SCIENTIFIC REPORTS

A joint matrix minimization approach for seismic wavefield recovery

Liping Wang[1] & Yanfei Wang[2,3]

Reconstruction of the seismic wavefield from sub-sampled data is important and necessary in seismic image processing; this is partly due to limitations of the observations which usually yield incomplete data. To make the best of the observed seismic signals, we propose a joint matrix minimization model to recover the seismic wavefield. Employing matrix instead of vector as weight variable can express all the sub-sampled traces simultaneously. This scheme utilizes the collective representation rather than an individual one to recover a given set of sub-samples. The matrix model takes the interrelation of the multiple observations into account to facilitate recovery, for example, the similarity of the same seismic trace and distinctions of different ones. Hence an $l_{2,p}(0 < p \leq 1)$-regularized joint matrix minimization is formulated which has some computational challenges especially when p is in (0, 1). For solving the involved matrix optimization problem, a unified algorithm is developed and the convergence analysis is accordingly demonstrated for a range of parameters. Numerical experiments on synthetic and field data examples exhibit the efficient performance of the joint technique. Both reconstruction accuracy and computational cost indicate that the new strategy achieves good performance in seismic wavefield recovery and has potential for practical applications.

Reconstruction of the seismic wavefield has recently attracted increasing attentions in geophysical community. This is due to the fact that seismic acquisition often violates the Shannon sampling theorem because of the restrictions of investment, topography, noise, bad traces and so on. The under-sampled data will bring aliasing and artifacts which will influence results of migration[1], de-noising[2], multiple elimination[3] and AVO analysis[4]. In addition, huge storage of the massive data is also a problem, lossless compression methods are desirable[5]. An important branch of these methods is the sparse transform based method combined with a regularization strategy[6,7]. For this method, seismic interpolation is treated as an inverse problem, and seismic events are assumed to be sparse in some transformed domain, such as the Fourier transform[1,8–11], or the linear Radon transform[12]. Usually the acquired geophysical data is subsampled due to the variations of landform[1,13,14], hence the seismic wavefield recovery is an ill-posed inverse problem. Therefore, a key issue is how to invert the mathematical model using only incomplete, sub-sampled data[1,13,14]. Variety of regularization methods has been developed to improve the quality of image and seismic wavefield recovery[6,15–17].

Previous methods for such a recovery problem are based on the l_q-norm minimization, e.g., the basis pursuit denoising (BPDN) criterion using (orthogonal) matching pursuit method[18,19] and the least absolute shrinkage and selection operator (LASSO)[20] for l_1-norm constrained minimization problems. Efficient optimization algorithms include conjugate gradient methods with preconditioning techniques[21] and gradient projection methods[22–26]. For solving the l_q-norm minimization problem, people usually convert the matrix form of the wavefield into the vector form and solve the corresponding matrix-vector equations. We observed that the seismic wavefield can be represented using matrix instead of vector as weight variable to express all the signals simultaneously, which takes the interrelation of the sampled observations into account. This is more reasonable as the seismic signals are correlated transversely. Therefore, in this paper we propose a matrix optimization model for the seismic wavefield recovery and study the related properties. The mixed matrix minimization models have been used in machine learning. Rakotomamonjy et al.[27] proposed to use the mixed matrix norm $l_{q,p}$ $(1 \leq q < 2, 0 < p \leq 1)$ in multi-kernel and multi-task learning. But the induced optimization problems in[27] have to be solved separately by

[1]Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, P. R. China. [2]Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, 100029, P. R. China. [3]University of the Chinese Academy of Sciences, Beijing, 100049, P. R. China. Correspondence and requests for materials should be addressed to Y.W. (email: yfwang@mail.iggcas.ac.cn)

different algorithms with respect to $p = 1$ and $0 < p < 1$. For grouped feature selection, Suvrit[28] addressed a fast projection technique onto $l_{1,p}$-norm balls particularly for $p = 2, \infty$. But the derived method in[28] does not match the proposed matrix optimization problem (11). Similar joint sparse representation has been used for robust multi-modal biometrics recognition in[29]. Sumit *et al.*[29] employed the traditional alternating direction method of multipliers to solve the involved optimization problem. Wang *et al.*[30] applied $l_{2,0+}$-norm to semi-supervised robust dictionary learning, while the optimization algorithm has not displayed definite convergence analysis[30].

Recently, matrix-minimization methods with nuclear norm have been developed for seismic wavefield recovery[31–34] which mainly considers the rank reduction as the sparse pattern in 2D cases. To avoid the expensive computations in solving the involved matrix completion optimization problems, a matrix factorization strategy was developed in[31,32]. This paper proposes a different matrix minimization approach based on $l_{2,q}-l_{2,p}$ norm which naturally generalizes the representative vector to matrix in joint distribution sense. A unified method is developed to solve the matrix optimization problem with mixed norm for any $q = 2$ and $0 < p \leq 1$. The innovations of this paper can be listed as follows:

1) A jointly sparse matrix minimization model is developed for seismic wavefield recovery. This approach employs matrix to expresses multiple signals simultaneously. The measurement of matrix row coefficients are expected to exhibit the compact priori of multiple observations which is different from the existed methods based on matrix nuclear-norm minimization[31–34].
2) A unified algorithm is developed to solve the mixed matrix optimization problem (7) for any $p \in (0, 1]$. This algorithm needs only matrix-vector operations but not matrix factorization which can be easily adapted to large-scale cases. The convergence analysis is also demonstrated.
3) Numerical experiments on synthetic and field data are carried out. The results on seismic wavefield recovery exhibit the efficient recovery performance of the joint sparse expression strategy.

Modeling

Given a set of seismic signals (traces) $x_1, x_2, \ldots, x_l$ in n-dimensional space, each signal $x_j (j = 1, 2, \ldots, l)$ is sensed by m sensors to yield seismic wavefield records as

$$d_{ij} = A^i x_j, \ i = 1, 2, \cdots, m, \tag{1}$$

where A^i is a row vector representing the impulse response of the i-th sensor. Denote $A = [(A^1)^T, (A^2)^T, \ldots, (A^m)^T]^T$, then the seismic observations $d_j = [d_{1j}, d_{2j}, \ldots, d_{mj}]^T \in R^m$ can be reformulated as $d_j = Ax_j (j = 1, 2, \ldots, l)$. Sparse expression is a popular strategy to restore x_j with m much less than n of the mapping operator A.

Suppose that the original seismic signal x_j can be spanned by a series of orthogonal bases $\{\Psi_k\}_{k=1}^K$ such that

$$x_j(t) = \sum_{k=1}^K m_j^k \Psi_k, \tag{2}$$

where $m_j^k = (x_j, \Psi_k)$. Denote Ψ the orthogonal matrix constituted by the orthogonal bases, then we have a more compact transformation $L = A\Psi \in R^{m \times K}$. Consequently the systems (1) and (2) can be incorporated to

$$L m_j = d_j, \qquad j = 1, 2, \cdots, l, \tag{3}$$

where $m_j = \Psi * x_j = \{m_j^k\}_{k=1}^K$ is the coefficient vector (weighting factor) corresponding to the seismic signal x_j. Usually, problem (3) is ill-posed due to the limitation of acquisition and violation of sampling requirements. Sparse regularization is preferred to restore the operation coefficients from the under-determined linear combination system (3). A general $l_q-l_p (q > 0, p > 0)$ model was presented in [16]

$$\min_{m_j} J_j^{(\alpha)}(m_j) = \|L m_j - d_j\|_q^q + \alpha_j \|m_j\|_p^p, \qquad q > 0, \ p > 0, \tag{4}$$

where $\left\|m_j\right\|_p^p = \sum_{k=1}^K |m_j^k|^p$ is the stabilizer bearing prior information with respect to d_j and $\alpha_j > 0$ is a regularization parameter. When $0 < p \leq 1$, the minimization model (4) tries to find a sparse recovery coefficient m_j with the least nonzero entries. However, the framework (4) recovers the weight factor m_j only using the j-th seismic trace record d_j independently which totally ignores the correlation with other sampled data $d_{\hat{j}} (\hat{j} \neq j)$. Generally, multiple seismic wavefield traces are related to each other. The similarity and difference hidden in the given group of seismic traces are expected to improve the recovery performance. To detailedly demonstrate the correlationship among multiple seismic traces, we randomly choose three trace observations from a seismogram generated from a seven layers geologic velocity model (see Experimental Section for details). Two neighboring traces are denoted by d_1 and d_2 while the third one d_3 is relatively far from them. We separately recover the representation coefficients m_j^*, $j = 1, 2, 3$ by solving

$$m_j^* = \arg \min_{m_j \in R^K} J(m_j) = \|L m_j - d_j\|_2^2 + \alpha \sum_{k=1}^K (|m_j^k|)^p, \quad j = 1, 2, 3, \ p = 0.5, \tag{5}$$

where m_j^k is k-th entry of m_j. The weight values of recovered coefficients are plotted in Fig. 1(a–c). The horizontal axis denotes the coordinates of the representation vector while the vertical axis shows the weight quantities of
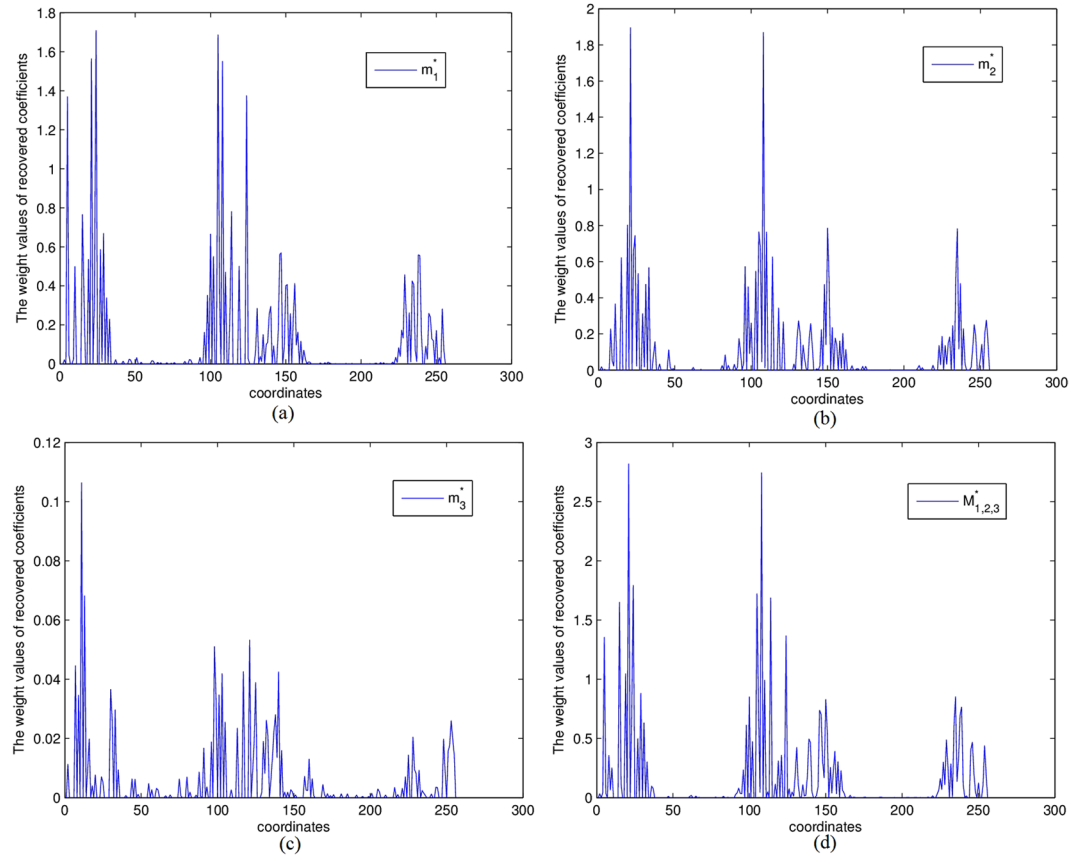
Figure 1. (**a**–**d**) denote the weight values of recovered coefficients of different traces.

representation coefficients, namely $\left|(m_j^*)^k\right|$, $k = 1, 2, \cdots, 256$; $j = 1, 2, 3$. The curves clearly display the similar clustering and sparse pattern of three recovered coefficients. The correlations inspire us to assume that the multiple traces coefficients share the same distribution. For comparison, we jointly recover three coefficients simultaneously from $D_{1,2,3} = [d_1, d_2, d_3] \in R^{m \times 3}$ by a matrix minimization problem

$$M_{1,2,3}^* = \arg \min_{M_{1,2,3} \in R^{K \times 3}} J(M_{1,2,3}) = \|LM_{1,2,3} - D_{1,2,3}\|_F^2 + \alpha \sum_{k=1}^{K}(\|m_{1,2,3}^k\|_2)^p, \quad p = 0.5, \tag{6}$$

where $M_{1,2,3}^* \in R^{K \times 3}$ and $m_{1,2,3}^k \in R^3$ is the k-th row of $M_{1,2,3}$. Since three vector minimizations as (5) are integrated to a matrix one (6), each entry m_j^k of representative vector is spanned to a row vector $m_{1,2,3}^k \in R^3$. Hence the absolute values of weight entries in (5) are naturally generalized to l_2 norm of row vector for its smoothness, that is $\left|m_j^k\right| \to \left\|m_{1,2,3}^k\right\|_2$. To illustrate the jointly recovered coefficient matrix $M_{1,2,3}^*$ of (6) also follows the similar variation as in Fig. 1(a–c), we measure the l_2 norm of each row vector in the joint sense corresponding to $\left|(m_j^*)^k\right|$,

$$|(m_j^*)^k| \to \|(M_{1,2,3}^*)^k\|_2 = (|(m_{1,2,3}^*)_1^k|^2 + |(m_{1,2,3}^*)_2^k|^2 + |(m_{1,2,3}^*)_3^k|^2)^{\frac{1}{2}}, \ k = 1, 2, \cdots, 256. \tag{7}$$

Clearly, the joint representation coefficients also exhibit similar sparse pattern and weight concentration to the individual models (see Fig. 1(d)).

Under the assumption that multiple seismic wavefield traces jointly share the similar weight parameter pattern, we propose to express all the sub-sampled observations over the same bases simultaneously as

$$LM = D, \tag{8}$$

where $D = [d_1, d_2, \ldots, d_l]$ is composed of l seismic observations and $M = [m_1, m_2, \ldots, m_l]$ denotes the corresponding coefficient matrix. As far as the columns are concerned, the equation (8) is an easy consequence of the equation (3). Figure 1 has demonstrated that the multiple seismic traces are related to each other, especially when the samples are obtained in the similar fields. We reasonably measure the joint compactness and correlation of the multiple observations in row sense. By reviewing $l_q - l_p (q > 0, p > 0)$ model (4), we notice that the expression errors $e_j = Lm_j - d_j, j = 1, 2, \ldots, l$ and the priori of representation coefficients are assumed to submit to the independent identically distribution,

3

$$P(e_j|m_j) \propto \prod_{k=1}^{K} \exp\left(-\left|e_j^k\right|^q\right),\ P(m_j) \propto \prod_{k=1}^{K} \exp\left(-\alpha_k\left|m_j^k\right|^p\right),\ \alpha_k > 0 \tag{9}$$

where m_j^k is the k-th entry of representation vector $m_j \in R^K$. The solution m_j^* to (4) can be rewritten as the maximum likelihood estimation

$$m_j^* = \arg\max_{m_j \in R^K} \ln P(m_j|e_j) = \arg\max_{m_j \in R^K}(\ln P(e_j|m_j) + \ln P(m_j)) \tag{10}$$

Because each coefficient component m_j^k in (3) is spanned to a row vector in the joint expression system (8), the absolute value of the scalar component is naturally replaced by a vector norm. Euclidean norm is preferred for its smoothness and easiness. Based on the analysis (9) and (10), the joint sparse priori of coefficient matrix M and fidelity error matrix $E = LM - D$ can be considered

$$P(E|M) \propto \prod_{k=1}^{K} \exp\left(-\|e^k\|_2^q\right),\ P(M) \propto \prod_{k=1}^{K} \exp\left(-\alpha_k\|m^k\|_2^p\right),\ \alpha_k > 0$$

where m^k, e^k are the k-th row vectors of $M \in R^{k \times l}$ and $E \in R^{m \times l}$ respectively. $\alpha_k > 0$ is a constant and $\|.\|_2$ stands for the Euclidean norm. In the similar relationship between (4) and (9), the joint matrix minimization approach for the ill-posed linear system (8) can be generally formulated as

$$\min_M J(M) = \left\|LM - D\right\|_{2,q}^q + \left\|\Lambda M\right\|_{2,p}^p,\qquad q > 0,\ p > 0, \tag{11}$$

where the $l_{2,p}$ norm of the priori matrix M is defined as

$$\|M\|_{2,p}^p = \sum_{k=1}^{K} \|m^k\|_2^p,\qquad p \in (0,1]. \tag{12}$$

Here $\left\|LM - D\right\|_{2,q} (q > 0)$ denotes the $l_{2,q}$ matrix norm of $LM - D$, $\Lambda = diag\{\alpha_k\}_{k=1}^K$ is a regularization matrix and its diagonal entry $\alpha_k > 0$ is the regularization parameter for the k-th row of M. Especially, if M contains only one column m_j, each $\|m^k\|_2$ is reduced to $|m_j^k|$ while $\left\|M\right\|_{2,p}$ is equivalent to $\left\|m_j\right\|_p$. When Λ takes scalar identity, the joint system (11) is exactly reduced to (4).

There are different choices of the parameter pair $q > 0$ and $p > 0$. Here we are interested in $q = 2$ and $p \in (0,1]$ for the practical purpose. Extensive studies have illustrated that the fractional norm l_p ($p \in (0,1)$) has better sparsity than l_1 norm[35–39]. But the l_p norm is neither Lipschitz nor convex which brings computational challenge. This paper presents a unified algorithm to solve the mixed $l_{2,p}$ regularized matrix minimization problem (11) for any $p \in (0,1]$. The computational results in seismic wavefield recovery validate the efficient performance of the joint matrix minimization approach. The convergence properties of our new algorithm are also analyzed.

Algorithms

In this section, a unified method will be developed to solve the $l_{2,q} - l_{2,p}$ matrix minimization problem for any $q = 2$ and $0 < p \leq 1$. Especially when p is fractional, (11) is neither convex nor Lipschitz continuous which brings many computational difficulties. Actually the unconstrained l_q-l_p minimization is strongly NP-hard for any $0 < q < q$ or $p < 1$[40]. Reweighed minimization algorithm[35,41,42] is an efficient algorithm for solving the l_2-l_p ($0 < p < 1$) vector minimization problem which has been extended by Wang *et al.*[43] to solve matrix minimization problem. Even the problem considered in[43] is the special case of (11) with $q = p \in (0,1]$, the idea motivates us to develop an iteratively quadratic algorithm for the generalized $l_{2,p}$ matrix minimization for $p \in (0,1]$. Moreover, the convergence analysis will be uniformly demonstrated.

After simple transformation, $\left\|\Lambda M\right\|_{2,p}^p$ can be rewritten as

$$\left\|\Lambda M\right\|_{2,p}^p = Tr(M^T H M), \tag{13}$$

where $Tr(\cdot)$ stands for the trace operation and

$$H = diag\left\{\frac{\alpha_1}{\|m^1\|_2^{2-p}},\ \frac{\alpha_2}{\|m^2\|_2^{2-p}},\ \cdots,\ \frac{\alpha_K}{\|m^d\|_2^{2-p}}\right\}, \tag{14}$$

where m^k ($k = 1, 2, \ldots, K$) is the k-th row vector of M.

Hence the objective function of (11) for $q = 2$, $p \in (0,1]$ can be reformulated as

$$\begin{aligned}
J(M) &:= \|LM - D\|_{2,2}^2 + \|\Lambda M\|_{2,p}^p \\
&= Tr((LM - D)^T(LM - D)) + Tr(M^T H M)
\end{aligned} \tag{15}$$

It is well known that the *KKT* point of the unconstrained optimization problem (11) is also the stationary point of $J(M)$[44]. Compute the derivative of $J(M)$ with respect to matrix M and set it to zero, we get the *KKT* equation of the problem (11) as follows

$$\frac{\partial J(M)}{\partial M} = 2L^T(LM - D) + pHM = 0. \tag{16}$$

Thus solving (11) is reduced to finding the solution of the nonlinear equation (16). If H is fixed and the matrix $N = L^TL + \frac{p}{2}H$ is invertible, equation (16) can be solved by

$$M = \left(L^TL + \frac{p}{2}H\right)^{-1} L^TD. \tag{17}$$

We notice that if some row of M is zero, the diagonal entries of H cannot be generated, nor can N. Then the iteration breaks down. In view of the seismic wavefield recovery, the zero row means the corresponding basis function has no contribution to reconstruct all the observed seismic traces. For example, if $m^k = 0$, then L_k (the k-th column of transformation matrix L) is nothing with the observations D in the representation system (8). To avoid the possible breakdown of the matrix N in (17) and reasonably explain this numerical behavior, we apply the Sherman-Morrison-Woodbury formula[45] to N^{-1}. Denote

$$G = \left(\frac{p}{2}H\right)^{-1} = \frac{2}{p}\text{diag}\left\{\frac{\|m^1\|_2^{2-p}}{\alpha_1}, \frac{\|m^2\|_2^{2-p}}{\alpha_2}, \cdots, \frac{\|m^K\|_2^{2-p}}{\alpha_K}\right\}, \tag{18}$$

then the formula (17) can be rewritten as

$$M = N^{-1}L^TD = [G - GL^T(I_m + LGL^T)^{-1}LG]L^TD, \tag{19}$$

where I_m is m-dimensional identity operator. If matrices G and M are computed alternatively corresponding to equations (18) and (19) respectively, then an iterative procedure can be naturally developed

$$G_t = \frac{2}{p}diag\left\{\frac{\|m_t^1\|_2^{2-p}}{\alpha_1}, \frac{\|m_t^2\|_2^{2-p}}{\alpha_2}, \cdots, \frac{\|m_t^K\|_2^{2-p}}{\alpha_K}\right\},$$

$$M_{t+1} = [G_t - G_tL^T(I_m + LG_tL^T)^{-1}LG_t]L^TD. \tag{20}$$

The iterative algorithm is outlined in Algorithm 1.

Algorithm 1. An iterative procedure for solving problem (16)

Step 1. Input $L \in R^{m \times K}$, $D \in R^{m \times l}$. Set the sparse parameter $p \in (0, 1]$ and diagonal matrix $\Lambda = diag\{\alpha_1, \alpha_2, \cdots, \alpha_K\} \succ 0$ (here $\succ$ refers to the positive definite). Given the stopping criterion $\epsilon > 0$.

Step 2. Set $t = 1$ and initialize $M_1 \in R^{K \times l}$.

Step 3. For $t = 1, 2, \ldots$ until $\rho_t \leq \epsilon$ do:

$$G_t = \frac{2}{p}diag\left\{\frac{\|m_t^1\|_2^{2-p}}{\alpha_1}, \frac{\|m_t^2\|_2^{2-p}}{\alpha_2}, \cdots, \frac{\|m_t^K\|_2^{2-p}}{\alpha_K}\right\};$$

$$M_{t+1} = [G_t - G_tL^T(I_m + LG_tL^T)^{-1}LG_t]L^TD;$$

$$\rho_t = \frac{\|M_{t+1} - M_t\|_F}{\|M_t\|_F}.$$

The m_t^k ($k = 1, 2, \ldots, K$) means the k-th row vector of M_t. Algorithm 1 aims to solve the fixed-point system (16) which is the stationary equation of the matrix function (15). Based on the iterative procedure of Algorithm 1, the iterative point M_k is the solution of the nonlinear equation (16) if and only if $M_t = [G_t - G_tL^T(I_m + LG_tL^T)^{-1}LG_t]L^TD$ which is equivalent to $M_k = M_{k+1}$. From this iteration on, the iteration point will not update which indicates that a stationary point has been found. Hence the stopping criterion of Algorithm 1 can be chosen as $\rho_t := \frac{\|M_{t+1} - M_t\|_F}{\|M_t\|_F} \leq \epsilon$, where $\|\cdot\|_F$ stands for the Frobenius norm[46].

Based on the definition (12) of $\|M\|_{2,p}$, the sparse parameter $p \in (0, 1]$ aims to find a solution with many zero row vectors of the $l_{2,p}$-regularized matrix minimization problem (11). This means that many basis functions have no contribution to reconstruct the seismic wavefields which accords with the prior knowledge. Therefore $(m_t)^k = 0$ might frequently occur during the iterations of Algorithm 1. We may formulate the following statement.

Remark. In Algorithm 1, if $m_{t_0}^k = 0$ happens for some iteration M_{t_0}, then $m_t^k = 0$ for $t \geq t_0$.

We give explanations of the above remark as follow. If $m_{t_0}^k = 0$ in the t_0-th iteration, then the diagonal entry of G_{t_0} is zero, namely $(G_{t_0})_{kk} = 0$. From the update formula $M_{t_0+1} = G_{t_0}[I_K - L^T(I_m + LG_{t_0}L^T)^{-1}LG_{t_0}]L^TD$, we know

Methods	No noise			Noise level 0.001			Noise level 0.01		
	err_{rel}	SNR	CPU(s)	err_{rel}	SNR	CPU(s)	err_{rel}	SNR	CPU(s)
Alg.3 (p = 0.5)	2.56e-4	71.83	4.56e-2	3.54e-4	69.02	1.7e-2	2.09e-2	33.62	3.25e-2
Alg.3 (p = 1)	5.08e-4	65.89	3.18e-2	6.67e-4	63.51	1.81e-2	2.48e-2	32.12	3.09e-2
SPG (p = 1)	2.62e-4	71.07	5.56e-2	3.75e-4	68.96	4.22e-2	2.54e-2	31.89	6.54e-2

Table 1. The experimental results of one-dimensional seismic wavefield reconstruction.

that $m_{t_0+1}^k = 0$ holds, so does $m_t^k = 0$ for $t \geq t_0$. After t_0 iterations with $m_{t_0}^k = 0$, the k-th column of the matrix L is unnecessary in the linear system (8) and the variational function $J(M)$ in (15). So we can discard the k-th column of the matrix L to reduce the system without any loss. The improvement of Algorithm 1 can be concluded as Algorithm 2.

Algorithm 2. Solving problem (16) for any $p \in (0, 1]$

Step 1. Input $L \in R^{m \times K}$, $D \in R^{m \times l}$. Set the sparse parameter $p \in (0, 1]$ and the diagonal matrix $\Lambda = diag\{\alpha_1, \alpha_2, \cdots, \alpha_K\} \succ 0$. Given stopping criterion $\epsilon > 0$.

Step 2. Set $t = 1$ and initialize $\hat{M}_1 \in R^{K \times l}$. Let $\Omega_0 = \{1, 2, \ldots, K\}$.

Step 3. For $t = 1, 2, \ldots$ until $\rho_t \leq \epsilon$ do:

$$\Omega_t = \Omega_{t-1} \setminus \{k: \|\hat{m}_t^k\|_2 = 0\};$$

$$M_t = \hat{M}_t(\Omega_t; :), \quad L_t = L(:;\Omega_t);$$

$$G_t = \frac{2}{p} diag\left\{\frac{\|m_t^k\|_2^{2-p}}{\alpha_k}\right\}_{k \in \Omega_t};$$

$$\hat{M}_{t+1} = [G_t - G_t L_t^T (I_m + L_t G_t L_t^T)^{-1} L_t G_t] L_t^T D;$$

$$\rho_t = \frac{\|\hat{M}_{t+1} - M_t\|_F}{\|M_t\|_F}.$$

In Algorithm 2, $M_t = \hat{M}_t(\Omega_t; :)$ means to keep the rows of $\hat{M}_t$ corresponding to the index set Ω_t while $L_t = L(:;\Omega_t)$ keeps the column of L corresponding to Ω_t. Compared with Algorithm 1, Algorithm 2 removes the zero rows of the approximation solution in each iteration and the corresponding columns of the bases matrix L. This technique iteratively reduces the inactive set of data.

Based on the procedure of Algorithm 2, $N_t = L_t^T L_t + \frac{p}{2} H_t$ is well defined and $\hat{M}_{t+1}$ is the solution of the linear system $N_t M = L_t^T D$. Since N_t is symmetric and positive definite, $\hat{M}_{t+1}$ is also the optimal matrix solution of the following quadratic subproblem

$$\min_M Q_t(M) := Tr((L_t M - D)^T (L_t M - D)) + Tr(M^T HM). \tag{21}$$

We would have $Q_t(\hat{M}_{t+1}) \leq Q_t(M_t)$, which is equivalent to

$$\|L_t \hat{M}_{t+1} - D\|_F^2 + \frac{p}{2} \sum_{k \in \Omega_t} \frac{\alpha_k \|\hat{m}_{t+1}^k\|_2^2}{\|m_t^k\|_2^{2-p}} \leq \|L_t M_t - D\|_F^2 + \frac{p}{2} \|\Lambda_t M_t\|_{2,p}^p. \tag{22}$$

It is noticed that $J(M_t) = \left\|L_t M_t - D\right\|_F^2 + \left\|\Lambda_t M_t\right\|_{2,p}^p$ and $J(M_{t+1}) = J(\hat{M}_{t+1})$. Using inequalities (A-2) (see the Appendix A) and (22), we can derive that

$$J(M_{t+1}) = J(\hat{M}_{t+1}) \leq J(M_t), \qquad p \in (0, 1], \tag{23}$$

which means $\{J(M_t)\}$ will decrease with respect to iterations for any $p \in (0, 1]$.

Once $J(M_{t+1}) = J(M_t)$ happens for some t, the equalities in (A-2) (see the Appendix A) and (22) hold simultaneously. From Proposition 2 of the Appendix A, we obtain $\|\hat{m}_{t+1}^k\|_2 = \|m_t^k\|_2$ for all $k \in \Omega_t$. Thus $G_{t+1} = G_t$ and $H_{t+1} = H_t$, which implies that $\hat{M}_{t+1}$ is a solution of the equation (17). Since the objective function sequence $\{J(M_t)\}$ for all t is strictly decreasing and lower bounded, any accumulation of the set $\{M_t\}$ is a stationary point of the equation (11). At the same time, the descending quantity of $\{J(M_t)\}$ measures the convergence precision of the matrix sequence $\{M_t\}$.

Once the nonzero set of the t-th iteration has been fixed, the subproblem (21) can be solved in a variety of ways such as preconditioned conjugate gradient methods[46], nonmonotone gradient descent methods[47,48], and so on. The framework can be concluded as Algorithm 3.
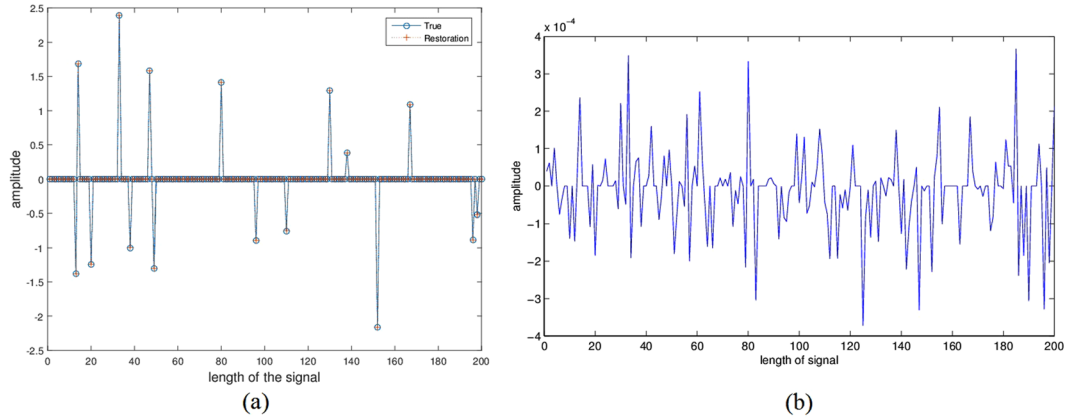
Figure 2. One-dimensional experimental results via model (11) with $q = 2$ and $p = 1/2$ for noisy data: (**a**) comparison of the original and the recovered signal; (**b**) difference between the recovered signal and the original signal.

	No noise			Noise level 0.001		
Methods	err$_{rel}$	SNR	CPU(s)	err$_{rel}$	SNR	CPU(s)
Alg.3 (p = 0.1)	0.0063	44.0585	0.326	0.0098	40.2009	0.5056
Alg.3 (p = 0.2)	0.0075	42.5209	0.3277	0.0092	40.7235	0.36
Alg.3 (p = 0.5)	0.0112	39.049	0.4796	0.0122	38.308	0.5044
Alg.3 (p = 1)	0.0115	38.7591	0.135	0.0121	38.3405	0.0973
SPG (p = 1)	0.0711	22.9599	13.121	0.0726	22.7818	12.9125

Table 2. Two-dimensional seismic wavefield reconstruction on 10% missing data.

	No noise			Noise level 0.001		
Methods	err$_{rel}$	SNR	CPU(s)	err$_{rel}$	SNR	CPU(s)
Alg.3 (p = 0.1)	0.0204	33.8269	0.2916	0.0209	33.5817	0.2451
Alg.3 (p = 0.2)	0.0179	34.9441	0.4101	0.0252	31.9641	0.4695
Alg.3 (p = 0.5)	0.0219	33.2043	0.3775	0.0283	30.9583	0.4783
Alg.3 (p = 1)	0.0314	30.075	0.0962	0.0346	29.228	0.0894
SPG (p = 1)	0.0953	20.4223	25.7615	0.0953	20.4138	21.894

Table 3. Two-dimensional seismic wavefield reconstruction on 25% missing data.

Algorithm 3. A unified algorithm for solving problem (16) for any $p \in (0, 1]$

Step 1. Input $L \in R^{m \times K}$, $D\Lambda = diag\{\alpha_1, \alpha_2, \cdots, \alpha_K\} \succ 0 \in R^{m \times l}$. Set the sparse parameter $p \in (0, 1]$ and the diagonal matrix . Given stopping criterion $\epsilon > 0$.

Step 2. Set $t = 1$ and initialize $\hat{M}_1 \in R^{K \times l}$. Let $\Omega_0 = \{1, 2, \ldots, K\}$.

Step 3. For $t = 1, 2, \ldots$ until $\rho_t \leq \epsilon$ do:

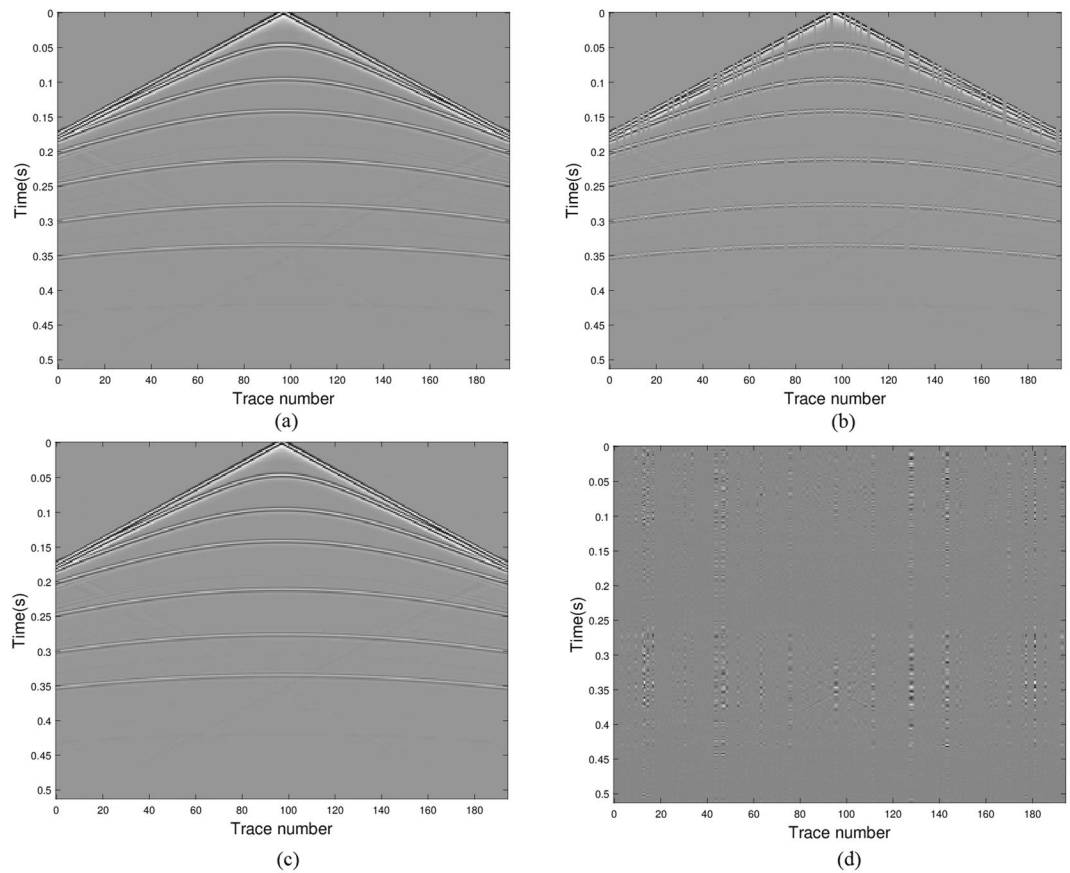$$\Omega_t = \Omega_{t-1} \setminus \{k | \|\hat{m}_t^k\|_2 = 0\};$$

$$M_t = \hat{M}_t(\Omega_t; :); \quad L_t = L(:;\Omega_t);$$

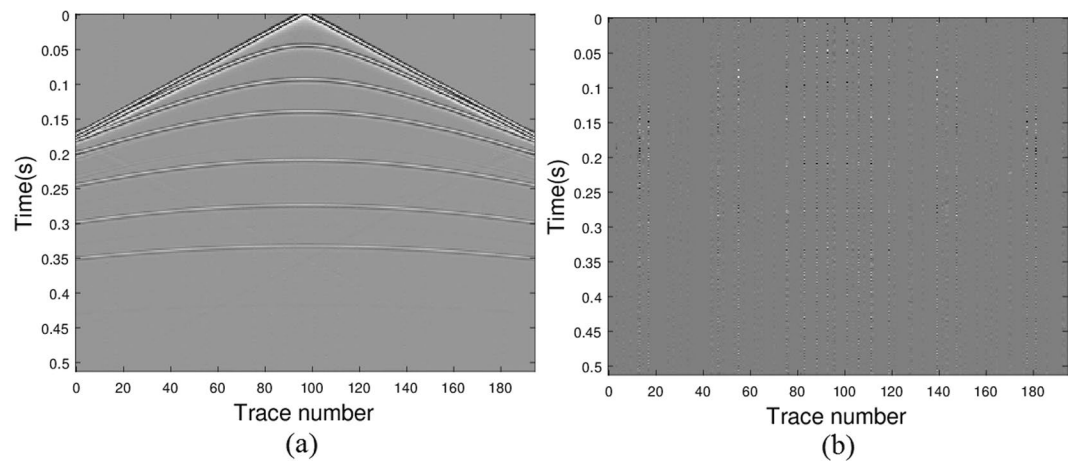$$H_t = diag\left\{\frac{\alpha_k}{\|m_t^k\|_2^{2-p}}\right\}_{k \in \Omega_t};$$

$$N_t = L_t^T L_t + \frac{p}{2} H_t;$$

Solve $N_t M = L_t^T D$ for the solution $\hat{M}_{t+1}$;

Figure 3. Seismic data results via model (11) with $q = 2$ and $p = 0.5$: (**a**) the real data; (**b**) the data with missing traces; (**c**) the recovered data; (**d**) error between the original and the recovered signals.
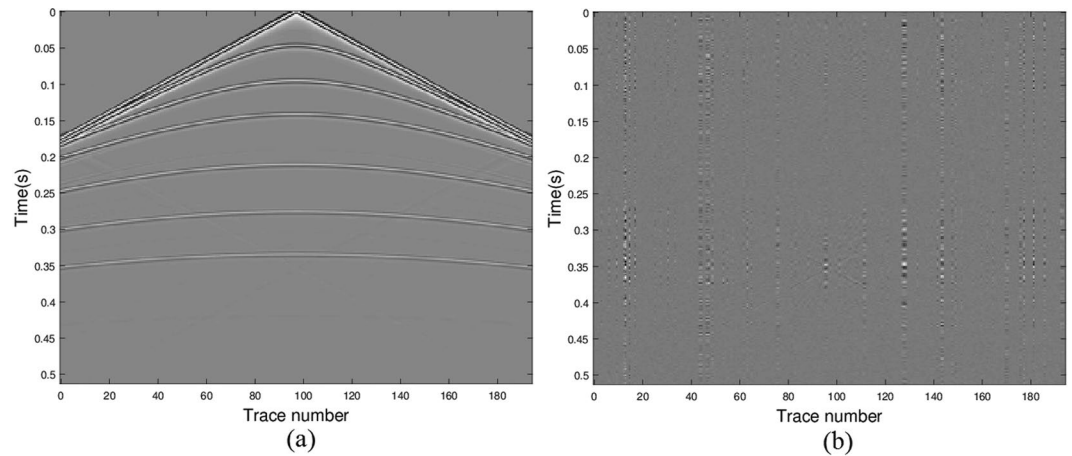

Figure 4. Seismic data results via SPG for l_1-regularized least square minimization: (**a**) the recovered data; (**b**) error between the original and the recovered signals.

$$\rho_t = \frac{\left\| \hat{M}_{t+1} - M_t \right\|_F}{\left\| M_t \right\|_F}.$$

Experimental results

To validate the efficiency of the joint matrix minimization approach and the unified algorithm for the problem (11), we perform three tests: (1) restoration of the input one-dimensional random signal with the randomly

Figure 5. Seismic data results in noisy case via model (11) with $q = 2$ and $p = 0.5$: (**a**) the recovered data; (**b**) error between the original and the recovered signals.
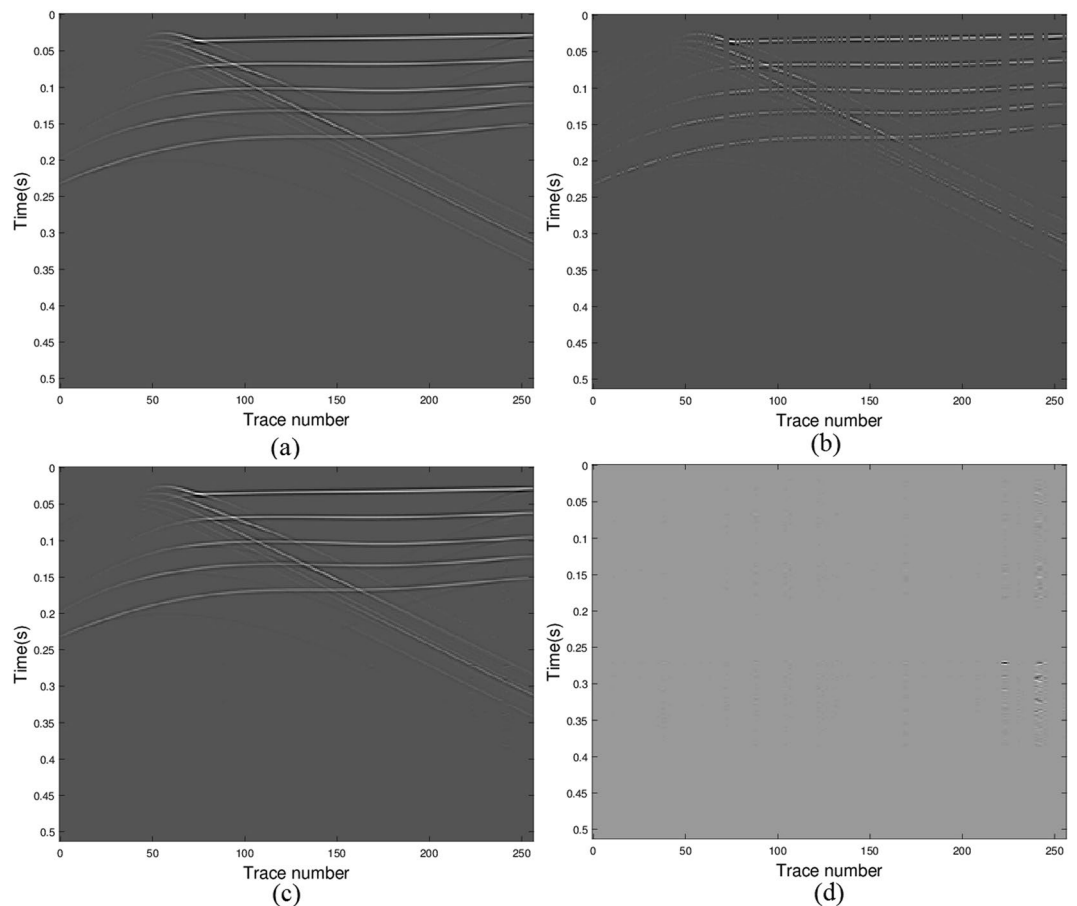
Figure 6. Seismic data results via model (11) with $q = 2$ and $p = 0.5$: (**a**) the real data; (**b**) the data with missing traces; (**c**) the recovered data; (**d**) error between the original and the recovered signals.

generated matrix L; (2) restoration of the synthetic seismic data with random loss of traces; (3) restoration of the field data.

One-dimensional signal reconstruction. We randomly take samples to generate the matrix L. For implementation, we try to restore the signal by the model (11) with $q = 2$ and $p \in (0, 1]$.
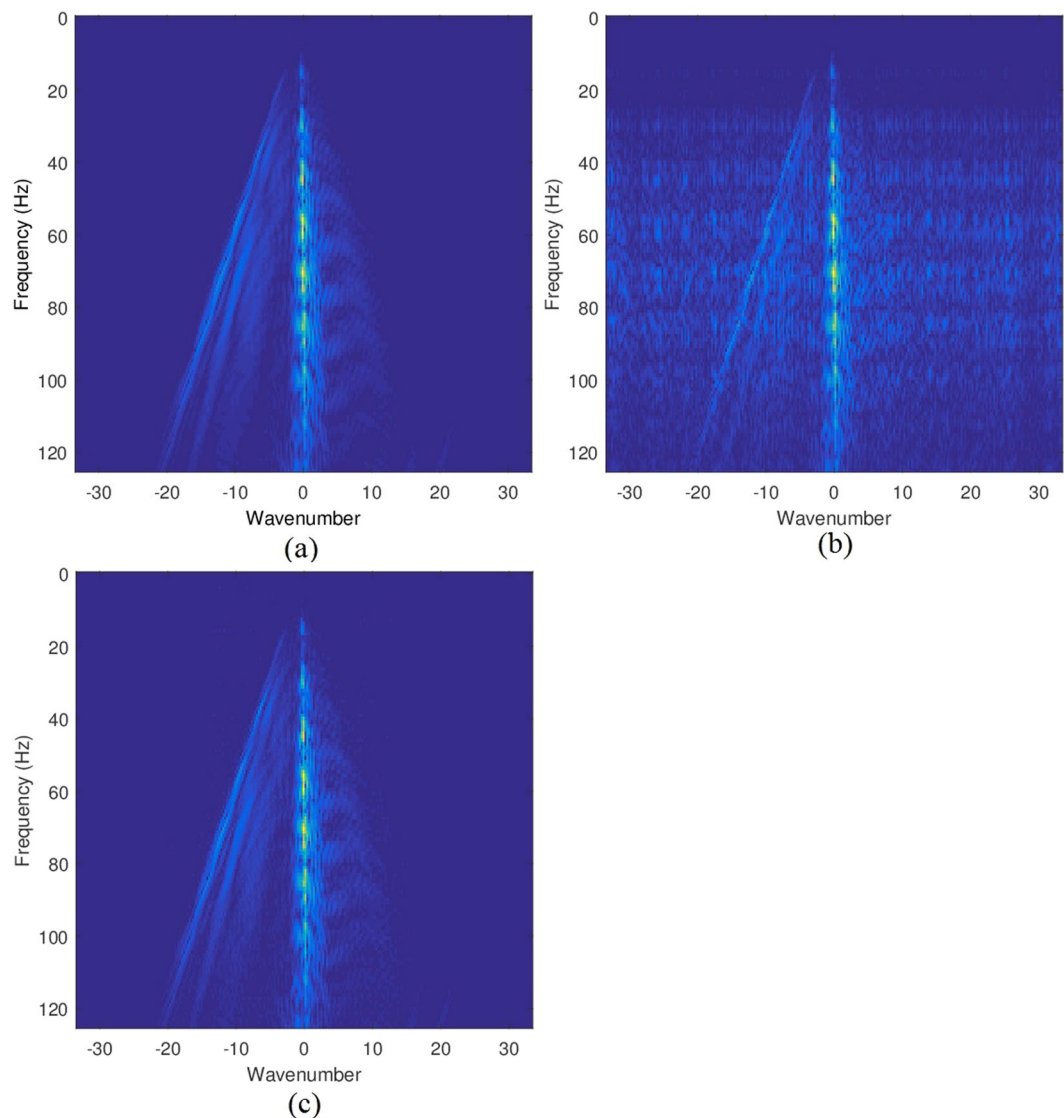
Figure 7. Frequency information: (**a**) frequency of the original data; (**b**) frequency of the sub-sampled data; (**c**) frequency of the restored data.

The stopping precision in Algorithm 3 is set to $\epsilon = 10^{-3}$. The sparse parameter p and regularization parameter α_k are typically chosen in (0, 1]. Results for other values of p are similar. The relative error of the recovered signal M_{rec} to the true (given) signal M_{true} is defined by

$$\text{err}_{\text{rel}} = \frac{\|M_{\text{rec}} - M_{\text{true}}\|_2}{\|M_{\text{true}}\|_2}.$$

To quantify the results, we define the signal-to-noise ratio (SNR) as $\text{SNR} = 10\log_{10}\frac{\|d_{\text{org}}\|_2^2}{\|d_{\text{org}} - d_{\text{rec}}\|_2^2}$, where d_{org} refers to the original data and d_{rec} is the restored data.

For the one-dimensional case, the matrix M is reduced to a vector, hence the unified Algorithm 3 can be used for solving (11). For comparison, we also apply spectral projected gradient (SPG) method[49] to solve the l_1-regularization problem. The code of SPG is downloaded from http://www.cs.ubc.ca/~mpf/spgl1/index.html. Two algorithms are carried out in the same environment and choose their best regularization parameters. The comparison items include err$_{\text{rel}}$ value, SNR and CPU running time (second). Each experiment is repeated five times and the average values are reported in Table 1. It indicates that both methods perform well for one-dimensional signal reconstruction problem.

Apart from the regular data, we also consider the noisy cases to show the robustness of two methods. Different noise levels are added to the simulated data. Noise level 0.001 means the noise is randomly generated with zero mean and 0.001 variance. The results of Algorithm 3 with sparse parameters $p = 1$ and $p = 0.5$ are displayed in Table 1. Compared with the l_1-regularized minimization model, the half-norm regularized minimization behaves better in
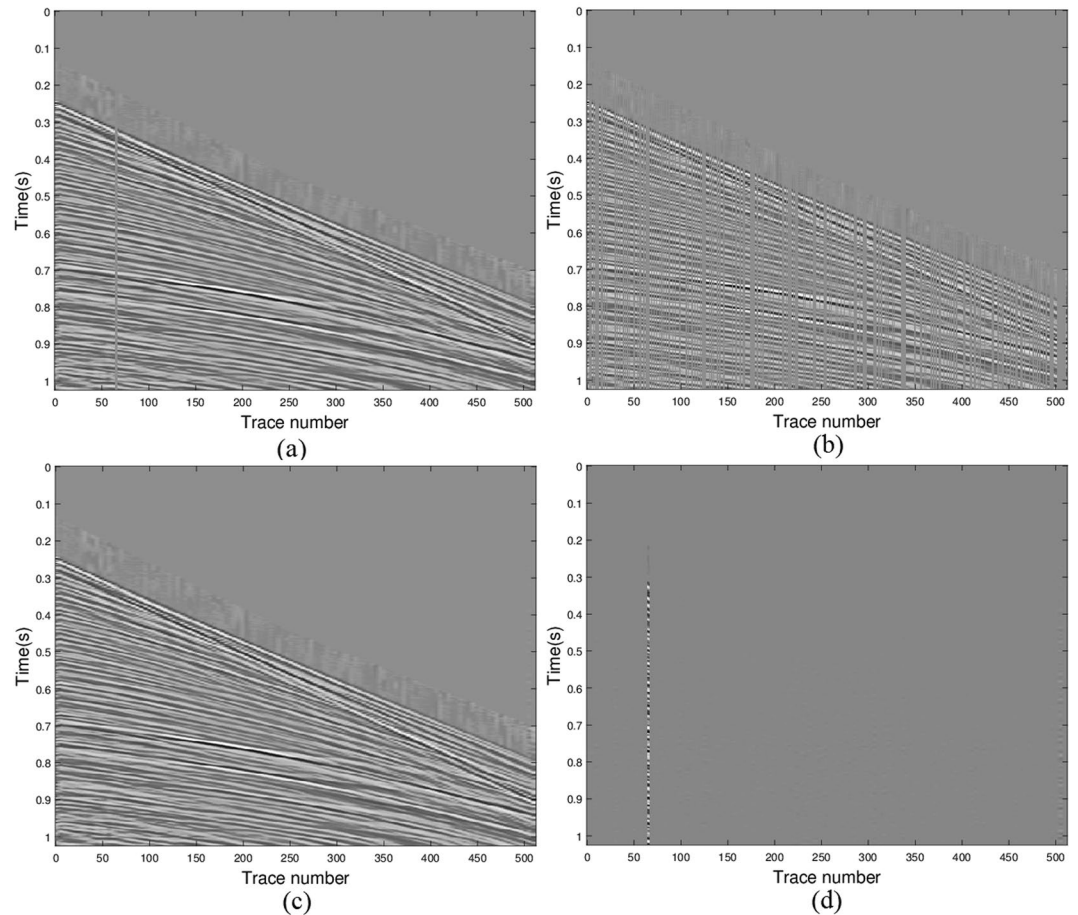
Figure 8. Seismic data results via model (11) with $q = 2$ and $p = 0.5$: (**a**) the real data, (**b**) the data with missing traces, (**c**) the recovered data and (**d**) error between the original and the recovered signals.

reconstruction. Figure 2 plots the recovery performance of the Algorithm 3 with $p = 0.5$ on noisy data. Figure 2(a) is the comparison of the real signal and the recovered signal, Fig. 2(b) illustrates the difference between the recovered signal and the input (true) signal. The recovery images of other cases are similar. The figures reveal that our model and algorithm perform well for one-dimensional seismic wavefield reconstruction problem even in noisy cases.

Reconstruction of seismograms from a layered model. Now we consider a seismogram generated from a seven layers geologic velocity model where the spatial sampling interval is 15 meters and the time sampling interval is 0.002 second. The velocity varies from 2500 m/s to 5500 m/s. The seismogram is generated using a source function given by a Ricker wavelet with central-frequency of 25 Hz. The dataset contains 256 traces with 256 time samples in each trace. Different percentages of missing traces in original data, 10%, 25% and 50%, are used to test the limitation of recovery methods. The joint matrix model (11) with Algorithm 3 is applied to reconstruct the seismic wavefield. Since the spectral projected gradient method only solves an l_1-regularized vector minimization problem, we decompose the matrix representation system (11) into the l_1-regularized vector minimization problem. Each column is considered as a subproblem to reconstruct its weight vector separately. Then all the solutions of the subproblems are sequentially aligned into a weighted matrix to evaluate the reconstruction performance. The experimental results on missing percentages 10% and 25% can be seen in Tables 2 and 3.

As for the data without noise but missing 50% traces, the reconstruction performance of joint matrix model with Algorithm 3 is much worse than missing percentages of 10% and 25%. The err_{rel} value is 0.5414 and SNR is around 5.1904dB, almost the same for any $p \in (0, 1]$. These results mean that our method may not completely recover the seismic wavefield well if the missing trace signals are more than 50%. Actually, the sub-sampled data missing 50% itself is a failed collection of seismic recodes.

The original shot gathers are shown in Fig. 3(a). The data with 25% traces missing are shown in Fig. 3(b). In forming the under-determined matrix L, a Haar wavelet orthogonal base is used to form the transform matrix Ψ. The unified Algorithm 3 is applied to solve the joint matrix minimization problems (11) with $q = 2$ and typical parameters $p \in (0, 1]$. Good recovery performance is observed and the result is demonstrated in Fig. 3(c). The error of the original and the recovered data shown in Fig. 3(d) illustrates the efficient recovery performance of joint matrix minimization approach. In displaying the results, the amplitude scale of the error map is the same as

the amplitude scale of the data. Of course, other values of the sparse parameter p can be chosen, the results in visualization are similar. So, we only list the quantitative results in Tables 2 and 3.

Comparatively, the recovery image of the SPG algorithm for the case of 25% traces missing is presented in Fig. 4. Figure 4(a) is the reconstruction and Fig. 4(b) displays the difference between the original and reconstructed seismic signals. It is noticed that SPG algorithm for the l_1-regularization vector minimization restores the seismic wavefield as accurate as the joint matrix approach with Algorithm 3. These results are obtained using the same code from http://www.cs.ubc.ca/~mpf/spgl1/index.html.

To show the anti-noise property of our algorithm, we add random noise with noise level 0.001 to the simulated data. The unified Algorithm 3 is applied to solve the joint matrix minimization problems. The err_{rel} value, SNR and CPU running time (second) are listed in Table 2 for 3 sparse parameters. The recovery image and the error of the original and the recovered data are shown in Fig. 5(a and b) respectively. The low relative error and high SNR indicate that our algorithm is stable for seismic data restoration.

To save memory requirement of large-scale data, we have observed the restoration behavior of our method on patch of the input synthetic data. We evenly partition the collection of trace signals D into several blocks, such as $D = [D_1, D_2, …, D_f]$, where $D_g \in R^{m \times l_g}$ and $\sum_{g=1}^{f} l_g = l$. Each D_g is input separately to recover the seismic signals by system (11). Then all the sub-solutions $M_g, g = 1, 2, …, f$ are combined into $M = [M_1, M_2, …, M_f]$. When the number of segments is two or three, the recovered err_{rel} values and SNR are almost the same as the integral case. When each column is considered as a segment, the joint matrix model is reduced to a sequence of vector recoveries, the recovery err_{rel} values and SNR are similar to the integral case but the computational time is around 50 times more.

Reconstruction of seismograms from a heterogeneous model.

Next we consider a seismogram generated from a velocity model varying both vertically and transversely (Wang et al.[5]). The original seismic wavefield, sub-sampled data (37% traces are randomly removed) and recovered data are shown in Fig. 6(a–c), respectively. The difference of the original data and the recovered data is illustrated in Fig. 6(d). In displaying the results, the amplitude scale of the error map is the same as the amplitude scale of the data. It illustrates that all the initial seismic energy is recovered with minor errors. Though the reconstruction is not perfect, most of the details of the wavefield are preserved. Again, to test the quality of our algorithm in seismic data restoration for complex structure, we calculate the signal-to-noise ratio and the relative error. From our calculation, for $p = 0.5$, the values of SNR and err_{rel} are 26.9792 and 0.0448, respectively; for $p = 1$, the values of SNR and err_{rel} are 25.6940 and 0.0519, respectively. The high value of SNR and low value of err_{rel} indicate our algorithm works for seismic data restoration even with complex structure.

To show the robustness of our algorithm to interference, we add random noise with level 0.001 and 0.01 to the simulated data respectively. The unified Algorithm 3 with $p = 0.5$ is applied to solve the joint matrix minimization problems. The values of SNR and err_{rel} for noise level equaling 0.001 are 26.9074 and 0.0451, and for noise level equaling 0.001 are 18.0355 and 0.1254, respectively.

In the noisy case, e.g., noise level equaling 0.01, the frequency information of the original data, sub-sampled data and the recovered data are shown in Fig. 7(a–c), respectively. Again, the aliasing of the sub-sampled data is reduced greatly in the recovered data.

Field data.

Finally, we examine the efficiency of the new method with field data. The seismic data is a marine shot gather shown in Fig. 8(a) which consists of 256 traces with spacing 25 m and time sampling interval 2 ms. There are damaged traces in the original gather. The subsampled gather is shown in Fig. 8(b) with 42% of the original traces randomly removed. This sub-sampled gather was used to restore the original gather with suitable solution methods. Again, the unified Algorithm 3 is applied to solve the joint matrix minimization problems (11) with $q = 2$ and $p = 0.5$. The recovery result is demonstrated in Fig. 8(c). The error of the original and the recovered data shown in Fig. 8(d) illustrates the efficient recovery performance of joint matrix minimization approach. In displaying the results, the amplitude scale of the error map is the same as the amplitude scale of the data. Comparing the subsampled image with the original image, the restored image can reconstruct most of the details. In addition the damaged trace in the original gather was restored as a good trace. Using the same definition of SNR as above, for $p = 0.5$, the value of SNR equals 19.7301; for $p = 1$ the value of SNR equals 19.7919. We only show figures for $p = 1$, since in visualization the results are similar for $p = 0.5$.

The frequency information of the original data, sub-sampled data and the recovered data are shown in Fig. 9(a–c), respectively. It indicates that the aliasing of the sub-sampled data is reduced greatly in the recovered data.

Conclusion

Sparse optimization has broad applications in seismic data processing. In this paper we focus on data restoration problem. Noticing that the seismic wavefield can be represented using matrix instead of vector as weight variable to express all the signals simultaneously, in this paper we propose a matrix optimization model to the seismic wavefield recovery. We first reformulate the data restoration problem using an $l_{2,p}$-norm constrained matrix minimization model for any $p \in (0, 1]$, which is a nonconvex and non-Lipschitz continuous minimization problem. Then we develop a unified algorithm to solve the mixed matrix optimization problem for any $p \in (0, 1]$. Convergence analysis of the new algorithm is also addressed. Numerical results on synthetic problems and the field data example indicate potential usage of our method for practical applications.
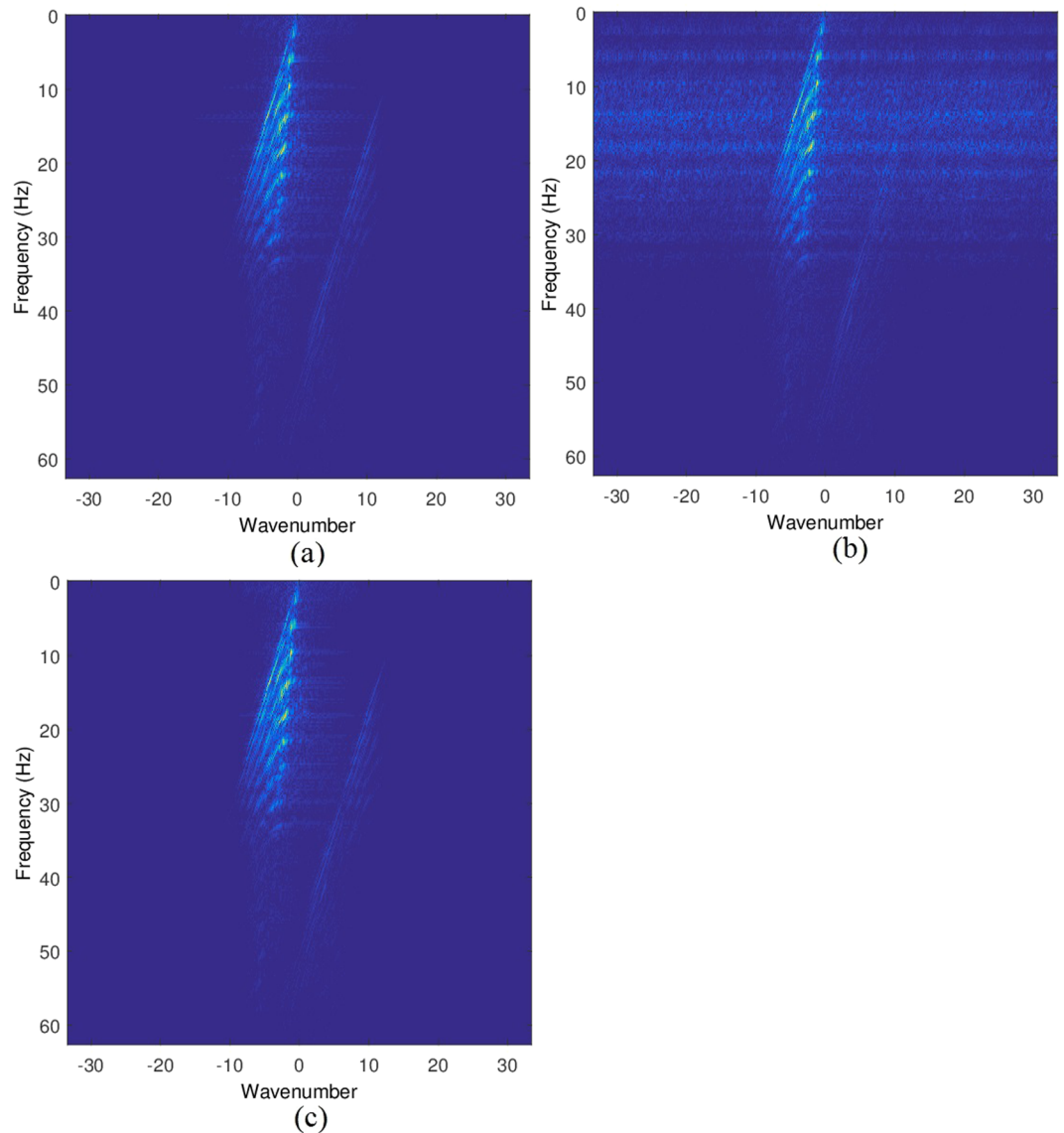
Figure 9. Frequency information: (**a**) frequency of the original data; (**b**) frequency of the sub-sampled data; (**c**) frequency of the restored data.

Appendix Properties of the new algorithms. In this section, we will analyze the convergence property of the Algorithm 2. The main result is that the objective function $J(M_t)$ strictly decreases with respect to iterations until the matrix sequence $\{M_t\}$ converges to a stationary point of $J(M)$.

Proposition 1. Let $\varphi(\tau) = \tau - a\tau^{\frac{1}{a}}$ be a function of the variable τ, where $a \in (0, 1)$. Then for any $\tau > 0$, $\varphi(\tau) \leq 1 - a$, and $\tau = 1$ is the unique maximizer.

To verify the above statements, let us take the derivative of $\varphi(\tau)$ and set it to be zero, that is

$$\varphi'(\tau) = 1 - \tau^{\frac{1}{a}-1} = 0,$$

then $\varphi'(\tau) = 0$ has the unique solution $\tau = 1$ for any $a \in (0, 1)$ which is just the maximizer of $\varphi(\tau)$ in $(0, +\infty)$.

Based on Proposition 1, for a given $a \in (0, 1)$,

$$\tau - a\tau^{\frac{1}{a}} \leq 1 - a \tag{A-1}$$

holds for $\tau \in (0, +\infty)$ and "=" is active if and only if $\tau = 1$. Let a takes special values such as $a = \frac{p}{2}$ ($p \in (0, 1]$), the inequality (A-1) will result in the following formula associated with $||M||_{2,p}^{p}$ ($0 < p \leq 1$).

Proposition 2. Suppose that M_t and $\hat{M}_{t+1}$ are generated in the t-th iteration by Algorithm 2, the following inequality holds,

$$\|\Lambda_t \hat{M}_{t+1}\|_{2,p}^p - \frac{p}{2}\sum_{k\in\Omega_t}\frac{\alpha_k\|\hat{m}_{t+1}^k\|_2^2}{\|m_t^k\|_2^{2-p}} \leq \left(1 - \frac{p}{2}\right)\|\Lambda_t M_t\|_{2,p}^p, \quad p\in(0,1\Big],$$
(A−2)

where $\Lambda_t = diag\{\alpha_k\}_{k\in\Omega_t}$. Moreover, the equality in (A-2) holds if and only if $\|\hat{m}_{t+1}^k\|_2 = \|m_t^k\|_2$ for $k\in\Omega_t$.

Consider the approximate value M_t. Since each $\|m_t^k\|_2 \neq 0$ for $k\in\Omega_t$, so we can r $\tau = \frac{\|\hat{m}_{t+1}^k\|_2^p}{\|m_t^k\|_2^p}$ and $a = \frac{p}{2}$ in Proposition 1. It can be obtained that

$$\frac{\|\hat{m}_{t+1}^k\|_2^p}{\|m_t^k\|_2^p} - \frac{p}{2}\frac{\|\hat{m}_{t+1}^k\|_2^2}{\|m_t^k\|_2^2} \leq 1 - \frac{p}{2}, \quad k\in\Omega_t \ .$$
(A−3)

Multiplying equation (A-3) by $\alpha_k\|m_t^k\|_2^p$, we have the following inequality

$$\alpha_k\|\hat{m}_{t+1}^k\|_2^p - \frac{p}{2}\frac{\alpha_k\|\hat{m}_{t+1}^k\|_2^2}{\|m_t^k\|_2^{2-p}} \leq \left(1 - \frac{p}{2}\right)\alpha_k\|m_t^k\|_2^p, \quad k\in\Omega_t.$$
(A−4)

Summing up $k\in\Omega_t$ in formula (A-4}), we can derive at (A-2).

Based on Proposition 1, $\tau=1$ is the unique minimizers for $\varphi(\tau)$ in $(0,\ +\infty)$ when $a = \frac{p}{2}$. Namely, $\|\hat{m}_{t+1}^k\|_2 = \|m_t^k\|_2 (k\in\Omega_t)$ is necessary and sufficient for equality holding in (A-4). Now, we can establish the following convergence property of the Algorithm 2.

Proposition 3. Suppose that $\{M_t\}$ is the matrix sequence generated by Algorithm 2. Then $J(M_t)$ strictly decreases with respect to t for any $0 < p \leq 1$ until $\{M_t\}$ converges to a stationary point of $J(M)$.

Proposition 4. Based on the derivation of Proposition 3, so long as the subproblem (16) is solved with $Q_t(\hat{M}_{t+1}) \leq Q_t(M_t)$, the convergence of Algorithm 3 will be guaranteed for any $p\in(0,1]$.

References

1. Liu, B. & Sacchi, M. D. Minimum weighted norm interpolation of seismic records. *Geophysics* **69**, 1560–1568 (2004).
2. Soubaras, R. Spatial interpolation of aliased seismic data. *Expanded Abstracts*, *74th Annual Meeting SEG*, Denver, USA (Denver, Oct. 2004), 1167–1170 (2004).
3. Naghizadeh, M. & Sacchi, M. D. Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data. *Geophysics* **75**, WB189–202 (2010).
4. Sacchi, M. D. & Liu, B. Minimum weighted norm wavefield reconstruction for AVA imaging. *Geophysical Prospecting* **53**, 787–801 (2005).
5. Wang, Y. F., Stepanova, I. E., Titarenko, V. N. & Yagola, A. G. *Inversion problems in geophysics and solution methods* (Higher Education Press, Beijing, 2011).
6. Wang, Y. F., Cao, J. J. & Yang, C. C. Recovery of seismic wavefields based on compressive sensing by l_1 an-norm constrained trust region method and the piecewise random subsampling. *Geophys. J. Int.* **187**, 199–213 (2011).
7. Wang, Y. F., Yang, C. C. & Cao, J. J. On Tikhonov regularization and compressive sensing for seismic signal processing. *Mathematical Models and Methods in Applied Sciences.* **22**, 1150008-1–1150008-24 (2012).
8. Sacchi, M. D. & Ulrych, T. J. Estimation of the discrete Fourier transform, a linear inversion approach. *Geophysics* **61**, 1128–36 (1996).
9. Sacchi, M. D., Ulrych, T. J. & Walker, C. J. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE Transactions on Signal Processing* **46**, 31–38 (1998).
10. Duijndam, A. J. W. & Schonewille, M. A. Non-uniform fast Fourier transform. *Geophysics* **64**, 539–551 (1999).
11. Xu, S., Zhang, Y., Pham, D. & Lambare, G. Anti-leakage Fourier transform for seismic data regularization. *Geophysics* **70**, V87–V95 (2005).
12. Trad, D., Ulrych, T. & Sacchi, M. Accurate interpolation with high-resolution time-variant Radon transforms. *Geophysics* **67**, 644–656 (2002).
13. Herrmann, F. J. & Hennenfent, G. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International* **173**, 233–248 (2008).
14. Sacchi, M. D., Verschuur, D. J. & Zwartjes, P. M. Data reconstruction by generalized deconvolution. *Expanded Abstracts 74th Annual Meeting SEG*, Denver, USA (Denver, Oct. 2004), 1989–1992 (2004).
15. Wang, Y. F., Cao, J. J., Yuan, Y. X., Yang, C. C. & Xiu, N. H. Regularization active set method for nonnegatively constrained ill-posed multichannel image restoration problem. *Appl. Opt.* **48**, 1389–1401 (2009).
16. Wang, Y. F. Sparse optimization methods for seismic wavefields recovery. *Proc. Inst. Math. Mech.* **18**(1), 42–55 (2011).
17. Cao, J. J. & Wang, Y. F. Seismic data restoration with a fast L1 norm trust region method. *J. Geophys. Eng.* \bf 11(4), 045010 (2015).
18. Chen, S., Donoho, D. & Saunders, M. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* **20**, 33–61 (1998).
19. Tropp, J. A. & Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory* **53**, 4655–4666 (2007).
20. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal Royal Statistical Society B* **58**, 267–288 (1996).
21. Kim, S.-J., Koh, K., Lustig, M., Boyd, S. & Gorinevsky, D. An interior-point method for large-scale l_1-regularized least squares. *IEEE Journal on Selected Topics in Signal Processing* **1**, 606–617 (2007).
22. Dai, Y. H. & Fletcher, R. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. *Numerische Mathematik* **100**, 21–47 (2005).
23. Wang, Y. F. & Ma, S. Q. Projected Barzilai-Borwein methods for large scale nonnegative image restorations. *Inverse Problems in Science and Engineering* **15**, 559–583 (2007).
24. Figueiredo, M. A. T., Nowak, R. D. & Wright, S. J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing* **1**, 586–597 (2007).

25. Ewout, V. B. & Michael, P. F. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* **31**, 890–912 (2008).

26. Cao, J. J., Wang, Y. F. & Wang, B. F. Accelerating seismic interpolation with a gradient projection method based on tight frame property of curvelet. *Exploration Geophysics* **46**, 253–260 (2015).

27. Rakotomamonjy, A., Flamary, R., Gasso, G. & Canu, S. l_p-l_p Penalty for sparse linear and sparse multiple kernel multitask learning. *IEEE Transactions on Neural Networks* **22**(8), 1307–1320 (2011).

28. Suvrit, S. Fast projection onto $l_{1,q}$-norm balls for grouped feature selection. *Proceeding of Machine Learning and Knowledge Discovery in Databases*, Athens, Greece (2011).

29. Sumit, S., Vishal, M. P., Nasser, M. N. & Rama, C. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Trans. PAMI* **36**(1), 113–126 (2014).

30. Wang, H., Nie, F. P., Cai, W. D. & Huang, H.Semi-supervised robust dictionary learning via efficient $l_{2,0+}$-norms minimizations. *IEEE International Conference on Computer Vision*, 1145–1152 (2013).

31. Aravkin, A., Kumar, R., Mansour, H., Recht, B. & Herrmann, F. J. Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation. *SIAM J. Sci. Comput.* **36**(5), S237–S266 (2014).

32. Kumar, R. *et al.* Efficient matrix completion for seismic data reconstruction. *Geophysics* **80**(5), V97–V114 (2015).

33. Rodriguez, I. V., Sacchi, M. D. & Gu, Y. J. Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory. *Geophys. J. Int.* **188**, 1188–1202 (2012).

34. Kreimer, N., Stanton, A. & Sacchi, M. D. Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction. *Geophysics* **78**(6), V273–V284 (2013).

35. Candés, E. J., Wakin, M. B. & Boyd, S. P. Enhancing sparsity by reweighed l_1 minimization. *Journal of Fourier Analysis and Applications* **14**(5), 877–905 (2008).

36. Chartrand, R. Exact reconstructions of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters* **14**(10), 707–710 (2007).

37. Chartrand, R. & Yin, W. Iteratively reweighed algorithms for compressive sensing. *33rd International Conference on Acoustics, Speech, and Signal Processing*, 3869–3872 (2008).

38. Chen, X. J., Xu, F. M. & Ye, Y. Y. Lower bound theory of nonzero entries in solutions of l_2-l_p minimization. *SIAM J. Scientific Computing* **32**(5), 2832–2852 (2010).

39. Xu, Z. B., Zhang, H., Wang, Y., Chang, X. Y. & Liang, Y. $L_{1/2}$ regularizer. *Science in China (Series F).* **52**(6), 1159–1169 (2010).

40. Chen, X. J., Ge, D. D., Wang, Z. Z. & Ye, Y. Y. Complexity of unconstrained L_2-L_p minimization. *Math. Program. (Ser. A)* **143**, 371–383 (2014).

41. Chen, X. J. & Zhou, W. J. Convergence of the reweighted l_1 minimization algorithm for l_2-l_p minimization. *Computational Optimization and Applications* **59**, 47–61 (2014).

42. Lu, Z. S. Iterative reweighted minimization methods for regularized unconstrained nonlinear programming. *Mathematical Programming* **147**, 277–307 (2014).

43. Wang, L. P., Chen, S. C. & Wang, Y. P. A unified algorithm for mixed $l_{2,p}$-minimizations and its application in feature selection. *Computational Optimization and Applications* **58**, 409–421 (2014).

44. Yuan, Y. X. *Numerical Methods for Nonlinear Programming* (Shanghai Science and Technology Publication, Shanghai, 1993).

45. Dai, H. *Matrix Theory* (Science Press, Beijing, 2004).

46. Golub, G. H. & Loan, C. F. *Matrix Computation* (The Johns Hopkins University Press (3rd), 1996).

47. Barzilai, J. & Borwein, J. M. Two-point step size gradient methods. *IMA Journal of Numerical Analysis* **8**, 141–148 (1988).

48. Wang, Y. F. & Yang, C. C. Accelerating migration deconvolution using a non-monotone gradient method. *Geophysics* **75**, S131–S137 (2010).

49. van den, Berg, E. & Friedlander, M. P. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* **31**(2), 890–912 (2008).

Acknowledgements

Author Contributions

Y.F. designed the study. Y.F. and W. conducted experiments. Y.F. and W. wrote the paper. All authors contributed to synthetic data interpretation and provided significant input to the final manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.