



Fast ℓ_1 -regularized Radon transforms for seismic data processing



Ali Gholami*, Toktam Zand

Institute of Geophysics, University of Tehran, Tehran, Iran

ARTICLE INFO

Article history:

Available online 11 September 2017

Keywords:

Sparse Radon transform
Parabolic Radon transform
Chirp z-transform
Tau-p transform
Slant-stack
Seismic multiple attenuation

ABSTRACT

The efficiency of the recently developed algorithms for sparse Radon transforms depends heavily on the efficiency of the inverse transformation and its adjoint. In this paper, we propose a fast algorithm for the implementation of these canonical transforms, which runs in complexity $O(N_f N \log N + N_x N_t \log N_t)$ for a signal of size $N_t \times N_x$, as opposed to $O(N_f N_x N_p + N_x N_t \log N_t)$ of the direct computation, where N depends on the maximum frequency and offset in the data set, the maximum curvature in the Radon space, and N_p . These transforms are utilized within the split Bregman iteration to solve the sparse Radon transform as an ℓ_2/ℓ_1 optimization problem. The computations involved in each iteration of the proposed algorithm, are carried out by a fast Fourier transform (FFT) algorithm. Furthermore, in the new algorithm, the amount of regularization is controlled by iteration number. We obtain a good estimate of predictive error via the generalized cross validation (GCV) analysis and automatically determine the optimum number of iterations at a negligible cost, thus leading to an automatic fast algorithm. This allows a time-domain sparse Radon transformation of large-scale data that are otherwise intractable. Numerical tests using simulated data confirm high efficiency of the proposed algorithm for processing large-scale seismic data.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Amongst the many transforms used within the seismological community, the time-invariant Radon transforms, e.g., linear Radon transform (LRT) and parabolic Radon transform (PRT), found special attention for use in a wide variety of applications.

Mathematically, the LRT was first introduced in 1917 by [1]. It has found successful applications in seismic data processing under the names of $\tau - p$ and slant stack transform, whereby, the data in the original observation time versus offset ($t - x$) domain are mapped into a new space of intercept time τ and horizontal ray parameter p , by summing the data along straight lines defined by these parameters; i.e. $t = \tau + px$. The $\tau - p$ transform was first used by [2], in a continuous form, for directional decomposition of seismic data and then by [3], in a discrete form, for seismic migration. Since then, much effort has been made by various researchers on the sampling and aliasing effects [4–6] and the invertibility issue [7–12]. Applications of the LRT have been well documented in seismic signal processing for, e.g., plane-wave decomposition and ground roll attenuation [4,13], multiple attenuation based on predictive deconvolution [14], inversion of common midpoint data [4],

and Rayleigh-wave mode separation and group velocity determination [15,16].

The PRT is also recognized as an accepted tool in seismic data processing. It was first introduced in [17], whereby, the data in the original $t - x$ domain, are mapped into a new space of intercept time τ and curvature parameter p , by summing the original data along parabolic paths defined by these parameters; i.e. $t = \tau + px^2$. The transform is mainly used to circumvent the computational load of hyperbolic Radon transform, defined by paths $t = \sqrt{\tau^2 + px^2}$ [18], either by an approximate normal-move-out (NMO) correction applied to the data [17] or by stretching the data along the time axis [19]. The transform has been used for, e.g., multiple attenuation [17,20,21] and offset interpolation [22–24].

The applicability of the Radon transforms discussed above rely heavily on the focusing of seismic events in the Radon domain. The discretization effects due to sampling and truncation effects due to finite spatial aperture of seismic data however causes the Radon basis functions to be highly correlated. Therefore, a simple adjoint solution causes a data loss that is not trivial. The conventional (non-sparse) Radon transforms are cast as a damped least-squares problem [17,9–12], requiring an extra computation step over the adjoint transform. A damped solution is small in the sense of Euclidean norm, having many unnecessary non-zero coefficients which decrease the resolution in the RT domain. The resolution is further improved by using more sophisticated but time-

* Corresponding author.

E-mail addresses: agholami@ut.ac.ir (A. Gholami), toktam.zand@ut.ac.ir (T. Zand).

consuming regularization tools. [18] employed a time-domain regularization to achieve a solution having minimum entropy (maximum sparseness) in velocity stack and slant stack inversion. [23] used a time stretching operator [19] to speed up the algorithm in [18] using PRT in the frequency-domain. To increase the efficiency, they applied the sparsity just along the curvature axis, through an iteratively re-weighted least squares (IRLS) algorithm [25,26]. The algorithm was further improved in [27] using a combination of conjugate-gradient (CG) and fast Fourier transform (FFT) algorithms in solving each re-weighted system.

Traditional algorithms [27] for the sparse RT are one-dimensional (1D) and perform in the frequency domain; that means, the sparsity is just forced along the p axis [23]. Furthermore, such algorithms are mainly based on IRLS methods and CG, and hence can be slow for processing large data sets. Another challenge in regard with sparse Radon transforms arises when dealing with noisy real world data. A balance between the amount of sparsity and fitness to data should be controlled by a regularization parameter. The choice of an appropriate value of regularization parameter involves performing the transform with multiple, and often many, values of the parameter. This adds an extra computation to the problem solution and is a common difficulty in performing all RTs [28].

In this paper, we develop a fast time-domain algorithm to calculate a sparse solution to RTs. The RT is formulated as an optimization problem, where the sparsity in the Radon plane is forced, in the sense of the ℓ_1 -norm, while requiring the inverse Radon transformation to match the observed data up to a predefined threshold, a problem known as basis pursuit denoising (BPDN) [29]. We employ the split Bregman iteration [30] to effectively and efficiently solve the generated optimization problem. We have three contributions to the computational efficiency of the Bregman iterations in solving the proposed sparse RTs:

First, a fast algorithm is constructed to effectively evaluate the inverse Radon transformation and its adjoint. The numerical complexity of the new algorithm is $O(N_f N \log N + N_x N_t \log N_t)$ for a signal of size $N_t \times N_x$, as opposed to $O(N_f N_x N_p + N_x N_t \log N_t)$ of the direct computation. Where N_f is the number of inverted frequencies, N_p is the number of Radon parameters, and N depends on the maximum frequency and offset in the data set, the maximum curvature in the Radon space, and N_p . The idea is constructing a super operator, which covers the ranges of all frequencies and Radon parameters. This frequency-independent super operator is used across all frequencies, and then the desired coefficients are selected from the computed results. The advantage of this approach is that the operator is extremely compressible in the Fourier domain and thus its implementation in that domain is very efficient. Furthermore, the desired coefficients can be directly computed by the chirp z -transform [31], which can be carried out by an FFT algorithm. Numerical tests show that the new algorithm can lead to speedup of several orders of magnitude.

Second, a fast strategy, of complexity $O(N_f N_p \log N_p)$, is also proposed to solve the normal equations arising in each iteration of the Bregman algorithm. The normal equations are exactly what appear in the least-squares (LS) Radon transform, which involves solving a set of N_f Toeplitz systems each of size $N_p \times N_p$. Traditional fast algorithms like Levinson's [32] for solving such a set of Toeplitz systems requires $O(N_f N_p^2)$ operations. A main property of the Bregman iteration for sparse RT is that it is still convergent when an approximate solution to the normal equations is used [33,30]. We thus approximate each Toeplitz system with the best circulant approximation of it in the sense of Frobenius norm. This approximation allows solving the replaced circulant systems with the FFT algorithm of computational complexity of $O(N_f N_p \log N_p)$. The solution to the resulting circulant systems is then used as approximate solution to the original Toeplitz systems.

Third, in the developed algorithm the amount of regularization is controlled by iteration number. We obtain a good estimate of predictive error via generalized cross validation (GCV) analysis and automatically determine the optimum number of iterations at a negligible cost. No parameter tuning is thus required in the new algorithm. This makes the proposed algorithm suitable for analyzing large-scale data. Numerical tests show very high performance of the new algorithm for calculating sparse solution to LRT and PRT.

2. Time-invariant Radon transform

In this section, we briefly review the basic formulations of the time-invariant RTs in two and three dimensions. We also discuss how our proposal can contribute in fast solution of sparse RTs via the split Bregman iteration.

2.1. Two-dimensional Radon transform

Let $\mathbf{u} \in \mathbb{R}^{N_t \times N_x}$, a discrete function of offset and time, be a two-dimensional (2D) seismic signal, including N_x traces (columns), each having N_t time samples (rows). Then the time-invariant RT models the signal as superposition of objects defined by events $t = \tau + p\varphi(x)$ as:

$$\mathbf{u}(t, x) = \sum_p \mathbf{v}(\tau = t - p\varphi(x), p) \quad (1)$$

where t is the time, x represents the offset (distance between the source and receiver), $\mathbf{v}(\tau, p)$ denotes the RT at intercept time τ and Radon parameter p . We assume that the Radon space is uniformly sampled into $N_\tau \times N_p$ bins and hence $\mathbf{v} \in \mathbb{R}^{N_\tau \times N_p}$. The strictly increasing function $\varphi(x)$ determines the shape of the objects defining certain seismic events which could be focused in the transform domain. Two main choices for $\varphi(x)$ are x (LRT) and x^2 (PRT). Here, we assume that the range of the Radon parameter is properly selected and sampled [10,34] and thus focus on the solution of (1). Let $\mathbf{x} = \text{vec}(\mathbf{x})$ where the vec operator vectorizes a matrix \mathbf{x} by stacking its columns, then (1) in matrix-vector form can be written as

$$\mathbf{u} = \mathbf{R}\mathbf{v} \quad (2)$$

where \mathbf{R} is the inverse RT operator. The adjoint solution to (2), defined as

$$\mathbf{v} = \mathbf{R}^T \mathbf{u} \quad (3)$$

or

$$\mathbf{v}(\tau, p) = \sum_x \mathbf{u}(t = \tau + p\varphi(x), x) \quad (4)$$

is known as low-resolution RT and suffers from smearing of energy in the Radon panel. The most commonly used solution to (2), other than (4), is to minimize the misfit between observed and predicted data subject to a constraint on the solution. The standard (non-sparse) RT is obtained by solving (2) via a damped LS approach [17,9–12]:

$$\mathbf{v} = (\mathbf{R}^T \mathbf{R} + \mu \mathbf{I})^{-1} \mathbf{R}^T \mathbf{u}, \quad (5)$$

where \mathbf{I} is an identity matrix and $\mu > 0$ is the damping parameter. A damped solution is, however, small in the sense of Euclidean norm, having many unnecessary non-zero coefficients which limit signal and noise separation. Due to the large dimensionality of the system of equations (5), traditional algorithms work in the frequency domain and replace this large system by a set of smaller systems of Toeplitz structure [17], corresponding to each frequency.

Algorithm 1: Split Bregman method for the sparse Radon transform.

```

1 Initialize:  $k = 0, \mathbf{u}^0 = \mathbf{u}, \mathbf{d}^0 = \mathbf{b}^0 = \mathbf{0}$ 
2 while  $\|\mathbf{u} - \mathbf{R}\mathbf{v}^k\|_2^2 > \varepsilon$  do
3    $\mathbf{v}^{k+1} = \mathbf{T}^{-1}(\alpha \mathbf{R}^T \mathbf{u}^k + \beta [\mathbf{d}^k - \mathbf{b}^k])$ 
4    $\mathbf{d}^{k+1} = \text{shrink}(\mathbf{v}^{k+1} + \mathbf{b}^k, 1/\beta)$ 
5    $\mathbf{b}^{k+1} = \mathbf{b}^k - [\mathbf{d}^{k+1} - \mathbf{v}^{k+1}]$ 
6    $\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{u} - \mathbf{R}\mathbf{v}^{k+1}$ 
7    $k = k + 1$ 
8 end

```

For a better separation between signal and noise, a sparsity promoting inversion is performed to restrict the Radon plane to the space of sparse signals while predicting the data [28]. In seismological community, the algorithms associated with sparse or high-resolution RT can be categorized into two main groups: time-domain algorithms, where the solution is forced to be jointly sparse in τ and p directions [18], and frequency domain algorithms, where each frequency slice is inverted independently and hence the solution is forced to be sparse only along the p axis [23,27]. Frequency-domain algorithms have been developed to circumvent the computational load of the time-domain approach by sacrificing the sparsity along the time axis [28]. In this paper, we consider a time-domain sparse RT by defining the Radon plane as a solution to the following BPDN problem:

$$\mathbf{v} := \arg \min_{\mathbf{v}} \|\mathbf{v}\|_1 \quad \text{subject to} \quad \|\mathbf{u} - \mathbf{R}\mathbf{v}\|_2^2 \leq \varepsilon, \quad (6)$$

where ε is an error bound in the data. In this paper, we develop our fast algorithm based on the split Bregman method to effectively and efficiently solve (6). The split Bregman iteration reduces complex convex optimization problem (6) into a series of smaller subproblems which are much easier to be solved (see [35,30] for more details):

$$\mathbf{v}^{k+1} = \arg \min_{\mathbf{v}} \alpha \|\mathbf{u}^k - \mathbf{R}\mathbf{v}\|_2^2 + \beta \|\mathbf{d}^k - \mathbf{v} - \mathbf{b}^k\|_2^2 \quad (7)$$

$$\mathbf{d}^{k+1} = \arg \min_{\mathbf{d}} \|\mathbf{d}\|_1 + \beta \|\mathbf{d} - \mathbf{v}^{k+1} - \mathbf{b}^k\|_2^2 \quad (8)$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k - [\mathbf{d}^{k+1} - \mathbf{v}^{k+1}] \quad (9)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{u} - \mathbf{R}\mathbf{v}^{k+1} \quad (10)$$

for $k = 0, 1, \dots$ and $\alpha, \beta > 0$ starting with $\mathbf{u}^0 = \mathbf{u}$ and $\mathbf{d}^0 = \mathbf{b}^0 = \mathbf{0}$. Sub-problems (7) and (8) have closed form solutions leading to a simple algorithm as presented in Algorithm 1. In Algorithm 1, shrink is the soft shrinkage filter, i.e., $\text{shrink}(x, \gamma) = \frac{x}{|x|} \max(|x| - \gamma, 0)$ which is applied to the elements of input vector independently, and

$$\mathbf{T} = \alpha \mathbf{R}^T \mathbf{R} + \beta \mathbf{I} \quad (11)$$

where \mathbf{I} is an identity matrix. The efficiency of this algorithm for solving sparse RTs will be explained and shown in the next sections.

2.2. Three-dimensional Radon transform

The azimuth independent three-dimensional (3D) Radon transform can be defined as a sum of 3D signal $\mathcal{U}(t, x, y)$ over objects defined by $t = \tau + p\varphi(\sqrt{x^2 + y^2})$ as

$$\underline{\mathbf{v}}(\tau, p) = \sum_x \sum_y \mathcal{U}(t = \tau + p\varphi(\sqrt{x^2 + y^2}), x, y). \quad (12)$$

Interestingly, all the formulations derived above depend upon the spatial variable x without requiring it to be uniformly sampled. Therefore, a surrogate 2D signal $\underline{\mathbf{u}}(t, r)$ can be constructed from

the 3D signal by sorting its traces according to their distance from the origin as $\underline{\mathbf{u}}(t, r = \sqrt{x^2 + y^2}) = \mathcal{U}(t, x, y)$. Accordingly, we have the following inverse and adjoint-forward transformations for 3D RTs which are the same as 2D formula given in (1) and (4).

$$\underline{\mathbf{u}}(t, r) = \sum_p \underline{\mathbf{v}}(\tau = t - p\varphi(r), p). \quad (13)$$

$$\underline{\mathbf{v}}(\tau, p) = \sum_r \underline{\mathbf{u}}(t = \tau + p\varphi(r), r). \quad (14)$$

Therefore, the algorithm presented in this paper can easily be used for computing 3D RTs.

2.3. Contributions

Up to now, we just formulated the sparse RT as a BPDN problem and employed the split Bregman iteration to solve the generated problem. In the sequel, we discuss the main contributions of the paper, which are as follows:

- i. Application of \mathbf{R} and \mathbf{R}^T on a given vector [computation of (1) and (4) in the frequency domain, as will be seen later] requires $O(N_f N_x N_p + N_x N_t \log N_t)$ flops. In this paper, we develop a fast algorithm, with numerical complexity $O(N_f N \log N + N_x N_t \log N_t)$, for application of these operators, which leads to speedup of several orders of magnitude.
- ii. Line #3 of the algorithm 1 requires inversion of \mathbf{T} in (11) which equals computation of an LS RT (5). As stated above frequency domain algorithms for this job require solving a set of N_f Hermitian Toeplitz systems, each of size N_p [9,27]. Conventional algorithms like Levinson's [32] for solving such a set of equations require $O(N_f N_p^2)$ operations. We approximate each Toeplitz system with the best circulant approximation of it in the sense of Frobenius norm. Then a solution to the resulting circulant system is used as an approximate solution to the original Toeplitz system while preserving the convergence properties of the algorithm, thus reducing the cost of inverting \mathbf{T} to $O(N_f N_p \log N_p)$ via an FFT algorithm.
- iii. The error bound ε is usually unknown when dealing with real world noisy data. Bregman iteration has a monotonic convergence behavior [30] thus leading to a semi-convergence behavior in the presence of noise [36,37]. We obtain a good estimate of predictive error via GCV analysis and automatically determine the optimum number of iterations, and hence ε , at a negligible cost.

Considering the above, our algorithm computes the time-domain sparse RT with total numerical complexity $O(N_f N \log N + N_x N_t \log N_t)$. It is sufficiently fast so that can easily be applied to large-scale seismic data sets. In the following we address each item described above in details.

3. Fast inverse transformation and its adjoint

Time-invariant RT is more conveniently applied in the Fourier domain because each frequency can be implemented independently. Consider computation of the adjoint problem (4). Applying the discrete Fourier transform (DFT) to both sides of it, along the time axis, gives:

$$\hat{\underline{\mathbf{v}}}(f, p) = \sum_x \hat{\underline{\mathbf{u}}}(f, x) e^{i2\pi f p \varphi(x)} \quad (15)$$

$$= \sum_x \hat{\underline{\mathbf{u}}}(f, x) \mathbf{S}_f(x, p) \quad (16)$$

where the hat represents frequency domain, f denotes frequency, $i^2 = -1$, and frequency-dependent kernel matrix $\mathbf{S}_f \in \mathbb{C}^{N_x \times N_p}$ is defined by its entries as

$$\mathbf{S}_f(x, p) = e^{i2\pi f p \varphi(x)}. \quad (17)$$

Obviously, computation of the f th frequency of the Radon matrix via equation (15) requires $N_x \times N_p$ complex multiplications and $(N_x - 1) \times N_p$ complex summations. Hence, the total cost for computation of all frequencies needs $O(N_f N_x N_p)$ flops. An initial Fourier transform of complexity $O(N_x N_t \log N_t)$ is also required for preparing data in the frequency-offset $(f - x)$ domain. Therefore, the complexity of the inverse transform and its adjoint is $O(N_f N_x N_p + N_x N_t \log N_t)$. We now develop a more efficient algorithm for computation of these transforms.

Let's assume regular sampling of the Radon parameter, such that $p_{j+1} = p_j + \Delta p$ for $\Delta p > 0$ and $j = 0, \dots, N_p - 2$, then a simple linear transformation allows defining $p_j = p'_0 + \gamma_j \Delta p$ for

$$\gamma_j \in \mathbb{Z}^{N_p} = -\lfloor \frac{N_p}{2} \rfloor, \dots, \lfloor \frac{N_p}{2} \rfloor - 1 \quad (18)$$

where $\lfloor \cdot \rfloor$ denotes the floor function and $p'_0 = p_0 - \gamma_0 \Delta p$. Furthermore, regular sampling of frequency, i.e., $f = k \Delta f$ for $\Delta f > 0$ and $k = 0, \dots, N_f - 1$, allows writing (17) for spatially sampled data as

$$\mathbf{S}_k(n, j) = \hat{\mathbf{a}}(k, n) e^{i2\pi k \Delta f \Delta p \varphi(x_n) \gamma_j}, \quad (19)$$

where $\hat{\mathbf{a}}(k, n) = e^{i2\pi k \Delta f \varphi(x_n) p'_0}$ is a shift operator along the Radon (p)-axis. Since $k = 0, \dots, N_f - 1$, then $k \gamma_j \in \lambda \forall j, k$, where

$$\lambda = 0, 1, \dots, \frac{M}{2} - 1, -\frac{M}{2}, -\frac{M}{2} + 1, \dots, -1 \quad (20)$$

with $M = 2 \lfloor \frac{N_p}{2} \rfloor (N_f - 1)$. Based on the definition of λ , a frequency-independent super matrix is constructed as

$$\mathcal{S}(n, j) = e^{i2\pi \Delta f \Delta p \varphi(x_n) \lambda_j} \quad j = 0, \dots, M - 1. \quad (21)$$

From (19) and (21) and the definition of λ , it is obvious that for all j and k the vector $\hat{\mathbf{a}}^*(k, n) \mathbf{S}_k(n, j)$ $n = 0, \dots, N_x - 1$ is a column of matrix \mathcal{S} , where the star shows conjugate. Therefore, all the coefficients of $\hat{\mathbf{y}}$ are computable from $(\hat{\mathbf{u}} \circ \hat{\mathbf{a}}) \mathcal{S}$, where \circ denotes the Hadamard (element-by-element) product operator. Explicit computation of $(\hat{\mathbf{u}} \circ \hat{\mathbf{a}}) \mathcal{S}$ in the current form is, however, impractical or even impossible because explicit computation of \mathcal{S} demands huge memory even for moderate size problems. Interestingly, the rows of matrix \mathcal{S} are discrete mono-frequency complex exponentials and hence extremely band limited. Applying the DFT along variable λ , after some algebraic manipulations, we get

$$\hat{\mathcal{S}}(n, l) = \frac{1}{M} \sum_{j=0}^{M-1} e^{i2\pi \Delta f \Delta p \varphi(x_n) \lambda_j} e^{-\frac{i2\pi j l}{M}} \quad (22)$$

$$= \text{Re}(n, l) + i \text{Im}(n, l) \quad (23)$$

where

$$\text{Re}(n, l) = \frac{(-1)^l}{M} \sin(M\pi \theta_n) \cot(\pi [\theta_n - l/M]) \quad (24)$$

and

$$\text{Im}(n, l) = \frac{(-1)^{l+1}}{M} \sin(M\pi \theta_n), \quad (25)$$

with $\theta_n = \Delta f \Delta p \varphi(x_n)$. Since $|\hat{\mathcal{S}}(n, l)|$ is the amplitude spectrum of complex exponentials the value one is an upper bound of it. Furthermore, M is a large number implying that $|\text{Im}(n, l)| \leq \frac{1}{M} \ll 1$. Therefore, $\hat{\mathcal{S}}$ is mostly dominated by the real part and its imaginary part is negligible, and thus

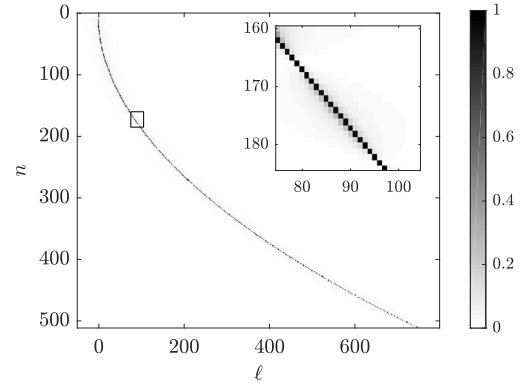


Fig. 1. The structure of matrix $\hat{\mathcal{S}}$ defined in (26) for PRT, $\varphi(x) = x^2$, and a set of parameters: $N_t = N_x = N_p = 512$, $0 \leq x \leq 2000$ m, $0 \leq t \leq 2$ s, $0 \leq p \leq 1.5 \times 10^{-6}$ s²/m².

$$\hat{\mathcal{S}}(n, l) \approx \frac{(-1)^l}{M} \sin(M\pi \theta_n) \cot(\pi [\theta_n - l/M]). \quad (26)$$

Obviously, (26) is valid for $l, x \neq 0$. If both of these variables are zero, then $\hat{\mathcal{S}}(n, l)$ is one. The band limited nature of $\hat{\mathcal{S}}$ allows efficient implementation of it within the spectral band, $-l_0 < l < M\theta_n + l_0$, where $l_0 \geq 0$ is a small integer for protection of the low and high spectral coefficients. As an example, Fig. 1 shows the structure of matrix $\hat{\mathcal{S}}$ for a set of parameters: $N_t = N_x = N_p = 512$, $\varphi(x) = x^2$, $0 \leq x \leq 2000$ m, $0 \leq t \leq 2$ s, $0 \leq p \leq 1.5 \times 10^{-6}$ s²/m². As seen, the matrix has a sparse structure and is band limited in l such that it is approximately zero for $l > 850$ while $M = 262144$. For the matrix shown in Fig. 1, 99% of its energy is contained in about 1% of the coefficients. Therefore, a thresholding filter can be applied to $\hat{\mathcal{S}}$ by zeroing its small coefficients, in order to implement it in a sparse format; that is,

$$\hat{\mathcal{S}}(n, l) = \begin{cases} \hat{\mathcal{S}}(n, l) & \text{if } |\hat{\mathcal{S}}(n, l)| \geq \epsilon \\ 0 & \text{if } |\hat{\mathcal{S}}(n, l)| < \epsilon \end{cases} \quad (27)$$

where ϵ is a small positive value. Numerical tests show that $\epsilon = 0.01$ provides satisfactory results. Given a set of parameters, matrix $\hat{\mathcal{S}} \in \mathbb{R}^{N_x \times N_l}$ for $N_l = M\theta_{N_x} + 2l_0$, can thus be easily computed and stored within its spectral band, then

$$\hat{\mathbf{y}} = (\hat{\mathbf{u}} \circ \hat{\mathbf{a}}) \hat{\mathcal{S}}, \quad (28)$$

is computed efficiently, because it involves a Hadamard product of two matrices $\hat{\mathbf{u}}$ and $\hat{\mathbf{a}}$ and a matrix production with sparse matrix $\hat{\mathcal{S}}$. From $\theta_n = \Delta f \Delta p \varphi(x_n)$ and $M = 2 \lfloor \frac{N_p}{2} \rfloor (N_f - 1)$, we get that $N_l = f_{\max} p_{\max} \varphi(x_{\max}) + 2l_0$ depends on the maximum frequency and offset in the data set and the maximum curvature in the Radon space. Note that in practice a good choice for N_p is $f_{\max} p_{\max} \varphi(x_{\max})$ [34,38] and hence $N_l \approx N_p$.

The main remaining problem is to compute $\hat{\mathbf{y}}$ from $\hat{\mathbf{y}}$ via an inverse DFT (IDFT). What we need is to compute N_p equally spaced samples of an M -point IDFT corresponding to each row of $\hat{\mathbf{y}}$. Since $\lambda = k \gamma_j \forall j, k$, the space between successive samples corresponding to the k th row is $\Delta \lambda = k$. The M -point IDFT of the sequence $\{\hat{\mathbf{y}}(k, l)\}_{l=0}^{N_l-1}$ is equivalent to the values of the z-transform of the sequence at M equally spaced points on the unit circle in the z-plane. According to (20) the points corresponding to positive/negative λ 's lay on the upper/lower half of the circle. Fig. 2 shows the distribution of the required samples for $N_p = 21$ and four different frequencies on the unit circle: $k = 0, \frac{N_f-1}{4}, \frac{N_f-1}{2}, N_f - 1$. As we will see in the next section, the chirp z-transform allows efficient computation of the desired samples without having to compute all M samples of the sequence.

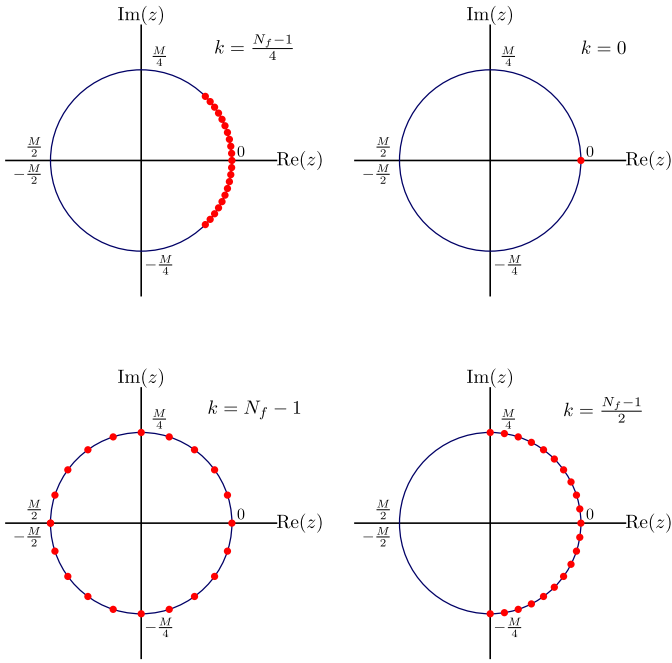


Fig. 2. Distribution of the required samples in the z -domain, corresponding to the IDFT of sequence $\{\hat{\mathbf{y}}(k, l)\}_{l=0}^{N_l-1}$ in (28) for $N_p = 21$ and four different frequencies: $k = 0, \frac{N_f-1}{4}, \frac{N_f-1}{2}, N_f - 1$.

Comparing to previous works, [34] proposed the variable transformation $\lambda = fp$ to speed up the computation of PRT, where a fixed and limited range of λ was used for all frequencies making the kernel independent of frequency, thus increasing the efficiency of the transform. The formulation, however, provide a modified Radon domain and an interpolation is required to obtain the usual Radon coefficients from the computed coefficients. In contrast, the algorithm in this paper solves the original RT. The used kernel matrix includes all possible values of f and p in λ . Therefore, no interpolation is required here and the original Radon domain is directly computed using the chirp z -transform.

3.1. Efficient computation via chirp z -transform

Chirp z -transform is an efficient tool for computing the z -transform as long as the angular distance between successive points in the z -domain is uniform [31]. It is mainly used for zooming the spectral content of time series. Recently, [11,12] employed the chirp z -transform for computing the so called pseudo-polar Fourier transform used for fast low-resolution LRT. Here, it is employed for computation of a general time-invariant Radon transform. Let $\hat{\mathbf{y}}(k, l)$, $l = 0, \dots, N_l - 1$, be the computed spectral sequence, according to (28), then the Radon coefficients are calculated as

$$\hat{\mathbf{y}}(k, \lambda) = e^{-\frac{i2\pi\lambda l_0}{M}} \sum_{l=0}^{N_l-1} \hat{\mathbf{y}}(k, l) e^{\frac{i2\pi\lambda l}{M}}, \quad (29)$$

for $\lambda = k\gamma_j$ and $j = 0, \dots, N_p - 1$. Therefore,

$$\hat{\mathbf{y}}(k, j) = e^{-\frac{i2\pi k\gamma_j l_0}{M}} \sum_{l=0}^{N_l-1} \hat{\mathbf{y}}(k, l) e^{\frac{i2\pi k\gamma_j l}{M}} \quad (30)$$

$$= e^{-\frac{i2\pi k\gamma_j l_0}{M}} \sum_{l=0}^{N_l-1} \hat{\mathbf{y}}(k, l) e^{\frac{i2\pi\gamma_j l_0}{M}} e^{\frac{i2\pi k\gamma_j l}{M}} \quad (31)$$

for $j = 0, \dots, N_p - 1$. Setting $jl = \frac{1}{2}[l^2 + j^2 - (j-l)^2]$ and introducing a linear chirp signal $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{h}}(k, l) = e^{-\frac{i\pi k l^2}{M}} \quad (32)$$

we obtain

$$\hat{\mathbf{y}}(k, j) = \hat{\mathbf{c}}(k, j) \sum_{l=0}^{N_l-1} \hat{\mathbf{y}}(k, l) \hat{\mathbf{h}}(k, l) \hat{\mathbf{h}}(j-l) \quad (33)$$

where

$$\hat{\mathbf{c}}(k, j) = e^{\frac{i2\pi k(-\gamma_j l_0 + j^2/2)}{M}} \quad (34)$$

and

$$\hat{\mathbf{h}}(k, l) = e^{\frac{i2\pi k(\gamma_0 l^2 + l^2/2)}{M}}. \quad (35)$$

Now, the summation in (33) has the form of a convolution

$$\hat{\mathbf{y}} = \hat{\mathbf{c}} \circ [\hat{\mathbf{y}} \circ \hat{\mathbf{b}} * \hat{\mathbf{h}}], \quad (36)$$

where $*_l$ denotes convolution with respect to l . (36) can now be computed efficiently via an FFT algorithm with an appropriate zero padding [31, pp. 544–549]. More precisely, we should select at least a DFT of size $N = N_l + N_p - 1$ in order to perform the convolution via an FFT without aliasing effects. Replacing $\hat{\mathbf{y}}$ from (28) into (36) we end up with our final formulation for the adjoint operator as

$$\hat{\mathbf{y}} = \hat{\mathbf{c}} \circ [(\hat{\mathbf{u}} \circ \hat{\mathbf{a}}) \hat{\mathbf{S}} \circ \hat{\mathbf{b}} * \hat{\mathbf{h}}] \quad (37)$$

$$= \hat{\mathbf{c}} \circ \text{IFFT} \left[\text{FFT} \left[(\hat{\mathbf{u}} \circ \hat{\mathbf{a}}) \hat{\mathbf{S}} \circ \hat{\mathbf{b}} \right] \circ \hat{\mathbf{h}} \right] \quad (38)$$

Where $\hat{\mathbf{h}}$ is obtained by applying a Fourier transform along the second dimension of $\hat{\mathbf{h}}$. (37) and (38) are compact forms of (15). Based on (38), the inverse operator can be computed simply from $\hat{\mathbf{u}}$ as

$$\hat{\mathbf{u}} = (\hat{\mathbf{b}}^* \circ \text{IFFT} \left[\text{FFT} \left[\hat{\mathbf{y}} \circ \hat{\mathbf{c}}^* \right] \circ \hat{\mathbf{h}}^* \right]) \hat{\mathbf{S}}^T \circ \hat{\mathbf{a}}^* \quad (39)$$

which is a compact form of the frequency domain representation of (1):

$$\hat{\mathbf{u}}(f, x) = \sum_p \hat{\mathbf{y}}(f, p) e^{-i2\pi fp\varphi(x)} \quad (40)$$

3.2. Computational complexity of forward/inverse transforms

As seen from the above algorithms, for a single data set, both inverse RT and its adjoint can be implemented by two FFT's, one IFFT, 4 complex multiplications, and a multiplication with $\hat{\mathbf{S}}$ or $\hat{\mathbf{S}}^T$. The FFT's and IFFT's are all of length N and thus can be performed in $O(N_f N \log N)$ flops, for all frequencies. The sizes of $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, and $\hat{\mathbf{h}}$ are respectively, $(N_f \times N_x)$, $(N_f \times N_l)$, $(N_f \times N_p)$, and $(N_f \times N)$. Therefore, the cost of complex multiplications is proportional to $O(N_f N_x + 2N_f N_l + 2N_f N_p)$. The cost of performing matrix $\hat{\mathbf{S}}$ or its transpose is negligible as it is significantly sparse. Some $O(N_x N_t \log N_t)$ computations are also required for preparing data in the frequency domain. Therefore, the total cost of the algorithm is roughly $O(N_f N \log N + N_x N_t \log N_t)$ flops. As mentioned above, in practice $N_l \approx N_p$ and thus $N \approx 2N_p$ which imply that the total cost of our algorithm is roughly $O(N_f N_p \log N_p + N_x N_t \log N_t)$ flops. Compared with the computational complexity of direct calculation, $O(N_f N_x N_p + N_x N_t \log N_t)$, the new method can be much faster, specifically for large data sets. The matrices $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, $\hat{\mathbf{h}}$, and

\hat{S} does not depend upon the input data, and so they can be pre-computed when dealing with multi-set data with the same geometry or when solving for a sparse solution. Thus, this algorithm just involves one FFT and one IFFT and thus the cost per data set or iteration is much more reduced.

4. Fast solution of normal equations

In this section we propose a fast algorithm for evaluation of the line #3 of Algorithm 1 which involves calculating a least-square RT

$$(\alpha \mathbf{R}^T \mathbf{R} + \beta \mathbf{I}) \mathbf{v} = \mathbf{y} \quad (41)$$

for a given right-hand-side vector $\mathbf{y} \in \mathbb{R}^{N_x N_p}$. From (1), (2), (3) and (4) it is easily confirmed that (41) in the frequency domain reads as

$$\alpha \sum_x \sum_p \hat{\mathbf{y}}(f, p) e^{-i2\pi f(p-q)\varphi(x)} + \beta \hat{\mathbf{y}}(f, q) = \hat{\mathbf{y}}(f, q) \quad (42)$$

or in matrix form

$$\hat{\mathbf{y}}(f, :) \mathbf{\Omega}_f = \hat{\mathbf{y}}(f, :) \quad (43)$$

where the frequency dependent kernel matrix $\mathbf{\Omega}_f$ is defined by its entries as

$$\mathbf{\Omega}_f(p, q) = \alpha \sum_x e^{-i2\pi f(p-q)\varphi(x)} + \beta \delta(p - q) \quad (44)$$

where $\delta(\cdot)$ is the Kronecker delta function. Obviously, $\mathbf{\Omega}_f$ is a Hermitian Toeplitz-structured matrix, as observed by [9] in slant stack problem. Therefore, line #3 of Algorithm 1 requires us to solve a set of N_f Hermitian Toeplitz systems of size N_p .

4.1. Computation of systems coefficients

The Hermitian Toeplitz matrix $\mathbf{\Omega}_f$ can be uniquely represented by its first row

$$\omega_f(j) = \alpha \sum_x e^{i2\pi f \Delta p j \varphi(x)} + \beta \delta(j) \quad (45)$$

$j = 0, \dots, N_p - 1$. For each frequency, direct computation of the summation term in (45) requires $O(N_x N_p)$ operations and hence the total cost for construction of the systems coefficients for all frequencies needs $O(N_f N_x N_p)$ flops. In contrast, these coefficients can easily be computed via the proposed method in the previous section. The summation in (45) is the same as the summation in (15) with $\hat{\mathbf{u}}$ replaced with an all ones matrix. Consequently, for all frequencies, it is computable with $O(N_f N \log N)$ flops via the algorithm in (38).

4.2. Approximate solution of Toeplitz systems

Having $\mathbf{\Omega}_f$, the conventional direct algorithms for solving (43), such as Gaussian elimination, require $O(N_p^3)$ operations, which are too slow for dealing with large problems. The traditional fast algorithms like Levinson's [32] requires $O(N_p^2)$ operations for this task while more recent complicated algorithms, under name of super-fast, reduced the cost to $O(N_p \log^2 N_p)$ [39]. However, the superiority of such super-fast algorithms is only seen when dealing with very large systems. The conjugate gradient methods have also been successfully used for solving Toeplitz systems [40]. In seismic data processing, it has been used for frequency-domain sparse PRT [27]. An important advantage of such iterative methods is that the complexity of solving a Toeplitz system can be reduced to $O(N_{cg} N_p \log N_p)$ where N_{cg} is the number of conjugate gradient iterations.

We use two important properties of Bregman iteration to effectively and efficiently solve this set of Toeplitz systems. Firstly, the algorithm is convergent for a fixed and finite Bregman parameters α and β . Therefore, they can be suitably chosen for conditioning the systems [30]. We choose $\alpha = 1/N_x$ and $\beta = 20$ which experimentally show a good balance between convergence and conditioning. Secondly, the Bregman iteration is still convergent for an approximate solution to the normal equations [33,30]. Indeed, a highly accurate solution can still be obtained by inaccurately solved normal equations [33]. Therefore, we don't need to exactly solve the Toeplitz systems (43). Toeplitz matrices are closely related to circulant matrices. Unlike Toeplitz matrices, circulant matrices are diagonalizable in the Fourier domain and hence more easily manipulable. A Toeplitz system can be extended into a circulant one which is twice the size. The extended system can then be solved via the CG algorithm by employing an FFT to carry out the matrix-vector operations [27]. In this paper, however, we follow a more efficient way to determine an approximate solution to the Toeplitz systems in (43). We approximate each Toeplitz system with the best circulant approximation of it in the sense of Frobenius norm. The resulting circulant system then is inverted by $O(N_p \log N_p)$ operations by using an FFT algorithm. Let \mathcal{F} be the DFT matrix defined as $\mathcal{F}(k, j) = \frac{1}{\sqrt{N}} e^{-i2\pi k j / N_p}$, $k, j = 0, \dots, N_p - 1$. Then the optimal circulant approximation to the Toeplitz matrix $\mathbf{\Omega}_f$ is defined as

$$\mathbf{\Omega}_f^c = \arg \min_{\mathbf{\Omega} \in \mathcal{C}} \|\mathbf{\Omega} - \mathbf{\Omega}_f\|_F \quad (46)$$

where $\|\cdot\|_F$ is the Frobenius norm, and

$$\mathcal{C} = \{\mathcal{F}^H \Sigma \mathcal{F} \mid \Sigma \text{ is any } N_p \times N_p \text{ diagonal matrix}\}. \quad (47)$$

Problem (46) has an explicit solution $\mathbf{\Omega}_f^c = \mathcal{F}^H \Sigma_f \mathcal{F}$ where the diagonal elements of Σ_f are the Fourier transform of the sequence

$$\hat{\omega}_f(j) = (1 - \frac{j}{N_p}) \omega_f(j) + \frac{j}{N_p} \omega_f^*(N_p - j) \quad (48)$$

for $j = 0, \dots, N_p - 1$ [41]. Matrix $\mathbf{\Omega}_f^c$ has been used as a successful preconditioner, known as T. Chan's preconditioner, for solving Toeplitz systems via conjugate gradient algorithm [41,42].

Having $\hat{\omega}_f$ an approximate solution to the system (43) is computed for all frequencies simultaneously as

$$\hat{\mathbf{v}} \approx \left(\frac{\hat{\mathbf{y}} \mathcal{F}^H}{\hat{\omega}_f \mathcal{F}^H} \right) \mathcal{F} \quad (49)$$

where with notation abuse the division operation is done componentwise. Applications of \mathcal{F} and \mathcal{F}^H is done via FFT algorithm, therefore, line #3 of the algorithm also needs $O(N_f N_p \log N_p)$ operations.

5. Determination of the stopping iteration

A common difficulty in performing all Radon transforms arises when dealing with noisy real world data $\mathbf{u} = \mathbf{u}_0 + \mathbf{e}$ where \mathbf{u}_0 is a noise free signal and \mathbf{e} is an additive noise. In this case, a balance between the amount of sparsity and fitness to data should be controlled by a regularization parameter which is to be determined by a posteriori rule. A main difficulty is that one needs multiple solutions of the problem for a range of regularization parameter in order to determine the optimum one. In the proposed algorithm, the iteration number K has the role of regularization parameter. We address automatic determination of this parameter via the GCV score analysis [43]. The GCV score has been successfully used by the first author for determination of the optimum number of Bregman iteration in solving seismic deconvolution [36,37]. Let \mathbf{v}^k be

the Radon coefficients obtained at the k th Bregman iteration via Algorithm 1, then the objective is determination of the optimum iteration number K as a minimizer of Predictive Mean-Squared Error (PMSE) defined as follows:

$$\text{PMSE}(k) = \frac{1}{N_t N_x} \|\mathbf{u}_0 - \mathbf{R}\mathbf{v}^k\|_2^2. \quad (50)$$

However, since \mathbf{u}_0 is not available in practice the GCV is used as an alternative score which asymptotically minimizes PMSE, and is defined as

$$\text{GCV}(k) = \frac{\|\mathbf{u} - \mathbf{R}\mathbf{v}^k\|_2^2}{(1 - (N_t N_x)^{-1} \text{freedom}(\mathbf{v}^k))^2} \quad (51)$$

where $\text{freedom}(\mathbf{v}^k)$ is a measure of the effective number of degrees of freedom in the solution for fitting the data. It is a measure of the sensitivities of the predicted data with respect to the observations and its evaluation is the main factor in performing GCV. Generally when \mathbf{u}_0 has a sparse Radon transformation, then a small amount of noise in it severely decreases the sparsity of the corresponding RT domain. In this case, the number of non-zero coefficients, $\text{nnz}(\mathbf{v}^k)$, provides a reasonable estimate of $\text{freedom}(\mathbf{v}^k)$ [36,37]. Here, we employ it to evaluate GCV for a fast sparse RT.

6. Experimental results

In this section, we present a set of experiments illustrating the performance of the proposed fast algorithm for time-invariant RTs. For all examples, we perform an $2N_t$ -point DFT to transform data to $f - x$ domain and choose $N_f = N_t$.

6.1. Inverse RT and adjoint solution

We present several numerical examples from simulated data to confirm the accuracy and efficiency of the proposed method for calculation of the inverse transform and its adjoint. All examples are tested for PRT by setting $\varphi(x) = x^2$, as it is more problematic than LRT. The algorithm is coded in MATLAB. All experiments were run on a 2.3 GHz Intel Core i7 machine with 7.7 GB of RAM under the Ubuntu 14.04 LTS operating system and MATLAB R2013a.

Fig. 3 shows a synthetic data in the $t - x$ domain containing events with parabolic moveout. For this data set, $N_t = N_x = 2048$, $\Delta t = 0.004$ s, and $\Delta x = 2$ m. Variable p was sampled with $N_p = 2048$ equally spaced samples in the range $[0, 4 \times 10^{-7}]$. Direct computation of the adjoint operator generated the PRT panel in Fig. 4(a) in 34.22 s while the proposed fast algorithm generated the PRT panel in Fig. 4(b) in 1.13 s. The PRT panels were transformed back to the $t - x$ domain via (40). The results of direct and fast computations are shown in Figs. 4(c) and 4(d), respectively. It is seen that the proposed algorithm significantly reduces the computational burden without affecting the accuracy.

The adjoint operator was applied to the data sets of different sizes in order to show the evolution of its run time with problem size and compare its sensitivity with respect to each parameter. Fig. 5a shows the evolution of the run time with the number of time-samples, $N_t (= N_f)$, while the number of traces, N_x , and the number of curvature bins, N_p , were kept fix, $N_x = N_p = 512$. Although, comparing with direct computation, the results show a decrease in the cost of the new method with respect to N_t , this improvement is not significant because the computational complexity of both algorithms is proportional to N_f or N_t . In another experiment, the run time was computed for different number of traces while the number of time samples and the number of curvature bins were kept fix, $N_t = N_p = 512$. The results are depicted in Fig. 5b, where, as expected, there is a significant advantage of the

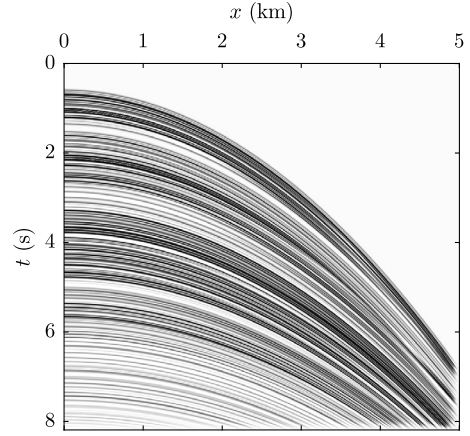


Fig. 3. A 2D seismic data containing events with parabolic moveout. Here, $N_t = N_x = 2048$.

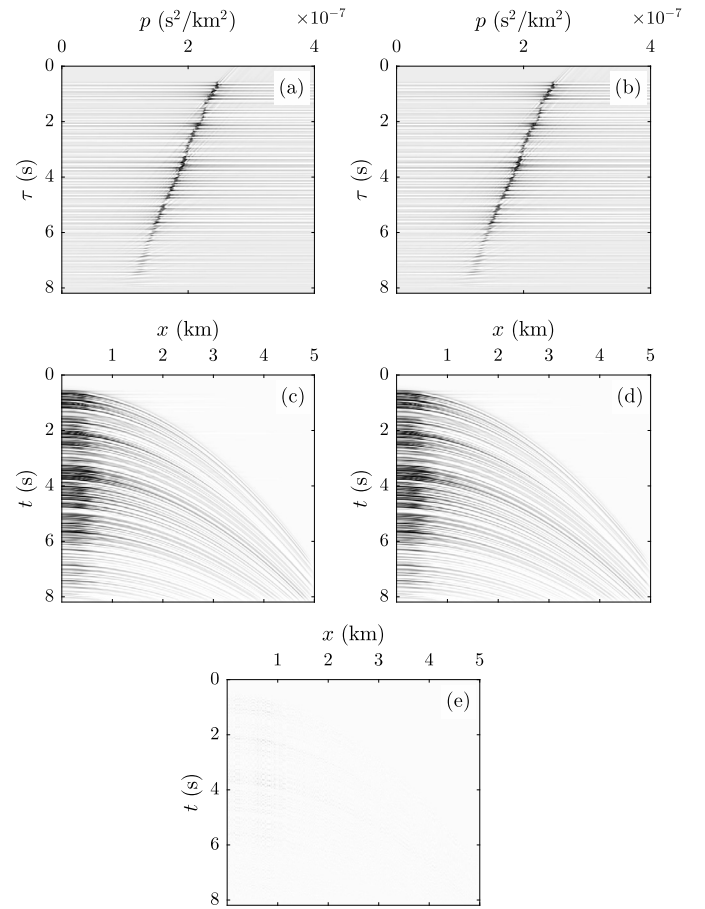


Fig. 4. The adjoint solution of PRT corresponding to the data shown in Fig. 3, obtained via (a) direct computation (CPU time = 34.22 s) and (b) fast computation (CPU time = 1.13 s). (c) The output of inverse PRT applied to (a). (d) The output of fast inverse PRT applied to (b). (e) Difference between (c) and (d) scaled by 10.

new method. Naturally, the next experiment was tested for different curvature bins while keeping fix the number of time samples and the number of traces, $N_t = N_x = 512$. The results are depicted in Fig. 5c, where, again, as expected, there is a significant decrease of the run time for the new method. The final experiment was to measure the run time while varying all three parameters and the results are shown in Fig. 5d. As seen the computation burden of direct calculation rises significantly with the problem size while the

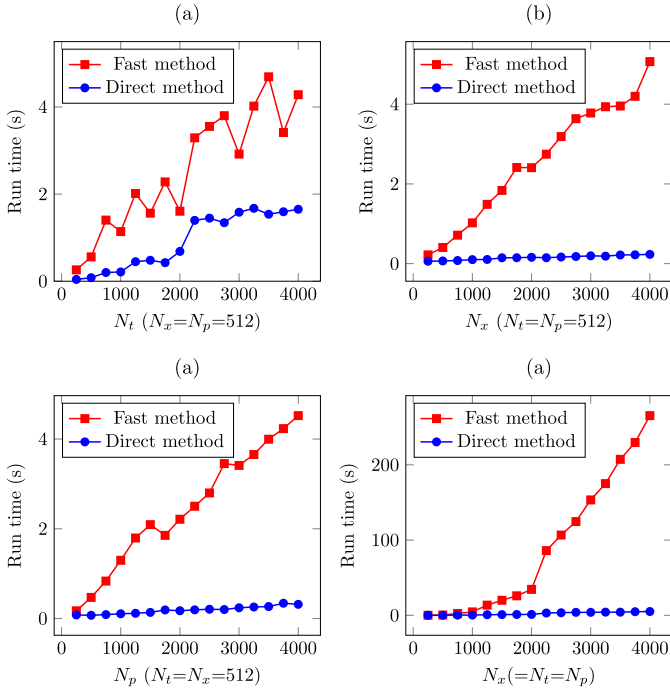


Fig. 5. The evolution of run time with the size of parameters for both direct and fast computations. Run time for (a) various N_t while keeping N_x and N_p fix, (b) various N_x while keeping N_t and N_p fix, (c) various N_p while keeping N_t and N_x fix, and (d) various N_t , N_x , N_p .

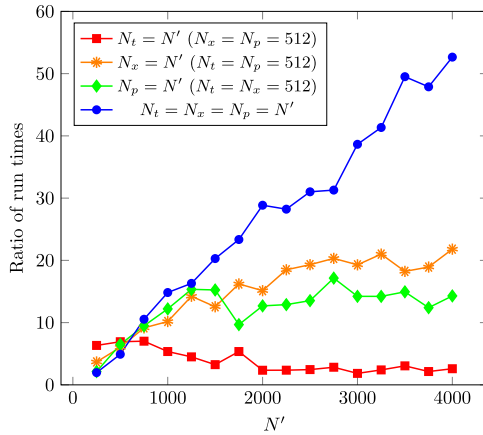


Fig. 6. The ratio of the run time of direct computation to the run time of the fast computation, with respect to the size of parameters.

cost of the fast method increases with an acceptable factor. Fig. 6, showing the ratio of the run time of direct computation to the run time of the proposed method, is depicted to show the factor of improvement in the run-time by the new method. As seen the improvement factor is a linear function of problem size which is a significant achievement.

6.2. Sparse RT of clean signals

The signal shown in Fig. 3 has been used for testing the performance of sparse PRT. We first used the Levinson's algorithm to exactly solve each Toeplitz system in the normal equations at each Bregman iteration. We performed 40 Bregman iterations and the resulting Radon plane, predicted data, and residual data are depicted in Figs. 7(a), 7(c), and 7(e), respectively. We also computed the Mean-Squared Error (MSE) between the original Radon plane, used for generating the signal, \mathbf{v}_{true} , and the estimated one

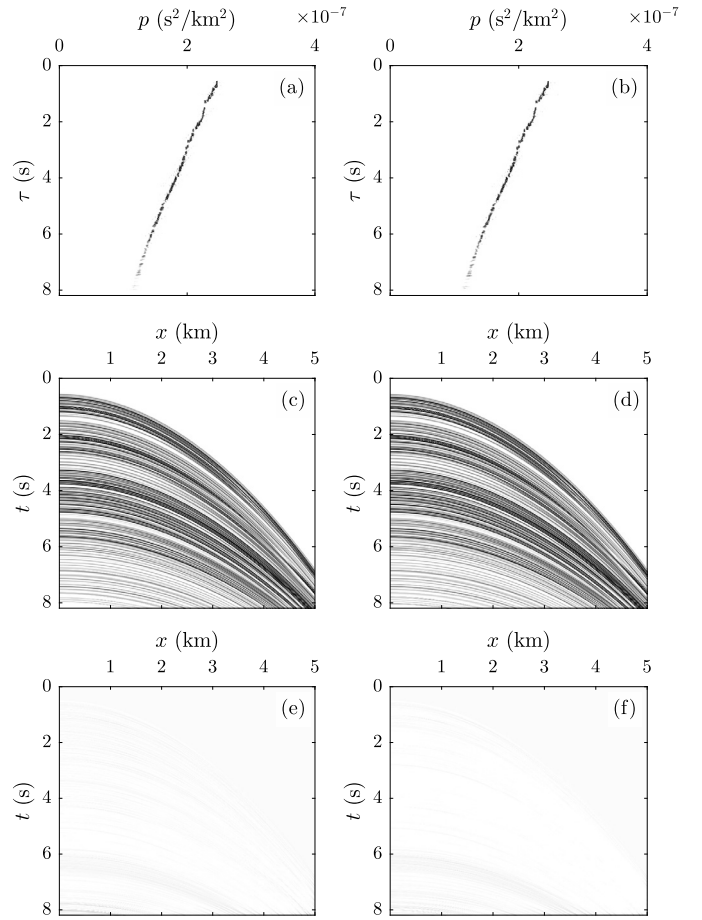


Fig. 7. The sparse solution of PRT corresponding to the data shown in Fig. 3, obtained by (a) exact solution of Toeplitz systems via Levinson's algorithm (CPU time = 27280 s) and (b) approximate solution of Toeplitz systems via circulant approximation and FFT (CPU time = 163 s). (c) and (d) The output of inverse transform applied to (a) and (b). (e) and (f) Residual data corresponding to (c) and (d). For each method, error versus iteration number is depicted in Fig. 8.

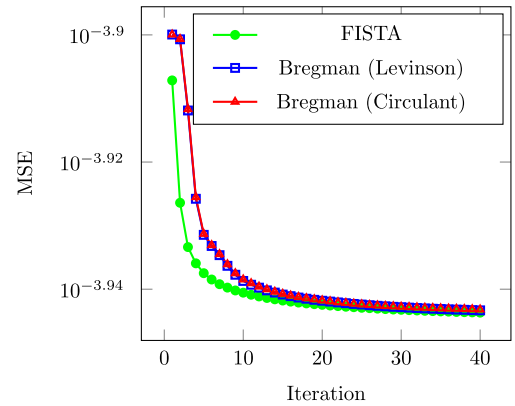


Fig. 8. MSE versus iteration for the sparse PRT corresponding to Fig. 7. The problem has also been solved by FISTA and the error is shown here for comparison.

at each iteration, \mathbf{v}^k , i.e. $MSE(k) = \frac{1}{N_t N_p} \|\mathbf{v}_{true} - \mathbf{v}^k\|_2^2$, and the result is shown in Fig. 8. Comparing to the adjoint solution in Fig. 4 the algorithm successfully generated a sparse Radon plane which is consistent with the signal. However, it took a large time (27280 s) by the Levinson's algorithm to exactly solve the Toeplitz systems and generate this solution. We next replaced the Levinson's algorithm with the proposed circulant approximation and solved the problem. The run time decreased to 163 s and the resulting Radon

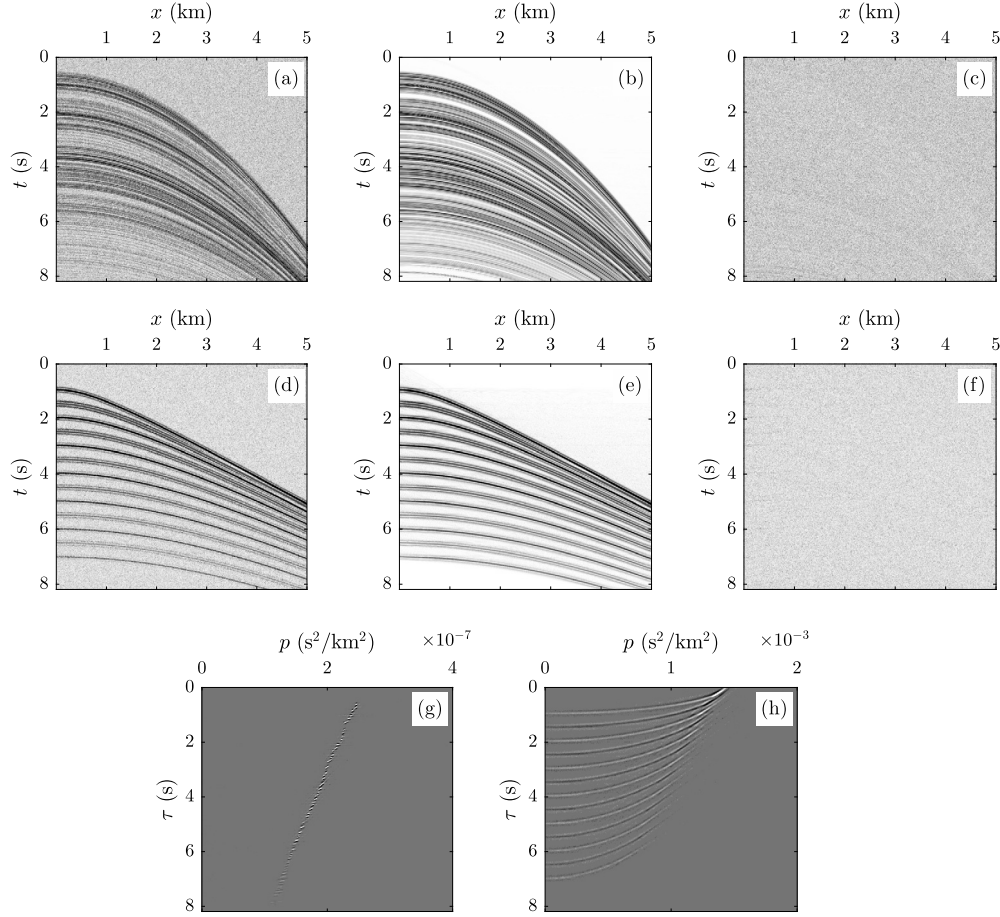


Fig. 9. Sparse RT of noisy signals. (a) The signal shown in Fig. 3 contaminated by random noise of SNR = 5 dB and the corresponding predicted data (b), predicted noise (c), and PRT panel (g). (d) A noisy seismic gather of SNR = 5 dB containing events with hyperbolic moveout and the corresponding predicted data (e), predicted noise (f), and LRT panel (h).

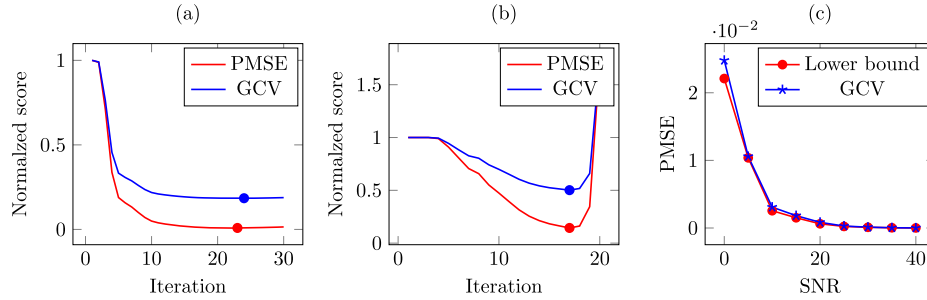


Fig. 10. (a) PMSE and GCV versus iteration corresponding to PRT of noisy signal in Fig. 9(a). (b) PMSE and GCV versus iteration corresponding to LRT of noisy signal in Fig. 9(d). (c) PMSE of the sparse PRT optimized by the GCV score versus SNR of input signal.

plane, predicted data, and residual data are depicted in Figs. 7(b) and 7(d), and 7(f), respectively. The corresponding MSE curve is also shown in Fig. 8. It is clearly seen that the result of exact and circulant solutions are not differentiable while the latter is calculated about 167 times faster. In order to confirm the correctness of the split Bregman method, we also replaced problems (7)–(9) (lines #3–5 of Algorithm 1) by

$$\mathbf{v}^{k+1} = \arg \min_{\mathbf{v}} \alpha \|\mathbf{u}^k - \mathbf{R}\mathbf{v}\|_2^2 + \|\mathbf{v}\|_1 \quad (52)$$

and solved it via FISTA [44,45] for $\alpha = 1/N_x$. For each k , a set of 50 FISTA iterations were performed, resulting in good convergence, and the MSE curve is shown in Fig. 8 for comparison. The obtained Radon plane and predicted data were very close to those shown in Fig. 7 and thus not shown here. It is seen from Fig. 8 that both al-

gorithms converge to the same solution but the run time for FISTA was about 50 times larger than that of the split Bregman.

6.3. Sparse RT of noisy signals

The signal shown in Fig. 3 was contaminated by white Gaussian noise of Signal-to-Noise Ratio (SNR) = 5 dB, as shown in Fig. 9(a), and the proposed algorithm was employed for a high resolution PRT analysis. The GCV was used to determine the optimum number of Bregman iterations and the results are shown in Figs. 9(b) and 9(g). The resulting GCV curve is also shown in Fig. 10(a) where the PMSE curve is also depicted for comparison. The minimum of each curve is pointed by circle. Note that GCV is used as a proxy for PMSE and it is seen that the generated GCV curve resembles the PMSE curve with a high degree of accuracy and that the pre-

dicted data [Fig. 9(b)] is very close to the original noise free signal [Fig. 3]. The resulting Radon panel is also similar to the result of clean signal [Fig. 7(b)].

To test the proposed algorithm for LRT, a noisy seismic gather of $\text{SNR} = 5$ dB containing events with hyperbolic moveout has been simulated [Fig. 9(d)]. An important property of LRT is that it maps a hyperbolic event in $t - x$ domain to an ellipse in the $\tau - p$ domain [46]. This property allows us to attenuate multiples in $\tau - p$ domain via predictive deconvolution filters because multiples are exactly periodic in the $\tau - p$ space while this is not the case in the $t - x$ domain [46–48]. The proposed sparse LRT was applied to the gather in Fig. 9(d) and the optimum results based on GCV score are shown in Figs. 9(e) and 9(h). The corresponding GCV and PMSE curves are also shown in Fig. 10(b). Again it is seen that the GCV successfully estimated the optimum result in the sense of PMSE. The resulting Radon panel [Fig. 9(h)] clearly contains high resolution elliptical events.

An extra test was conducted to better see the performance of the GCV score in determination of the optimum Bregman iteration in the presence of different noise levels. Fig. 10(c) shows the attained PMSE by GCV versus SNR corresponding to sparse PRT. For each noise level a set of 6 signals, containing 100 parabolic events with random curvatures, were generated and analysed, each with 6 different noise realizations, and the observed errors were averaged. The lower bound of error is also shown in Fig. 10(c) for comparison. It is clearly seen that the GCV score is nearly optimal in determination of the number of Bregman iterations while its computation adds no extra cost to the algorithm. This property when combined with its accuracy brings a significant advantage for analysis of practical large seismic data sets.

6.4. Multiple attenuation

Separation of primary and multiple wavefields is a challenging problem in seismic data processing. The PRT is accepted as a successful tools in seismological community for this task. A primary and multiple arriving at the same time usually have different velocities. Therefore, applying normal moveout (NMO) corrections using a velocity between the stacking velocities of the primaries and those of the multiples causes the corrected primaries to have negative curvature values, whereas the multiples have positive curvatures. Furthermore, the parabolic approximation of hyperbolic events is better after NMO correction than before, especially for the events that have a small curvature after correction [17]. Consequently, primaries and multiples can be separated in the PRT domain of NMO corrected CMP gathers. Fig. 11(a) shows a simulated noisy post-NMO CMP gather (of $\text{SNR} = 5$ dB) containing primaries and multiples having curvatures around zero and thus a proper mute application can lead to separation of primaries and multiples following their curvatures, provided the events are well focused in the RT domain. The sparse PRT has been applied to the gather and the resulting PRT panel well focused as shown in Fig. 11(b). The events with positive/negative curvatures were separated and inverse transformed to the spatial domain and the results are shown in Figs. 11(c) and 11(d). It is seen that the random noises are suppressed and the primaries and multiples are well separated.

As a real data example, we tested our method on a real CMP gather for multiple attenuation. Fig. 12(a) shows a field CMP gather after NMO correction applied using the velocity of primary reflections. Fig. 12(b) shows the associated sparse parabolic Radon panel generated by the proposed method. As can be seen, the energy of primary reflections are well clustered around p value zero and separated from the multiples energy, enclosed by the rectangle in Fig. 12(b). We just transformed back the multiples energy (Fig. 12(c)) and subtracted it from the original data to get the pri-

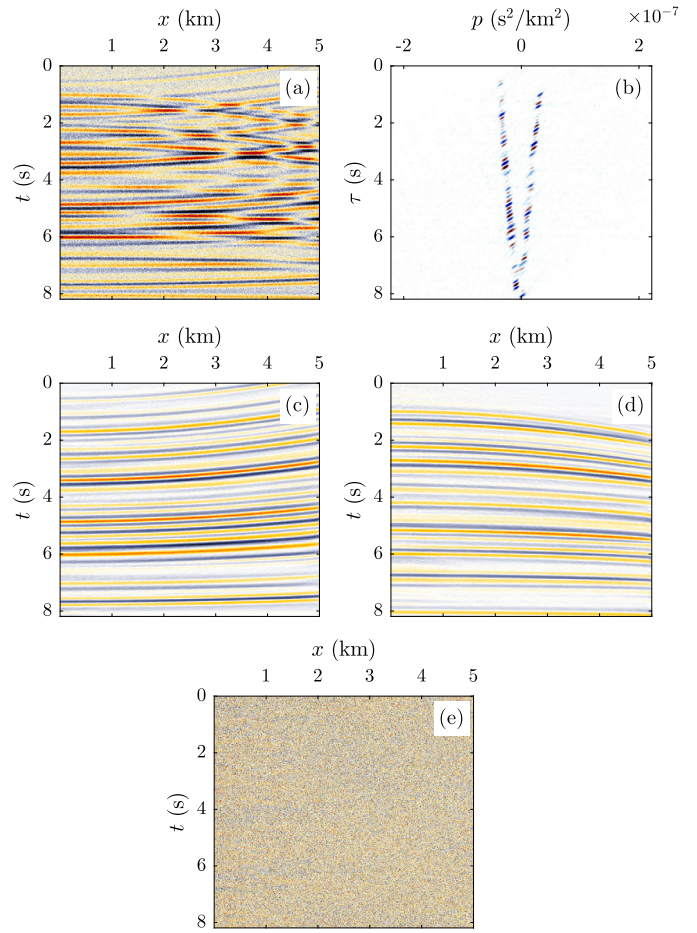


Fig. 11. The sparse PRT used for multiple attenuation. (a) A seismic gather containing primary and multiple waves contaminated with random noise of $\text{SNR} = 5$ dB. (b) The sparse PRT of (a). (c) Predicted primary wavefield corresponding to negative curvatures. (d) Predicted multiple wavefield corresponding to positive curvatures. (e) Predicted noise.

mary only wavefield as shown in Fig. 12(d). It can be seen that the multiples have been attenuated over the full offset range.

As a final example, we tested our algorithm in the presence of missing traces in the input gather. Fortunately, the proposed algorithm explicitly depends upon the spatial coordinates of the traces without requiring x to be uniformly sampled, therefore, it can easily handle irregularly sampled data. When the Radon coefficients have been estimated, the interpolated data can easily be constructed by the inverse Radon transformation to any desired grid points. In order to show this we used 50 percent of the field traces shown in Fig. 12(a) for Radon transformation. Decimated traces and the corresponding Radon coefficients are shown in Figs. 13(a) and (b), respectively. Generated coefficients were used to estimated multiples and primaries as shown in Figs. 13(c) and (d), respectively. It can be seen that the missing traces have little effect on the reconstructed wavefields.

7. Conclusion and discussion

The speed of inverse Radon transformation and its adjoint, solution of a set of Toeplitz systems arising in LS type Radon transforms, and regularization parameter tuning are three main factors affecting the efficiency of the sparse Radon transforms. In this paper we have addressed these factors and described a fast algorithm for calculation of sparse solution to large-scale Radon transforms. A new algorithm was developed for calculation of inverse RT and its adjoint which runs in $O(N_f N \log N + N_x N_t \log N_t)$, as opposed

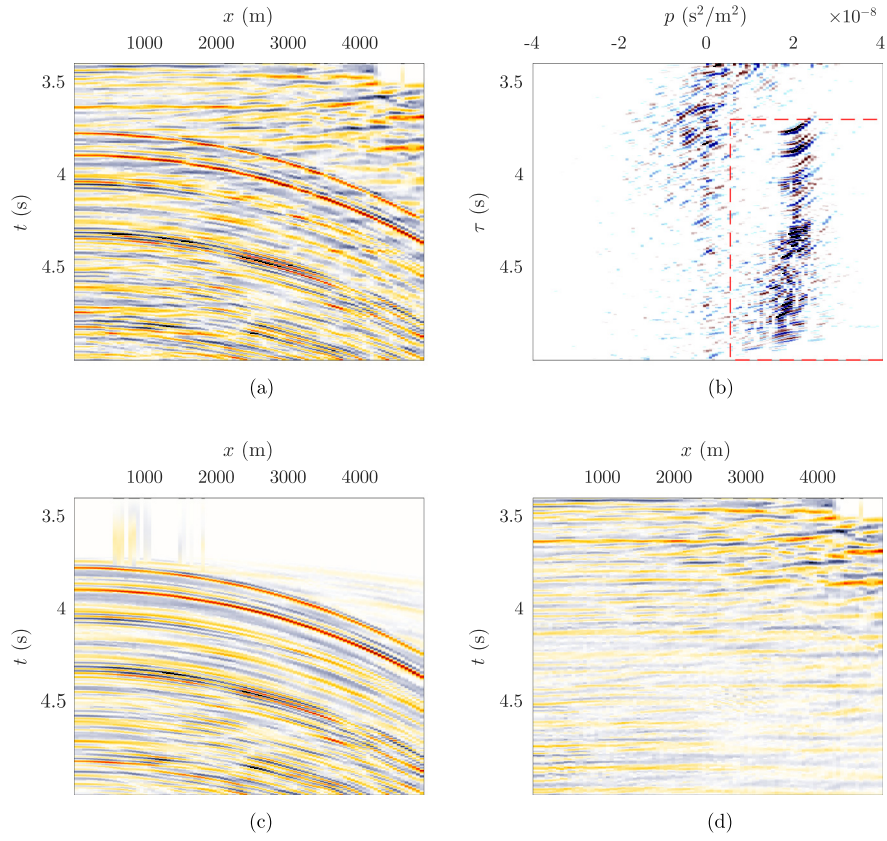


Fig. 12. (a) A field CMP gather, NMO corrected using the velocity of primaries, and (b) associated sparse parabolic Radon transform, generated by the proposed method. (c) Separated multiples, corresponding to the coefficients enclosed by the rectangle in (b). (d) Separated primaries, obtained by subtraction of (c) from (a).

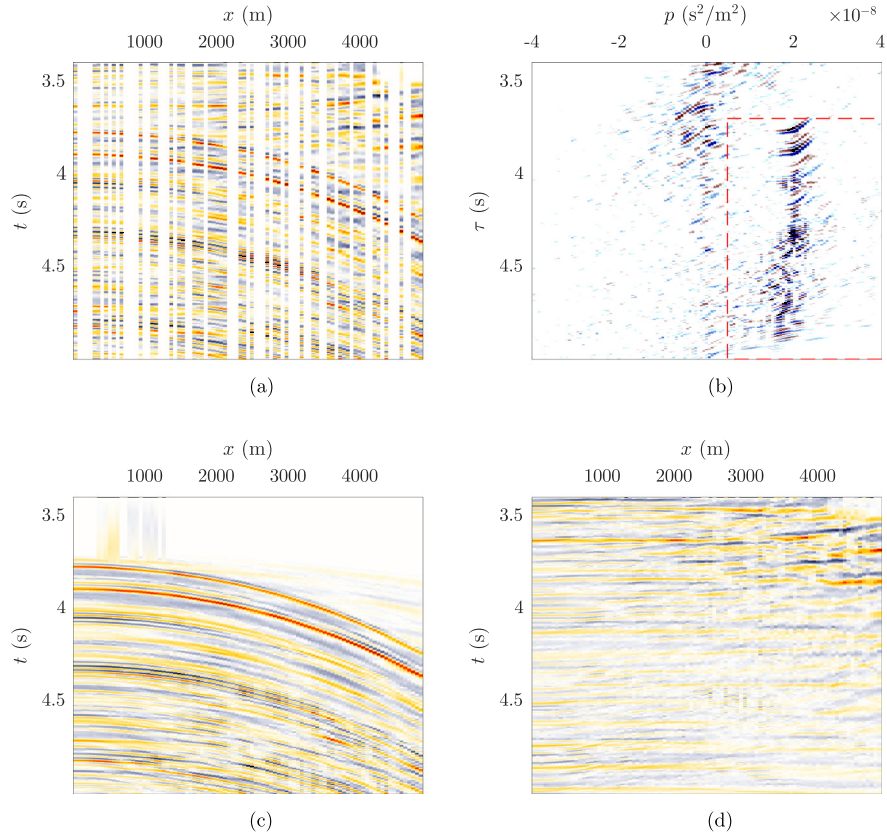


Fig. 13. (a) Decimated field data shown in Fig. 12, generated by randomly removing 50 percent of the traces, and (b) associated sparse parabolic Radon transform, generated by the proposed method. (c) Separated multiples, corresponding to the coefficients enclosed by the rectangle in (b). (d) Separated primaries.

to $O(N_f N_x N_p + N_x N_t \log N_t)$ of the direct computation, where N depends on the maximum frequency and offset in the data set, the maximum curvature in the Radon space, and N_p . These canonical transforms have been utilized within the split Bregman iteration to solve a BPDN problem for sparse RTs. Error forgetting property of the Bregman iteration alongside with its good properties for conditioning the quadratic sub-problem allowed circulant approximation of a Toeplitz system and efficient solution of it by an FFT algorithm.

In practical RTs, the choice of an appropriate value of regularization parameter involves performing the transform with multiple, and often many, values of the parameter. More often this limits their usage for processing large-scale seismic data. In the new algorithm, amount of regularization is controlled by iteration number. We obtained a good estimate of PMSE score via GCV analysis and automatically determined the optimum number of iterations at a negligible cost, leading to a fast time-domain algorithm for sparse RT. Numerical experiments confirmed the accuracy and efficiency of the proposed algorithm.

References

- [1] J. Radon, Über die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten, *Ber. Sachs. Akad. Wiss.* 69 (1917) 262–277.
- [2] F. Rieber, A new reflection system with controlled directional sensitivity, *Geophysics* 1 (1936) 97–106.
- [3] P. Schult, J. Claerbout, Velocity estimation and downward continuation by wavefront synthesis, *Geophysics* 43 (1978) 691–714.
- [4] P. Stoffa, P. Buhl, J. Diebold, F. Wenzel, Direct mapping of seismic data to the domain of intercept time and ray parameter – a plane-wave decomposition, *Geophysics* 46 (1981) 255–267.
- [5] G. Turner, Aliasing in the tau-p transform and the removal of spatially aliased coherent noise, *Geophysics* 55 (1990) 1496–1503.
- [6] E. Maeland, Sampling, aliasing, and inverting the linear Radon transform, *Geophysics* 69 (3) (2004) 859–861.
- [7] C. Chapman, Generalized Radon transforms and slant stacks, *Geophys. J. R. Astron. Soc.* 66 (1981) 445–453.
- [8] R. Tatham, Multidimensional filtering of seismic data, *Proc. IEEE* 72 (10) (1984) 1357–1369.
- [9] K. Kostov, Toeplitz structure in slant-stack inversion, in: 60th SEG Annual Meeting, 1990, pp. 1618–1621.
- [10] G. Beylkin, Discrete Radon transform, *IEEE Trans. Acoust. Speech Signal Process.* 35 (1987) 162–172.
- [11] A. Averbuch, R. Coifman, D.L. Donoho, M. Israeli, Y. Shkonisky, A framework for discrete integral transformation I – The pseudo-polar Fourier transform, *SIAM J. Sci. Comput.* 30 (2) (2008) 764–784.
- [12] A. Averbuch, R. Coifman, D.L. Donoho, M. Israeli, Y. Shkonisky, I. Sedelnikov, A framework for discrete integral transformation II – The 2D discrete Radon transform, *SIAM J. Sci. Comput.* 30 (2) (2008) 785–803.
- [13] R. Tatham, J. Keeney, I. Noponen, Application of the tau-p transform (slant stack) in processing seismic reflection data, in: 52nd Annual International SEG Meeting, Dallas, 1982.
- [14] S. Treitel, P. Gutowski, D. Wagner, Plane-wave decomposition of seismograms, *Geophysics* 47 (1982) 1375–1401.
- [15] Y. Luo, J. Xia, R. Miller, Y. Xu, J. Liu, Q. Liu, Rayleigh-wave mode separation by high-resolution linear Radon transform, *Geophys. J. Int.* 179 (2009) 254–264.
- [16] R. Askari, S. Hejazi, Estimation of surface-wave group velocity using slant stack in the generalized S-transform domain, *Geophysics* 80 (4) (2015) EN83–EN92.
- [17] D. Hampson, Inverse velocity stacking for multiple elimination, *J. Can. Soc. Explor. Geophys.* 22 (1) (1986) 44–55.
- [18] J. Thorson, J. Claerbout, Velocity-stack and slant-stack stochastic inversion, *Geophysics* 50 (1985) 2727–2741.
- [19] Ö. Yilmaz, Velocity-stack processing, *Geophys. Prospect.* 37 (1989) 357–382.
- [20] D. Foster, C. Mosher, Suppression of multiple reflections using the Radon transform, *Geophysics* 57 (1992) 386–395.
- [21] P. Herrmann, T. Mojesky, M. Magesan, P. Hugonnet, De-aliased, high-resolution Radon transforms, in: SEG Technical Program Expanded Abstracts, 2000, pp. 1953–1956.
- [22] D. Darche, Spatial interpolation using a fast parabolic transform, in: SEG Annual Meeting, San Francisco, 1990.
- [23] M. Sacchi, T. Ulrych, High-resolution velocity gathers and offset space reconstruction, *Geophysics* 4 (1995) 1169–1177.
- [24] D. Trad, D. Ulrych, M. Sacchi, Accurate interpolation with high-resolution time-variant Radon transforms, *Geophysics* 67 (2002) 644–656.
- [25] C.L. Lawson, Contributions to the Theory of Linear Least Maximum Approximation, Ph.D. thesis, University of California, Los Angeles, 1961.
- [26] J.A. Scales, A. Gersztenkorn, S. Treitel, Fast ℓ_p solution for large, sparse linear systems: application to seismic travel time tomography, *J. Comput. Phys.* 75 (1988) 314–333.
- [27] M.D. Sacchi, M. Porsani, Fast high resolution parabolic Radon transform, in: SEG Technical Program Expanded Abstracts, 1999, pp. 1477–1480.
- [28] D. Trad, D. Ulrych, M. Sacchi, Latest views of the sparse Radon transform, *Geophysics* 68 (2003) 386–399.
- [29] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1) (1998) 33–61.
- [30] T. Goldstein, S. Osher, The split Bregman method for l1 regularized problems, *SIAM J. Imaging Sci.* 2 (2) (2009) 323–343.
- [31] J. Proakis, D. Manolakis, Digital Signal Processing, Prentice Hall, 2006.
- [32] N. Levinson, The Wiener RMS error criterion in filter design and prediction, *J. Math. Phys.* 25 (1947) 261–278.
- [33] W. Yin, S. Osher, Error forgetting of Bregman iteration, *J. Sci. Comput.* 54 (2013) 684–695.
- [34] M. Schonewille, A. Duijndam, Parabolic Radon transform, sampling and efficiency, *Geophysics* 66 (2) (2001) 667–678.
- [35] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for l1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.* 1 (2008) 143–168.
- [36] A. Gholami, M.D. Sacchi, A fast and automatic sparse deconvolution in the presence of outliers, *IEEE Trans. Geosci. Remote Sens.* 50 (10) (2012) 4105–4116.
- [37] A. Gholami, M.D. Sacchi, Fast 3D blind seismic deconvolution via constrained total variation and GCV, *SIAM J. Imaging Sci.* 6 (4) (2013) 2350–2369.
- [38] A. Gholami, M.D. Sacchi, Time-invariant Radon transform by generalized Fourier slice theorem, *Inverse Probl. Imaging* 11 (3) (2017) 501–519.
- [39] C. Turnes, D. Balcan, J. Romberg, Superfast Tikhonov regularization of Toeplitz systems, *IEEE Trans. Signal Process.* 62 (15) (2014) 3809–3821.
- [40] R. Chan, M. Ng, Conjugate gradient methods for Toeplitz systems, *SIAM Rev.* 38 (1996) 427–482.
- [41] T. Chan, An optimal circulant preconditioner for Toeplitz systems, *SIAM J. Sci. Stat. Comput.* 9 (1988) 766–771.
- [42] M. Cai, X. Jin, A note on T. Chan's preconditioner, *Linear Algebra Appl.* 376 (2004) 283–290.
- [43] G. Wahba, Practical approximate solution of linear operator equations when the data are noisy, *SIAM J. Numer. Anal.* 14 (4) (1977) 651–667.
- [44] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2 (2009) 183–202.
- [45] X. Zhang, M. Burger, X. Bresson, S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM J. Imaging Sci.* 3 (3) (2010) 253–276.
- [46] A. Alam, J. Austin, Suppression of multiples using slant stacks, in: 51st SEG Technical Program Expanded Abstracts, 1981, pp. 3225–3257.
- [47] Ö. Yilmaz, Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data, Vol. I, Society of Exploration Geophysics, 2008.
- [48] H. Masoomzadeh, Processing Techniques for Wide-angle Seismic Data, Ph.D. thesis, University of Cambridge, 2006.

Ali Gholami is an Associate Professor at the Institute of Geophysics, University of Tehran, Iran. His interests are in the area of statistical signal processing in geophysics, inverse problems, seismic data processing, seismic inversion and imaging. His work involves the development of fast techniques for solving large-scale problems in geophysics.

Toktam Zand received a B.Sc. degree in physics and M.Sc. degree in geophysics from the University of Tehran, Iran. She is now a PhD student in geophysics at the Institute of Geophysics, University of Tehran.