Adv Comput Math (2012) 36:353–376
DOI 10.1007/s10444-011-9207-2

Preconditioning non-monotone gradient methods for retrieval of seismic reflection signals

Y. F. Wang

Received: 6 May 2010 / Accepted: 3 March 2011 /
Published online: 23 September 2011

Abstract The core problem in seismic exploration is to invert the subsurface reflectivity from the surface recorded seismic data. However, most of the seismic inverse problems are ill-posed by nature. To overcome the ill-posedness, different regularized least squares methods are introduced in the literature. In this paper, we developed a preconditioning non-monotone gradient method, proved it converges with R-superlinear rate and applied it to seismic deconvolution and imaging. Numerical examples demonstrate that the method is efficient. It helps to improve the resolution of the seismic inversions.

Keywords Ill-posed problem · Regularization · Non-monotone gradient · Preconditioning · Seismic inversion

Mathematics Subject Classifications (2010) 65J20 · 65J22 · 65K10

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 speculate the spatial distribution of underground strata structure and physical parameter

Communicated by Yuesheng Xu and Hongqi Yang.

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

by using seismic wave propagation law. Meanwhile, restoring the seismic reflectivity is a main problem. After linearization and by ignoring the source and receiver signatures, the forward model for representing a seismic imaging process is usually written as [1, 24]

$$Lm = d, \tag{1}$$

where L is the forward modeling kernel operator; m is the (singular) fluctuations in the earth's acoustic properties with respect to an appropriately chosen smoothly varying background velocity model (e.g., a constant velocity in respective layers). These fluctuations are referred to as the model or reflectivity and seismic imaging aims to recover both the locations and the relative amplitudes of the velocity fluctuations from seismic data d.

Equation 1 can be solved by finding a least-squares error solution with smallest norm, i.e., solving a problem

$$\min_m J[m] := \frac{1}{2}\|Lm - d\|^2, \tag{2}$$

such that the norm $\|m\|$ takes the minimum. This is equivalent to solving a normal equation

$$L^*Lm = L^*d, \tag{3}$$

where L^* is the adjoint operator of L defined by $(x, Ly) = (L^*x, y)$ for any functions x and y.

A naive approach for finding the reflectivity model m is by

$$m = (L^*L)^{-1}L^*d. \tag{4}$$

It is evident that if L^* is an approximate inverse to the forward operator L, then the reflectivity model can be obtained by $m = L^*d$. This process in seismic exploration is called the migration [37]. However, for seismic imaging problems, due to limited bandwidth and limited acquisition spaces, the seismic images obtained are blurred, and then direct inversion may cause distortion on the low-frequency component and the high-frequency component as well. The long-standing seismic migration methods give distorted images of the subsurface, even with an accurate velocity model because of limitations in bandwidth, recording time, and aperture of the seismic-reflection experiment. Furthermore, in the Hadamard's sense, it is known that deconvolution and inversion are ill-posed problems, and hence the least-squares results inherit intrinsic instability and multiplicity [25].

In addition, we must keep in mind another drawback of the least-squares method: huge computational cost; for a large scale problem, unless powerful methods are found, it is difficult to solve the inverse problem. Therefore we need much effort to improve the solvability of the problem.

A quick survey from literature shows that there exists a large amount of methods to solve different geophysical inverse problems, e.g., the damped least-squares solution method [12, 15], the trust region methods [35], Lie group method [13], Bayesian statistical inference in seismic data processing

[24, 28, 29], the singular value decomposition [1, 11], the iterative algorithms for implementation of least-squares problems [19, 20, 26], the nonstationary seismic deconvolution via the Gabor transform for source waveform [14], the mollification methods, Backus–Gilbert method [2, 3, 17], and also the long standing migration problem with its improvements [18, 23, 30, 41]. However, it is observed that the convergence of the iterative methods may be very slow [31] and a suitable optimization algorithm is needed and must be applied many times [27]. Though the efficiency of a particular iterative method exists, it may not be feasible for large scale seismic inversion problems. As argued in [22], efficient numerical optimization methods for large scale problems are urgently needed. Furthermore, as mentioned in AGU General Assembly that high fidelity algorithm and high-end computing will be the task for next generation seismic imaging [5]. This motivates us to devote ourselves to studying powerful methods for solving large scale numerical seismic inversion problems.

In geophysical applications, e.g., migration imaging, simple and fast convergent methods are always desirable. It is well known that the gradient-based method, e.g., the steepest descent method, is simple in iteration and low memory because of the second-order derivative information of the objective function of the regularized functional is not required. To improve the performance of this simple method and applied it to practical problems, much more attention has been paid in recent ten years [4, 6, 7, 33, 34, 43]. This motivates us to consider applying the gradient-based method to seismic signal recovery problem. In particular, we develop a non-monotone gradient method with preconditioning technique which is demonstrated to have the advantages of fast convergence and easy to perform for large scale ill-posed inversion in seismic imaging.

The outline of the paper is as follows. In Section 2, we briefly mention the regularization model. Then in Section 3, we apply the non-monotone gradient descent method to solve the regularizing minimization model. Convergence analysis of the non-monotone gradient descent method in Rayleigh type is performed. Preconditioning technique is addressed in detail in Section 4. In Section 5, we give several numerical simulations. Examples on multichannel image deblurring, layered velocity model and migration deconvolution are illustrated. Potential usage of our methods and issues for field applications are discussed in Section 6. Finally in Section 7, a conclusion is made.

2 Regularization

A powerful method for solving the ill-posed problem is the Tikhonov regularization [25], which refers to solve a minimization of a "regularized functional"

$$J^{\alpha}[m] := \frac{1}{2}\|Lm - d\|^2 + \alpha\Omega[m], \tag{5}$$

where $\Omega[\cdot]$ is the so-called Tikhonov stabilizer, that is usually preassigned by users, and $\alpha \in (0, 1)$ is the regularization parameter. If $\Omega[m] = \frac{1}{2}\|Dm\|^2$, where

D is a positive (semi-)definite bounded scale operator, then the minimizer of $J^\alpha[m]$ is

$$m = \left(L^*L + \alpha D^*D\right)^{-1} L^*d, \tag{6}$$

the solution of the Euler equation $(L^*L + \alpha D^*D)m = L^*d$. For solving (6), several methods such as LU decomposition, singular value decomposition and QR decomposition can be used [32, 39]. However any direct methods should be avoided. Instead, iterative solvers are preferred for geophysical inversion problems as the scale of the problems is usually large. And the fast non-monotone gradient method [33, 34] are iterative methods that can be employed to find the minimizer of the regularized minimization problem (5).

3 Iterative regularization: non-monotone gradient iteration

Iterative regularization methods are more adaptable for large scale problems in applied sciences. The well-know iterative regularization methods are Fridman–Landweber iterative methods [10, 38]. And many optimization methods can be regularized for solving practical inverse problems, e.g., regularizing Gauss–Newton method, trust region method, steepest descent method and conjugate gradient method [32, 35]. We recall that the gradient-based method is simple in iteration and low memory because of the second-order information of the objective function of the regularized functional is not required. Therefore, we pay attention to the gradient method and study the improvement of its performance for geophysical exploration problems in this paper.

In the following, we assume that the problem is formulated in finite spaces and accordingly the adjoint of the operator is transferred to transpose of the operator. There are several highly cited gradient methods in the literature. Perhaps the simplest and the easiest one is the steepest descent method

$$m_{k+1} = m_k + \omega_k s_k, \tag{7}$$

where $s_k = -g_k$, $g_k = g(m_k)$, $g(m)$ is the gradient of the function $J^\alpha[m]$ and ω_k the steplength which can be obtained by line search, i.e., optimal ω_k^* satisfies

$$\omega_k^* = \operatorname{argmin}_\omega J^\alpha\left(m_k + \omega s_k\right),$$

where the notation argmin denotes minimizing a function with specific argument. However, the steepest descent method is slow in convergence and zigzagging after several iterations [42]. The poor behavior is due to the optimal choice of the step size and not to the choice of the steepest descent direction g_k. If we restrict the steplength ω_k to a constant value in the interval $(0, \|L\|^{-2})$ at each iteration, we obtain a special gradient method which is known as the aforementioned Fridman–Landweber iteration

$$m_{k+1} = m_k + \omega s_k. \tag{8}$$

However this method is quite slow in convergence and is difficult to be used for most of practical problems [10, 32, 38]. The poor behavior lies in that both the step size ω and the gradient descent direction g_k are not optimizing.

Instead of using the negative gradient in each iteration, the non-monotone gradient methods were developed recently [6, 7, 43], meanwhile the Barzilai–Borwein method was one of the most well-known. This method was first proposed for solving the unconstrained optimization problem [4]. In this paper, we propose to apply the nonmonotone gradient method to ill-posed seismic signal retrieval problems by solving the following minimization problem

$$\min \ J^{\alpha}[m]. \tag{9}$$

The non-monotone gradient method is aimed to accelerate the convergence of the steepest descent method and requires few storage locations and inexpensive computations. Barzilai and Borwein incorporated the quasi-Newton property with the gradient method for obtaining the second order information of the objective function $J^{\alpha}[m]$. Specifically, they approximated the Hessian $\nabla^2 J^{\alpha}[m_k]$ by $\nu_k I$ and based on the secant condition, they considered two minimization problems $\nu_k = \text{argmin}_{\nu} \| y_{k-1} - \nu I z_{k-1} \|$ and $\nu_k = \text{argmin}_{\nu} \| \nu I y_{k-1} - z_{k-1} \|$, where $y_{k-1} = g_k - g_{k-1}$ and $z_{k-1} = m_k - m_{k-1}$. This leads to the two choices of the stepsize ν_k

$$\nu_k^1 = \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, A g_{k-1})} \tag{10}$$

and

$$\nu_k^2 = \frac{(g_{k-1}, A g_{k-1})}{(g_{k-1}, A^T A g_{k-1})}, \tag{11}$$

where $A = L^T L + \alpha D^T D$. It is evident that the two choices of stepsizes inherit the information of the gradient information from the former iteration instead of the current iteration. Hence the stepsizes inherit the regularized spectrum of the regularization model. Therefore we believe that this method is very efficient for solving ill-posed convex quadratic programming problem [33].

Let $\{m_k\}$ be the sequence generated by the above method from initial vectors m_0 and m_1. Then the gradient of the object function $J^{\alpha}[m]$ at m_k is $g_k = A m_k - b$, where A is mentioned above and $b = L^T d$. We have for all $k \geq 1$,

$$g_{k+1} = \nu_k \left(\frac{1}{\nu_k} I - A \right) g_k. \tag{12}$$

To analyze the convergence of the Barzilai–Borwein method, we can assume without loss of generality that an orthogonal transformation is made that transforms A to a diagonal matrix of eigenvalues $\text{diag}(\lambda_i)$. Moreover, if there are any eigenvalues of multiplicity $M > 1$, then we can choose the corresponding eigenvectors so that $g_k^{(i)} = 0$ for at least $M - 1$ corresponding indices of g_k.

It follows from (12) and using spectrum representation $A = \mathrm{diag}(\lambda_i)$ that

$$g_{k+1}^{(i)} = \nu_k \left(\frac{1}{\nu_k} - \lambda_i \right) g_k^{(i)}. \tag{13}$$

Using the recurrence, Barzilai and Borwein prove an R-superlinear convergence result for the particular choice of the stepsize ν_k.

Note that from (10) and (11) the inverse of the scalar ν_k is the Rayleigh quotient of A or $A^T A$ at the vector g_{k-1}. More choices for the step-lengths can be obtained by setting a hybrid of the two stepsizes, e.g., the mean values of any two Rayleigh ratios, and the convergence properties were obtained [44]. We consider a linear combination of the two stepsizes, i.e.,

$$\nu_k^{\text{Rayleigh}} = \beta_1 \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, Ag_{k-1})} + \beta_2 \frac{(g_{k-1}, Ag_{k-1})}{(g_{k-1}, A^T Ag_{k-1})}, \tag{14}$$

where β_1 and β_2 are two positive parameters. An investigation of (10) and (11) reveals that (10) is better than (11), since ν_k^1 has small jumps than ν_k^2 has when A is ill-conditioning. However, there is no reason that one should discard the stepsize given by ν_k^2. To make a trade-off, we choose the parameter β_2 geometrically, i.e.,

$$\beta_2 = \beta_0 \xi^{k-1}, \ \ \beta_0 > 0, \ \ \xi \in (0, 1), \ \ k = 1, 2, \cdots. \tag{15}$$

Another parameter β_1 can be set to $1 - \beta_2$. This choice of parameters assigns more weights to ν_k^1 than ν_k^2, however both step information inherits into the next iteration.

It can be shown that the iteration points generated by the above method converge to the minimal solution of $J^\alpha[m]$. Similar to [44], we have the following theorem.

Theorem 1 *Let $J^\alpha[m]$ be given in (5) with $\Omega[m] = \frac{1}{2}\|Dm\|^2$ and D a positive definite bounded scale operator and let $\{m_k\}$ be generated by the above non-monotone gradient method in Rayleigh type with stepsize ν_k satisfying (14). Then the sequence $\{m_k\}$ converges to the minimal solution of $J^\alpha[m]$.*

Proof It is evident that $A = L^T L + \alpha D^T D$ is positive definite for $\alpha > 0$. Therefore, the orthogonal decomposition of A can be written as

$$A = Q \Lambda Q^T,$$

where $Q = [q_1, q_2, \cdots, q_n]$ is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ satisfying

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n.$$

For any initial vector m_0, the gradient $g_0 = Am_0 - b$ can be written as

$$g_0 = \sum_{i=1}^{n} \rho_i^{(0)} q_i,$$

where $\rho_i^{(0)}$ are constants, $i = 1, 2, \cdots, n$. Using (13) we obtain for any $k \geq 0$,

$$g_k = \sum_{i=1}^{n} \rho_i^{(k)} q_i,$$

where $\rho_i^{(k)}$ $(i = 1, 2, \cdots, n)$ satisfy

$$\rho_i^{(k+1)} = (1 - \nu_k \lambda_i)\, \rho_i^{(k)} = \prod_{j=0}^{k} \left(1 - \nu_j \lambda_i\right) \rho_i^{(0)}. \tag{16}$$

The above equation indicates that

$$\left|\rho_1^{(k)}\right| = \left|(1 - \nu_{k-1}\lambda_1)\, \rho_1^{(k-1)}\right| \leq \left|\left(1 - \frac{\lambda_1}{\lambda_n}\right) \rho_1^{(k-1)}\right| \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right)^k \left|\rho_1^{(0)}\right|.$$

The above inequality reveals that $\lim_{k\to\infty} \rho_1^{(k)} = 0$. Now it suffices to proves the following equality

$$\lim_{k\to\infty} \rho_i^{(k)} = 0, \quad i = 1, 2, \cdots, n. \tag{17}$$

On the contrary, there exists a positive number δ and an integer $l \in [1, n-1]$ such that (17) holds for $i = 1, 2, \cdots, l$ and

$$\limsup_{k\to\infty} \rho_{l+1}^{(k)} > \delta > 0. \tag{18}$$

For any given positive number $\hat{\delta}$ we have that

$$\begin{aligned}
\lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, Ag_{k-1})} &= \lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{\sum_{i=1}^{n} \left(\rho_i^{(k-1)}\right)^2}{\sum_{i=1}^{n} \left(\rho_i^{(k-1)}\right)^2 \lambda_i} \\
&\leq \lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{\sum_{i=1}^{l+1} \left(\rho_i^{(k-1)}\right)^2}{\sum_{i=1}^{l+1} \left(\rho_i^{(k-1)}\right)^2 \lambda_i} \\
&= \frac{1}{\lambda_{l+1}}
\end{aligned}$$

and

$$\begin{aligned}
\lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{(g_{k-1}, Ag_{k-1})}{\left(g_{k-1}, A^T Ag_{k-1}\right)} &= \lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{\sum_{i=1}^{n} \left(\rho_i^{(k-1)}\right)^2 \lambda_i}{\sum_{i=1}^{n} \left(\rho_i^{(k-1)}\right)^2 \lambda_i^2} \\
&\leq \lim_{|\rho_{l+1}^{(k)}|\geq\hat{\delta}, k\to\infty} \frac{\sum_{i=1}^{l+1} \left(\rho_i^{(k-1)}\right)^2 \lambda_i}{\sum_{i=1}^{l+1} \left(\rho_i^{(k-1)}\right)^2 \lambda_i^2} \\
&= \frac{1}{\lambda_{l+1}}.
\end{aligned}$$

Let us recall (17) holds for $i = 1, 2, \cdots, l$, $|\rho_{l+1}^{(k)}| \geq \hat{\delta} \geq \frac{\lambda_1}{\lambda_{l+1}}\hat{\delta}$, hence there exists a sufficiently large integer $\hat{k}$ such that for all $k \geq \hat{k}$

$$\nu_k^{\text{Rayleigh}} \leq \frac{11}{10}\frac{1}{\lambda_{l+1}}$$

and

$$\left|\rho_{l+1}^{(k)}\right| \geq \frac{\lambda_1}{\lambda_{l+1}}\frac{\delta}{2}. \tag{19}$$

Therefore, if (19) holds for any $k \geq \hat{k}$, we have that

$$\begin{aligned}\left|\rho_{l+1}^{(k+1)}\right| &= \left|1 - \nu_k^{\text{Rayleigh}}\lambda_{l+1}\right|\left|\rho_{l+1}^{(k)}\right| \\ &\leq \max\left\{1 - \frac{\lambda_{l+1}}{\lambda_n}, \left|\lambda_{l+1}\frac{11}{10}\frac{1}{\lambda_{l+1}} - 1\right|\right\}\left|\rho_{l+1}^{(k)}\right| \\ &\leq \max\left\{1 - \frac{\lambda_{l+1}}{\lambda_n}, 0.1\right\}\left|\rho_{l+1}^{(k)}\right|. \end{aligned} \tag{20}$$

Conversely, if (19) does not hold, we have from (16) that

$$\left|\rho_{l+1}^{(k+1)}\right| \leq \max\left\{1 - \frac{\lambda_{l+1}}{\lambda_n}, \frac{\lambda_{l+1}}{\lambda_1} - 1\right\}\left|\rho_{l+1}^{(k)}\right| \leq \frac{\lambda_{l+1}}{\lambda_1}\left|\rho_{l+1}^{(k)}\right| \leq \frac{\delta}{2}. \tag{21}$$

Equations 20 and 21 reveals that

$$\limsup_{k\to\infty}\left|\rho_{l+1}^{(k)}\right| \leq \frac{\delta}{2}.$$

which contradicts the assertion (18). This shows the theorem to be true. □

Barzilai and Borwein proved an *R*-superlinear convergence result for the particular choice of the stepsize ν_k. Since their method is a special case of choosing the stepsize ν_k with formula (14), therefore, it is ready to see that the nonmonotone gradient method in Rayleigh type is *R*-superlinear convergent. We have the following results.

Theorem 2 *Let $J^\alpha[m]$ be given in (5) with $\Omega[m] = \frac{1}{2}\|Dm\|^2$ and D a positive definite bounded scale operator and let $\{m_k\}$ be generated by the above nonmonotone gradient method in Rayleigh type with stepsize ν_k satisfying (14). Then the sequence $\{m_k\}$ converges to the minimal solution of $J^\alpha[m]$ with R-superlinear convergence.*

Proof When $\beta_1 = 0$ or $\beta_2 = 0$, the Rayleigh type method reduces to the Barzilai and Borwein's method. Since the Barzilai and Borwein's algorithm is *R*-superlinear convergent, hence the algorithm of Rayleigh type is *R*-superlinear convergent. □

Remark 1 One may argue that why the non-monotone gradient method works well for ill-posed inverse problems. It is easy to see that $\nu_k^1 \geq \nu_k^2$. Therefore, it would be favorable to choose a longer step ν_k^1 instead of ν_k^2. However, the shorter step ν_k^2 yields a smaller $\|g_{k+1}\|$, which indicates that the shorter step ν_k^2 would be efficient for obtaining an accurate solution of a large scale and ill-conditioned problem. It is clear that the choice of ν_k^{Rayleigh} possesses both the suitable length of the step and the reasonable approximation of an accurate solution. So, it is no doubt that proposed method works efficiently for seismic inversion and imaging.

Remark 2 In iterative regularization methods, the stopping rule is also an important issue. Since the noise for practical seismic data is hard to be identified, it is infeasible to use the commonly adopted discrepancy principle as the stopping rule [38]. Therefore, for gradient descent methods, the stoping rule is based on the values of the norm of the gradient g_k. We preassign a tolerance $\epsilon > 0$. Once $\|g_k\| \leq \epsilon$ is reached, the iterative process will be stopped. Smaller ϵ yields better approximate solution, however, induces more CPU time. Empirical values of ϵ is in the interval $(10^{-3}, 10^{-5})$.

4 Preconditioning

4.1 Preconditioning non-monotone gradient method

The convergence property of iterative gradient methods depends on the conditioning of the operators. Preconditioning is a technique to improve convergence by lowering the condition number or increasing the eigenvalue clustering. This technique applied to gradient descent methods has been considered sufficiently in literature, e.g., [8, 9, 16, 21]. The idea is to solve a modified problem

$$P^{-1}L^T Lm = P^{-1}L^T d, \tag{22}$$

where P is a symmetric positive definite matrix. If the condition number of $P^{-1}L^T L$ is less than that of $L^T L$ or the eigenvalues of $P^{-1}L^T L$ are more clustered than that of $L^T L$, a higher rate of convergence will be reached.

Let C be a nonsingular matrix and define the factorization of P as $P = CC^T$, then solving (22) is equivalent to solving

$$C^{-1}L^T LC^{-T}z = C^{-1}L^T d, \tag{23}$$

$$z = C^T m. \tag{24}$$

Note that minimization of the objective function in (5) is equivalent to minimizing a quadratic programming problem

$$Q^{\alpha}[m] = \frac{1}{2}m^T\big(L^T L + \alpha D^T D\big)m - d^T Lm, \tag{25}$$

therefore the preconditioning problem can be written as minimizing a new quadratic programming problem

$$\tilde{Q}^{\alpha}[z] = \frac{1}{2} z^T A z - b^T z, \tag{26}$$

where $\tilde{A} = C^{-1}(L^T L + \alpha D^T D)C^{-T}$, $\tilde{b} = C^{-1} L^T d$ and $z = C^T m$.

The gradient of $\tilde{Q}^{\alpha}$ can be evaluated as

$$\tilde{g}(z) = \tilde{A} z - \tilde{b} = C^{-1}\left(L^T L + \alpha D^T D\right) C^{-T} z - C^{-1} L^T d. \tag{27}$$

The iterations of the steepest descent method are described by

$$\tilde{g}_k = \tilde{A} z_k - \tilde{b}, \tag{28}$$

$$z_{k+1} = z_k - \nu_k \tilde{g}_k, \tag{29}$$

$$\nu_k = \frac{\tilde{g}_k^T \tilde{g}_k}{\tilde{g}_k^T \tilde{A} \tilde{g}_k}, \tag{30}$$

Straightforward calculation yields the equivalent iterative formula

$$m_{k+1} = m_k - \tilde{\nu}_k h_k, \tag{31}$$

$$\tilde{\nu}_k = \frac{g_k^T h_k}{h_k^T (L^T L + \alpha D^T D) h_k}, \tag{32}$$

$$h_k = P^{-1} g_k, \tag{33}$$

$$g_{k+1} = g_k - \tilde{\nu}_k \left(L^T L + \alpha D^T D\right) h_k. \tag{34}$$

Similarly, the iterations of the non-monotone gradient method are described by

$$m_{k+1} = m_k - \tilde{\nu}_k h_k, \tag{35}$$

$$\tilde{\nu}_k = \begin{cases} \tilde{\nu}_k^1, & \text{or} \\ \tilde{\nu}_k^2, & \text{or} \\ \tilde{\nu}_k^{\text{Rayleigh}} \end{cases}, \tag{36}$$

$$h_k = P^{-1} g_k, \tag{37}$$

$$g_{k+1} = g_k - \tilde{\nu}_k \left(L^T L + \alpha D^T D\right) h_k, \tag{38}$$

where

$$\tilde{\nu}_k^1 = \frac{g_{k-1}^T h_{k-1}}{h_{k-1}^T \left(L^T L + \alpha D^T D\right) h_{k-1}},$$

$$\tilde{\nu}_k^2 = \frac{h_{k-1}^T \left(L^T L + \alpha D^T D\right) h_{k-1}}{h_{k-1}^T \left(L^T L + \alpha D^T D\right) P^{-1} \left(L^T L + \alpha D^T D\right) h_{k-1}}$$

and $\tilde{\nu}_k^{\text{Rayleigh}}$ is the linear combinations of $\tilde{\nu}_k^1$ and $\tilde{\nu}_k^2$.

For any positive definite bounded scale operator D, J^α is strictly convex. Since P is nonsingular matrix, therefore similarly to Theorem 1, we can establish the following convergence results of the preconditioning non-monotone gradient method.

Theorem 3 *Let $J^\alpha[m]$ be given in (5) with $\Omega[m] = \frac{1}{2}\|Dm\|^2$ and D a positive definite bounded scale operator and let $\{m_k\}$ be generated by the preconditioning nonmonotone gradient method. Then the sequence $\{m_k\}$ converges to the minimal solution of $J^\alpha[m]$.*

4.2 Preconditioners

Many preconditioners have been developed since the resuscitation of the conjugate gradient method. Widely referred preconditioners are Jacobian, Gauss–Seidel's and incomplete factorization LU proconditioners [32]. The Jacobian preconditioner uses the diagonal of $L^TL + \alpha D^TD$ and has been shown to be useful if the diagonal elements are relatively different. The Gauss–Seidel's preconditioner originates from Gauss–Seidel's iterative method for solving linear matrix-vector equations. Incomplete factorization preconditioning uses an approximation to $L^TL + \alpha D^TD$ which is easy to invert. These preconditioners are efficient for well-posed linear problems. However for ill-posed problems, their advantages are not so obvious.

We apply a symmetric successive over relaxation preconditioner. We assume that the matrix $S = L^TL + \alpha D^TD$ can be decomposed as

$$S = M - N,$$

where

$$M = \frac{1}{\omega(2-\omega)}\left[(K - \omega C_l)K^{-1}(K - \omega C_u)\right],$$

$$N = \frac{1}{\omega(2-\omega)}\left[(1-\omega)K + \omega C_l\right]K^{-1}\left[(1-\omega)K + \omega C_u\right],$$

and K, C_l and C_u are the diagonal, lower triangular parts and upper triangular parts of S, respectively. Then we choose P as

$$P = (K - \omega C_l)K^{-1}(K - \omega C_u),$$

where ω is a real scalar parameter within $(0, 2)$. The optimal choice of the ω is not easy to do, as it requires very complicated eigenvalue analysis. And in many cases, such eigenvalue analysis becomes unavailable because the matrix L is not explicitly given, e.g., seismological problems, and it has to be estimated iteratively. In the current paper, we only approximate it by the degree of ill-posedness of the problem.

According to Perron–Frobenius Theorem [40], the spectrum radius of S denoted by $\rho(S)$ is greater than 0. It is ready to see that $\rho(M^{-1}N) = \frac{\rho(S^{-1}N)}{1+\rho(S^{-1}N)} < 1$.

Since $M^{-1}S = I - M^{-1}N$ and $0 < \rho(M^{-1}N) < 1$, hence $0 < \rho(M^{-1}S) < 1$. So, the above choice of the preconditioner P is sufficient to guarantee the acceleration of convergence of the iterative methods.

5 Numerical results

5.1 Seismic signal deconvolution

We show our method works for seismic deconvolution problem. In simulation, the operator L is formulated by the wavelet matrix with length $L+1$ expressed by the wavelet function w as

$$L_{(N+L)\times N} = \begin{bmatrix} w_0 & 0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & 0 & \cdots & 0 \\ 0 & w_1 & w_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ w_L & w_{L-1} & \cdots & w_1 & w_0 & 0 \\ 0 & w_L & w_{L-1} & \cdots & w_1 & w_0 \\ 0 & 0 & w_L & w_{L-1} & \cdots & w_1 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & w_L & w_{L-1} \\ 0 & 0 & 0 & \cdots & 0 & w_L \end{bmatrix}.$$

The reflectivity function is expressed as $m = [m_0, m_1, \cdots, m_{N-1}]^T$ and the recorded seismic data is generated through (1): $d = [d_0, d_1, \cdots, d_{N-1}]^T$.

5.1.1 Multichannel seismic image deblurring

Numerical simulation is a necessary step to prove the efficiency of our proposed method before applying it to practical application problems. It is evident that one cannot expect that a computational method would be feasible for practical problems if it fails to work on synthetic seismic data.

The simulation consists of two steps. First, a simulated seismic signal is generated by computer according to (1) for a given seismic reflectivity function $m^{\text{true}}(t)$. Then, the generated seismic signal is processed through our algorithm, and the retrieved distribution signal is compared with the input one. Note that d is the observation containing noise, hence an additive noise with level δ is added. Our example is a multichannel seismic signal deblurring example. The synthetic seismic deconvolution is very important in seismic interpretation. We consider a synthetic convolution data generated by Ricker wavelet convolved with an input signal consists of two layers' reflection, the data are added random noise with error level $\delta = 0.001$. The central frequency of the Ricker

wavelet is 20 Hz, the sampling interval is two milliseconds. The condition number of the discrete operator L is 3.2241×10^{16}. A value of ω equaling to 0.2 is used in constructing the preconditioner. The regularization parameter α is chosen as an a priori value 0.001. The stabilizer is chosen as $\frac{1}{2}\|Dm\|^2$ and D is a tridiagonal matrix in the form

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

where h_t is the time sampling interval. We refer to [36] for detailed discussion of choosing the stabilizer for geophysical inverse problems. Our preconditioning non-monotone gradient algorithm converges at five iterations, while the algorithm without preconditioning requires 18 iterations to yield convergence. For both algorithms, the root mean square error is 0.0400. The root mean square error specifies the average deviation of the recovered signal from the true signal, which is defined as $\sqrt{\sum_{i=1}^{N} \frac{(d_{\text{true}}(:)-d_{\text{estimate}}(:))^2}{N}}$, where N is the length of

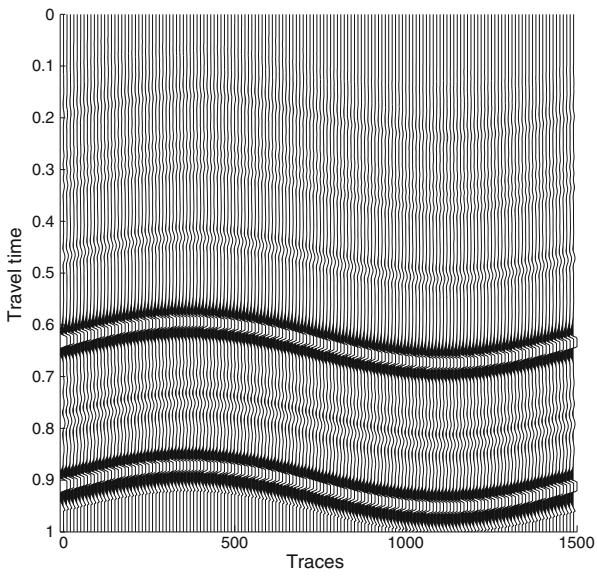

Fig. 1 A simulated multichannel seismic signal

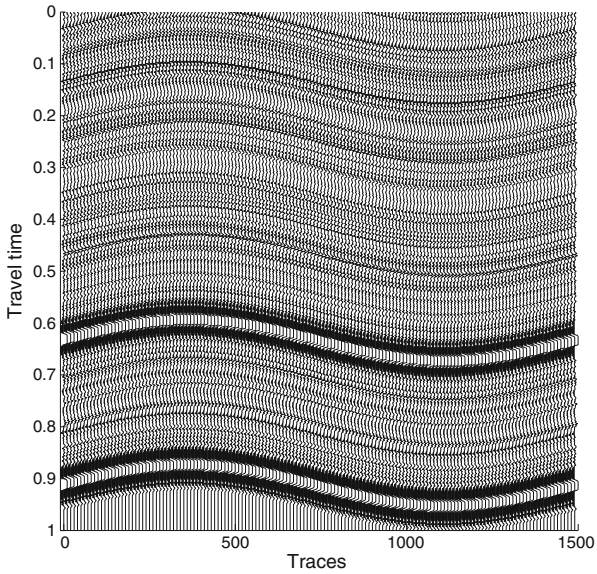

Fig. 2 The noisy multichannel seismic signal

measurements d_{true}, $d_{\text{true}}(:)$ and $d_{\text{estimate}}(:)$ mean that the operation is for each value of d_{true} and d_{estimate}. We also make test for other noise levels, the results reveal that preconditioning technique with fast non-monotone iteration yields good performances. The input reflectivity function is plotted in Fig. 1. The synthetic data are plotted in Fig. 2. The deblurred seismic signal is plotted in Fig. 3. It is clear that the noises are greatly reduced.

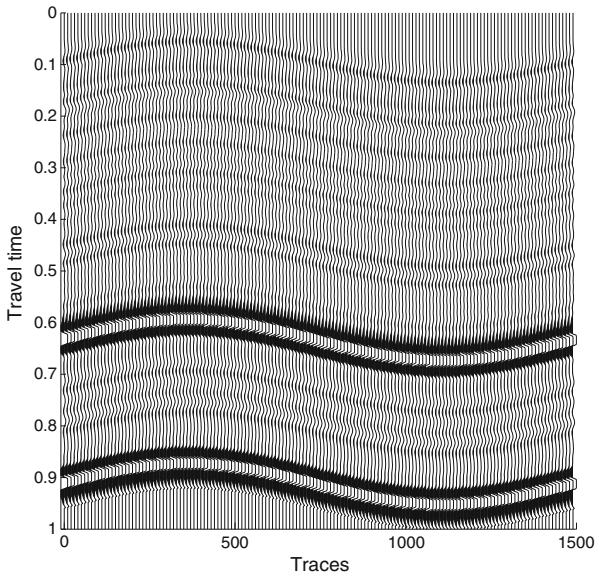

Fig. 3 The recovered multichannel seismic signal

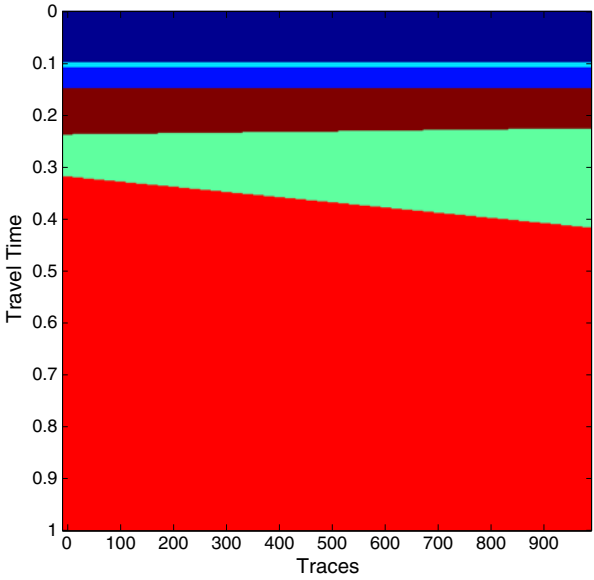

Fig. 4 Layered velocity model

5.1.2 Layered velocity model

In the following we consider a 6 layers velocity model, see Fig. 4. To generate the seismogram, a Ricker wavelet with central frequency equaling 30 Hz and sampling interval equaling 2 ms is used to perform a convolution. The noisy

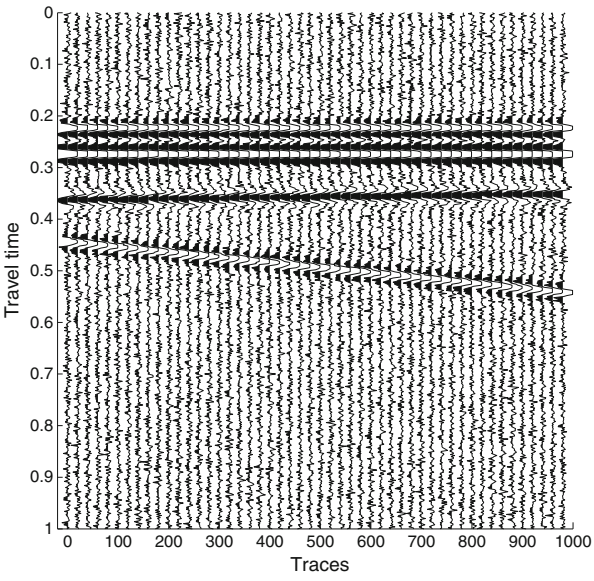

Fig. 5 The blurred seismogram

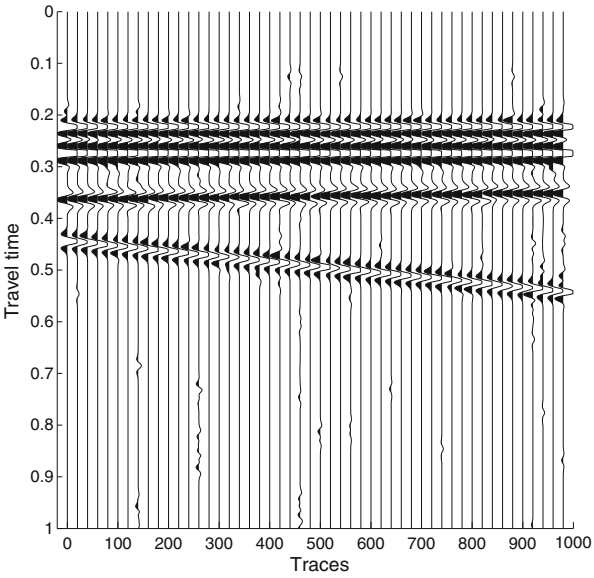

Fig. 6 The deblurred seismogram

seismogram with noise level $\delta = 0.001$ is shown in Fig. 5. The regularization parameter α is chosen as 0.001. The same stabilizer and preconditioner are applied as before. By our algorithm, deblurred seismogram is shown in Fig. 6. It is clear that the noise parts (which may be multiples in practice) are removed greatly. With the deblurred data, the reflectivity model can be retrieved. A comparison of the true reflectivity model and the retrieved reflectivity

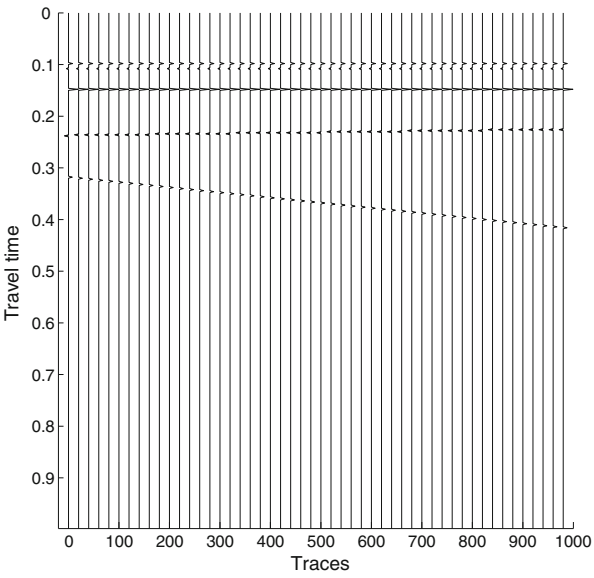

Fig. 7 The true reflectivity signal

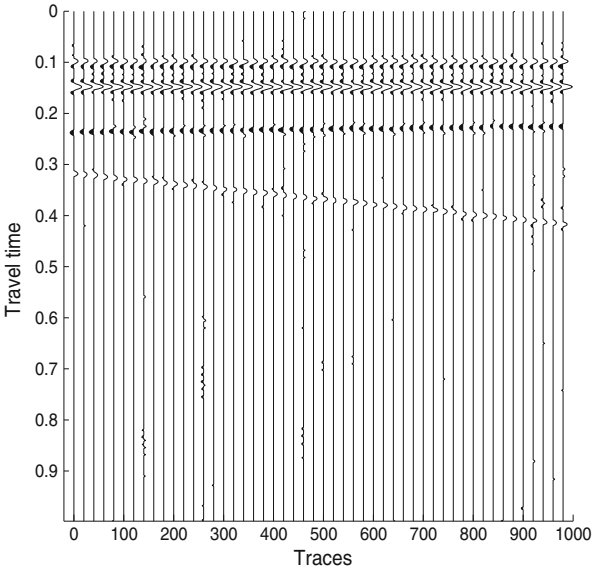

Fig. 8 The recovered reflectivity signal

model are illustrated in Figs. 7 and 8, respectively. To show more deals of the results, we select 11 traces from Figs. 5–8, respectively. The comparison results are illustrated in Fig. 9. The illustrations indicate that our method is very stable for numerical inversion and can generate results with high resolution. Figure 10 plots the least squares errors of the solution by the non-monotone gradient descent method at each iteration. This figure vividly shows

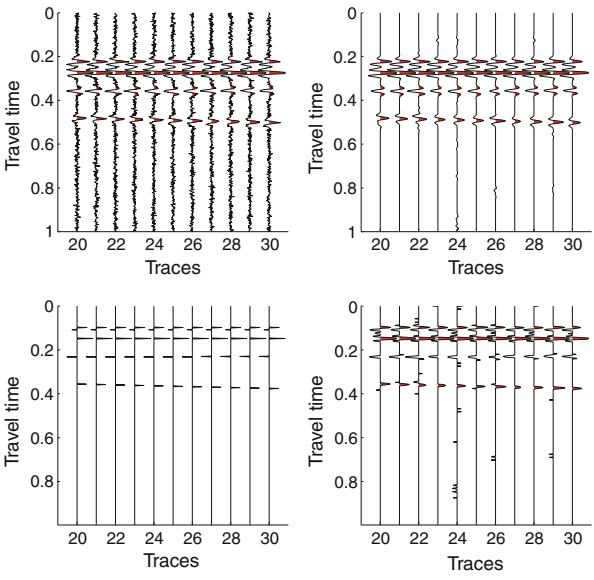

Fig. 9 Comparison of selected traces: noisy data (*upper left*); denoised data (*upper right*); true reflectivity (*lower left*); recovered reflectivity (*lower right*)

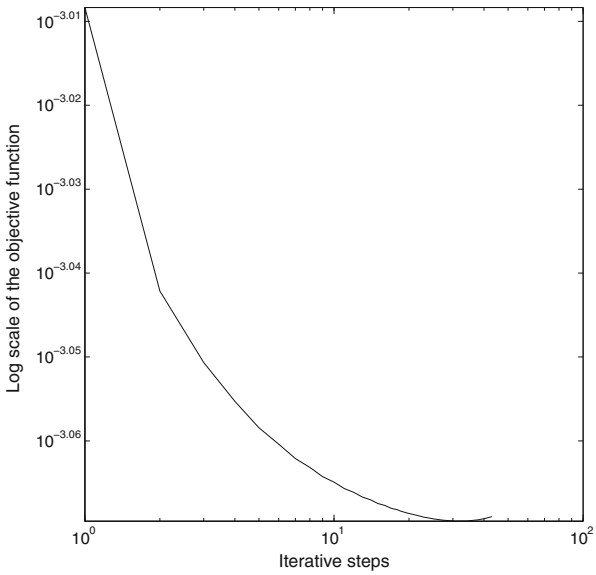

Fig. 10 Error plot of the least squares errors of the gradient descent method

us the nonmonotonicity and speediness of the non-monotone gradient descent method. It is well known that the gradient descent method and the conjugate gradient method possess linear convergence rate [42]. And with proceeding of iterations, zigzagging phenomenon would occur for these two methods [33, 34]. This phenomenon does not occur for non-monotone gradient descent method. The R-superlinear convergence property ensures its efficiency in inversion as Figs. 6–10 illustrated.

5.2 Seismic migration imaging

Deconvolution is an approach to improve the resolution of migration imaging because it uses knowledge of the resolution kernel of the seismic experiment to compensate for the effects mentioned in [18, 23, 34, 37]. The model for migration is in the form of (1) except that the operator L is approximated by the integral operator

$$L := \omega^2 \int_V G_0(g|x) G_0(x|s) dV, \tag{39}$$

where $G_0(g|x)$ and $G_0(x|s)$ are the wave equation Green's function satisfying the acoustic Helmholtz equation for a specified background medium, g is the location of receiver, x the imaging point and s the source position, dV means volume integral [39]. The migration imaging refers to solve the (1) to get the reflector with true amplitude and right location.

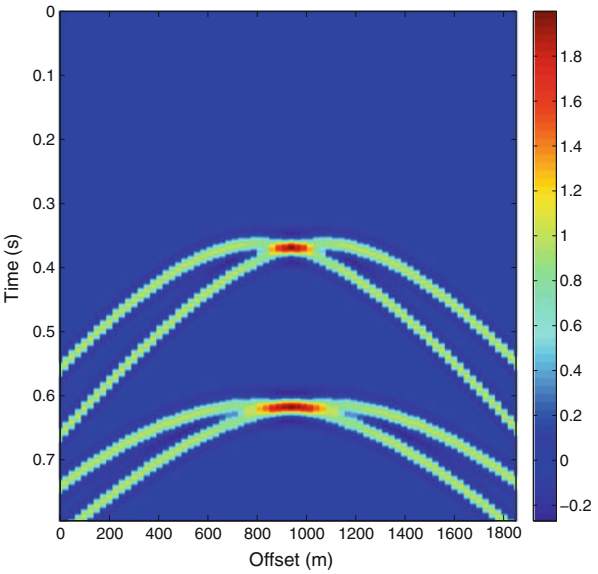

Fig. 11 Seismic data of point diffraction scatterers with large gap

We perform experiments on a six-point scatterers diffraction model. These scatterer models are buried at a depth of 625–1,625 m. A source wavelet with central frequency 20 Hz and time sample rate 2 ms is used to generate the data. We assume that 75 receivers are uniformly distributed on a survey line with maximum length of 1,875 m. The sampling interval of the survey line and the depth gridpoint spacing are both 25 m. This yields the grid dimensions of the

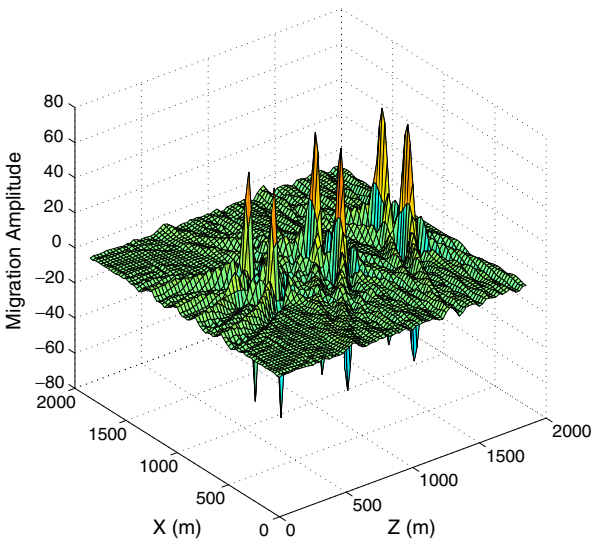

Fig. 12 The standard migration image for the data in Fig. 11

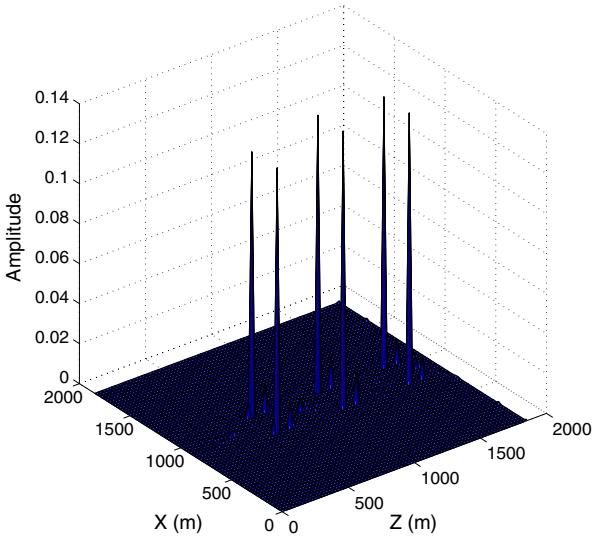

Fig. 13 The gradient descent migration deconvolution image for the data in Fig. 11

reflectivity model are $1{,}875 \times 1{,}875$ points. The background velocity is homogeneous with $c = 2{,}000$ m/s, and the time sampling interval is $dt = 2$ ms. First, we assume that the reflectors are well separated, i.e., the reflectors are in a large distances in offset. The seismogram, the standard migration and gradient descent migration deconvolution images are illustrated in Figs. 11, 12 and 13, respectively. Next, we assume that the reflectors are near from each other, i.e.,

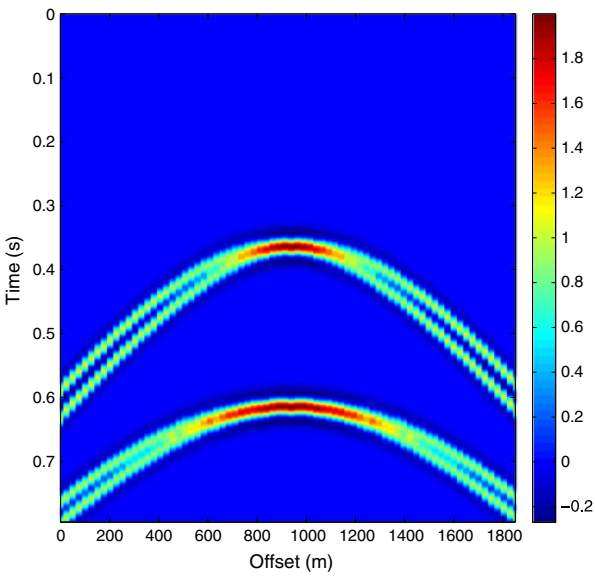

Fig. 14 Seismic data of point diffraction scatterers with small gap

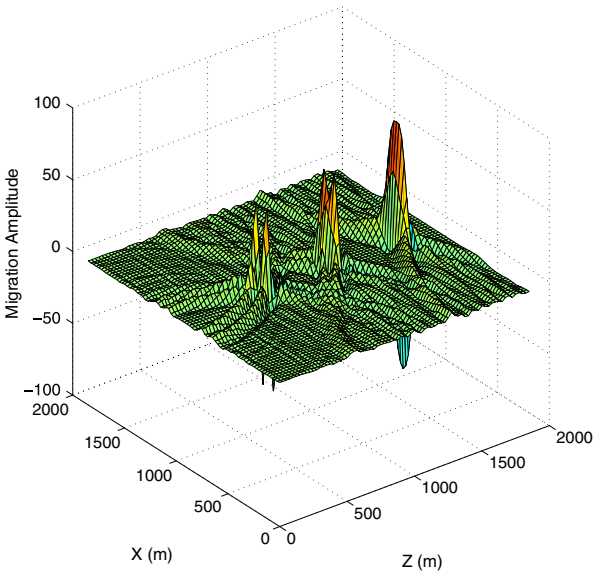

Fig. 15 The standard migration image for the data in Fig. 14

the reflectors are in a small distances in offset. The seismogram, the standard migration and gradient descent migration deconvolution images are illustrated in Figs. 14, 15 and 16, respectively. It is clear from the illustrations that the non-monotone gradient descent method yields better resolution imaging results than the standard migration tool. Comparison of the Figs. 12–16, it reveals that the non-monotone gradient descent method possesses better resolution ability

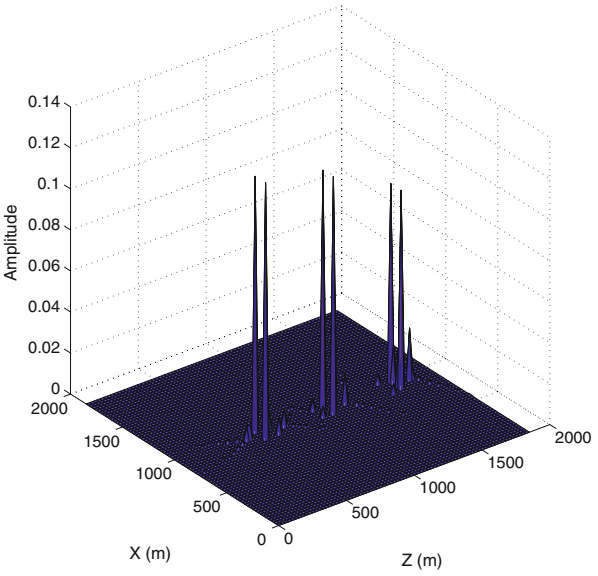

Fig. 16 The gradient descent migration deconvolution image for the data in Fig. 14

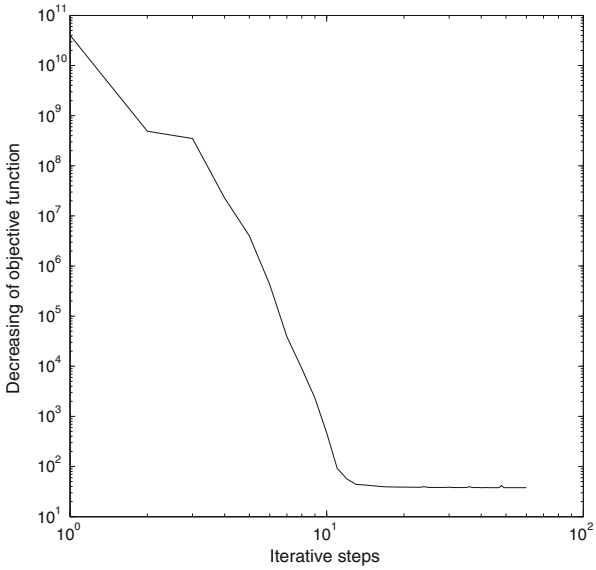

Fig. 17 Error plot of the least squares errors of the gradient descent method

than the standard migration method. Figure 17 plots the least squares errors of the solution by the non-monotone gradient descent method at each iteration. The nonmonotonicity and speediness of the method are clearly illustrated.

6 Discussions

We apply preconditioning techniques to the gradient iteration. A symmetric successive over relaxation preconditioner is performed in this study. We emphasize that this kind of choice of preconditioners is not optimal since we know that these preconditioners are originally designed for well-posed problems, and sometimes it is not suitable for ill-posed problems [34]. Moreover, the choice of the preconditioner is closely related to the structure of the model of the problems. Therefore, how to choose proper preconditioning matrices for ill-posed seismic inversion and imaging problem is still to be an interesting issue.

Applying the above method to real data applications is much more complicated. For example, for retrieval of the reflectivity model of real data, one needs to estimate the source wavelet, this usually involves well logging information. For migration deconvolution, since the earth medium is usually anisotropic, explicit formula of the wave equation Green's function is hard to obtain. Therefore one has to resort to other tools, e.g., ray tracing technique, to compute the travel time and estimate the Green's function. In that case, finding preconditioning technique will be a delicate work.

7 Conclusion

In this paper, we developed a non-monotone gradient method and applied it to seismic inversion, especially to the seismic deconvolution and least-square migration. We introduced a regularization technique to reduce the ill-posedness and designed a preconditioner to speed up the convergence. We have proved that the method converges with R-superlinear rate. Numerical examples demonstrate that the new method is efficient and improve the resolution of the inversion results.

Acknowledgements I am grateful to anonymous reviewers' and Profs. Yuesheng Xu and Hongqi Yang's helpful comments and suggestions on the paper. This work is supported by National Natural Science Foundation of China under grant numbers 10871191, 40974075 and Knowledge Innovation Programs of Chinese Academy of Sciences KZCX2-YW-QN107.

References

1. Aki, K., Richards, P.G.: Quantitative Seismology: Theory and Methods. Freeman, San Francisco (1980)
2. Backus, G., Gilbert, J.: Numerical applications of a formalism for geophysical inverse problems. Geophys. J. R. Astron. Soc. **13**, 247–276 (1967)
3. Backus, G., Gilbert, J.: Numerical applications of a formalism for geophysical inverse problems. Geophys. J. R. Astron. Soc. **16**, 169–205 (1968)
4. Barzilai, J., Borwein, J.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**, 141–148 (1988)
5. Bevc, D., Ortigosa, F., Guitton, A., Kaelin, B.: Next generation seismic imaging: high fidelity algorithms and high-end computing. In: AGU General Assembly, Acapulo, Mexico (2007)
6. Dai, Y.H.: On the nonmonotone line search. J. Optim. Theory Appl. **112**, 315–330 (2002)
7. Fletcher, R.: On the Barzilai–Borwein Method. University of Dundee Report NA/207 (2001)
8. Greenbaum, A.: Iterative Methods for Solving Linear Systems. SIAM, Philadelphia (1997)
9. Kelley, C.T.: Iterative Methods for Linear and Nonlinear Equations. SIAM, Philadelphia (1995)
10. Kirsch, A.: An Introduction to the Mathematical Theory of Inverse Problems. Springer, New York (1996)
11. Lanczos, C.: Linear Differential Operators. Van Nostrand, New York (1961)
12. Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Q. Appl. Math. **2**, 164–166 (1944)
13. Liu, H., He. L.: Pseudo-differential operator and inverse scattering of multi-dimensional wave equation. In: Wang, Y.F., Yagola, A., Yang, C. (eds.) Optimization and Regularization for Computational Inverse Problems and Applications, pp. 301–325. Springer/Higher Education Press, Berlin/Beijing (2010)
14. Margrave, G.F., Lamoureux, M.P., Grossman, J.P., Iliescu, V.: Gabor deconvolution of seismic data for source waveform and Q correction. In: SEG Extended Abstracts, pp. 2190–2193 (2002)
15. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear inequalities. SIAM J. Appl. Math. **11**, 431–441 (1963)
16. Molina, B., Raydan, M.: Preconditioned Barzilai–Borwein method for the numerical solution of partial differential equations. Numer. Algorithms **13**, 45–60 (1996)
17. Murio, D.A.: The Mollification Method and the Numerical Solution of Ill-Posed Problems. Wiley, New York (1993)
18. Nemeth, T., Wu, C.J., Schuster, G.T.: Least-squares migration of incomplete reflection data. Geophysics **64**(1), 208–221 (1999)

19. Nolet, G.: Solving or resolving inadequate noisy tomographic systems. J. Comput. Phys., **61**, 463–482 (1985)
20. Nolet, G., Snieder, R.: Solving large linear inverse problems by projection. Geophys. J. Int. **103**, 565–568 (1990)
21. Saad, Y.: Iterative Methods for Sparse Linear Systems. SIAM, Philadelphia (2000)
22. Sacchi, M.D., Wang, J., Kuehl, H.: Regularized migration/inversion: new generation of imaging algoritihms. CSEG Recorder **31**(special edition), 54–59 (2006)
23. Sjøerg, T., Gelius, L.-J., Lecomte, I.: 2D deconvolution of seismic image blur. In: Expanded Abstracts, SEG 73nd Annual Meeting, Dallas (2003)
24. Tarantola, A.: Inverse Problems Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier, Amsterdam (1987)
25. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-posed Problems. Wiley, New York (1977)
26. 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)
27. Treitel, S., Lines, L.R.: Past, present, and future of geophysical inversion-a new millennium analysis. Geophysics **66**, 21–24 (2001)
28. Ulrych, T.J., Sacchi, M.D., Grau, M.: Signal and noise separation: art and science. Geophysics **64**, 1648–1656 (1999)
29. Ulrych, T.J., Sacchi, M.D., Woodbury, A.: A Bayesian tour to inversion. Geophysics **66**, 55–69 (2000)
30. Valenciano, A., Biondi, B., Guitton, A.: Target oriented wave-equation inversion. Geophysics **71**, A35–A38 (2006)
31. VanDecar, J.C., Snieder, R.: Obtaining smooth solutions to large linear inverse problems. Geophysics **59**, 818–829 (1994)
32. Wang, Y.F.: Computational Methods for Inverse Problems and Their Applications. Higher Education Press, Beijing (2007)
33. Wang, Y.F., Ma, S.Q.: Projected Barzilai–Borwein methods for large scale nonnegative image restorations. Inverse Probl. Sci. Eng. **15**(6), 559–583 (2007)
34. Wang, Y.F., Yang, C.C.: Accelerating migration deconvolution using a non-monotone gradient method. Geophysics **75**(4), S131–S137 (2010)
35. Wang, Y.F., Yuan, Y.X.: Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. Inverse Probl. **21**, 821–838 (2005)
36. 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)
37. Wang, Y.F., Yang, C.C., Duan, Q.L.: On iterative regularization methods for seismic migration inversion imaging. Chin. J. Geophys. **52**(3), 1615–1624 (2009)
38. Xiao, T.Y., Yu, S.G., Wang, Y.F.: Numerical Methods for Solution of Inverse Problems. Science Press, Beijing (2003)
39. Wang, Y.F., Stepanova, I.E., Titarenko, V.N., Yagola, A.G.: Inverse Problems in Geophysics and Solution Methods. Higher Eduacaiton Press, Beijing (2011)
40. Xu, S.F.: Theory and Methods for Matrix Computation. Beijing University Press, Beijing (1995)
41. Yu, J.H., Hu, J.X., Schuster, G.T., Estill, R.: Prestack migration deconvolution. Geophysics **71**(2), S53–S62 (2006)
42. Yuan, Y.X.: Numerical Methods for Nonliear Programming. Shanghai Science and Technology Publication, Shanghai (1993)
43. Yuan, Y.X.: Step-sizes for the gradient method. In: Proceedings of the Third International Congress of Chinese Mathematicians, AMS/IP Studies in Advanced Mathematics, pp. 785–796 (2008)
44. Yuan, Y.X.: Gradient methods for large scale convex quadratic functions. In: Wang, Y.F., Yagola, A., Yang, C. (eds.) Optimization and Regularization for Computational Inverse Problems & Applications, pp. 141–155. Springer/Higher Education Press, Berlin/Beijing (2010)