J. Inverse Ill-Posed Probl., Ahead of Print
DOI 10.1515/jip-2012-0030 © de Gruyter 2012

Data regularization using Gaussian beams decomposition and sparse norms

Yanfei Wang, Peng Liu, Zhenhua Li, Tao Sun, Changchun Yang and Qingsheng Zheng

Dedicated to Academician M. M. Lavrentiev on the occasion of the 80th anniversary of his birthday

Abstract. We consider seismic data regularization problems in this paper. Gaussian beam decomposition model is proposed for seismic data representation. To solve the representation problem, an l_0 quasi-norm minimization model with different smooth approximations is proposed. To solve the l_0 quasi-norm minimization problem, a projected gradient method with nonmonotone choice of iterative steps is developed. Numerical simulations on one-dimensional and two-dimensional seismic imaging problems are performed to verify the feasibility of our methods.

Keywords. Data regularization, Gaussian beams, l_0 optimization, sparse inversion.

2010 Mathematics Subject Classification. 65J22, 86-08, 86A22.

1 Introduction

In seismic exploration, the process of acquisition records the continuous wavefield 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 be at least twice of the maximum frequency of original signal. However, 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. This means that the data are usually incomplete. A direct effect of the limitations of the acquisition is 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 [26, 28, 35].

The research is supported by National Science and Technology Major Project 2011zx05057-001-003, National Natural Science Foundation of China under grant number 40974075 and Knowledge Innovation Programs of Chinese Academy of Sciences KZCX2-YW-QN107.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

In order to remove the influence of sub-sampled data and to restore the full data from incomplete data, the seismic data regularization technique is often used, e.g., *f-x* predictive method [19, 20], *f-k* domain method [12], interpolation methods relying on the *t-x* domain predictive filters [5, 6, 10], sparse optimization and regularization [28, 29, 35], linear Radon transform method [24] and curvelet transform method [13, 18]. Let us denote by m the original seismic wavefield, d the sampled data, and L the sampling operator, the data regularization can be written as (see [1])

$$Lm = d. \tag{1.1}$$

Our purpose is to restore m from the sampled data d. Since d is usually incomplete and L is an underdetermined operator, this indicates that there are infinite solutions satisfying the seismic imaging equation (1.1). Hence, seismic data regularization is an ill-posed inverse problem. Usually, we expect the signal m can be represented by linear combination of some functions, e.g., $m = \Psi c$, where c is a coefficient vector and Ψ the representing matrix with each column a transform vector. Then, solving (1.1) is equivalent to solving

$$L\Psi c = d. \tag{1.2}$$

Therefore, if the representing matrix Ψ exists, the coefficient vector c can be obtained through solving the equation (1.2), hence the wavefield could be restored from sampled data.

2 Gaussian beams decomposition

An important step for seismic data regularization is proper representation of the wavefields using some transforms, e.g., Radon transform, curvelet transform and so on. Then the wavefields can be represented with proper bases. In the far field assumption, the recorded wavefield can be considered as plane waves. Therefore, the wavefields interpolation could be decomposed with Gaussian beams [2, 14].

A Gaussian beam has the form

$$\psi(x,t) = c\exp[i2\pi\omega(x - x_c + p_x t)]\exp[-c_0(\omega/w)^2(x - x_c + p_x t)^2], \tag{2.1}$$

where c is the amplitude, x_c is the central location, c_0 is a constant, ω is the wave frequency, p_x is the slowness and w is the width of the Gaussian beam.

In two dimensions, a Gaussian beam has the form

$$\begin{aligned}\psi(x,t) = {} & c(t)\exp[i2\pi\omega(x - x_c + p_x t)] \\ & \times \exp[-c_0(\omega/w)^2(x - x_c + p_x t)^T M(t)(x - x_c + pt)],\end{aligned} \tag{2.2}$$

where $c(t)$ is the amplitude, x and x_c are in two dimensions, $M(t)$ is a two-dimensional matrix and other parameters are as mentioned above.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

Following the Gaussian beam function, the initial wave field can be given as

$$\psi(x) = \exp[i2\pi\omega(x - x_c)]\exp[-c_0(\omega/w)^2(x - x_c)^2]. \tag{2.3}$$

Again, x_c is the central location, c_0 is a constant, ω is the wave frequency and w is the width of the packet. The field $\psi(x)$ is called a Gaussian wave packet in one-dimensional case.

In two dimensions, a practical parameterized Gaussian wave packet can be formulated as follows (see [2]):

$$\psi_\gamma(x) = \exp[i2\pi\kappa^T(x - x_c)]\exp[-(x - x_c)^T R_\theta \Lambda(\alpha, \beta, \kappa) R_\theta^{-1}(x - x_c)], \tag{2.4}$$

where x, x_c and κ are in two dimensions, R_θ with θ the angle of orientation is the rotation matrix defined by

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

γ is the set of parameters given as $\gamma = (x_c, \kappa, \alpha, \beta, \theta)$, $\Lambda(\alpha, \beta, \omega)$ is a diagonal matrix defined by

$$\Lambda(\alpha, \beta, \omega) = \begin{bmatrix} \ln(16)\|\omega\|^2 & 0 \\ 0 & \ln(16)\|\omega\|^2/(\alpha^2\beta^2) \end{bmatrix},$$

where the parameters α and β define the number of oscillations within a half-width and the ratio of the Gaussian widths perpendicular to and along the direction of oscillation. Figures 1 and 2 illustrate the Gaussian beams in one dimension and two dimension cases, respectively. Figure 3 shows us a collection of Gaussian wave packets in different rotation angles.

Given a wavefield data $d(x)$ ($x = [x_1\ x_2]^T$) and a collection of wave packets ψ_γ, we aim to decompose the data in the following form:

$$d(x) = \sum_\gamma c_\gamma \psi_\gamma(x), \tag{2.5}$$

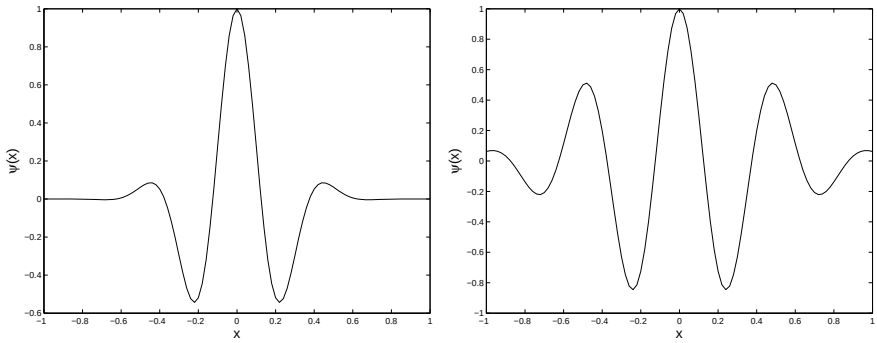
where c_γ for all chosen parameters $\gamma \in \Gamma$ is a collection of coefficients, Γ is the set for all possible parameters $\gamma = (x_c, \kappa, \alpha, \beta, \theta)$.

We could write the above equation into a compact form

$$d = Ac, \tag{2.6}$$

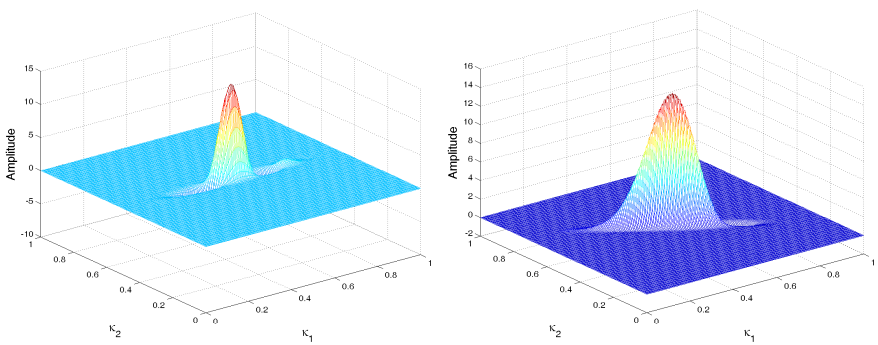
where A is the discrete operator forming by the wave packets $\psi_\gamma(x)$, $c = \{c_\gamma\}$ and d is the data with each column a vector. Now, a key issue in decomposition of data is solving the vector coefficient c given data d and proper forming of the operator A.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM



(a) (b)

Figure 1. A Gaussian beam in one dimension: (a) one oscillation within a half-width; (b) two oscillations within a half-width.



(a) (b)

Figure 2. (a) and (b) represent Gaussian beams in two dimensions in different angles.

3 Regularization model

Solving for the coefficient vector $c = \{c_\gamma\}$, the proper inversion model is required. It is evident that the Tikhonov regularization model in general form ([15, 22, 23])

$$J^\nu(c) := \frac{1}{2}\|Ac - d\|_{l_2}^2 + \nu\Omega(c) \tag{3.1}$$

with particular choice of the stabilizer $\Omega(c)$ works for this purpose, where $\nu > 0$ is the regularization parameter. For some special mathematical physics problems, the above regularization model follows the Lavrentiev's regularization [16]. The above

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

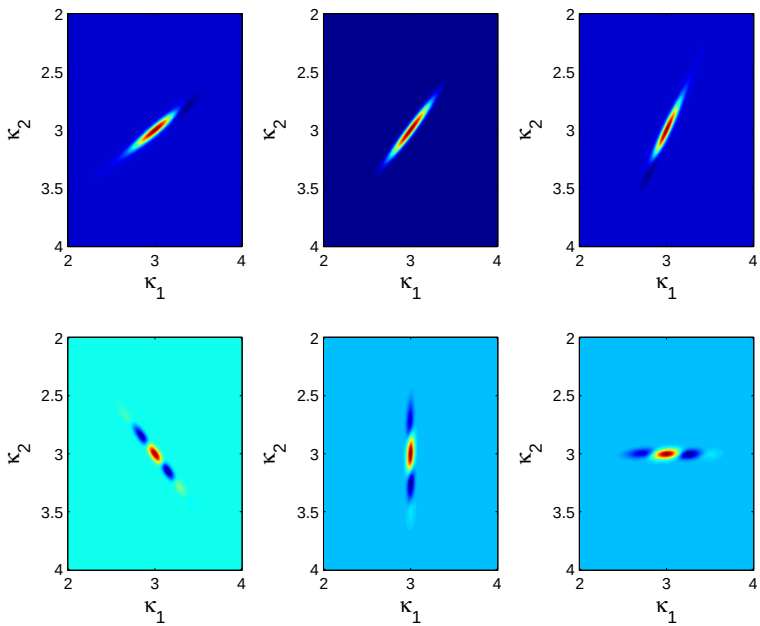


Figure 3. A collection of Gaussian wave packets in different rotation angles.

equation is equivalent to solving an equality-constrained minimization problem

$$\begin{aligned} &\min \Omega(c) \\ &\text{s.t. } Ac - d = 0 \end{aligned} \tag{3.2}$$

for noiseless data d or a constrained minimization problem

$$\begin{aligned} &\min \Omega(c) \\ &\text{s.t. } \|Ac - d\|_{l_2} \leq \Delta \end{aligned} \tag{3.3}$$

for inexact data d. In the following, we consider some specific sparse regularization models which will be used for regularizing data in data space.

4 Sparseness constraints

4.1 l_p-l_q norms constraint

The l_p norm in vector space $\mathbb{R}^N$ for an vector v is defined as

$$\|v\|_{l_p} = \left(\sum_i |v_i|^p\right)^{\frac{1}{p}}, \quad 0 \leq p \leq \infty.$$

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

The energy function

$$m(v,u) = \sum_i |v_i - u_i|^p$$

defines a metric. Different metric yields different scale of the functions.

Both the energy of the residual $r = Ac - d$ and the energy of the coefficient of c could be properly scaled using different norms. For example, the regularization model in general form could be formulated as an l_p-l_q model [30]

$$J^\alpha[c] := \frac{1}{2}\|Ac - d\|_{l_p}^p + \frac{\alpha}{2}\|D(c - c^0)\|_{l_q}^q \to \min, \tag{4.1}$$

for $p, q \geq 0$, c^0 a priori value and D a scale operator. If $p = q = 2$, (4.1) is the standard Tikhonov regularization in finite spaces; if $p = 2$ and $q = 1$, (4.1) is the commonly adopted sparse (spiky) regularization, especially frequently refereed in geophysical community, and also the LASSO problem in linear regression where they refer it as the l_1-constrained fitting for statistics and data mining [21]. Clearly, people favor l_2-l_1 model because that the objective function is convex and derivatives of J^α can be easily made, so solving methods could be easily developed. However, the l_1 constraint is not the sparsest "norm".

4.2 l_0 quasi-norm constraint

A best model to satisfy the "sparsity" requirement is the equality constrained minimization model with l_0 quasi-norm:

$$\|c\|_{l_0} \to \min, \text{ s.t } Ac = d. \tag{4.2}$$

This corresponds to choosing $\Omega(c) = \|c\|_{l_0}$ in (3.2), where $\|\cdot\|_{l_0}$ is defined as

$$\|v\|_{l_0} = \lim_{p\to 0} \|v\|_{l_p} = \lim_{p\to 0}\left(\sum_i |v_i|^p\right)^{\frac{1}{p}}.$$

Assuming that $0^0 = 0$, we could define

$$\|v\|_{l_0} = \{\text{num}(v \neq 0), \text{ for all } v \in \mathbb{R}^N\},$$

in which, $\text{num}(v \neq 0)$ denotes the cardinality of nonzero components of the vector v. In (4.2), the equality constraint is used, which means that the equation is consistent even with noisy data d. This is because practically the true signals are mixed with noises simultaneously and it is difficult to separate them sometimes. The equality constrained minimization model with l_0 quasi-norm could be relaxed into an unconstrained minimization problem

$$\tilde{c} = \arg\min_c \|c\|_{l_0} + \lambda\|Ac - d\|_{l_2}^2, \tag{4.3}$$

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

where $\lambda \geq 0$ is the Lagrangian multiplier. Minimization of $\|c\|_{l_0}$ means the number of nonzero values of c to be minimal. It is well known that minimization of $\|c\|_{l_0}$ is an NP-complex problem in combinatorial optimization, which indicates that optimization algorithms solving the problem cannot be finished in polynomial times. Hence this model is doomed to be infeasible in practice.

To overcome this NP-complex problem, people favor approximation of the l_0 quasi-norm minimization problem by an l_1 norm as discussed in Section 4.1, i.e., instead of (4.2) and (4.3), one solves

$$\|c\|_{l_1} \to \min, \text{ s.t } Ac = d \tag{4.4}$$

or

$$\tilde{c} = \arg\min_c \|c\|_{l_1} + \lambda \|Ac - d\|_{l_2}^2. \tag{4.5}$$

Solving (4.4) or (4.5) is easy as the objective function is convex, but in other words, solving the true l_0 quasi-norm optimization is superior to the l_1 norm optimization though both methods can yield sparse solutions. To maintain this property, we consider approximation of l_0 quasi-norm minimization problem. Let $f(t)$ be an one-variable derivable and convex function satisfying

$$\lim_{t\to 0} f(t) = 0 \text{ or } 1, \tag{4.6}$$

$$\lim_{t\to \infty} f(t) = 1 \text{ or } 0. \tag{4.7}$$

If one limit of the above condition is satisfied, then another condition can be easily satisfied by setting a new $f(t)$ as $1 - f(t)$. Therefore, for a single variable t, its l_0 quasi-norm can be defined via $f(t)$

$$\|t\|_{l_0} = f(t) \quad \text{or} \quad \|t\|_{l_0} = 1 - f(t). \tag{4.8}$$

For a given vectorized function $t \in \mathbb{R}^N$, its l_0 quasi-norm can be defined via $f(t_i)$, $i = 1, 2, \ldots, N$,

$$\|t\|_{l_0} = \sum_{i=1}^{N} f(t_i) \quad \text{or} \quad \|t\|_{l_0} = \sum_{i=1}^{N} (1 - f(t_i)). \tag{4.9}$$

To realize the above definition of l_0 quasi-norm, we usually introduce other parameters to specify some specific computable functions. A familiar function is the reverse Gaussian function:

$$f_\sigma(t) = 1 - \exp(-t^2/(2\sigma^2)),$$

where the parameter σ controlling the width of the Gaussian wave. This function

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

satisfies the following properties:

(a) $f_\sigma(t)$ is derivable and monotone,

(b) $f_\sigma(t)$ tends to the l_0 quasi-norm when σ tends to 0, i.e.,

$$\lim_{\sigma\to 0} f_\sigma(t) = \begin{cases} 0, & t = 0, \\ 1, & t \neq 0. \end{cases} \tag{4.10}$$

A plot of the function is shown in Figure 4a. Thus, we can construct a continuous function to approximate the l_0 quasi-norm, and then solve the optimal solution. In this way, problem (4.2) is approximated by

$$\min J_\sigma(c) := \sum_{i=1}^{N} f_\sigma(c_i), \quad \text{s.t. } Ac = d. \tag{4.11}$$

Other functions could be also applied. For example, solution of the differential equation (DE)

$$dg(t) = g(t)(1 - g(t))dt \tag{4.12}$$

may serve for our purpose. It is ready to see that a solution of the above DE can be written as $g(t) = e^t/(e^t + \text{const})$. If we set the boundary conditions to the DE (4.12) to be $g(0) = 1/2$, we obtain const $= 1$. So, the function $g(t)$ could be

$$g(t) = \frac{1}{1 + e^{-t}}, \tag{4.13}$$

which is a sigmoid curve, satisfying $\lim_{t\to\infty} g(t) = 1$. Therefore, we can construct a sigmoid-like curve function $f_\sigma(t)$ satisfying the conditions (4.6) and (4.10)

$$f_\sigma(t) = \frac{1}{1 + e^{-|t|/\sigma}}. \tag{4.14}$$

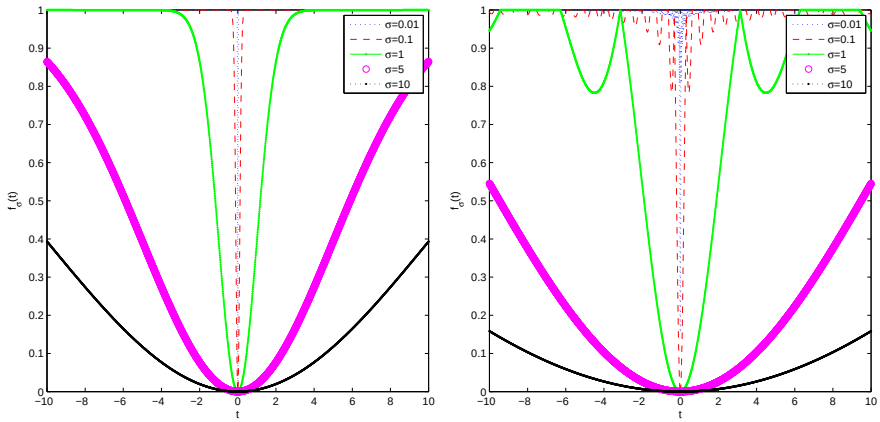
When σ approaches zeros, the above function approximates $\|t\|_{l_0}$ sufficiently. A plot of the function is shown in Figure 5a. It is immediately to see that

$$\tilde{g}(t) = 2 - \frac{1}{g(t)} = 1 - e^{-t} \tag{4.15}$$

satisfies $\lim_{t\to\infty} \tilde{g}(t) = 1$ and $\tilde{g}(0) = 0$. Therefore, this function may work better than $g(t)$ in the construction of the approximation function of l_0 quasi-norm. Therefore, we could formulate a new function $\tilde{f}_\sigma(t)$ as

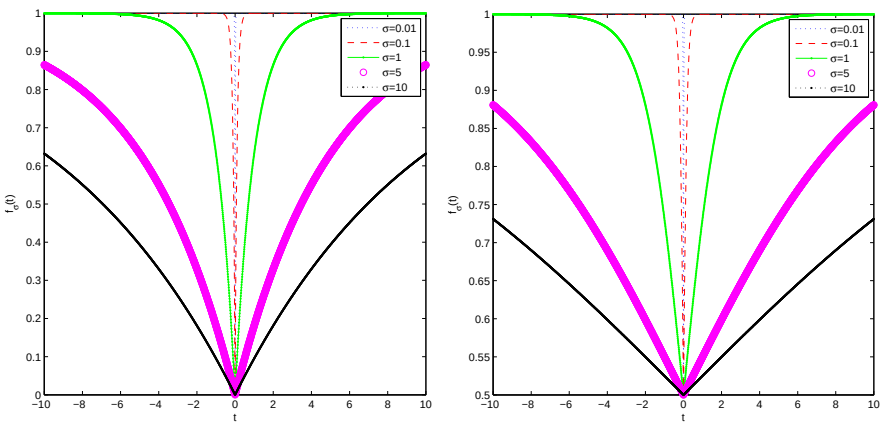
$$\tilde{f}_\sigma(t) = 1 - e^{-|t|/\sigma}. \tag{4.16}$$

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM



(a) (b)

Figure 4. Approximation of l_0 quasi-norm using different functions with an increasing value of σ: (a) $f_\sigma(t) = 1 - \exp(-t^2/2\sigma^2)$ and (b) $f_\sigma(t) = 1 - g(t/\sigma) = 1 - \sin(t/\sigma)/(t/\sigma)$.



(a) (b)

Figure 5. Approximation of l_0 quasi-norm using different functions with an increasing value of σ: (a) $f_\sigma(t) = \frac{1}{1+e^{-|t|/\sigma}}$ and (b) $f_\sigma(t) = 1 - e^{-|t|/\sigma}$.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

A plot of the function is shown in Figure 5b. It is illustrated from Figures 5a and 5b that these two functions shares similar approximation degree to the l_0 quasi-norm.

Another quite simple example satisfying conditions (4.6) and (4.7) is the function $g(t) = \sin t/t$, which is quite similar to a Dirac δ function. It is evident that

$$g(t) = \begin{cases} 1, & \text{as } t \to 0, \\ 0, & \text{as } t \to \infty. \end{cases}$$

We introduce $f_\sigma(t) = 1 - g(t/\sigma) = 1 - \sin(t/\sigma)/(t/\sigma)$ with the controlling parameter $\sigma > 0$ to realize the l_0 quasi-norm, i.e.,

$$f_\sigma(t) \approx \begin{cases} 0, & \text{as } \sigma \to 0 \text{ and } t \ll \sigma, \\ 1, & \text{as } \sigma \to 0 \text{ and } t \neq 0. \end{cases}$$

A plot of the function is shown in Figure 4b. However there are small oscillations as $\sigma \to 0$, this may cause problems in computation.

5 Solving methods

5.1 Previous solution methods for finding a sparse solution

There are a lot of numerical methods for finding a sparse solution of under-determined problems. Meanwhile, algorithms based on l_1-norm minimization take a big deal. For example, the basis pursuit denoising (BPDN) criterion [4,25] and the least absolute shrinkage and selection operator (LASSO) [21] are well studied in literature. The BPDN problem with $\|Ac - d\|_{l_2}^2 = \delta = 0$ (δ is the upper bound of the norm of the misfit) is equivalent to an equality-constrained linear programming problem, a particular method called the interior point (IP) solution method can be employed [31, 35]. However the IP solutions may be physically meaningless for some geophysical problems [31]. In [9], the authors express the minimization of the l_2-l_1 regularization problem with a nonnegativity-constrained quadratic program and solve it by projected gradient methods. Recently, the spectral gradient-projection method was developed for solving an l_1 minimization problem with l_2 norm constraints [8]. The method relies on projected gradient descent step (including nonmonotone gradient step [7, 32]) and root-finding of the parameter σ through solving the nonlinear convex, monotone and differential equation $\Gamma(\sigma) = \epsilon$. This method can handle both noisy and noiseless cases. However, it is clear that the root-finding method is just the famous "discrepancy principle" in regularization theory for ill-posed problems [33]. Relations between sparse recovery problems involving compressive sensing and regularization theory are discussed

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

in [35]. In addition, (orthogonal) matching pursuit method, a popular method in engineering, can also be used for solving a sparse recovery problem [4, 25]. 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. Because this method is not optimized, it may cause a lot of CPU time for convergence. Methods based on non-convex objectives are studied in [11, 17]. In [17], the l_0 quasi-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 [36] and fix point method are also developed. The article [3] makes a general review of different methods which may be applied in seismic data regularization. Recently, a globally convergent trust region method with sparse constraint is developed for data regularization [29].

5.2 Projected gradient methods for l_0 minimization

Solving the minimization problem (4.11) requires that the solution must be in the feasible set $S_c = \{c : Ac = d\}$. Therefore, a good initial guess value of the constrained minimization problem (4.11) is the one chosen from S_c. Since the null space of A is not empty, there are infinite number of solutions in S_c. To choose a particular solution, we consider the vector c can be represented by $c = A^T x$. This indicates that

$$x = (AA^T)^{-1} d, \tag{5.1}$$

where AA^T is full rank hence its inverse exists. This indicates that the solution $A^T(AA^T)^{-1}d$ is a good initial value in S_c of the problem (4.11). Now, we try to minimize the object function $J_\sigma(c)$. The Lagrangian function of (4.11) is

$$L(c, \lambda) = J_\sigma(c) - \lambda^T (Ac - d), \tag{5.2}$$

where $\lambda \geq 0$ is the Lagrangian multiplier. The object function $J_\sigma(c)$ is differentiable, thus the gradient and Hessian of $L(c, \lambda)$ can be evaluated as

$$\nabla_c L = \nabla_c J_\sigma - A^T \lambda = \left[\frac{\partial f_\sigma}{\partial c_1}, \frac{\partial f_\sigma}{\partial c_2}, \ldots, \frac{\partial f_\sigma}{\partial c_N} \right]^T - A^T \lambda, \tag{5.3}$$

$$\nabla_\lambda L = d - Ac \tag{5.4}$$

and

$$\nabla^2 L = \begin{bmatrix} H_c & -A^T \\ -A & 0 \end{bmatrix}, \tag{5.5}$$

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

respectively, in which, $H_c = \frac{\partial \nabla_c L}{\partial c}$. Now it is clear that the solution satisfies the following matrix-vector equations:

$$\nabla_c J_\sigma - A^T \lambda = 0, \tag{5.6}$$

$$Ac - d = 0. \tag{5.7}$$

However, both the gradient and the Hessian contain nonlinear functions. Therefore, for finding an optimized solution, we consider the gradient descent method for minimization of $J_\sigma(c)$ subject to equality constraints. The iterative formula reads as

$$c_{k+1} = c_k - \mu_k \nabla_{c_k} J_\sigma, \tag{5.8}$$

where μ_k is the step-length along the search direction $-\nabla_{c_k} J_\sigma$.

When we apply the gradient method to large scale problems, the most important issue is which step-length will give a fast convergence rate. Therefore it is vitally important to find what choices of μ_k require less number of iterations to reduce the gradient norm to a given tolerance. We consider the nonmonotone gradient methods which are proved to be effective both in theory and in seismic imaging [27, 32, 34, 37], i.e., choices for the step-length μ_k are based on two formulas:

$$\mu_k^1 = \frac{(s_{k-1}, s_{k-1})}{(s_{k-1}, y_{k-1})}, \quad \mu_k^2 = \frac{(s_{k-1}, y_{k-1})}{(y_{k-1}, y_{k-1})}, \tag{5.9}$$

where $y_{k-1} = \nabla_{c_k} J_\sigma - \nabla_{c_{k-1}} J_\sigma$, $s_{k-1} = c_k - c_{k-1}$, and our choice of the step-length is given by

$$\mu_k = \beta_1 \mu_k^1 + \beta_2 \mu_k^2, \tag{5.10}$$

where β_1 and β_2 are two positive parameters assigned by users.

To be sure that the solution is in the feasible set S_c, we use projection, which is defined by

$$\tilde{c}_{k+1} = P_{S_c}(c_{k+1}), \tag{5.11}$$

where $P_{S_c}(x)$ is defined by $x := x - A^T(AA^T)^{-1}(Ax - d)$.

Note that to maintain the approximation of $J_\sigma(c)$ to the l_0 quasi-norm, we require $\sigma \to 0$. We observe that the object function $J_\sigma(c)$ varies with changing of the parameter σ: the smaller value of σ, the closer behavior of $J_\sigma(c)$ to the $\|c\|_{l_0}$. For small values of σ, $J_\sigma(c)$ is highly non-smooth and contains a lot of local minimum, hence its minimization is not easy. On the other hand, for larger values of σ, $J_\sigma(c)$ is smoother and contains less local minimum, and its minimization is easier. Practically, we use a decreasing sequence values of σ: for minimizing $J_\sigma(c)$ for each value of σ, the initial value of the minimization algorithm is the minimum of $J_\sigma(c)$ for the previous value of σ. Then we apply a projected nonmonotone gradient method to solve for the coefficient vector c.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

6 Numerical experiments

6.1 Performance on the solution method for l_0 minimization

We consider a sparse signal $m \in \mathbb{R}^N$, which is measured 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. In what follows, $M < N$ means the number of measurements is smaller than the length of the signal, thus it is an under-determined discrete ill-posed problem. 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. Since the measurement is random, the data is randomly recorded.

We test three examples based on approximation of l_0 quasi-norm:

- the reverse Gaussian function $f_\sigma(t) = 1 - \exp(-t^2/2\sigma^2)$,
- sigmoid-like curve $f_\sigma(t) = \frac{1}{1+e^{-|t|/\sigma}}$,
- the function $f_\sigma(t) = \sin(t/\sigma)/(t/\sigma)$.

The original sparse random signals are shown in Figures 6–8 with legend "o" lines, respectively. Using our nonmonotone gradient descent algorithm, the restoration results ("+" lines) comparing with the original signal is shown in Figures 6–8, respectively. 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 using random measurement matrix. Therefore, it would be a reliable and stable method for potential practical problems.

Remark 6.1. Our experimental experience infers us that the most stable function approximating the l_0 quasi-norm is the reverse Gaussian function, although the above three functions yield similar recovery results. During testing on the function $f_\sigma(t) = \sin(t/\sigma)/(t/\sigma)$, the recovery results are not always stable. This is because that there are oscillations of this function before σ approaching zero (see Figure 4b). These oscillations may lead to multiples just like random noise. Therefore, this function is not recommended for applications.

6.2 One-dimensional seismic signal decomposition with Gaussian beams

The one-dimensional synthetic seismic simulations are very important in seismic interpretation. We consider a simple synthetic seismogram d generated by a Ricker wavelet convolved with an input signal with 5 peaks and 5 valleys (see Figure 9). The theoretical Ricker wavelet

$$w(t) = (1 - 2\pi^2 f_m^2 t^2)\exp(-(\pi f_m t)^2)$$

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

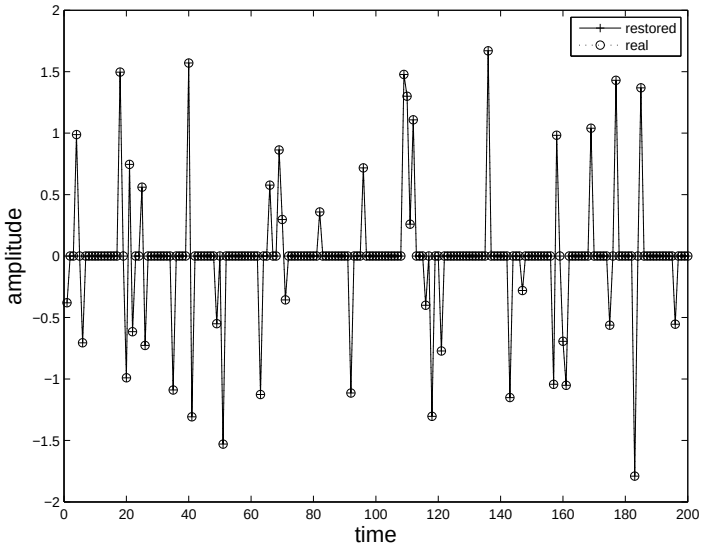


Figure 6. Comparison of the true and recovered random signals using our solving method for l_0 minimization: $f_\sigma(t) = 1 - \exp(-t^2/2\sigma^2)$.

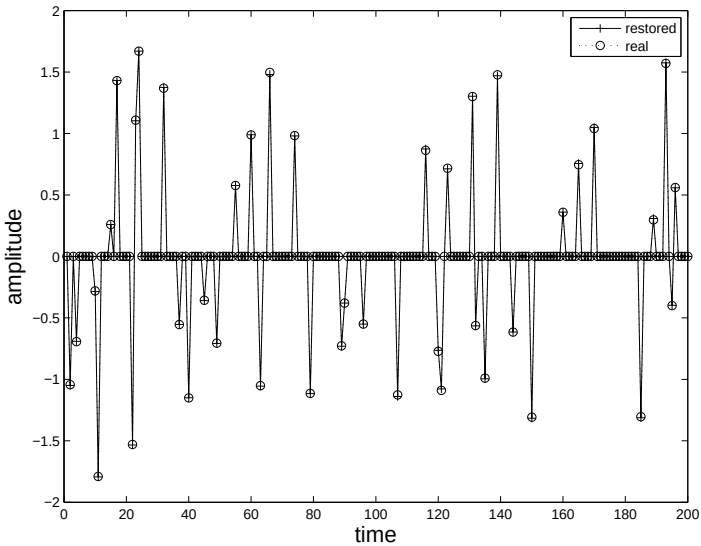


Figure 7. Comparison of the true and recovered random signals using our solving method for l_0 minimization: $f_\sigma(t) = \frac{1}{1+e^{-|t|/\sigma}}$.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

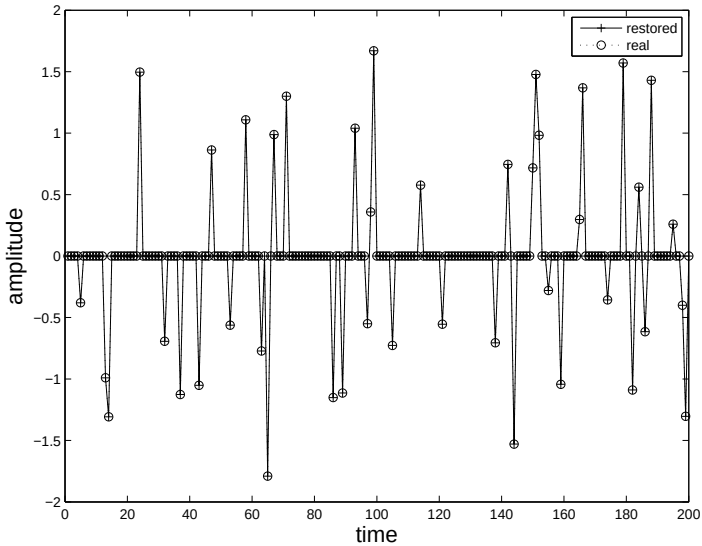


Figure 8. Comparison of the true and recovered random signals using our solving method for l_0 minimization: $f_\sigma(t) = \sin(t/\sigma)/(t/\sigma)$.

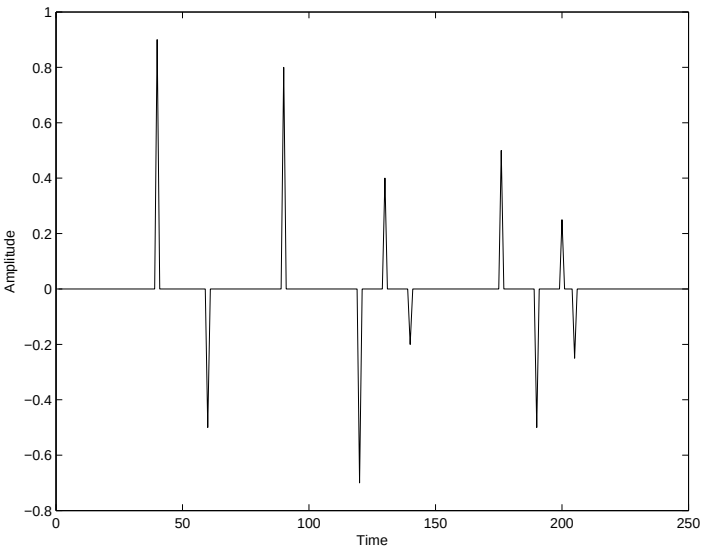


Figure 9. Input reflectivity function.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

is used to perform a convolution, where f_m represents the central frequency. The central frequency of the Ricker wavelet is 15 Hz, the sampling interval is 0.001 second. The seismogram is given in Figure 10.

The data $d(x)$ is decomposed by a collection of wave packets ψ_γ, i.e., $d(x) = \sum_\gamma c_\gamma \psi_\gamma(x)$, where c_γ for all chosen parameters $\gamma \in \Gamma$ is a collection of coefficients, Γ is the set for all possible parameters $\gamma = (x_c, \omega, w)$. The Gaussian beams used in this test are shown in Figure 11.

The restored data using Gaussian beams representation is shown in Figure 12. Comparing the original signal (Figure 10) with the restored signal using Gaussian beams (Figure 12) reveals that original signal can be fairly well represented using Gaussian beams.

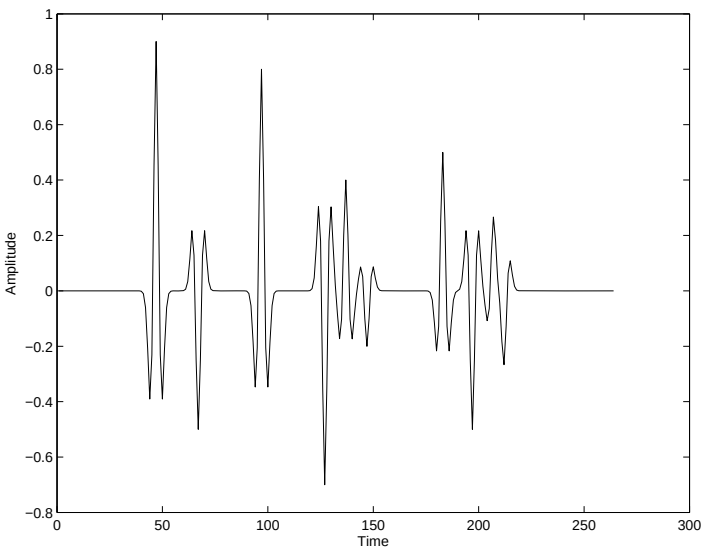


Figure 10. Synthetic seismic data.

6.3 Layered wavefield regularization

Figure 13 is a seismogram of layered model with angle of inclination. We assume again that there are abnormal traces, e.g., some traces are lost or no signal at all. This gives the image in Figure 14. We aim to regularize the data to restore the missed traces.

The data $d(x)$ ($x = [x_1\ x_2]^T$) is decomposed by a collection of wave packets ψ_γ, i.e., $d(x) = \sum_\gamma c_\gamma \psi_\gamma(x)$, where c_γ for all chosen parameters $\gamma \in \Gamma$ is a collection of coefficients, Γ is the set for all possible parameters $\gamma = (x_c, \kappa, \alpha, \beta, \theta)$. Collection of Gaussian beams is shown in Figure 15.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

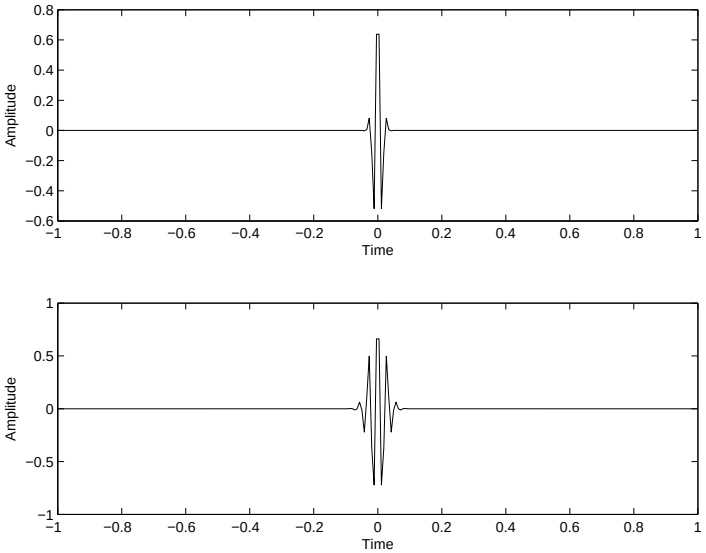


Figure 11. Gaussian beams.

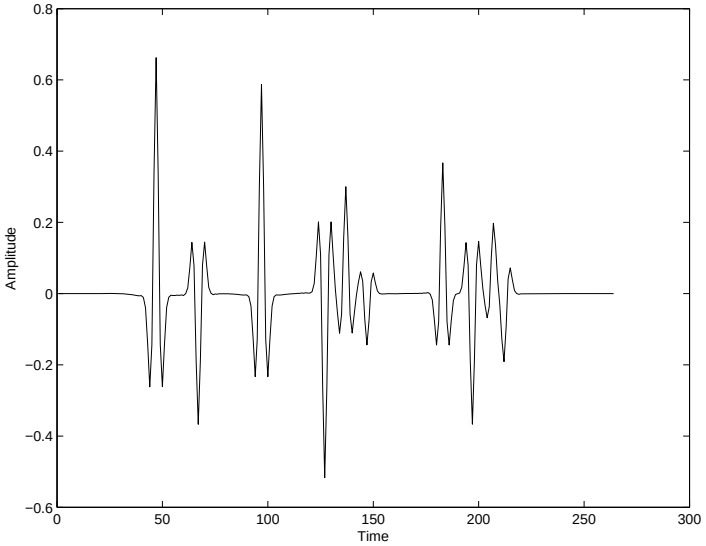


Figure 12. Restoration of original data using Gaussian beams decomposition.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

Using the l_0 minimization model and our algorithm, the restored result is shown in Figure 16. Evidently, the continuation of the data is guaranteed well. Comparing the original data (Figure 13) with the restored data using Gaussian beams (Figure 16) reveals that the data (especially the reflective wave) can be reasonably well represented using Gaussian beams, though the results is not significantly satisfactory. In our future research, we will optimize choosing the Gaussian beams and try to improve our algorithms to get better results.

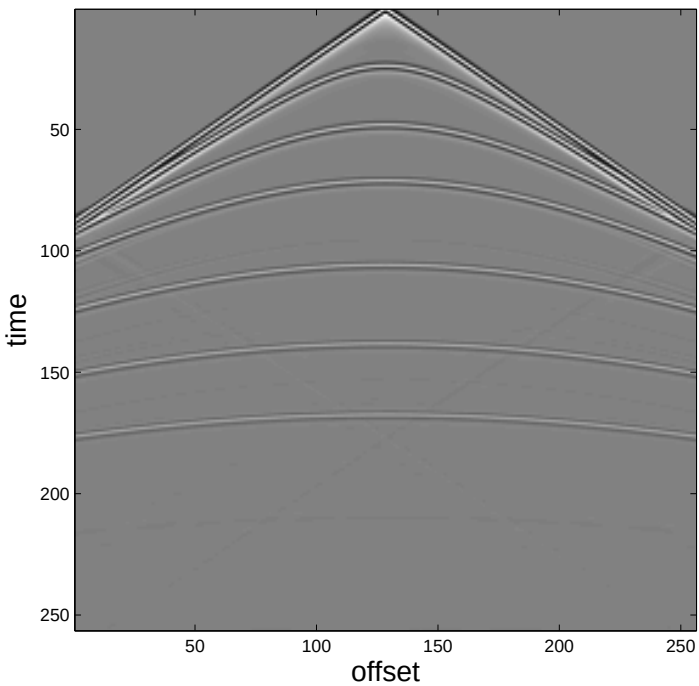


Figure 13. Seismogram of layered model.

7 Conclusion

We develop an l_0 quasi-norm constrained regularization model for seismic data regularization problems. Several smooth functions approximating l_0 quasi-norm is discussed. In building the regularization model, the Gaussian beams are collected to represent the wavefields. Solving methods based on projected nonmonotone gradient descent method are proposed. Finally, numerical simulations for one-dimensional and two-dimensional seismic imaging problems are performed. The results indicate the potential applications of our methods in the future.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

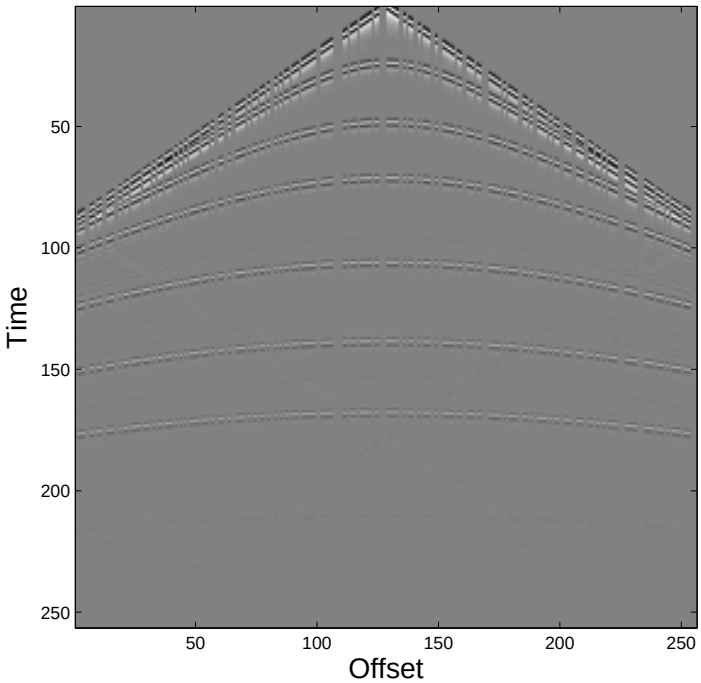


Figure 14. Random loss traces of data.

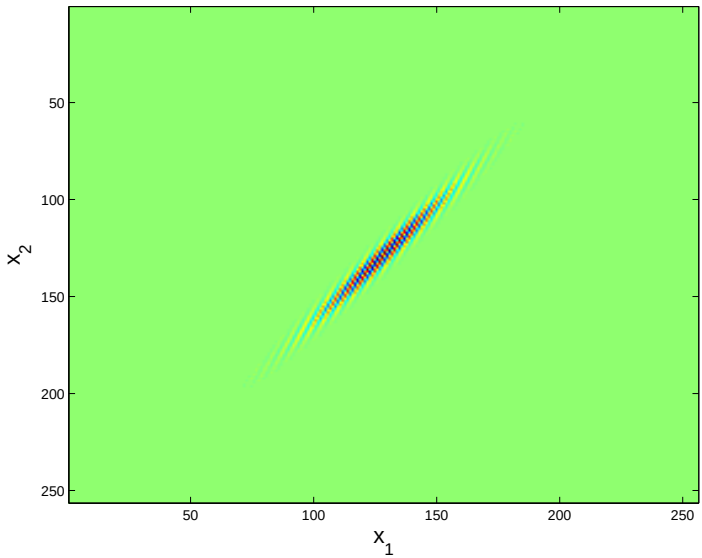


Figure 15. A collection of Gaussian beams used for representation of wavefields.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

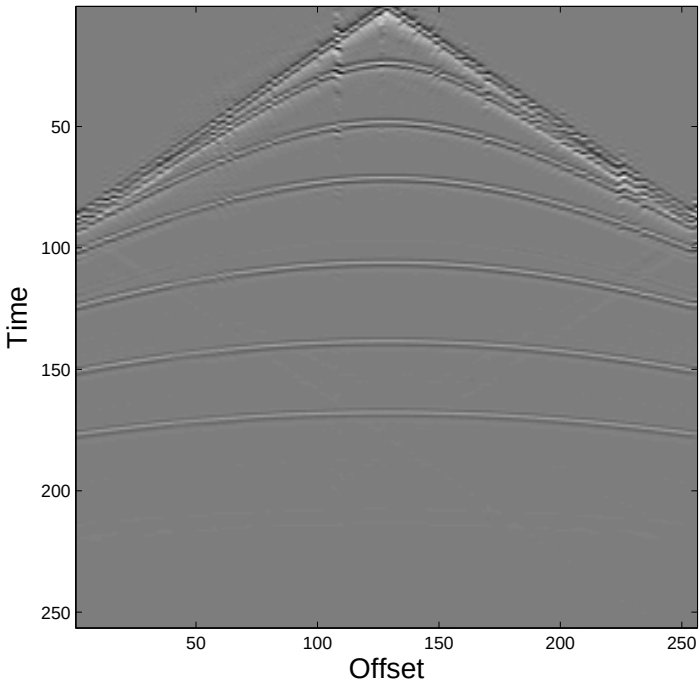


Figure 16. Restored data after regularization.

Bibliography

[1] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods*, W. H. Freeman, San Francisco, 1980.

[2] F. Andersson, M. Carlssony and L. Tenorioz, On the representation of functions with Gaussian wave packets, *J. Fourier Anal. Appl.* **18** (2012), 146–181.

[3] J. J. Cao, Y. F. Wang, J. T. Zhao and C. C. Yang, A review on restoration of seismic wavefields based on regularization and compressive sensing, *Inverse Probl. Sci. Eng.* **19** (2011), 679–704.

[4] S. Chen, D. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* **20** (1998), 33–61.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

[5] J. Claerbout, *Earth Soundings Analysis: Processing Versus Inversion*, Blackwell Science, Cambridge, 1992.

[6] S. Crawley, *Seismic trace interpolation with non-stationary predictionary-error filters*, PhD thesis, Stanford University, 2000.

[7] Y. H. Dai and R. Fletcher, Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numer. Math.* **100** (2005), 21–47.

[8] V. B. Ewout and P. F. Michael, Probing the pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.* **31** (2008), 890–912.

[9] M. A. T. Figueiredo, R. D. Nowak and S. J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signal Process.* **1** (2007), 586–597.

[10] S. Fomel, Application of plane-wave destruction filters, *Geophysics* **67** (2002), 1946–1960.

[11] 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), 4686–4698.

[12] N. Gulunay, Seismic interpolation in the Fourier transform domain, *Geophysics* **68** (2003), 355–369.

[13] G. Hennenfent and F. J. Herrmann, Simply denoise: Wavefield reconstruction via jittered undersampling, *Geophysics* **73** (2008), V19–V28.

[14] N. R. Hill, Prestack Gaussian-beam depth migration, *Geophysics* **66** (2001), 1240–1250.

[15] S. I. Kabanikhin, Definitions and examples of inverse and ill-posed problems, *J. Inverse Ill-Posed Probl.* **16** (2008), 317–357.

[16] S. I. Kabanikhin, V. G. Romanov and V. V. Vasin, Pioneering papers by M. M. Lavrentiev, *J. Inverse Ill-Posed Probl.* **15** (2007), 441–450.

[17] 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), 289–301.

[18] M. Naghizadeh and M. D. Sacchi, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data, *Geophysics* **75** (2010), WB189.

[19] M. J. Porsani, Seismic trace interpolation using half-step prediction filters, *Geophysics* **64** (1999), 1461–1467.

[20] S. Spitz, Seismic trace interpolation in the *F-X* domain, *Geophysics* **56** (1991), 785–794.

[21] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58** (1996), 267–288.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

[22] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York, 1977.

[23] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Kluwer, Dordrecht, 1995.

[24] D. O. Trad, T. J. Ulrych and M. D. Sacchi, Accurate interpolation with high-resolution time-variant Radon transforms, *Geophysics* **67** (2002), 644–656.

[25] J. A. Tropp and A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory* **53** (2007), 4655–4666.

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

[27] Y. F. Wang, Seismic impedance inversion using l_1-norm regularization and gradient descent methods, *J. Inverse Ill-Posed Probl.* **18** (2011), 823–838.

[28] Y. F. Wang, Sparse optimization methods for seismic wavefields recovery, *Trudy Inst. Mat. i Mekh. UrO RAN*, **18** (2012), 42–55.

[29] Y. F. Wang, J. J. Cao and C. C. Yang, Recovery of seismic wavefields based on compressive sensing by an l_1-norm constrained trust region method and the piecewise random sub-sampling, *Geophys. J. Int.* **187** (2011), 199–213.

[30] Y. F. Wang, J. J. Cao, Y. X. Yuan, C. C. Yang and N. H. Xiu, Regularizing active set method for nonnegatively constrained ill-posed multichannel image restoration problem, *Appl. Optics* **48** (2009), 1389–1401.

[31] 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), 885–901.

[32] Y. F. Wang and S. Q. Ma, Projected Barzilai-Borwein methods for large scale nonnegative image restorations, *Inverse Probl. Sci. Eng.* **15** (2007), 559–583.

[33] Y. F. Wang and T. Y. Xiao, Fast realization algorithms for determining regularization parameters in linear inverse problems, *Inverse Problems* **17** (2001), 281–291.

[34] Y. F. Wang and C. C. Yang, Accelerating migration deconvolution using a nonmonotone gradient method, *Geophysics* **75** (2010), S131–S137.

[35] Y. F. Wang, C. C. Yang and J. J. Cao, On Tikhonov regularization and compressive sensing for seismic signal processing, *Math. Models Methods Appl. Sci.* **22** (2012), 1150008-1–1150008-24.

[36] Y. L. Wang and W. T. Yin, Sparse signal reconstruction via iterative support detection, Technical Report, Rice University, 2009.

[37] Y. Yuan, Gradient methods for large scale convex quadratic functions, in: *Optimization and Regularization for Computational Inverse Problems and Applications*, edited by Y. F. Wang, A. G. Yagola and C. C. Yang, Springer-Verlag, Berlin (2010), 141–155.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM

Received May 8, 2012.

Author information

Yanfei Wang, 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`

Peng Liu, Lanzhou Center for Oil and Gas Resources, Institute of Geology and Geophysics, Chinese Academy of Sciences, Lanzhou, 730000, P. R. China; and University of Chinese Academy of Sciences, Beijing 100049, P. R. China.

Zhenhua Li, Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, P. O. Box 9825, Beijing 100029, P. R. China; and University of Chinese Academy of Sciences, Beijing 100049, P. R. China.

Tao Sun, Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, P. O. Box 9825, Beijing 100029, P. R. China; and University of Chinese Academy of Sciences, Beijing 100049, P. R. China.

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.

Qingsheng Zheng, Zhuzhihaihui Petroleum Company of Science and Technology, Beijing 100083, P. R. China.

Brought to you by | Technische Universitaet Berlin
Authenticated | 141.23.107.250
Download Date | 8/20/12 11:58 AM