ORIGINAL PAPER

Hybrid regularization methods for seismic reflectivity inversion

Yanfei Wang · Yan Cui · Changchun Yang

Abstract   In this paper, we consider new solution methods for seismic deconvolution problem in seismic data processing. We study regularization methods and introduce a hybrid regularization scheme for seismic deconvolution. Fast gradient descent methods are considered for solving the hybrid regularization model. An l_1 minimization model based on primal-dual interior point methods is presented and used for comparison. Numerical experiments for synthetic signal simulations and field data applications are given. The results show the potentials for applications of the developed method.

Keywords   Numerical solutions · Computational seismology · Regularization · Image processing · Numerical approximations

Mathematics Subject Classification (2000)   65J22 · 86-08 · 86A22

Y. Wang (✉) · C. Yang
Key Laboratory of Petroleum Resources Research, Institute of Geology
and Geophysics, Chinese Academy of Sciences, P.O. Box 9825,
Beijing 100029, People's Republic of China
e-mail: yfwang@mail.iggcas.ac.cn

Y. Cui
Geoscience Research Institute of Shengli Oilfield, SINOPEC, Dongying,
257015 Shandong Province, People's Republic of China

1 Introduction

In the Earth's subsurface, different layers have different impedances. The reflection seismic exploration is a method of exploration geophysics that uses the principles of seismology to estimate the properties of these layers from reflected seismic waves. The purpose of seismic inversion is to determine the spatial distribution of underground strata structure and physical parameter by using seismic wave propagation law. Meanwhile, a key step for reflectivity inversion is deconvolution (Robinson 1967; Aki and Richards 1980; Yilmaz 1987; Huang 1992). By deconvolution, we mean that we attempt to recover the reflectivity function from the seismic records. So far, the inversion and deconvolution technique has experienced the development process from direct inversion to model-based inversion, from post-stack inversion to pre-stack inversion, and from linear inversion to nonlinear inversion (Levenberg 1944; Marquardt 1963; Aki and Richards 1980; Tarantola 1987; Nolet 1987; Nolet and Snieder 1990; Trampert 1990; Dimri 1992; Bernabini and Cardarelli 1997). In recent ten years, nonlinear inversion methods have been utilized effectively. Practice shows that the nonlinear inversion is much closer to the real situation than the linear inversion. Nonlinear optimization methods are based on gradient computation, including the steepest descent method, Newton's method and the conjugate gradient method (Zhdanov 2002; Wang 2007; Wang et al. 2010); statistical methods include the Bayesian inference (Buland and Omre 2003; Buland et al. 2003; Duijndam 1988a,b; Gouveia and Scales 1997, 1998; Sen and Stoffa 1996), the neural network method, genetic algorithm (GA), simulated annealing algorithm (SA) and the random search algorithm (Ulrych et al. 1999, 2000; Wang 2002, 2007; Zhang et al. 2007; Mosegaard and Tarantola 1995; Mosegaard 1998). To improve the solution precision and accelerate the convergence speed, many hybrid optimization and regularization algorithms were developed (Yang 2002; Wang 2007). Roughly speaking, there are three classes of commonly used methods (Goldberg 1989). The first class of methods is based on local differential characteristics of the objective function, represented by the conjugate gradient method and variable metric methods. Since impedance inversion is quasi-linear, this kind of iterative methods is feasible, but the results strongly depend on the choice of the initial model. The second class of methods is randomized algorithm based on pure random searching, represented by SA. It can obtain global optimal solution, but has difficulty in determining temperature parameter and need large amount of calculation. The last kind of methods is intelligent algorithm which incorporates randomness and inheritance, represented by GA. One major disadvantage of GA is that it cannot invert too many parameters, so it is usually used to invert some characteristic parameters such as interval velocity and reflection depth.

Though a lot of research works have been done, they still cannot completely satisfy the practical requirements. The results of different methods may lead to differences and may result in incorrect geological explanation. The reasons may come from the low quality of seismic data, inaccurate wavelet extraction, and errors between normal incidence assumption and real situation. In addition, two main influences should be accounted: the first is the band-limited property of seismic data, hence direct inversion can only obtain the mid-frequency component of impedance model and lack the low-frequency and high-frequency component; second, in the Hadamard sense,

deconvolution and inversion are ill-posed problems, so the results inherit intrinsic instability and nonuniqueness (Tikhonov and Arsenin 1977; Zhdanov 2002; Wang 2007).

Nowadays, the most effective way to solve band-limited problem is the constrained inversion which can be generalized as the Tikhonov regularization (Tikhonov and Arsenin 1977), and the most effective way to solve ill-posed problem is the regularization methods assisted with proper optimization techniques (Wang 2007). However, classical regularization methods may lead to over-regularization (e.g., too smooth regularization) (Pilkington and Todoeschuck 1992; VanDecar and Snieder 1994) or insufficient/under regularization (e.g., too sparse regularization) (Wang et al. 2007). Robust methods should balance these two effects; this is what we studied in this paper. In particular, we present a smooth and nonsmooth hybrid nonlinear regularization method for retrieval of seismic reflectivity function. The hybrid regularization model can make a trade-off between the over-regularization and insufficient regularization. We also discuss about choices of proper smoothing scale operators and efficient solution methods. A fast gradient descent method based on Rayleigh quotient is developed. An l_1 minimization model based on primal-dual interior point methods is also considered in this paper. Finally, we demonstrate our method with both synthetic and field data examples.

2 Reflection model

The time series responses of the layers recorded by geophones to an impulse seismic source as the seismic signal are the effectively impulsive response of the layers. The signals recorded by geophones can be considered as the convolution of the reflectivity (model) and the system kernel

$$d_\delta(t) = d(t) + \delta \cdot n(t), \ \ d(t) = \sum_j k_j(t - \tau_j) r_j, \tag{1}$$

where r_k is the reflectivity (model), τ_k is the time delay, k_j is the kernel function, $d_\delta(t)$ is the seismic data, $\delta \in (0, 1)$ is the noise level and $n(t)$ is the additive noise. The system kernel $k(t)$ can be written as

$$k(t) = s(t) \star f_e(t) \star f_g(t),$$

where $\star$ refers to convolution, $s(t)$ defines the source kernel (wavelet), $f_e(t)$ defines the ground filter and $f_g(t)$ defines the receiver filter. Note that in practice, $k(t)$ has to be estimated since we only know the data, source signature and the geometric spreading of the receivers.

To estimate the kernel $k(t)$ from the data, we assume an uncorrelated random time series $x(t)$ which convolved with $k(t)$ yields the data $d(t)$. Denote $\phi_{dd}(t)$, $\phi_{kk}(t)$ and $\phi_{xx}(t)$ the autocorrelation of $d(t)$, $k(t)$ and $x(t)$, respectively, we have that (Huang 1992)

$$\phi_{xx}(\tau) = \sum_t x_t x_{t+\tau} = E(x_t x_{t+\tau}), \quad \phi_{kk}(\tau) = \sum_t k_t k_{t+\tau} = E(k_t k_{t+\tau}),$$

where E is the expectation. Since $x(t)$ is an uncorrelated series, hence $\phi_{xx}(t) = E(x_t x_t) = \sigma_x^2$ for $t = 0$ and $\phi_{xx}(t) = 0$ for $t \neq 0$, where σ_x^2 is the variance of the series $x(t)$. Thus we obtain

$$\phi_{dd}(\tau) = E(d_t d_{t+\tau}) = \sigma_x^2 \sum_t k_t k_{t+\tau} = \sigma_x^2 \phi_{kk}(\tau). \tag{2}$$

Equation (2) indicates that the kernel k of Eq. (1) can be estimated using the autocorrelation of the data d_δ.

3 Classical solution methods

Let us rewritten the Eq. (1) into the compact matrix-vector form

$$Kr + \delta \cdot n = d_\delta, \tag{3}$$

where K is the convolution operator, r is the seismic reflectivity and d_δ is the contaminated data with noise level $\delta \in (0, 1)$. In ideal case, i.e., the recording has no noise, (3) reduces to the equation

$$Kr = d. \tag{4}$$

The inverse problem is to determine r given d or d_δ and K. In the present paper, we do not consider the problems of blind deconvolution and source signature estimation.

A naive method for finding an approximate solution $\tilde{r}$ would be applying the inverse operator K^{-1} of the operator K to the Eq. (3) to find

$$\tilde{r} = K^{-1}d_\delta = K^{-1}d + \delta K^{-1}n = r + \delta K^{-1}n. \tag{5}$$

However, because of ill-conditioning property, large jump occurs for $K^{-1}n$. For non-invertible operator K, a least squares error method may be applied, which solves a residual minimization problem

$$\|Kr - d_\delta\|_{l_2}^2 \longrightarrow \min. \tag{6}$$

Denote the singular value decomposition of K as

$$K = V \Sigma U^T,$$

where $V = [v_1, v_2, \ldots]$ and $U = [u_1, u_2, \ldots]$ are orthogonal matrices, Σ is a diagonal matrix with nonzero elements in the main diagonal line. The solution of (6) can be

expressed as

$$\tilde{r} = \sum_k \frac{(d_\delta, v_k)}{\sigma_k} u_k = \sum_k \left[\frac{(d, v_k)}{\sigma_k} u_k + \delta \frac{(n, v_k)}{\sigma_k} u_k \right],$$

which is unstable for ill-posed problems since $\frac{(n, v_k)}{\sigma_k}$ may be astonishingly large for random noise n (Nashed 1976).

To tackle the ill-posed nature, regularization techniques were developed. The complete theory is established by Tikhonov and his colleagues (Tikhonov and Arsenin 1977). Tikhonov regularization method refers to solve a minimization of a "regularized functional" if we regard r and d_δ as functions

$$J^\alpha[r] := \frac{1}{2} \| Kr - d_\delta \|_{l_2}^2 + \frac{1}{2} \alpha \Omega[r], \tag{7}$$

where $\Omega[\cdot]$ is the so-called Tikhonov stabilizer, which can be preassigned by users; $\alpha \in (0, 1)$ is the regularization parameter, which plays a role in finding a trade-off to balance the instability and over regularization.

Many methods developed in literature are based on (7) or its variations. We notice that if $\Omega[r]$ is chosen as a smooth function, the result of regularization leads to a smooth solution; if $\Omega[r]$ is chosen as a nonsmooth (sparse) function, the result of regularization leads to a nonsmooth solution. Both choices may lead to over regularization (too smooth) or under regularization (too sparse). Our approaches try to balance these two aspects. Since the Earth can be considered to consist of reflectors on smooth and nonsmooth (subsurfaces), the seismic reflectivity has smooth (regular) as well as nonsmooth (irregular) components. In exploration geophysics, robust estimates of model parameters using norm usually refer to the norm possesses both good fitting of the huge data and efficient identifying the outliers. Under such requirements, the l_2 norm is not sufficient. In this paper, we not only consider the absolute value of deviation of the model (reflectivity) to the data, but also consider the absolute value of deviation of the variations of the model (reflectivity) to the data. This indicates an l_1 estimator.

4 Hybrid regularization

4.1 Smooth and nonsmooth hybrid regularization

A well used smooth stabilizer is $\Gamma(r) = \| D^{\frac{1}{2}} r \|_{l_2}^2$, where D is a scale operator, which can be homogeneous or nonhomogeneous. The nonsmooth regularization is important for problems when their model parameters are nonsmooth, or non-differentiable, or sparse. We adopt the widely used nonsmooth stabilizer $TV(r)$, where TV is the abbreviation of the total variation. Total variation regularization is not a new thing. Delicate descriptions of the methods in Hilbert spaces and the convergence analysis are given in (Nashed and Scherzer 1998; Ito and Kunisch 1999, 2000). Our choice for the Tikhonov stabilizer $\Omega[r]$ is the combination of the smooth stabilizer and the

nonsmooth stabilizer with different weight, i.e.,

$$\Omega[r] = \beta\Gamma(r) + \alpha DTV(r),$$

where $DTV(r)$ is the discretization of $TV(r)$. Using total variation stabilizer for noise removal has been discussed adequately in literature, e.g., (Rudin et al. 1992; Vogel and Oman 1998; Vogel 2002; Chan and Kang 2001; Wang et al. 2004; Wang 2007) for different applications. The definition of total variation of a continuous function f defined on the interval $[0, 1]$ is given by

$$TV(f) := \int_0^1 \left|\frac{df}{dt}\right| dt \qquad (8)$$

or

$$TV(f) := \sup_{\phi\in\Phi} \int_0^1 f(t)\frac{d\phi}{dt} dt, \qquad (9)$$

where $\Phi = \{\phi \in C^1(0, 1) : |\phi(t)| \le 1, t \in [0, 1], \phi(0) = \phi(1) = 0\}$. This corresponds to a sparse representation of variations of the function f in L_1 space. Details about the total variation related with L_1-norm minimization in general case can be found in (Ziemer 1989). Note that the unknowns may be sparse, or smooth or combining both, therefore, only one choice of the stabilizer $\Omega[r] = DTV(r)$ is insufficient. Considering the nonsmoothness of the stabilizer $TV(r)$ and practical computer simulations, we propose a hybrid regularization model as follows

$$J(r) = \frac{1}{2}\|Kr - d_\delta\|_{l_2}^2 + \alpha DTV(r) + \frac{1}{2}\beta\Gamma(r), \qquad (10)$$

where α and β are two regularization parameters, $\Gamma(r)$ is mentioned above and $DTV(r)$ can be approximated by discretization of the smooth function $M_\zeta(r) = \int_0^1 \sqrt{\left(\frac{dr}{dt}\right)^2 + \zeta^2} dt$ ($\zeta > 0$). The discretization of $M_\zeta(r)$ can be realized by setting $h_t = \frac{1}{n}$, then the parameter r can be discretized as $L_i r = \frac{r_i - r_{i-1}}{h_t}$, $i = 1, 2, \ldots, n$ and L_i for all i form an $n \times (n + 1)$ matrix. For solving the above minimization problem, one may try Gauss-Newton method (Yuan 1993) since the gradient and Hessian can be easily computed. For simplicity of notation, we set $\psi(r) := \frac{1}{2}\|Kr - d_\delta\|_{l_2}^2$. Denote the gradient of $\psi(r)$ as $\nabla\psi(r)$, $DTV(r)$ as $\nabla M(r)$, $\Gamma(r)$ as $\nabla\Gamma(r)$ and the gradient of $J(r)$ as $\nabla J(r)$, and denote the Hessian of $\psi(r)$, $DTV(r)$, $\Gamma(r)$ and $J(r)$ as $H_{\psi,r}$, $H_{M,r}$, $H_{\Gamma,r}$ and $H_{J,r}$, respectively, then the solution in each iteration can be written as

$$r_{k+1} = r_k - H_{J,r_k}^{-1}\nabla J(r_k), \qquad (11)$$

where H_{J,r_k}^{-1} is the inverse of H_{J,r_k}, and $\nabla J(r_k)$ and H_{J,r_k} can be written as

$$\nabla J(r_k) = \nabla \psi(r_k) + \alpha \nabla M(r_k) + \beta \nabla \Gamma(r_k)$$

and

$$H_{J,r_k} = H_{\psi,r_k} + \alpha H_{M,r_k} + \beta H_{\Gamma,r_k},$$

respectively. Details about calculation of $\nabla J(r)$ and $H_{J,r}$ are given in Appendix A. If we choose $\Gamma(r)$ as $\|D^{\frac{1}{2}}r\|_{l_2}^2$ with D symmetric and positive (semi-)definite, then $\nabla \Gamma(r_k)$ can be computed easily. Note that the derivative of the function $\phi(x) = 2\sqrt{x + \zeta^2}$ is greater than zero for $x \geq 0$, H_{J,r_k} must be symmetric and definite, hence there is a unique solution in the above equation.

We argue that, for large scale problems, solving the Gauss-Newton equation is quite expensive since in each iteration it requires decomposition of the Hessian matrix with amount of computation $O(n^3)$. This is inefficient for seismic data processing with many traces. Therefore, we consider gradient methods in Sect. 4.3.

4.2 Choosing the regularization parameters and the scale operator D

There are several ways for choosing the operator D. A simple way is choosing the D as the identity, i.e., the weight imposed to each element is identical. In Phillips-Twomey's formulation of regularization (Wang et al. 2006), the matrix D is created by the norm of the second differences, $\sum_{i=2}^{N-1}(x_{i-1} - 2x_i + x_{i+1})^2$, which leads to the following form of matrix D

$$
D = \begin{bmatrix}
1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
-2 & 5 & -4 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1
\end{bmatrix}.
$$

However, the matrix D is badly conditioned and thus the solution to minimize the functional $J(r)$ with D as the smooth constraint is observed to have some oscillations (Wang et al. 2006). Another option is the negative Laplacian (Wang 2007):

$$Lx := -\sum_{i=1}^{n} \frac{\partial^2 x}{\partial \tau_i^2},$$

for which the scale matrix D for the discrete form of the negative Laplacian Lx is

$$D = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

where we assume the discretization step-length is as 1. The scale matrix is positive semi-definite but not strictly positive definite and hence the minimization problem may not work efficiently for severely ill-posed inverse problems. It is reported in (Wang et al. 2006) and (Wang et al. 2008) that, the choice of $\Gamma(r)$ from discretization of a Sobolev norm function $\|x\|_{W^{1,2}}^2$ has better conditioning state than others and is especially useful for solving inverse problems of Earth sciences. Here the Sobolev $W^{1,2}$ space is defined as the set of functions which are continuous and differentiable with the bounded norms of themselves and their generalized derivatives in L_2, i.e., $W^{1,2}(\Pi) := \{x(t) : x(t) \in C(\Pi), x(t) \in L_2(\Pi), \frac{dx}{dt} \in L_2(\Pi), \int_{\Pi} \left(x^2 + \sum_{i=1}^{n} \left(\frac{dx}{dt_i} \right)^2 \right) dt_1 dt_2 \cdots dt_n < \infty \}$, where $C(\Pi)$ denotes the continuous space. The inner product of two functions $x_1(\tau)$ and $x_2(\tau)$ in $W^{1,2}$ space is defined by

$$(x_1(\tau), x_2(\tau))_{W^{1,2}} := \int_{\Omega} \left(x_1(\tau) x_2(\tau) + \sum_{i=1}^{n} \frac{\partial x_1}{\partial \tau_i} \frac{\partial x_2}{\partial \tau_j} \right) d\tau_1 d\tau_2 \cdots d\tau_n, \quad (12)$$

where Π is the assigned interval of the definition.

Assume that the variation of $x(\tau)$ is flat near the boundary of the integral interval Π. In this case, the derivatives of x are zeros at the boundary of Π. Let h_t be the step size of the grids in Π, then after discretization of $\|x\|_{W^{1,2}}^2$ using the finite difference technique, D is a tridiagonal matrix in the form

$$D = \begin{bmatrix} 1 + \frac{1}{h_t^2} & -\frac{1}{h_t^2} & 0 & \cdots & & 0 \\ -\frac{1}{h_t^2} & 1 + \frac{2}{h_t^2} & -\frac{1}{h_t^2} & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & -\frac{1}{h_t^2} & 1 + \frac{2}{h_t^2} & -\frac{1}{h_t^2} \\ 0 & \cdots & 0 & -\frac{1}{h_t^2} & 1 + \frac{1}{h_t^2} \end{bmatrix},$$

where h_t is the step size which equals the time sampling length.

Remark 1 For the choice of the two regularization parameters α and β, we suggest using a priori values, i.e., in computation, we set values of α and β to be in the interval $(0, 1)$. Statistically, it needs a large amount of experiments to simulate suitable values of α and β. However, according to our experience, choice of parameters $\beta < \alpha \leq 0.01$ works for our problem.

Of course, the a posteriori choice of regularization parameters is better than the a priori way. However, it is difficult to do since two optimal parameters α^* and β^* involving to be solved. In addition, the a posteriori choice of regularization parameters involves much more complex calculation and much more CPU time.

4.3 Gradient descent solution using Rayleigh quotient

The gradient method is one of the simplest methods for solving (10). Given an iteration point r_k, the gradient method chooses the next iteration point r_{k+1} in the following form:

$$r_{k+1} = r_k - \tau_k g_k, \tag{13}$$

where $g_k = \nabla J(r_k)$ is the gradient at r_k and $\tau_k > 0$ is a step-length. The gradient method has the advantages of being easy to program and suitable for large scale problems. Different step-lengths τ_k give different gradient algorithms. If $\tau_k = \tau^*$ where τ^* satisfies

$$J(r_k - \tau_k^* g_k) = \min_{\tau > 0} J(r_k - \tau g_k), \tag{14}$$

the gradient method is then the steepest descent method, which is also called the Cauchy's method. However, the steepest descent method, though it uses the "best" direction and the "best" step-length, turns out to be quite slow in convergence, particularly for ill-conditioned problems. The steepest descent method, however, is a simple but basic method, and has played an important role in the development of nonlinear optimization and many other fields. 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 τ_k require less number of iterations to reduce the gradient norm to a given tolerance.

Recently, nonmonotone gradient methods are much popular, see, e.g., (Yuan 2010; Wang and Ma 2007). We recall a well-known such kind of method, developed by Barzilai and Borwein (1988), which lies in the two choices for the step-length τ_k:

$$\tau_k^{BB1} = \frac{(s_{k-1}, s_{k-1})}{(s_{k-1}, y_{k-1})}, \quad \tau_k^{BB2} = \frac{(s_{k-1}, y_{k-1})}{(y_{k-1}, y_{k-1})}, \tag{15}$$

where $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = r_k - r_{k-1}$. This method initially designs for well-posed convex quadratic programming problems. However, it reveals that the method is also applicable for ill-posed problems and non quadratic programming problems provided that the deviation of the non quadratic model is not far away from the quadratic model (Wang and Ma 2007; Wang 2008).

Let us consider the quasi-Newton equation of the minimization problem (10). It is easy to deduce that the quasi-Newton equation satisfies

$$H_{J,r_{k+1}} s_k = y_k, \tag{16}$$

Noting that $s_k = -\tau_k g_k$, we have that

$$\tau_k^{BB1'} = \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, H_{J,r_k} g_{k-1})}, \quad \tau_k^{BB2'} = \frac{(g_{k-1}, H_{J,r_k} g_{k-1})}{(g_{k-1}, H_{J,r_k}^T H_{J,r_k} g_{k-1})}. \quad (17)$$

This indicates that the two step-lengths inherit different information from the kernel. In literature, people usually favor τ_k^{BB1} or $\tau_k^{BB1'}$. It is readily to see that $\tau_k^{BB1'} > \tau_k^{BB2'}$, i.e., the BB1 step-length is usually larger than the BB2 step-length. However, there is no reason to disregard τ_k^{BB2} or $\tau_k^{BB2'}$ since the method using BB2 step-length would be efficient if we want to obtain a very accurate solution of a very large-scale and ill-conditioned problem (Yuan 2010).

Further, we observe that the inverse of the scalar τ_k is the Rayleigh quotient of H_{J,r_k} or $H_{J,r_k}^T H_{J,r_k}$ at the vector g_{k-1}. This indicates that more choices can be obtained by combining Rayleigh quotients. To make a trade-off, we develop a new choice of the step-length by

$$\tau_k^{\text{Raleigh}} = \beta_1 \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, H_{J,r_k} g_{k-1})} + \beta_2 \frac{(g_{k-1}, H_{J,r_k} g_{k-1})}{(g_{k-1}, H_{J,r_k}^T H_{J,r_k} g_{k-1})}, \quad (18)$$

where β_1 and β_2 are two positive parameters assigned by users. In our opinion, we prefer $\beta_1 > \beta_2$, i.e., larger step-length is taken than shorter step-length in each iteration.

Remark 2 Let $J(r)$ be given in (10) with $DTV(r)$ the discretization of the smooth function $M_\zeta(r) = \int_0^1 \sqrt{\left(\frac{dr}{dt}\right)^2 + \zeta^2} dt$ ($\zeta > 0$) and $\Omega[m] = \frac{1}{2}\|Dm\|_{l_2}^2$, where D a positive (semi-)definite bounded matrix and let $\{r_k\}$ be generated by the above non-monotone gradient method with stepsize τ_k satisfying (18). It is easy to verify that H_{J,r_k} will be a positive definite matrix. Therefore, similar to (Yuan 2010), the sequence $\{r_k\}$ converges to the minimal solution of $J(r)$.

4.4 Comparison with the linear programming method

We recall that the l_1-norm minimization method (sometimes it is referred as sparse-spiky regularization in geophysics community) is quite important for seismic inversion (Claerbout and Muir 1973; Contreras et al. 2006). This is because that the l_1-norm could penalize outliers and large amplitude anomalies. Thus, l_1-norm is more robust than l_2-norm for seismic noise. However, l_1-norm has a singular problem when the values of residual vanish. Even if the values of residuals are not zero, the numerical inversion process goes to failure at very small residual. To tackle this difficulty, we reformulate the l_1-norm minimization model as

$$\|r\|_{l_1} \longrightarrow \min, \quad \text{subject to } Kr = d. \quad (19)$$

It is clear that Eq. (19) is equivalent to a constrained linear programming problem (B1) (see Appendix B). Solving a linear programming problem is much easier now.

Therefore, many methods can be applied to solve the problem, e.g., primal-dual interior point methods, second-order cone program and Newton methods for logarithm barrier problem (Yuan 1993). An outline of solving the l_1-norm minimization problem by linear programming is given in Appendix B.

Remark 3 Since the primal-dual interior point methods using the central path to finding a solution of the l_1-norm problem. Therefore, the solution is just picked one from the solution space, which sometimes may not be the wanted solution for specific problems, e.g., atmospheric aerosol particle size distribution (Wang et al. 2007), where the l_1-norm solution served as an a priori estimate of the true solution. However, the linear programming method is suitable for geophysical problems, e.g., impedance inversion (Levy and Fullagar 1981; Oldenburg et al. 1983) and inverse gravity problems (Huestis 1986; Cuer and Bayer 1980; Safon et al. 1977).

Remark 4 Note that if additive noise is considered, (19) can be formulated by combining l_1-norm and l_2-norm

$$\|r\|_{l_1} \longrightarrow \min, \quad \text{subject to } \|Kr - d_\delta\|_{l_2} \leq \epsilon, \tag{20}$$

where ϵ is a nonnegative real parameter. Problems (19) and (20) are closely related to the following two problems:

$$\|Kr - d_\delta\|_{l_2}^2 + \lambda\|r\|_{l_1} \longrightarrow \min \tag{21}$$

and

$$\|Kr - d_\delta\|_{l_2}^2 \longrightarrow \min, \quad \text{subject to } \|r\|_{l_1} < \sigma, \tag{22}$$

where λ is the Lagrangian multiplier and σ is a nonnegative real parameter. With appropriate parameter choices of ϵ, λ and σ, the solutions of (20), (21) and (22) coincide, and these problems are in some sense equivalent. Performance of the l_1-norm and l_2-norm hybrid method was shown in (Wang 2011).

5 Numerical results

5.1 1d signal simulation

Numerical simulation is a necessary step to prove the efficiency of our proposed method before applying it to practical application problems. The simulation consists of two steps. First, a simulated signal is generated by computer according to Eq. (3) for a given reflectivity function $r^{\text{true}}(t)$. Then, the generated signal is processed through our algorithm, and the retrieved distribution signal is compared with the input one.

We first generate a reflectivity signal which consists of two sparse nonsmooth parts (rectangular signals) and two smooth parts (sine signals). Then we use this signal to
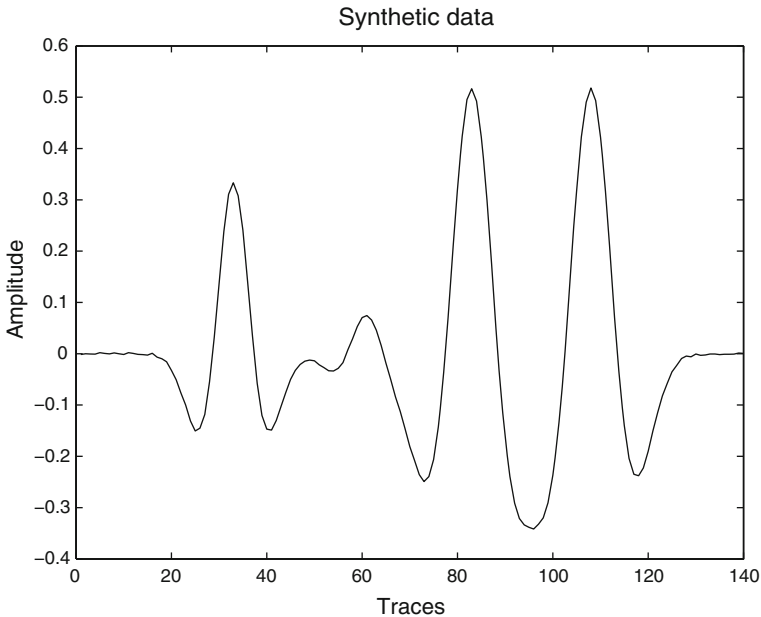
Fig. 1 Synthetic data

generate data. The operator K is generated using Ricker wavelet

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

with central frequency equaling $f_m = 27.5$ Hz and time sampling interval equaling
2 ms. The simulated data is generated using Eq. (3) with additive Gaussian random
noise, see Fig. 1. The true reflectivity signal is shown in Fig. 2 (dashed lines). Com-
parison of the input and the retrieved reflectivity signal is shown in Fig. 2 ("+" lines
refer to retrieved results). To illustrate the matching, we plot the synthetic data and the
recovered data using the retrieved reflectivity signal in Fig. 3. It reveals that the fit is
quite good.

We also apply the sparse-spiky l_1-norm minimization method to the above prob-
lem. Numerical results are obtained using the primal-dual interior point method,
see Fig. 4. We also make tests for smooth-constrained or nonsmooth-constrained
minimization problem (10). For the former, we set $\alpha = 0$; while for the latter, we
set $\beta = 0$. The numerical results are shown in Figs. 5 and 6, respectively. We also
apply the above three methods for restoring the data. The results are shown in
Figs. 7, 8 and 9, respectively. It is evident that the fittings using l_1-norm minimi-
zation based on linear programming are not as good as others. To make a compar-
ison, we also zoom the selected area for the restoration results using hybrid reg-
ularization, smooth regularization and nonsmooth regularization, see Figs. 10, 11
and 12, respectively. It reveals that the hybrid regularization method yield the best
results.
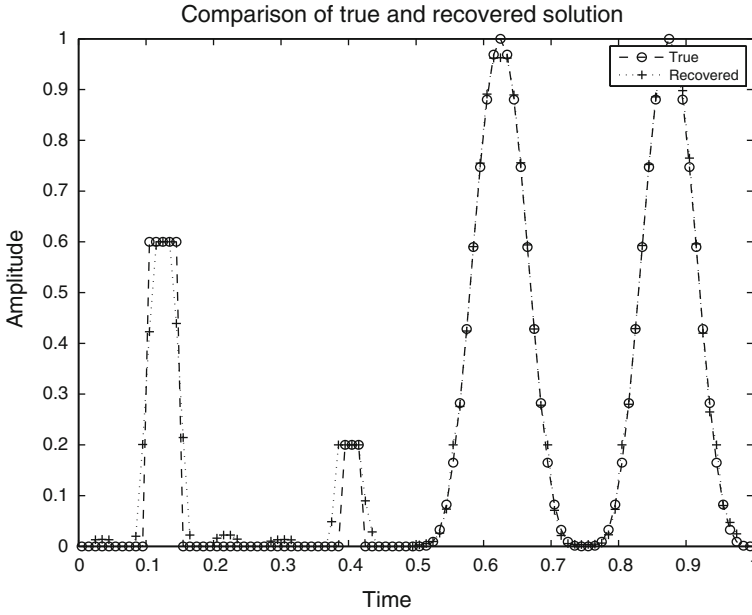
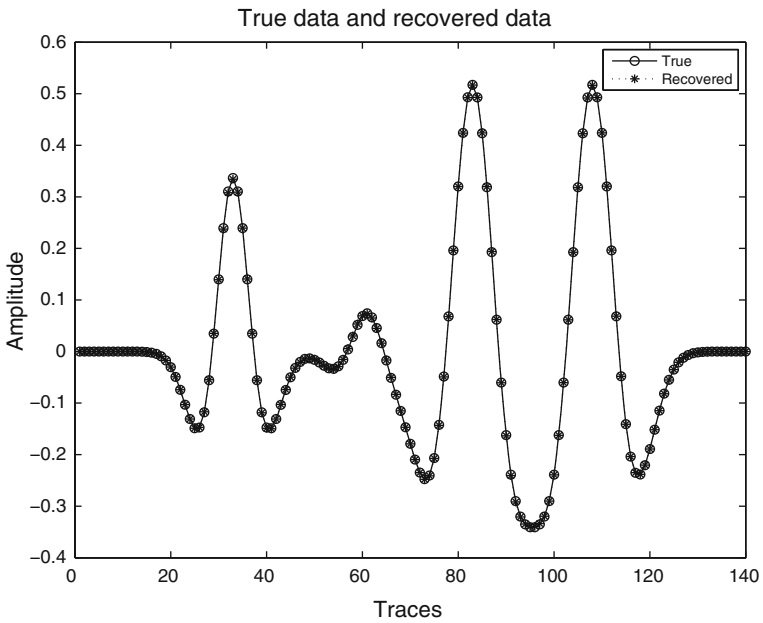**Fig. 2** Comparison of the input and recovered reflectivity signals

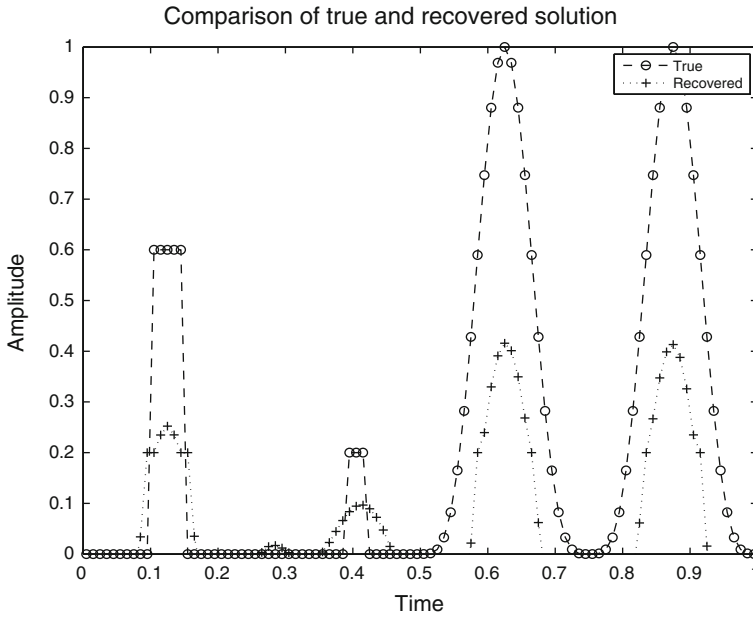**Fig. 3** Comparison of the synthetic and recovered data

Comparison of true and recovered solution

Fig. 4 Comparison of the input and recovered reflectivity signals using l_1 norm minimization
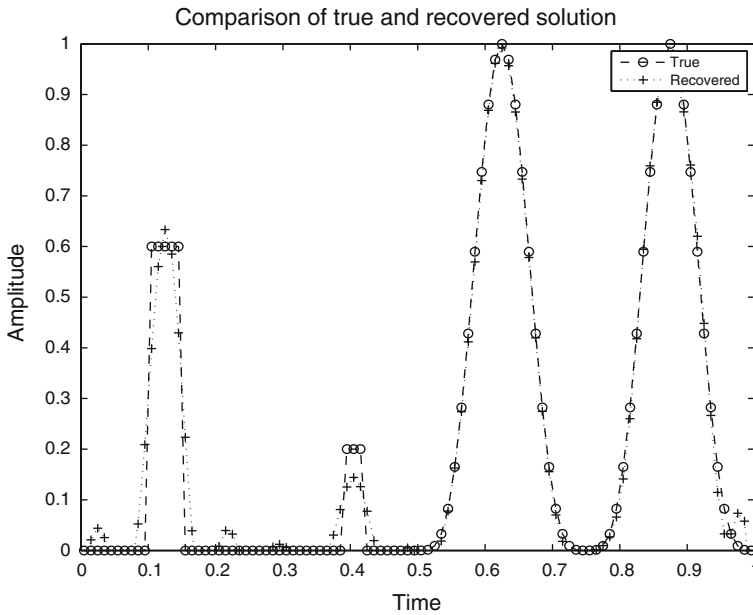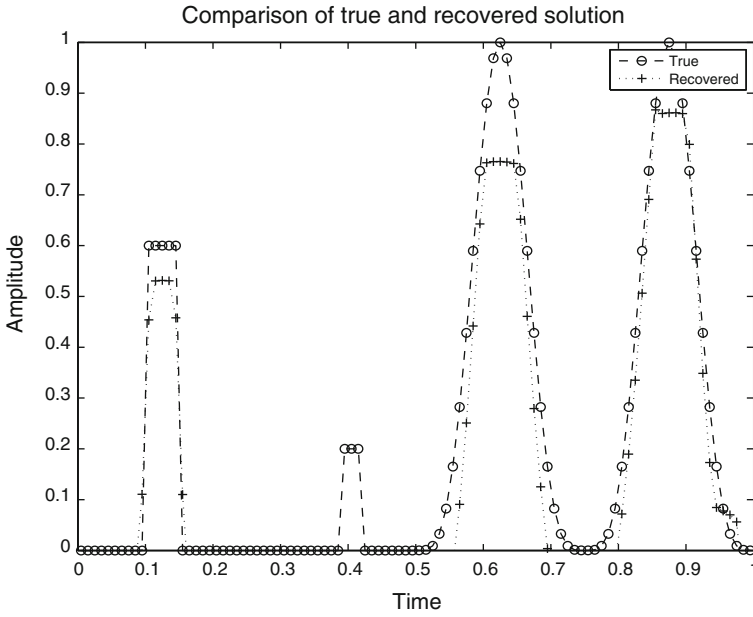
Comparison of true and recovered solution

Fig. 5 Comparison of the input and recovered reflectivity signals using smooth regularization

Fig. 6 Comparison of the input and recovered reflectivity signals using nonsmooth regularization
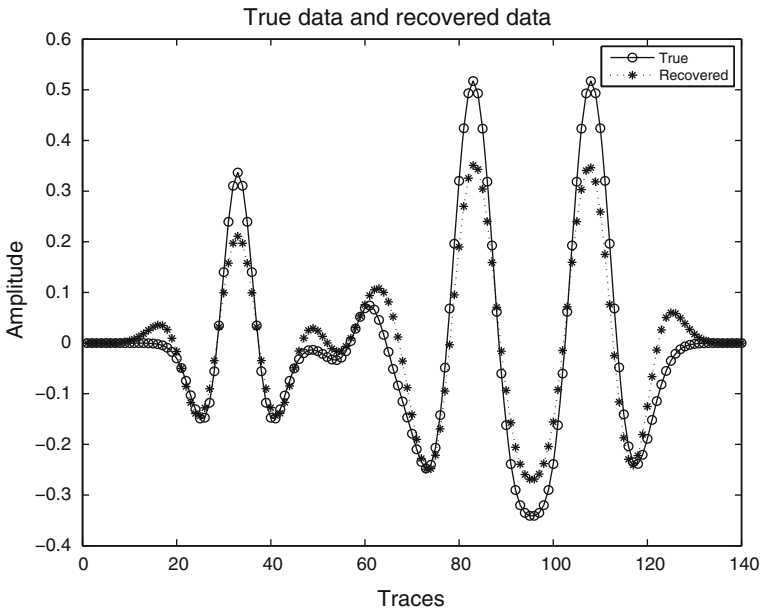
Fig. 7 Comparison of the synthetic and recovered data using l_1 norm minimization
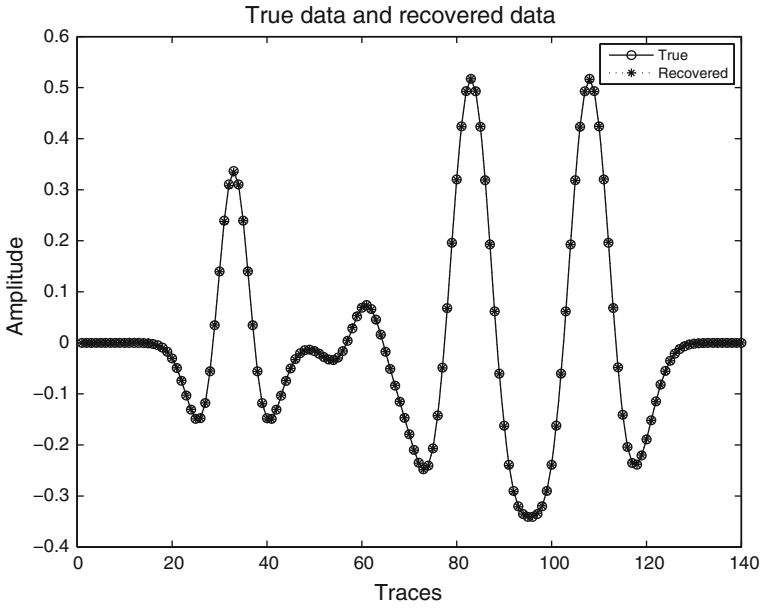
Fig. 8 Comparison of the synthetic and recovered data using smooth regularization
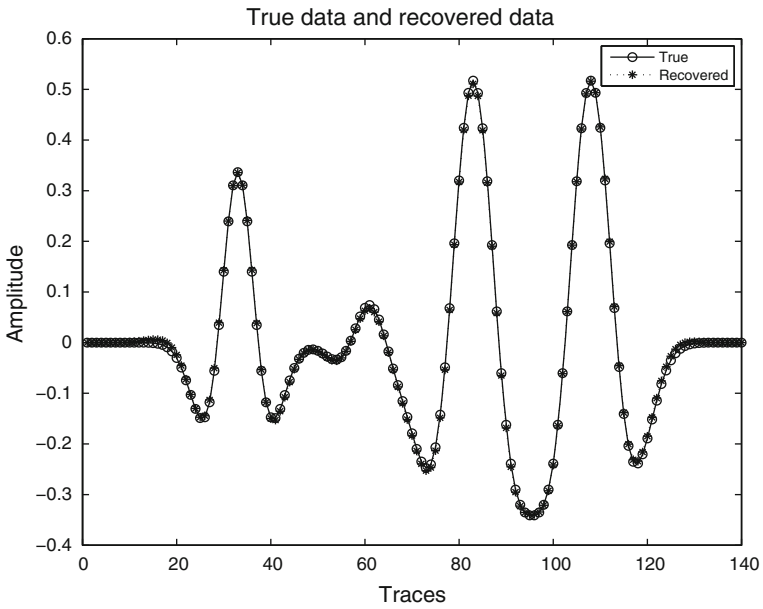
Fig. 9 Comparison of the synthetic and recovered data using nonsmooth regularization
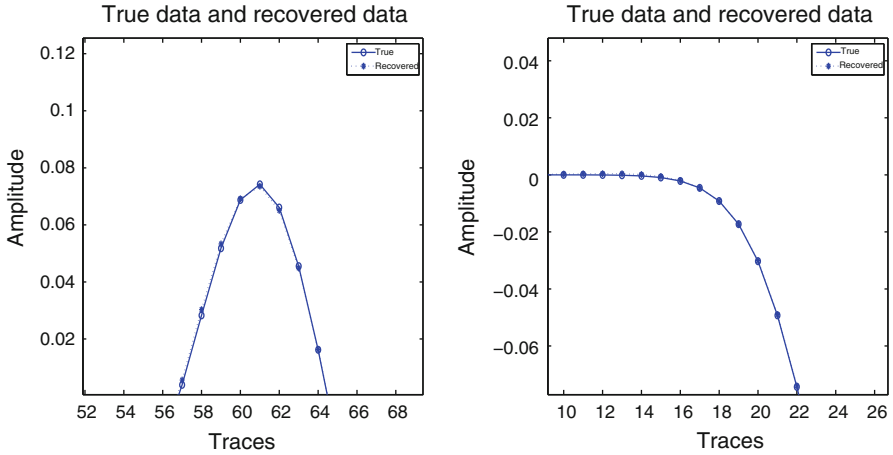
True data and recovered data

Fig. 10 Comparison of selected area of the synthetic and recovered data using hybrid regularization
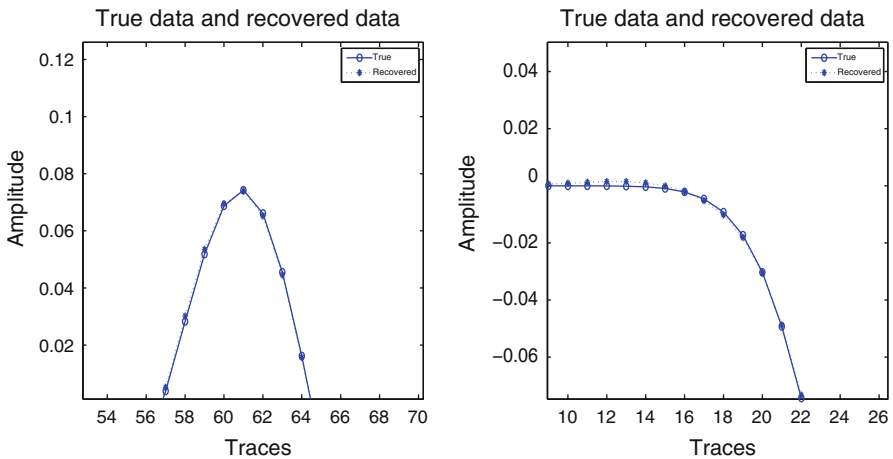
True data and recovered data

Fig. 11 Comparison of selected area of the synthetic and recovered data using smooth regularization

5.2 Layered model simulations

In the following we consider a simple 4 layers velocity model, see Fig. 13. To generate the seismogram, a Ricker wavelet with central frequency equaling 60 Hz and sampling interval equaling 4 ms is used to perform a convolution. The noisy seismogram with additive Gaussian random noise is shown in Fig. 14. By our algorithm, the retrieved reflectivity model is illustrated in Fig. 15. The inversion results indicate that the reflectivity is well established.
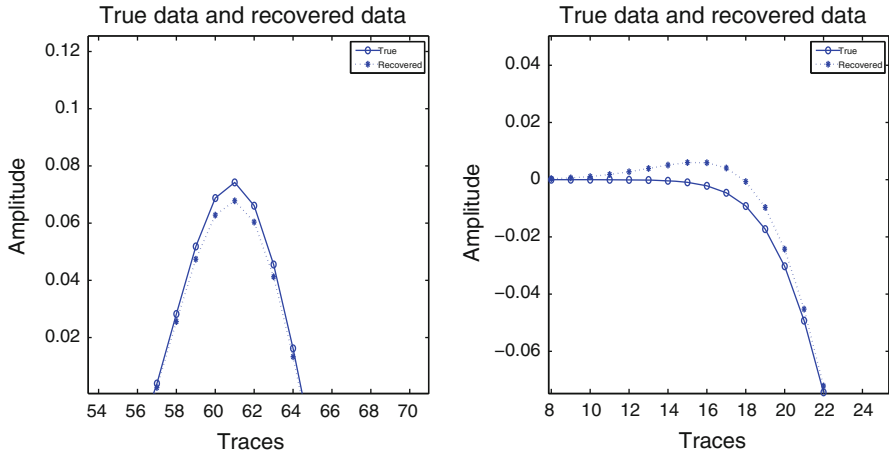
Fig. 12 Comparison of selected area of the synthetic and recovered data using nonsmooth regularization
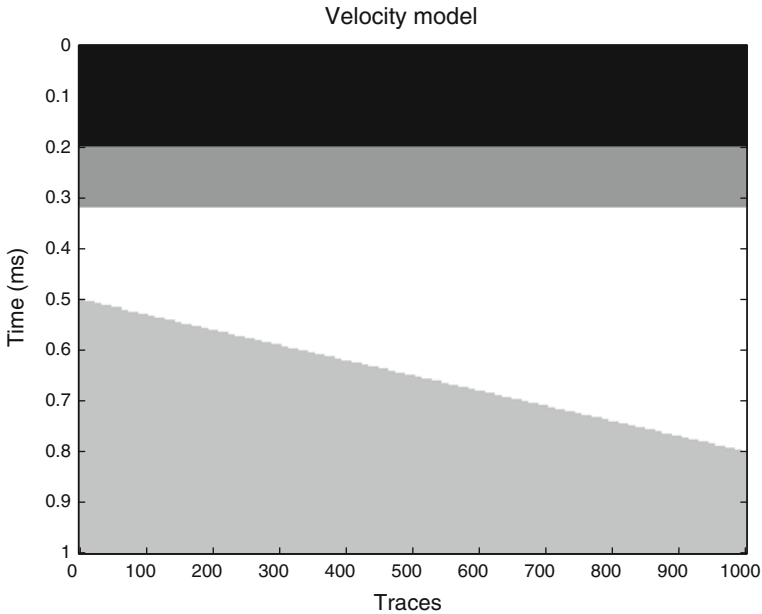

Fig. 13 Velocity model

5.3 Field applications

We apply our method to the field data. The data is taken from the Tarimu Oilfield. The time sampling is 2 ms and the spacing of traces is 20 m. Figure 16 plots the original seismic section. The wavelet function was extracted using the autocorrelation method mentioned in Sect. 2. Then we apply our algorithm to the seismic data. The retrieved
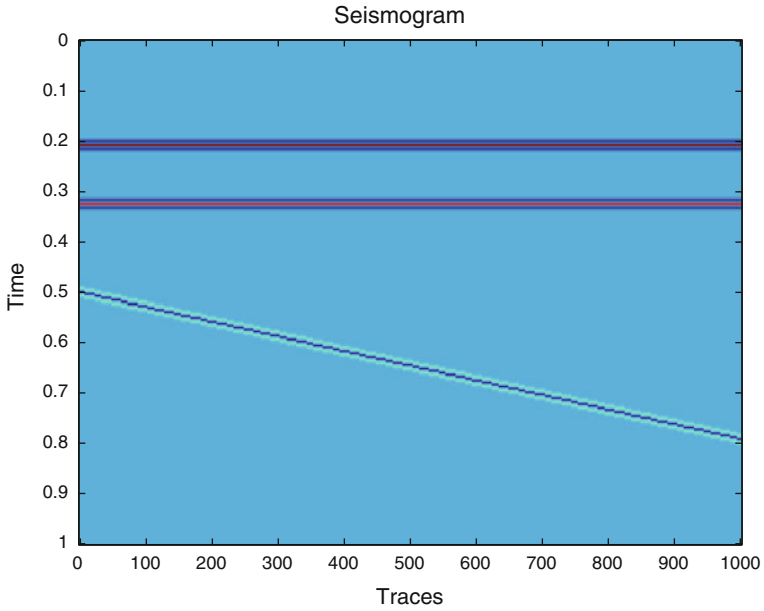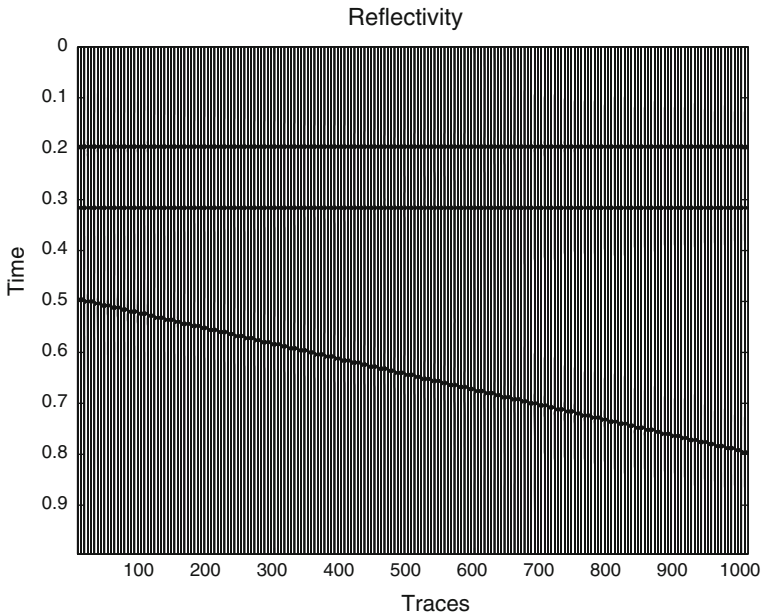
Fig. 14 The synthetic seismogram

Fig. 15 The retrieved reflectivity model

reflectivity function is shown in Fig. 17. It is illustrated from the inversion results that the recovered section well represents reflection of layers and much noise is reduced using our algorithms.
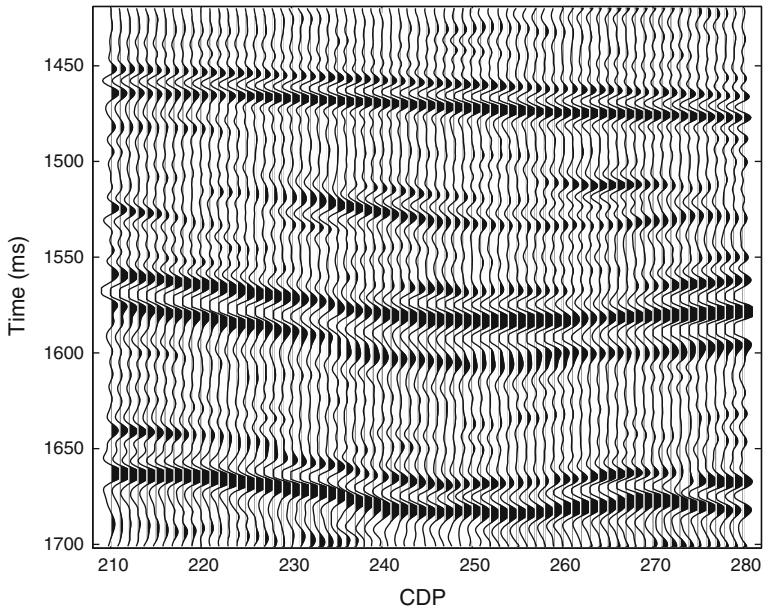
Fig. 16 Seismic section
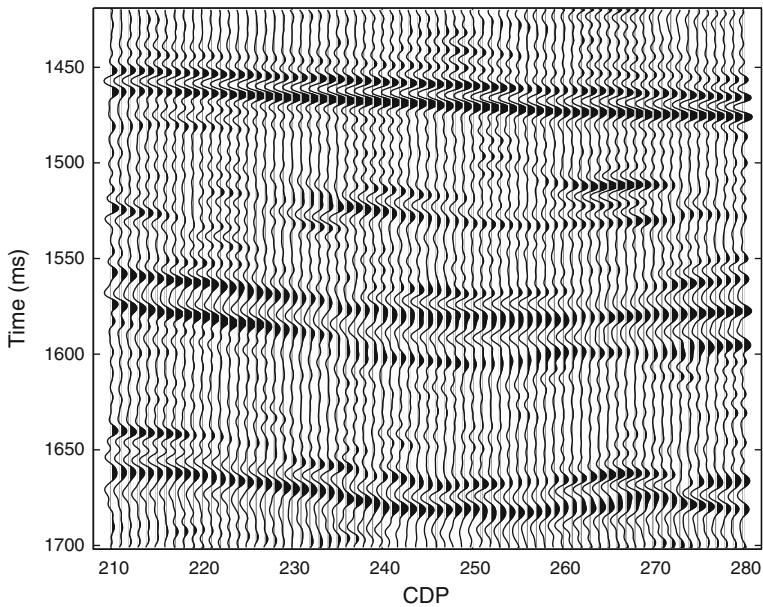

Fig. 17 Recovered reflectivity function

6 Discussion and conclusion

In this paper, we develop a smooth and nonsmooth hybrid regularization method for
solving ill-posed and band-limited seismic deconvolution problem. In solving the

minimization model, a fast gradient descent method based on Rayleigh quotient is used. Comparison of this method with the linear programming method for l_1-norm minimization is performed. It reveals that the hybrid regularization model with gradient descent method outperforms the linear programming method and can: (1) balance the smooth stabilizer (e.g., l_2 norm or $W^{1,2}$ norm) and nonsmooth stabilizer (e.g., sparseness); (2) yield fast convergent results; (3) adequately remove noise, suppress ill-posedness and restore stability.

We argue that for solving the minimization of the nonlinear hybrid regularization problem, a key point is the proper fast economic algorithms. This is particularly desirable for large scale problems. To enhance the fast convergence of gradient type of methods, many techniques developed in literature are based on preconditioning (Golub and Van Loan 1996). For our hybrid regularization problem, because of the nonlinearity and indifferentiability of the objective function, how to incorporate proper preconditioning techniques deserves our further investigation.

Finally, we want to mention that in (Meyer 2001), the author proposed a form of hybrid regularization form

$$\min \|A\rho - z\|_{L_2}^2 + \alpha \Gamma[u] + \beta G[v], \quad \text{subject to } u + v = \rho,$$

for ill-posed problems $A\rho = z$, where u and v are two parameters matching the unknown parameter ρ, $\Gamma[u]$ and $G[v]$ are smooth function and nonsmooth function, respectively. If we consider the reflectivity function can be decomposed into two parts, e.g., smooth parts and nonsmooth parts, we may use his method, this could be an interesting topic.

Appendix A: Gradient and Hessian Calculations

In discrete form, the $M_\zeta(r)$ can be written as

$$M_\zeta(r) = \frac{1}{2} \sum_{i=1}^{n} \phi((L_i r)^2) h_t. \tag{A1}$$

The gradient of $M_\zeta(r)$ can be obtained by for any $u \in \mathbb{R}^{n+1}$,

$$\frac{d}{d\tau} M_\zeta(r + \tau u)|_{\tau=0} = \frac{1}{2} \frac{d}{d\tau} \sum_{i=1}^{n} \phi((L_i(r + \tau u))^2) h_t|_{\tau=0}$$

$$= (h_t L^T \text{diag}(\phi'(r)) L r, u), \tag{A2}$$

that

$$\nabla M(r) = M(r)r, \tag{A3}$$

where $\mathrm{diag}(\phi'(r))$ denotes the $n \times n$ diagonal matrix whose i-th diagonal entry is $\phi'((L_i r)^2)$, L is the $n \times (n+1)$ matrix whose i-th row is L_i and

$$M(r) = h_t L^T \mathrm{diag}(\phi'(r))L. \tag{A4}$$

Now we go back to the gradient of $\psi(r) = \frac{1}{2}\|Kr - d_\delta\|_{l_2}^2$ and $\Gamma(r) = \frac{1}{2}\|D^{\frac{1}{2}}r\|_{l_2}^2$. Note that by the Euclidean inner product, we have

$$\|Kr - d_\delta\|_{l_2}^2 = (Kr - d_\delta, Kr - d_\delta), \quad \left\|D^{\frac{1}{2}}r\right\|_{l_2}^2 = (Dr, r).$$

So we have

$$\nabla \psi(r) = K^T(Kr - d_\delta), \tag{A5}$$
$$\nabla \Gamma(r) = Dr. \tag{A6}$$

Therefore, the gradient of $J(r)$ becomes

$$\begin{aligned}
\nabla J(r) &= \nabla \psi(r) + \alpha \nabla M(r) + \beta \nabla \Gamma(r) \\
&= K^T(Kr - d_\delta) + \alpha M(r)r + \beta Dr.
\end{aligned} \tag{A7}$$

The Hessian of $J(r)$ would be complicated in computation, it consists of the Hessians of $\psi(r)$, $\Gamma(r)$ and $M_\zeta(r)$. One may readily see that

$$H_{\Psi,r} = \mathrm{Hess}[\psi(r)] = K^T K, \tag{A8}$$
$$H_{\Gamma,r} = \mathrm{Hess}[\Gamma(r)] = D. \tag{A9}$$

To obtain the Hessian of $M_\zeta(r)$, we need to calculate

$$\begin{aligned}
&\frac{\partial^2}{\partial \tau \partial \xi} M_\zeta(r + \tau u + \xi v)|_{\tau,\xi=0} \\
&= \frac{1}{2} \frac{\partial^2}{\partial \tau \partial \xi} \sum_{i=1}^{n} \phi((L_i(r + \tau u + \xi v))^2) h_t|_{\tau,\xi=0}.
\end{aligned} \tag{A10}$$

Straightforward calculation yields that

$$\mathrm{Hess}[M_\zeta(r)] = M'(r)r + M(r), \tag{A11}$$

where $M(r)$ is given in (A4) and $M'(r)$ is an operator defined by

$$M'(r)r = h_t L^T \mathrm{diag}(2(Lr)^2 \phi''(r))L,$$

where $\mathrm{diag}(2(Lr)^2\phi''(r))$ denotes the $n \times n$ diagonal matrix whose i-th diagonal entry is $2(L_i r)^2\phi''((L_i r)^2)$.

Note that $\phi''(x) < 0$ for $x \geq 0$, hence nonnegativity may be lost in $\mathrm{Hess}[M_\zeta(r)]$. Therefore, we drop the term $M'(r)r$ in $\mathrm{Hess}[M_\zeta(r)]$ and obtain an approximate Hessian of $M_\zeta(r)$

$$H_{M,r} = M(r),$$

hence an approximate Hessian of $J(r)$ would be

$$H_{J,r} = K^T K + \alpha M(r) + \beta D.$$

Appendix B: l_1-norm Minimization

The l_1-norm minimization problem (19) can be written as a constrained optimization problem

$$\min_r c_0^T r,$$
$$\text{subject to } f_i(r) \leq 0, \ \ Kr = d, \tag{B1}$$

where c_0 is a vector with all components equaling 1, f_i is a linear function in the form of $f_i(r) = c_i^T r + e_i$ for some $c_i \in \mathbb{R}^n$, $e_i \in \mathbb{R}$. Details about linear programming problems and their solution methods can be found in (Ye 1997).

Using logarithmic barrier, the problem (B1) can be approximated by

$$\min_r c_0^T r - (1/\gamma) \sum_{i=1}^{n} \log(-f_i(r)),$$
$$\text{subject to } Kr = d, \tag{B2}$$

where $\gamma > 0$ is a constant. Define $\psi(r) = -\sum_{i=1}^{n} \log(-f_i(r))$ and suppose $r^*(\gamma)$ for $\gamma > 0$ is the optimal solution of the problem

$$\min_r \gamma c_0^T r + \psi(r),$$
$$\text{subject to } Kr = d. \tag{B3}$$

We can define the central path as $\{r(\gamma) : \gamma > 0\}$. Therefore $r = r^*(\gamma)$ if there exists a vector u such that

$$\gamma c_0 + \nabla\psi(r) + K^T u = 0, \quad Kr = d. \tag{B4}$$

This indicates that $r^*(\gamma)$ minimizes the Lagrangian function

$$L(x, \lambda^*(\gamma), \nu^*(\gamma)) := c_0^T r + \sum_{i=1}^{n} \lambda_i^*(\gamma) f_i(r) + \nu_i^*(\gamma)^T (Kr - d), \qquad \text{(B5)}$$

where $\lambda_i^*(\gamma) = -1/(\gamma f_i(r^*(\gamma)))$ and $\nu^*(\gamma) = u/\gamma$. Therefore, the solution of the problem (B1) can be obtained by solving the modified KKT (Karush-Kuhn-Tucker) equations

$$
\begin{aligned}
& c_0 + \sum_{i=1}^{n} \lambda_i \nabla f_i(r) + K^T \nu = 0, \\
& -\lambda_i f_i(r) = 1/\gamma, \ i = 1, 2, \ldots, n, \\
& Kr = d, \\
& \lambda \geq 0, \\
& f_i(r) \leq 0.
\end{aligned}
\qquad \text{(B6)}
$$

The primal-dual interior point method refers to choosing a triplet (r, λ, ν) to satisfy the modified KKT conditions. This can be done via Newton iterative methods. To do so, we defines the residuals of the primal problem, dual problem and central path as

$$
\begin{aligned}
& \mathrm{res}_{\mathrm{primal}}(r, \lambda, \nu) = Kr - d, \\
& \mathrm{res}_{\mathrm{dual}}(r, \lambda, \nu) = c_0 + \sum_{i=1}^{n} \lambda_i \nabla f_i(r) + K^T \nu, \\
& \mathrm{res}_{\mathrm{central}}(r, \lambda, \nu) = -\Lambda f - 1/\gamma c_0, \ i = 1, 2, \ldots, n,
\end{aligned}
\qquad \text{(B7)}
$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\mathrm{diag}(\cdot)$ denotes a diagonal operator, and $f = (f_1(r), f_2(r), \ldots, f_n(r))^T$. The solution of the triplet (r, λ, ν) can be obtained by solving a nonlinear equation

$$
F(r, \lambda, \nu) = \begin{bmatrix} \mathrm{res}_{\mathrm{primal}}(r, \lambda, \nu) \\ \mathrm{res}_{\mathrm{dual}}(r, \lambda, \nu) \\ \mathrm{res}_{\mathrm{central}}(r, \lambda, \nu) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\qquad \text{(B8)}
$$

via Newton's iterative method

$$(r^{k+1}, \lambda^{k+1}, \nu^{k+1}) = (r^k, \lambda^k, \nu^k) - \nabla F(r^k, \lambda^k, \nu^k)^{-1} F(r^k, \lambda^k, \nu^k). \qquad \text{(B9)}$$

Details about the implementation of the method and a practical algorithm can be found in (Ye 1997; Wang et al. 2007).

References

Aki, K., Richards, P.G.: Quantitative seismology: theory and methods. W. H. Freeman and Company, San Francisco (1980)

Barzilai, J., Borwein, J.: Two-point step size gradient methods:. IMA J. Numer. Anal. **8**, 141–148 (1988)

Bernabini, M., Cardarelli, E.: Variable damping factors in travel time tomography. J. Appl. Geophys. **38**, 131–141 (1997)

Buland, A., Omre, H.: Bayesian linearized AVO inversion. Geophysics **68**, 185–198 (2003)

Buland, A., Kolbjørnsen, O., Omre, H.: Rapid spatially coupled AVO inversion in the Fourier domain. Geophysics **68**, 824–836 (2003)

Chan, T.F., Kang, S.H.: Total variation denoising and enhancement of color images based on the CB and HSV color models. J. Visual Comm. Image Rep. **12**, 422–435 (2001)

Claerbout, J.F., Muir, F.: Robust modeling with erratic data. Geophysics **38**, 820–844 (1973)

Contreras, A., Torres-Verdín, C., Fasnacht, T.: AVA simultaneous inversion of partially stacked seismic amplitude data for the spatial delineation of lithology and fluid units of deepwater hydrocarbon reservoirs in the central Gulf of Mexico. Geophysics **71**, E41–E48 (2006)

Cuer, M., Bayer, R.: Fortran routines for linear inverse problems. Geophysics **45**, 1706–1719 (1980)

Dimri, V.: Deconvolution and inverse theory—applications to geophysical problems. Elsevier, Amsterdam (1992)

Duijndam, A.J.W.: Bayesian estimation in seismic inversion-part I: Principles. Geophys. Prospect. **36**, 878–898 (1988a)

Duijndam, A.J.W.: Bayesian estimation in seismic inversion-part II: Uncertainty analysis. Geophys. Prospect. **36**, 878–898 (1988b)

Goldberg, D.E.: Genetic algorithms in search. Optimization & machine learning. Addison–Wesley, Boston (1989)

Golub, G.H., Van Loan, C.F.: Matrix computations. The Johns Hopkins University Press, Baltimore (1996)

Gouveia, W.P., Scales, J.A.: Resolution of seismic waveform inversion: Bayes versus Occam. Inverse Probl. **13**, 323–349 (1997)

Gouveia, W.P., Scales, J.A.: Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis. J. Geophys. Res. **103**, 2759–2779 (1998)

Huang, X.D.: Deconvolution and Seismic Traces Inversion. Petroleum Industry Press, Beijing (1992)

Huestis, S.P.: Uniform norm minimization in three dimensions. Geophys. J. Roy. Astr. Soc. **58**, 249–260 (1986)

Ito, K., Kunisch, K.: An active set strategy based on the augmented Lagrangian formulation for Image restoration. Math. Model. Numer. Anal. **33**, 1–21 (1999)

Ito, K., Kunisch, K.: Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. Nonlinear Anal. **41**, 591–616 (2000)

Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Qart. Appl. Math. **2**, 164–166 (1944)

Levy, S., Fullagar, P.K.: Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. Geophysics **46**, 1235–1243 (1981)

Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear inequalities. SIAM J. Appl. Math. **11**, 431–441 (1963)

Meyer, Y.: Oscillating patterns in image processing and nonlinear evolution equations, University Lecture Series, vol. 22, American Mathematical Society, Providence (2001)

Mosegaard, K., Tarantola, A.: Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. **100**, 12431–12447 (1995)

Mosegaard, K.: Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse Probl. **14**, 405–426 (1998)

Nashed, M.Z.: Perturbations and approximations for generalized inverses and linear operators equations. In: Nashed, M.Z. (ed.) Generalized inverses and applications, pp. 325–396. Academic Press, New York (1976)

Nashed, M.Z., Scherzer, O.: Least squares and bounded variation regularization with nondifferentiable functionals. Numer. Funct. Anal. Optim. **19**, 873–901 (1998)

Nolet, G.: Seismic tomography. Reidel Publishing, Boston (1987)

Nolet, G., Snieder, R.: Solving large linear inverse problems by projection. Geophys. J. Int. **103**, 565–568 (1990)

Oldenburg, D.W., Scheuer, T., Levy, S.: Recovery of the acoustic impedance from reflection seismo-grams. Geophysics **48**, 1318–1337 (1983)

Pilkington, M., Todoeschuck, J.P.: Natural smoothness constraints in cross-hole seismic tomography. Geo-phys. Prospect. **40**, 227–242 (1992)

Robinson, E.A.: A predictive decomposition of time series with application to seismic exploration. Geo-physics **32**, 418–484 (1967)

Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenom. **60**, 259–268 (1992)

Safon, L., Vasseur, G., Cuer, M.: Some applications of linear programming to the inverse gravity prob-lem. Geophysics **42**, 1215–1229 (1977)

Sen, M.K., Stoffa, P.L.: Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. Geophys. Prospect. **44**, 313–350 (1996)

Tarantola, A.: Inverse problems theory: methods for data fitting and model parameter estimation. Elsevier, Amsterdam (1987)

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

Trampert, J., Leveque, J.J.: Simultaneous iterative reconstruction technique: physical interpretation based on the generalized least squares solution. J. Geophys. Res. **95**, 12553–12559 (1990)

Ulrych, T.J., Sacchi, M.D., Graul, M.: Signal and noise separation: art and science. Geophysics **64**, 1648–1656 (1999)

Ulrych, T.J., Sacchi, M.D., Woodbury, A.: A Bayesian tour to inversion. Geophysics **66**, 55–69 (2000)

VanDecar, J.C., Snieder, R.: Obtaining smooth solutions to large linear inverse problems. Geophys-ics **59**, 818–829 (1994)

Vogel, C.R., Oman, M.E.: A fast, robust algorithm for total variation based reconstruction of noisy, blurred images. IEEE Trans. Image Process. **7**, 813–824 (1998)

Vogel, C.R.: Computational methods for inverse problems, SIAM, Philadelphia (2002)

Wang, J.Y.: Inversion theory for geophysics. Higher Education Press, Beijing (2002)

Wang, Y.W., Wang, Y.F., Xue, Y., Gao, W.: A new algorithm for remotely sensed image texture classification and segmentation. Int. J. Remote Sens. **25**, 4043–4050 (2004)

Wang, Y.F., Fan, S.F., Feng, X., Yan, G.J., Guan, Y.N.: Regularized inversion method for retrieval of aerosol particle size distribution function in $W^{1,2}$ space. Appl. Opt. **45**, 7456–7467 (2006)

Wang, Y.F.: Computational methods for inverse problems and their applications. Higher Education Press, Beijing (2007)

Wang, Y.F., Ma, S.Q.: Projected barzilai-borwein methods for large scale nonnegative image restora-tions. Inverse Probl. Sci. Eng. **15**, 559–583 (2007)

Wang, Y.F., Fan, S.F., Feng, X.: Retrieval of the aerosol particle size distribution function by incorporating a priori information. J. Aerosol Sci. **38**, 885–901 (2007)

Wang, Y.F., Yang, C.C., Li, X.W.: A regularizing kernel-based brdf model inversion method for ill-posed land surface parameter retrieval using smoothness constraint. J. Geophys. Res. **113**, D13101 (2008)

Wang, Y.F.: An efficient gradient method for maximum entropy regularizing retrieval of atmospheric aerosol particle size distribution function. J. Aerosol Sci. **39**, 305–322 (2008)

Wang, Y.F., Yagola, A.G., Yang, C.C. (eds.): Optimization and Regularization for Computational Inverse Problems and Applications. Higher Education Press and Springer, Beijing/Berlin (2010)

Wang, Y.F.: Seismic impedance inversion using l_1-norm regularization and gradient descent methods. J. Inv. Ill-Posed Problems **18**, 823–838 (2011)

Yang, W.C.: On inversion methods for nonlinear geophysics. Prog. Geophys. **17**, 255–261 (2002)

Ye, Y.Y.: Interior point algorithms: theory and analysis. John Wiley and Sons, New York (1997)

Yilmaz, O.: Seismic Data Processing, Investigations in Geophysics No. 2, Society of Exploration Geophysi-cists, Tulsa (1987)

Yuan, Y.X.: Numerical Methods for Nonliear Programming. Shanghai Science and Technology Publica-tion, Shanghai (1993)

Yuan, Y.X.: Gradient methods for large scale convex quadratic functions. In: Wang, Y.F., Yagola, A.G., Yang, C.C. (eds.) Optimization and regularization for computational inverse problems and applica-tions, pp. 141–155. Springer, Berlin (2010)

Zhang, H.B., Shang, Z.P., Yang, C.C.: A non-linear regularized constrained impedance inversion. Geophys. Prospect. **55**, 819–833 (2007)

Zhdanov, M.S.: Geophysical inverse theory and regularization problems. Elsevier, Amsterdam (2002)

Ziemer, W.P.: Weakly differentiable functions. Springer, New York (1989)