of the sub-sampled data

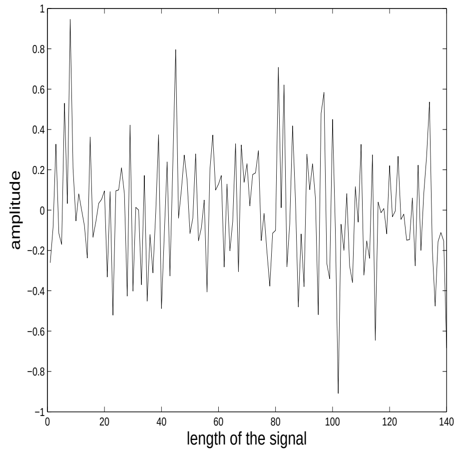

Figure 3. Random measurement data.

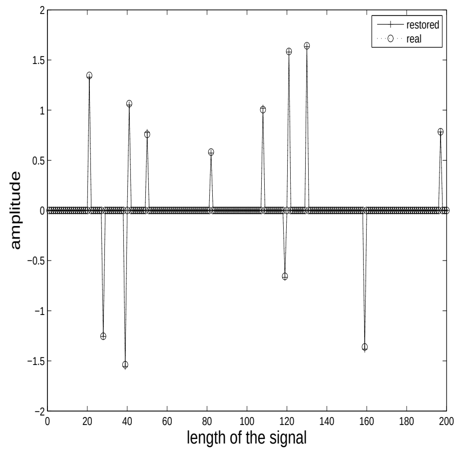

Figure 4. Comparison of input and restored random signals.

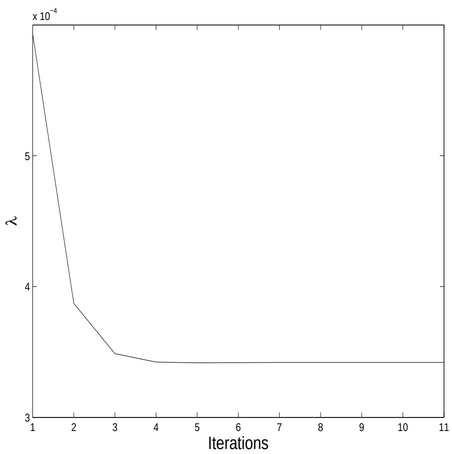

Figure 5. Variations of the Lagrangian parameters λ.

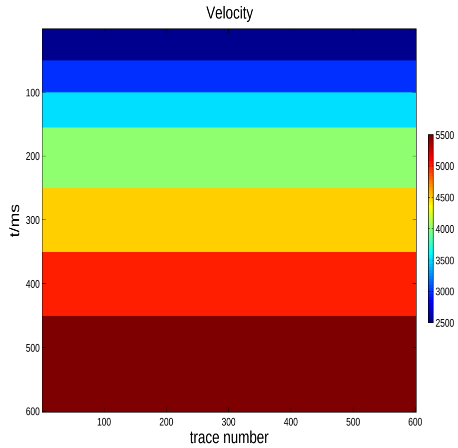

Figure 6. Velocity model.

is reduced greatly in the recovered data. The difference of the original data and the recovered data is illustrated in Figure 12. Virtually, all the initial seismic energy is recovered with minor errors. Though there are still the artifacts such as vertical stripes, we consider it might be caused by ill-posed nature of the inversion process and insufficient iterations. We also plot the variations of the Lagrangian parameter at each iteration in Figure 13. Again, we find that that this parameter is in a decreasing tendency with process of iterations.

5.3 Reconstruction of inhomogeneous media

Next we consider a more complicated inhomogeneous media. The data is generated using a velocity model varying both vertically and transversely, see Figure 14. The original data, sub-sampled data and recovered data are shown in Figures 15, 16 and 18, respectively. The frequency information of the sub-sampled data and the recovered data are shown in Figures 17 and 19, respectively. Again, the aliasing of the sub-sampled data is reduced greatly in the recovered data. The difference of the original data and the recovered data is illustrated in Figure

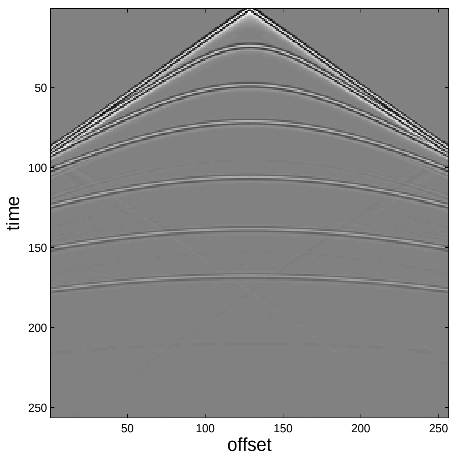

Figure 7. Seismogram.

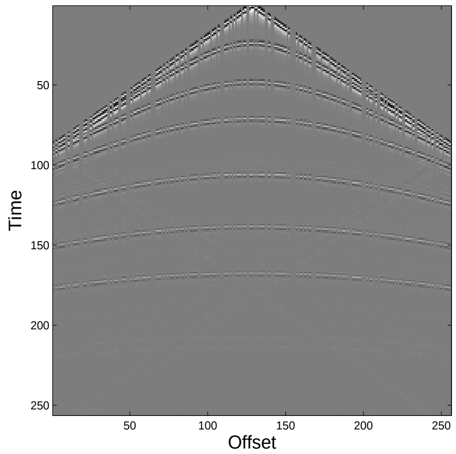

Figure 8. Incomplete data.

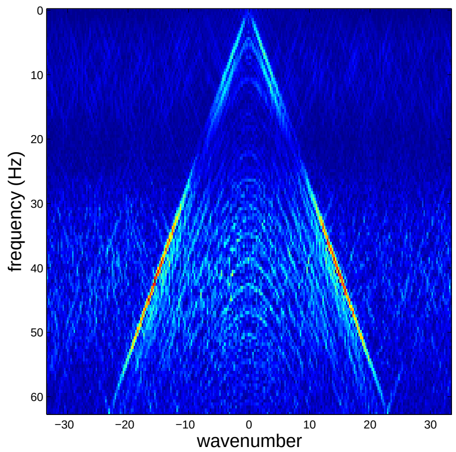

Figure 9. Frequency of the sub-sampled data.

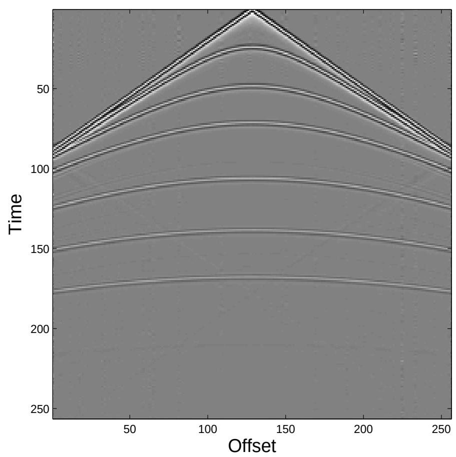

Figure 10. Recovery results.

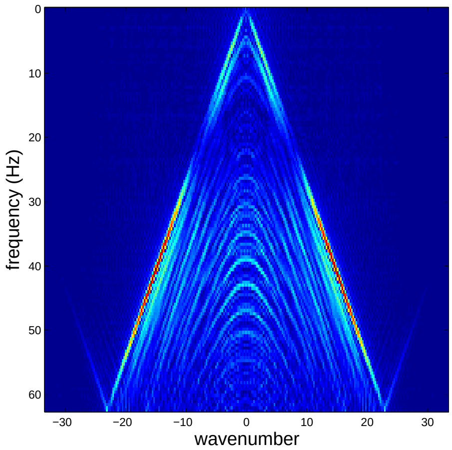

Figure 11. Frequency of the restored data.

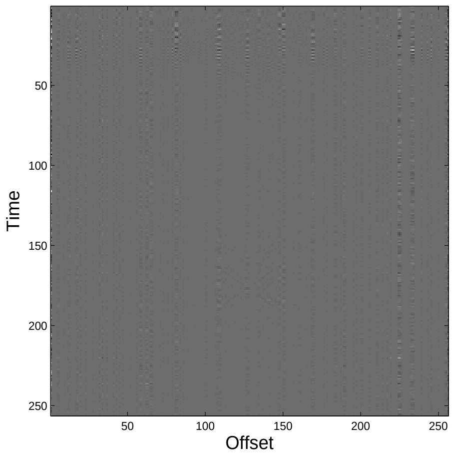

Figure 12. Difference between the restored data and the original data.

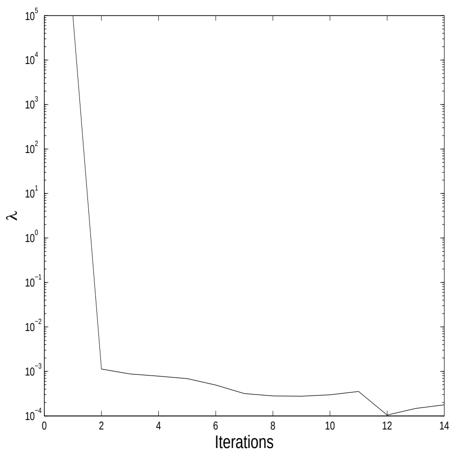

Figure 13. Variations of the Lagrangian parameters λ.

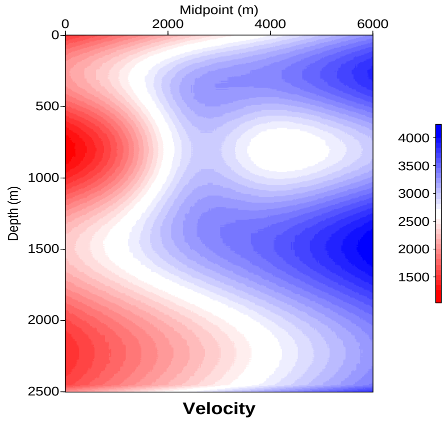

Figure 14. Velocity model.

20. It illustrates that all the initial seismic energy is recovered with minor errors. Though the reconstruction is not perfect, most of the details of the wavefield are preserved. Decreasing tendency of the Lagrangian parameter at each iteration is shown in Figure 21.

6 CONCLUSION

In this paper, we consider using trust region methods for solving the compressive sensing problem in seismic imaging. Due to limitations of local convergence of gradient descent methods in literature, we consider a global convergent method in this paper. Particularly, we propose an l_1 constrained trust region method for solving the compressive sensing problem. In solving the trust region subproblem, an exact solution method with the determination of the Lagrangian parameter by Newton's method is developed. In numerical tests, a piecewise random sub-sampling technique for wavefield reconstruction is also developed.

We argue that the trust region subproblem can be also solved in an exact way, e.g., using gradient type of methods. This deserves further investigation. In addition, to tackle the ill-posedness of the reconstruction problem, proper regularization is necessary, e.g., choosing a proper

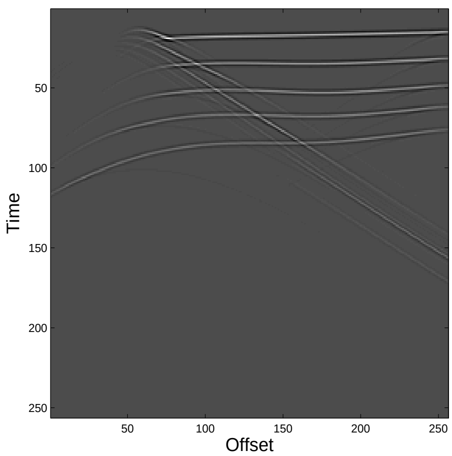

Figure 15. Seismogram.

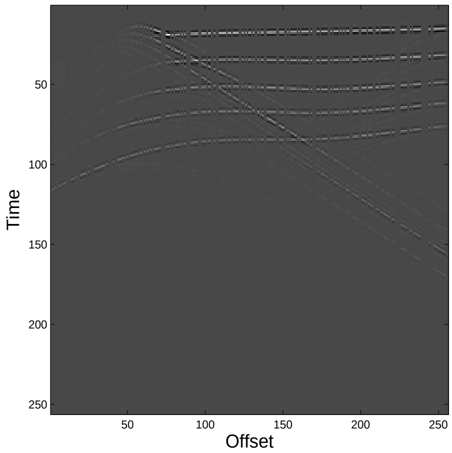

Figure 16. Incomplete data.

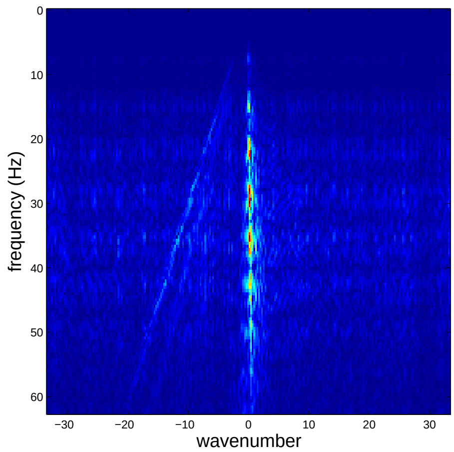

Figure 17. Frequency of the sub-sampled data.

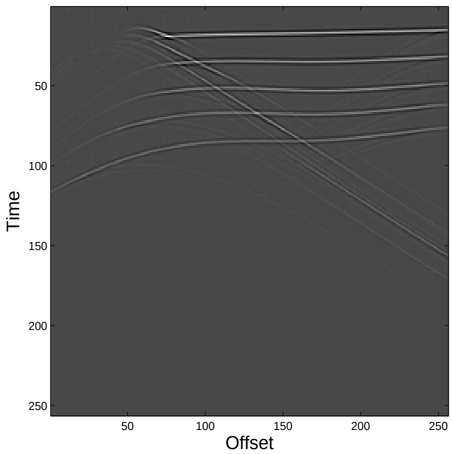

Figure 18. Recovery results.

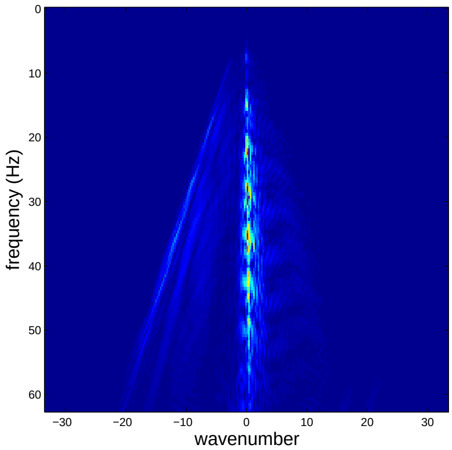

Figure 19. Frequency of the restored data.

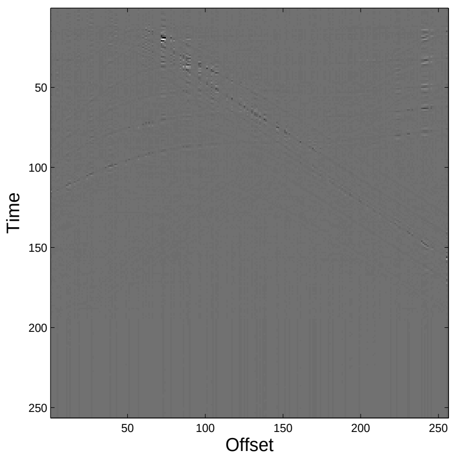

Figure 20. Difference between the restored data and the original data.

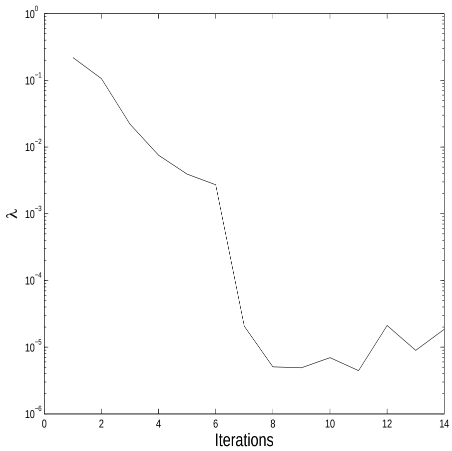

Figure 21. Variations of the Lagrangian parameters λ.

regularization parameter α. In this paper, we use the *a priori* approach. Clearly this kind of choice is not optimal. It is desirable to find a better regularization parameter α by *a posteriori* techniques.

ACKNOWLEDGMENTS

We are grateful to Prof. Andrew Curtis for his help in editing of our paper. We quite thank Ali Özbek and an anonymous referee's helpful suggestions and their editing on the paper. This work is supported by National Natural Science Foundation of China under grant numbers 10871191, 40974075 and Knowledge Innovation Programs of Chinese Academy of Sciences KZCX2-YW-QN107.

REFERENCES

Candes, E. J., Romberg, J. & Tao, T., 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. on Information Theory*, **52**, 489-509.

Candes, E. J. & Wakin, M. B., 2008. An introduction to compressive sampling, *IEEE Signal Processing Magazine*, **25**, 21-30.

Cao, J. J., Wang, Y. F., Zhao J. T. & Yang, C. C., 2011. A review on restoration of seismic wavefields based on regularization and compressive sensing, *Inverse Problems in Science and Engineering*, Vol. 19, No. 6, 2011. (in press)

Chen, S., Donoho, D. & Saunders, M. 1998. Atomic decomposition by basis pursuit, *SIAM Journal of Scientific Computation*, **20**, 33-61.

Dai, Y. H. & Fletcher, R., 2005. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numerische Mathematik*, **100**, 21-47.

Donoho, D., 2006. Compressed sensing, *IEEE Trans. on Information Theory*, **52**, 1289-1306.

Ewout, V. B. & Michael, P. F., 2008. Probing the pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.*, **31**, 890-912.

Figueiredo, M. A. T., Nowak, R. D. & Wright, S. J., 2007. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE Journal of Selected Topics in Signal Processing*, **1**, 586-597.

Hennenfent, G. & Herrmann, F. J., 2008. Simply denoise: wavefield reconstruction via jittered undersampling, *Geophysics*, **73**: v19-v28.

Herrmann, F. J. & Hennenfent, G., 2008. Non-parametric seismic data recovery with curvelet frames, *Geop. J. Int.*, **173**, 233-248.

Herrmann, F. J., Wang, D. L., Hennenfent, G. & Moghaddam, P. P., 2008. Curvelet-based seismic data processing: a multiscale and nonlinear approach, *Geophysics*, **73**, A1-A6.

Kim, S.-J., Koh, K., Lustig, M., Boyd, S. & Gorinevsky D., 2007. An interior-point method for large-scale l_1-regularized least squares, *IEEE Journal on Selected Topics in Signal Processing*, **1**, 606-617.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso, *Journal Royal Statistical Society B*, **58**, 267-288.

Tikhonov, A. N. and Arsenin, V. Y., 1977. *Solutions of Ill-posed Problems*, John Wiley and Sons, New York.

Tropp, J. A. & Gilbert, A. C., 2007. Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, **53**, 4655-4666.

Wang, Y. F. & Yuan, Y. X., 2005. Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems, *Inverse Problems*, **21**, 821-838.

Wang, Y. F., 2007. *Computational Methods for Inverse Problems and Their Applications*, Higher Education Press, Beijing.

Wang, Y. F. & Ma, S. Q., 2007. Projected Barzilai-Borwein methods for large scale nonnegative image restorations, *Inverse Problems in Science and Engineering*, **15**, 559-583.

Wang, Y. F., Fan S. F. & Feng, X., 2007. Retrieval of the aerosol particle size distribution function by incorporating *a priori* information, *Journal of Aerosol Science*, **38**, 885-901.

Wang, Y. F., Cao, J. J., Yuan, Y. X., Yang, C. C. & Xiu, N. H., 2009. Regularizing active set method for nonnegatively constrained ill-posed multichannel image restoration problem, *Applied Optics*, **48**, 1389-1401.

Wang, Y. F., Yang, C. C. & Cao, J. J., 2011. On Tikhonov regularization and compressive sensing for seismic signal processing, *Mathematical Models and Methods in Applied Sciences*. (accepted)

Wang, Y. F., 2011. Sparse optimization methods for seismic wavefields recovery, *Proceedings of the Institute of Mathematics and Mechanics* (IMM, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia). (to appear)

Yuan, Y. X., 1993. *Numerical Methods for Nonliear Programming*, Shanghai Science and Technology Publication, Shanghai.