

## Sparse Radon transform with dual gradient ascent method

Yujin Liu<sup>\*†</sup>, Zhimin Peng<sup>†</sup>, William W. Symes<sup>†</sup> and Wotao Yin<sup>†</sup>, <sup>\*</sup>China University of Petroleum (Huadong), <sup>†</sup>Rice University

### SUMMARY

The Radon transform suffers from the typical problems of loss of resolution and aliasing that arise as a consequence of incomplete information, such as limited aperture and discretization. Sparseness in Radon domain, which is equivalent to assuming smooth amplitude variation in the transition between known and unknown (missing) data, is a valid and useful prior information (Trad et al., 2003). The most commonly used method to solve sparsity-promotion inverse problems in geophysics is reweighted least-squares inversion (IRLS) method. As IRLS method needs to compute the weighting function iteratively at the outer loop of conjugate gradient iteration, the computational cost is very expensive. In this abstract, we adopt the dual gradient ascent methods, developed in compressive sensing into geophysics and compare them with an updated version of IRLS, namely conjugate guided gradient method (CGG). Numerical tests show that the dual gradient ascent method with Nesterov's acceleration (DGAN) can provide results with higher resolution than CGG method after a few iterations, which is also of great potential in other seismic applications.

### INTRODUCTION

The Radon transform (RT) is widely used in geophysics, such as signal/noise separation (Harlan et al., 1984; Hampson, 1986; Sava and Guitton, 2005), interpolation (Sacchi and Ulrych, 1995; Trad et al., 2002), velocity analysis (Thorson and Claerbout, 1985) and so on. According to the integral formula, it can be divided into several classes, including linear RT (also known as slant stack), parabolic RT, hyperbolic RT (also known as velocity stack) et al. There are two kinds of implementations of RT, i.e. time domain (Stoffa et al., 1981; Cary, 1998) and frequency domain (Beylkin, 1987; Sacchi and Ulrych, 1995; Jingwei Hu, 2012). In this abstract, we will focus on time-domain hyperbolic RT.

As RT is not orthogonal like Fourier transform, wavelet transform and curvelet transform, it's not trivial to apply forward and inverse transform. One most commonly used way is least-squares inversion (Hampson, 1986; Beylkin, 1987). Due to the limited information we can obtain, the resolution is often decreased if we use Tikhonov regularization, which assumes the solutions be of Gaussian distribution (Tarantola, 2005). In order to promote sparsity of RT, other kinds of norm and criterion to measure the solution was proposed by geophysicists, like  $\ell_1$  norm (Claerbout and Muir, 1973), Huber norm (Guitton and Symes, 2003), Hybrid norm (Bube and Langan, 1997), Cauchy criterion (Sacchi and Ulrych, 1995) and so on.

One of the most commonly used algorithms to solve sparse Radon transform is Iteratively Reweighted Least-Squares (IRLS)

method. As the weighting function is related with model, the original linear problem is changed into non-linear problem, IRLS needs to calculate the weighting function iteratively after several inner iterations of conjugate gradient method, so the computational cost is high if we want to get the result with high resolution. Another variant of IRLS, namely CGG method, guides the gradient vector during the iteration and only need performing one times of operator and its adjoint, so it's more efficient than IRLS (Ji, 2006). In this abstract, we introduce dual gradient ascent method developed in compressive sensing into solving this problem. Numerical tests show that, compared with CGG method, our proposed method can provide a higher resolution after a few iterations.

### METHOD

In this section, we will discuss how to apply dual gradient ascent method to hyperbolic Random Transform (HRT). The formula of HRT can be expressed as:

$$m(\tau, v) = \sum_{x=x_{min}}^{x_{max}} d(t^2 = \tau^2 + \frac{x^2}{v^2}, x)$$

where  $d(t, \mathbf{x})$  denotes common-midpoint (CMP) gather and  $m(\tau, v)$  denotes stack velocity spectrum. The velocity-time pair  $(\tau, v)$  are the coordinate axes of the stack velocity spectrum and the offset-time pair  $(t, x)$  are the coordinate axes of the CMP gather. The adjoint of HRT is:

$$d(t, x) = \sum_{v=v_{min}}^{v_{max}} m(\tau^2 = t^2 - \frac{x^2}{v^2}, v)$$

which can be transformed into a compact form as:

$$\mathbf{d} = \mathbf{L}\mathbf{m}$$

In order to improve the resolution of RT,  $\ell_1$  norm is commonly used to regularize the potential velocity  $\mathbf{m}$ , which results in the following Basis Pursuit (BP) problem:

$$\min_{\mathbf{m}} \{ \|\mathbf{m}\|_1 : \mathbf{L}\mathbf{m} = \mathbf{d} \} \quad (1)$$

One way to solve (1) is IRLS method, which introduces weighting matrix  $\mathbf{W}_{\mathbf{m}} = \text{diag}(m_i^{-\frac{1}{2}})$  and solves an equivalent form of (1):

$$\min_{\mathbf{m}} \{ \|\mathbf{W}_{\mathbf{m}}\mathbf{m}\|_2^2 : \mathbf{L}\mathbf{m} = \mathbf{d} \} \quad (2)$$

As the weighting matrix is related with model itself, we need to compute it iteratively at the outer loop of inner conjugate gradient iterations, which makes it expensive. Conjugate guided gradient (CGG) method is another variant of IRLS method (Ji, 2006) which puts the calculation of weighting function into the inner loop of CG, so that there is only one iteration loop and one calculation of  $\mathbf{L}$  and  $\mathbf{L}^T$  is needed at each iteration.

## Sparse Radon transform with dual gradient ascent method

Instead of solving (1) directly, we solve the following the  $\ell_1\ell_2$  problem:

$$\min_{\mathbf{m}} \{ \|\mathbf{m}\|_1 + \frac{1}{2\alpha} \|\mathbf{m}\|_2^2 : \mathbf{L}\mathbf{m} = \mathbf{d} \} \quad (3)$$

and use its solution to approximate the solution of (1). In fact, as long as the smooth parameter  $\alpha$  is greater than a certain value, the solutions for (1) and (3) are identical. Notice that (3) is still a non-smooth constrained optimization problem, even though the smooth term  $\frac{1}{2\alpha} \|\mathbf{m}\|_2^2$  is added to the non-smooth objective function. In fact, the advantage of (3) over (1) is that its dual problem:

$$\min_{\mathbf{y}} \{ g(\mathbf{y}) = -\mathbf{d}^T \mathbf{y} + \frac{\alpha}{2} \|\mathbf{L}^T \mathbf{y} - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y})\|_2^2 \} \quad (4)$$

is first order differentiable and its gradient is:

$$\nabla g(\mathbf{y}) = -\mathbf{d} + \alpha \mathbf{L}(\mathbf{L}^T \mathbf{y} - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y})) \quad (5)$$

where  $\text{Proj}_{[-1,1]^n}(\mathbf{x})$  projects  $\mathbf{x}$  into  $[-1,1]^n$ . An example of  $g(\mathbf{y})$  and  $\nabla g(\mathbf{y})$  is shown in Fig. 1.

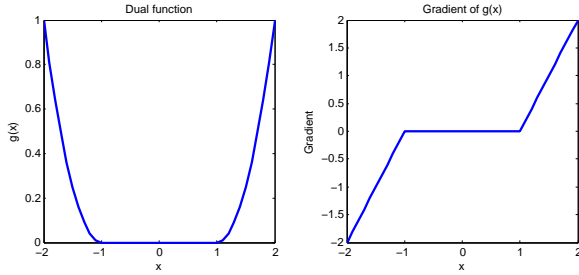


Figure 1: Dual objective function  $g(\mathbf{y})$  (left); and its derivative  $\nabla g(\mathbf{y})$  (right)

Note that objective function  $g(\mathbf{y})$  is a convex function and its gradient  $\nabla g(\mathbf{y})$  is not smooth, hence, we can only apply first order methods to solve (4) and derive the following update scheme:

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \delta \nabla g(\mathbf{y}^k) \quad (6)$$

where  $\delta > 0$  is the step size. After the sequence  $\{\mathbf{y}^k\}$  generated by (6) converges to the optimal solution  $\mathbf{y}^*$ , the first-order optimal condition  $\nabla g(\mathbf{y}^*)$  should be satisfied, i.e.,  $\alpha \mathbf{L}(\mathbf{L}^T \mathbf{y}^* - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y}^*)) = \mathbf{d}$ . Let  $\mathbf{m}^* = \mathbf{L}^T \mathbf{y}^* - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y}^*)$ , then we have  $\mathbf{L}\mathbf{m}^* = \mathbf{d}$ , thus  $\mathbf{m}^*$  is a feasible solution to the original problem, it has been proved that the objective value of the primal problem given by  $\mathbf{m}^*$  matches the optimal value of the dual objective given by  $\mathbf{y}^*$  (Yin, 2010). Hence,  $\mathbf{m}^*$  is also optimal. To conclude, we have the following fixed step size dual gradient ascent algorithm:

---

### Algorithm 1 DGA with fixed stepsize

---

- 1: **for**  $k = 0, 1, \dots, niter$  **do**
  - 2:    $\mathbf{y}^{k+1} = \mathbf{y}^k + \delta(\mathbf{d} - \mathbf{L}\mathbf{m}^k)$
  - 3:    $\mathbf{m}^{k+1} = \alpha(\mathbf{L}^T \mathbf{y}^{k+1} - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y}^{k+1}))$
  - 4: **end for**
- 

It is natural to speedup fixed stepsize gradient descent method by incorporating methods such as line search, quasi-Newton

methods and Nesterov's acceleration scheme (Nesterov, 2007), all of which requires only gradient computations. Here we will focus on Nesterov's method. Instead of only using information from previous iteration, Nesterov's method use the usual projection-like step, evaluated at an auxiliary point which is constructed by a special linear combination of the previous two points  $\mathbf{z}^{k-1}, \mathbf{z}^k$ , the accelerated algorithm is described as:

---

### Algorithm 2 DGA with Nesterov's acceleration

---

- 1:  $\theta_0 = 1, h > 0$
  - 2: **for**  $k = 0, 1, \dots, niter$  **do**
  - 3:    $\beta_k = \frac{(1-\theta_k)(\sqrt{\theta_k^2+4}-\theta_k)}{2}$
  - 4:    $\mathbf{z}^{k+1} = \mathbf{y}^k + \delta(\mathbf{d} - \mathbf{L}^T \mathbf{m}^k)$
  - 5:    $\mathbf{y}^{k+1} = \mathbf{z}^{k+1} + \beta_k(\mathbf{z}^{k+1} - \mathbf{z}^k)$
  - 6:    $\mathbf{m}^{k+1} = \alpha(\mathbf{L}^T \mathbf{y}^{k+1} - \text{Proj}_{[-1,1]^n}(\mathbf{L}^T \mathbf{y}^{k+1}))$
  - 7:    $\theta_{k+1} = \theta_k \frac{\sqrt{\theta_k^2+4}-\theta_k}{2}$
  - 8: **end for**
- 

## NUMERICAL TESTS

### Synthetic data

To examine the performance of the proposed DGAN method, a synthetic CMP data set with various types of noise and missing traces is used. Figure 2(a) shows the original synthetic data modeled by the adjoint of HRT with five spikes. Three of them simulate primaries and two of them simulate multiples. Moreover, Gaussian noise is added in the original data and 60% of traces are removed randomly. The input data is shown in figure 2(b).

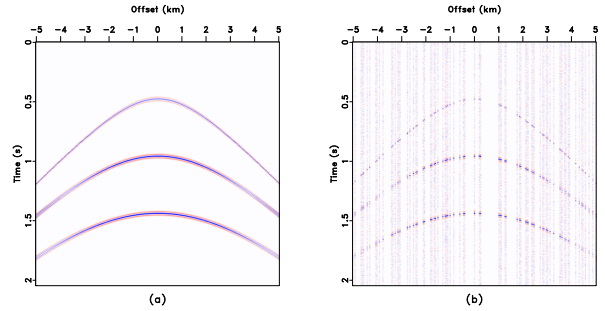


Figure 2: (a) Original data; (b) Input data

As we can see from figure 2(b), it's very hard to discriminate primaries from multiples when the data is under-sampled and the S/N is low. However, if we assume that all the events are consistent with the formula of HRT, which is true in our synthetic test, we can make use of sparsity of Radon transform. CGG and DGAN are used to solve this sparsity-promoting optimization problem respectively.

In our first test, the iteration is performed 50 times. From the relative model error ( $\frac{\|\mathbf{m}^k - \mathbf{m}^*\|}{\|\mathbf{m}^*\|}$ ,  $k$  is the iteration number) curves of CGG and DGAN method (figure 3), we can see that

## Sparse Radon transform with dual gradient ascent method

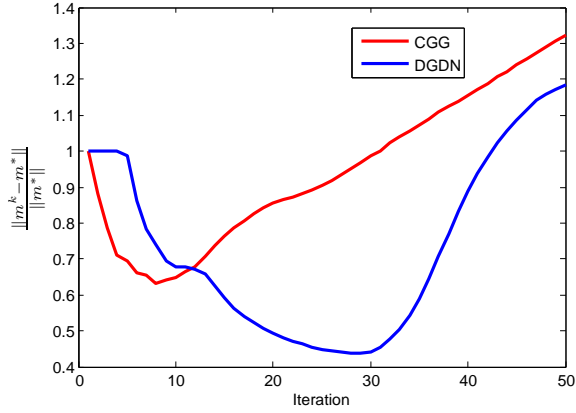


Figure 3: Relative model error curve of CGG method and DGDN method

both of them are decreased at the first stage of iterations and increased afterwards. That's because both of them are over-fitted after too many iterations. However, we should also note that the model error of DGDN is smaller than CGG after about 10 iterations and it can achieve a much smaller value than CGG method, which demonstrates that the proposed method is more accurate than CGG method in sparsity promotion after certain number of iterations.

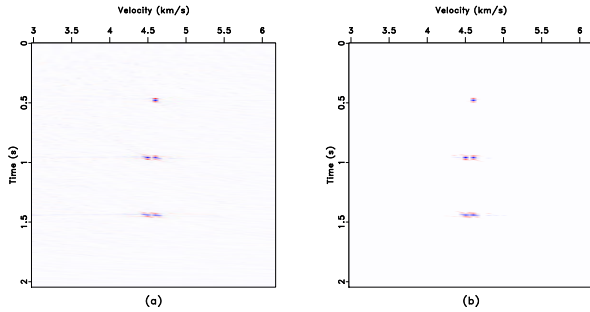


Figure 4: Inversion result with (a) CGG method; (b) DGDN method

Figure 4 shows the inversion results of CGG and DGDN at their best iteration number from the relative model error curve, i.e., 8 iterations for CGG and 30 iterations for DGDN respectively. From these figures, we can see that the best solution of DGDN is more sparse than CGG. After picking a line in the middle of two different velocity and muting the low velocity part in the Radon domain, which often corresponds to multiples, we can reconstruct the CMP gather using the adjoint of HRT.

From the reconstructed results of CGG and DGDN as shown in Figure 5(a) and 5(b), we can see that both of these two methods can suppress random noise, eliminate multiples and interpolate missing traces successfully, but CGG method also introduces some coherent noise, which is due to the lower sparse solution it obtained.

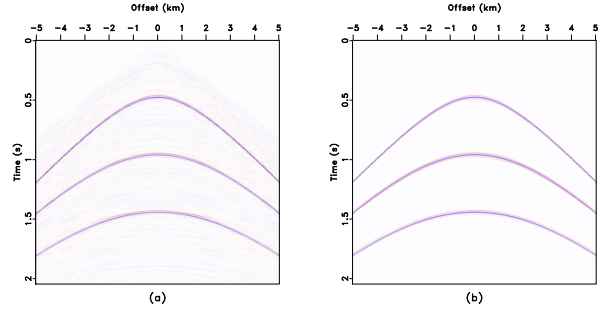


Figure 5: Reconstructed data with (a) CGG method; (b) DGDN method

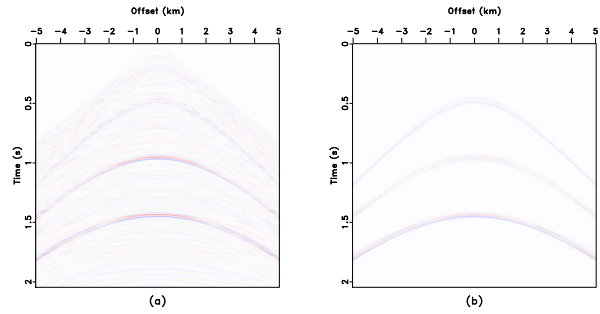


Figure 6: Residual with (a) CGG method; (b) DGDN method

Figure 6 is the difference in the same scale between reconstructed results and the correct "answer", which can be modeled by the adjoint of HRT with the three spikes at the right side. From these figures, we can also see that the proposed method provides the better result after denoising and interpolation.

### Real data

We test the proposed method on a real data set that contains various types of noise and missing traces. The data set is a super CMP gather from a marine survey. The trajectories of the events in the data set look "hyperbolic" enough to be tested with HRT. Figure 7 shows the real data set after data binning. From this figure, we can see that there are lots of missed traces and multiples.

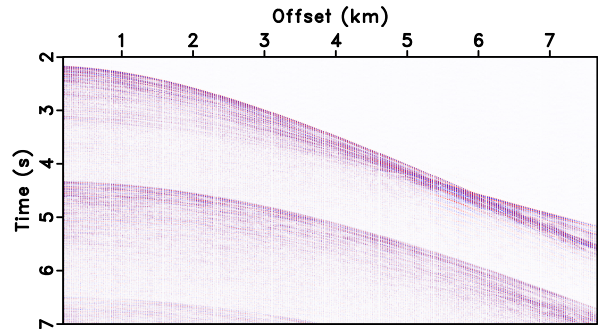


Figure 7: CMP gather after data binning

## Sparse Radon transform with dual gradient ascent method

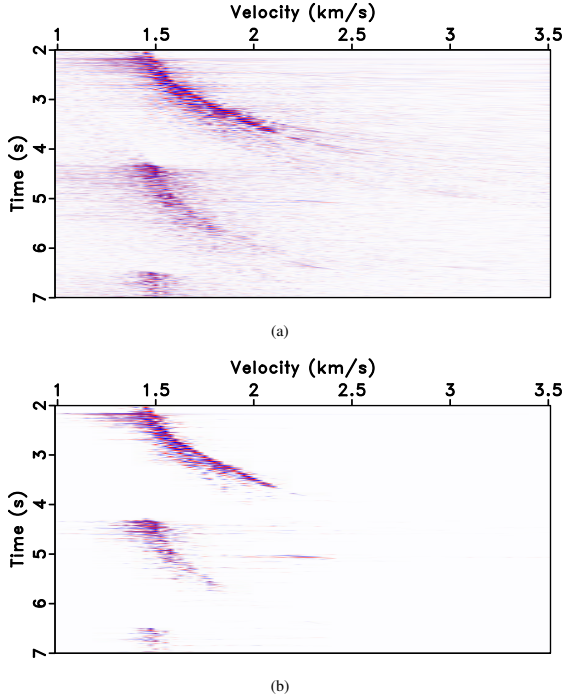


Figure 8: Inversion result with (a) CGG method; (b) DGAN method

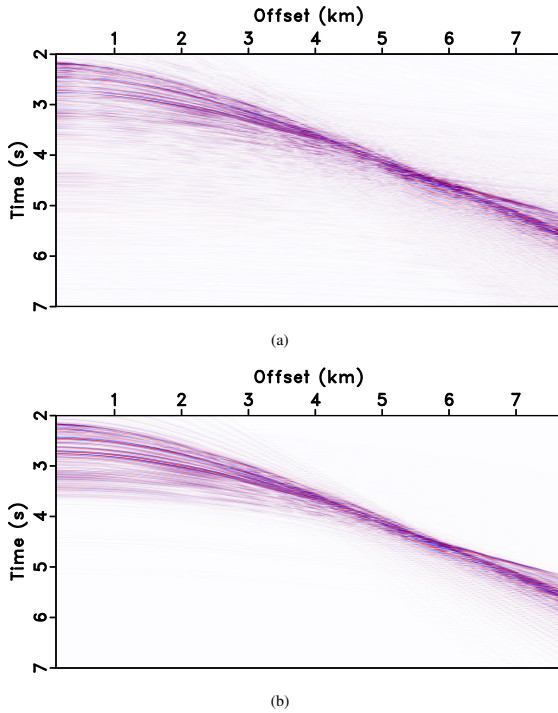


Figure 9: Denoising and interpolation result with (a) CGG method; (b) DGAN method

Figure 8(a) and Figure 8(b) shows the velocity spectrum calculated by HRT with CGG and DGAN method after 30 iterations respectively. From these two figures, we can see that both of them can get a sparse solution and we can separate primaries and multiples visually in Radon domain. Moreover, the results obtained by DGAN method is more sparse than CGG method and has fewer non-zeros.

Figure 9(a) and Figure 9(b) are the denoising and interpolation results after mutting in the Radon transform and applying adjoint of HRT. From these two figures, we can see that multiples have been suppressed more completely using DGAN method. However, we should note that some events have been smoothed by these processes as sparse constraint is equivalent to smoothness along trajectories defined by RT operator.

## CONCLUSION

With the help of different transformation, most signals, including seismic wave, can be expressed in a sparse form, which implies that sparsity is a very important prior information in many applications. Dual gradient ascent, which was recently developed in compressive sensing, is introduced into Radon transform in this abstract. We incorporate Nesterov's acceleration scheme to the fixed stepsize dual gradient ascent method and obtain a faster algorithm namely DGAN. Compared to CGG method, DGAN outperforms it in the following aspects:

- The sparsity level of the solution is higher;
- Reconstruction results have much less coherent noise;
- More accurate solution can be obtained with a few iterations.

However, there are still lots of challenges in this field, such as how to choose an appropriate sparse representation of seismic wave. More work needs to do in order to apply this method in seismic imaging and velocity analysis.

## ACKNOWLEDGMENTS

We want to thank Hui Zhang, Ming Yan for helpful discussion. The work of the first author is supported by China Scholarship Council during his visit to Rice University. YL thanks TRIP (The Rice Inversion Project), CAAM Department, Rice University for hosting him. WWS and YL thank the sponsors of TRIP for their support.