Irregular seismic data reconstruction using a percentile-half-thresholding algorithm[a]

[a]Published in Journal of Geophysics and Engineering, 11, no. 6, 065001, (2014)

Yangkang Chen[1], Keling Chen[2], Peidong Shi[3] and Yanyan Wang[3]

ABSTRACT

In this paper, a percentile-half-thresholding approach is proposed in the transformed domain thresholding process for iterative shrinkage thresholding (IST). The percentile-thresholding strategy is more convenient for implementing than the constant-value, linear-decreasing, or exponential-decreasing thresholding because it's data-driven. The novel half-thresholding strategy is inspired from the recent advancement in the researches on optimization using non-convex regularization. We summarize a general thresholding framework for IST and show that the only difference between half thresholding and the conventional soft or hard thresholding lays in the thresholding operator. Thus it's straightforward to insert the existing percentile-thresholding strategy to the half-thresholding iterative framework. We use both synthetic and field data examples to compare the performances using soft thresholding or half thresholding with constant threshold or percentile threshold. Synthetic and field data show consistent results that apart from the threshold-setting convenience, the percentile thresholding also has the possibility for improving the recovery performance. Compared with soft thresholding, half thresholding tends to have a more precise reconstructed result.

INTRODUCTION

Seismic data interpolation plays a fundamental role in seismic data processing, which provides the regularly sampled seismic data for the following jobs like 3D SRME and wave equation-based migrations (Xu et al., 2010). There has been a number of effective methods to recover missing seismic traces, these methods can be generally divided into three types. The first kind is based on a convolution operator, which utilizes a prediction filter computed from the low-frequency parts to predict the high-frequency components (Spitz, 1991; Porsani, 1999; Wang, 2002). However, the predictive filtering method can only be applied to regularly sampled seismic data (Naghizadeh and Sacchi, 2007). The second type is a transformed domain method, which is based on a sparsity assumption and makes use of the theory of compressive sensing or compressive sampling (CS) (Candès et al., 2006; Donoho, 2006; Mallat, 2009; Herrmann, 2010) to achieve a successful recovery using highly incomplete available data (Sacchi et al., 1998; Wang, 2003; Zwartjes and Gisolf, 2007; Zwartjes and Sacchi, 2007; Yang et al., 2012, 2013b). Naghizadeh and Sacchi (2007) proposed a multistep autoregressive strategy which combines the first two types of methods to reconstruct irregular seismic data. The third type is based on the integral of continuation operators. This integral is performed based on the traveltime calculation from a velocity model, thus it depends on the known velocity model, which also becomes its limitation (Canning and Gardner, 1996; Bleistein and Jaramillo, 2000; Stolt, 2002; Fomel, 2003).

A recently very popular way for interpolating irregularly sampled seismic data is by using iterative thresholding, such as projection onto convex sets (POCS) and iterative shrinkage thresholding (IST), mainly because of their simple formulations and convenient implementations. Different thresholding approaches will lead to different performances for interpolation. Because of the advancement of the non-convex optimization algorithms, a half-thresholding algorithm has been proposed both in signal-processing and exploration geophysics communities (Yang et al., 2013a; Xu et al., 2012). In this paper, we summarize a general thresholding framework for IST algorithm. The only difference among soft, hard, and half thresholding lays in the thresholding operator. Thus, a percentile thresholding strategy, which is particularly useful in practice, can be straightforwardly applied to the case of half thresholding. We use both synthetic and field data examples to demonstrate the performance of our proposed percentile-half-thresholding algorithm, compared with other three existing approaches.

THEORY

Seismic interpolation

The basic target of seismic interpolation is to solve the following equation:

$$\mathbf{d}_{obs} = \mathbf{M}\mathbf{d}, \tag{1}$$

where $\mathbf{d}_{obs}$ is the observed data which is regularly or irregularly sampled, $\mathbf{d}$ is the unknown data we want to reconstruct and $\mathbf{M}$ is the sampling matrix. The sampling operator has a diagonal structure, which is composed by zero and identity matrix:

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} & & & & & \\ & \mathbf{O} & & & & \\ & & \mathbf{I} & & & \\ & & & \mathbf{I} & & \\ & & & & \ddots & \\ & & & & & \mathbf{I} \end{bmatrix}. \tag{2}$$

Each $\mathbf{I}$ in equation 2 corresponds to sampling a trace, and each $\mathbf{O}$ corresponds to missing a trace.

As equation 1 is under-determined, additional constraint is required in order to solve the equation. By applying a regularization term, we get a least-squares minimization solution for solving equation 1:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \|\mathbf{d}_{obs} - \mathbf{M}\mathbf{d}\|_2^2 + \mathbf{R}(\mathbf{d}), \tag{3}$$

where $\mathbf{R}$ is a regularization operator and $\|\cdot\|_2^2$ denotes the square of L_2 norm.

Iterative shrinkage thresholding

In order to reconstruct the missing traces in the seismic data, one can use a sparsity-promoting transform to precondition $\mathbf{d}$ in equation 1, that is:

$$\mathbf{d} = \mathbf{A}\mathbf{x}, \tag{4}$$

where $\mathbf{A}$ is a tight frame such that $\mathbf{x} = \mathbf{A}^H\mathbf{d}$ and $\mathbf{A}^{-1} = \mathbf{A}^H$, and $[\cdot]^H$ denotes adjoint. A common selection for $\mathbf{A}$ is the Fourier transform. Inserting equation 4 into equation 1, and let $\mathbf{K} = \mathbf{M}\mathbf{A}$, we obtain

$$\mathbf{d}_{obs} = \mathbf{K}\mathbf{x}. \tag{5}$$

Correspondingly, equation 3 turns to:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{d}_{obs} - \mathbf{K}\mathbf{x}\|_2 + \mathbf{R}'(\mathbf{x}) \tag{6}$$

The well-known IST algorithm is used for solving equation 6 with a sparsity constraint:

$$\mathbf{x}_{n+1} = \mathbf{T}_{\gamma(\tau,p)}[\mathbf{x}_n + \mathbf{K}^H(\mathbf{d}_{obs} - \mathbf{K}\mathbf{x}_n)]. \tag{7}$$

Here $\mathbf{T}_{\gamma(\tau,p)}$ corresponds to a thresholding operator performed element-wise with threshold $\gamma(\tau,p)$ (Yang et al., 2013a). When $p=1$, $\gamma(\tau,1)=\tau$, $\mathbf{R}'(\cdot) = \tau\|\cdot\|_1$, where τ is a regularization parameter which controls the weight between misfit and constraint in the minimization problem, $\mathbf{T}_{\gamma(\tau,1)}$ corresponds to a soft-thresholding operator:

$$\mathbf{T}_{\gamma(\tau,1)}[v(\mathbf{x})] = \begin{cases} v(\mathbf{x}) - \gamma\frac{v(\mathbf{x})}{|v(\mathbf{x})|} & \text{for} \quad |v(\mathbf{x})| > \gamma(\tau,1) \\ 0 & \text{for} \quad |v(\mathbf{x})| \leq \gamma(\tau,1) \end{cases}, \tag{8}$$

where $v(\mathbf{x})$ denotes the amplitude of each position-coordinate vector $\mathbf{x}$. When $p=0$, $\gamma(\tau,0)=\sqrt{2\tau}$, $\mathbf{R}'(\cdot)=\tau\|\cdot\|_0$, $\mathbf{T}_{\gamma(\tau,0)}$ corresponds to a hard-thresholding operator:

$$\mathbf{T}_{\gamma(\tau,0)}[v(\mathbf{x})] = \begin{cases} v(\mathbf{x}) & \text{for} \quad |v(\mathbf{x})| > \gamma(\tau,0) \\ 0 & \text{for} \quad |v(\mathbf{x})| \leq \gamma(\tau,0) \end{cases}, \tag{9}$$

Percentile half thresholding

Recent research suggests that it's possible to reconstruct the seismic data using non-convex L_p-norm minimization, $0 < p < 1$. Particularly, iterative half thresholding has been developed both in signal processing and exploration geophysics communities (Xu et al., 2012; Yang et al., 2013a). The difference between half thresholding and the conventional soft thresholding is just the thresholding operator. When $p = 1/2$, $\gamma(\tau, 1/2) = \frac{3}{2}\tau^{2/3}$, $\mathbf{R}'(\cdot) = \tau\|\cdot\|_{1/2}^{1/2}$, $\mathbf{T}_{\gamma(\tau,1/2)}$ becomes a half-thresholding operator:

$$\mathbf{T}_{\gamma(\tau,1/2)}[v(\mathbf{x})] = \tag{10}$$

$$\begin{cases} \frac{2}{3}v(\mathbf{x})\left(1+\cos(\frac{2}{3}\pi - \frac{2}{3}\arccos(\frac{\tau}{8}(\frac{|v(\mathbf{x})|}{3})^{-\frac{3}{2}}))\right) \\ \qquad \text{for} \quad |v(\mathbf{x})| > \gamma(\tau, 1/2) \\ 0 \\ \qquad \text{for} \quad |v(\mathbf{x})| \leq \gamma(\tau, 1/2) \end{cases}. \tag{11}$$

The threshold $\gamma(\tau, p)$ can be constant, linear-decreasing (Abma and Kabir, 2006) and exponential-decreasing (Gao et al., 2010). However, all of these defining criterion are based on a prior knowledge about the data and is often not easy to choose. Instead, a processing-convenient percentile-thresholding strategy can be selected to overcome this inconvenience (Wang et al., 2008). $\gamma(\tau, p)$ is selected as the kth largest absolute value among all N values in the transformed domain, where the predefined percentile threshold $pc = k/N$ such that $\gamma(\tau, p) = prctile(|v(\mathbf{x})|, pc)$. Here, $prctile$ returns percentile pc of the values in $|v(\mathbf{x})|$.

Signal-to-noise ratio

In order to test the convergence performance, we use the widely used measurement(Hennenfent and Herrmann, 2006; Liu et al., 2009; Mahdad et al., 2011):

$$SNR = 10\log_{10}\frac{\|\mathbf{d}\|_2^2}{\|\mathbf{d}-\hat{\mathbf{d}}\|_2^2}, \tag{12}$$

where the unit for SNR is dB, $\mathbf{d}$ and $\hat{\mathbf{d}}$ denote the true and estimated data, respectively.

EXAMPLES

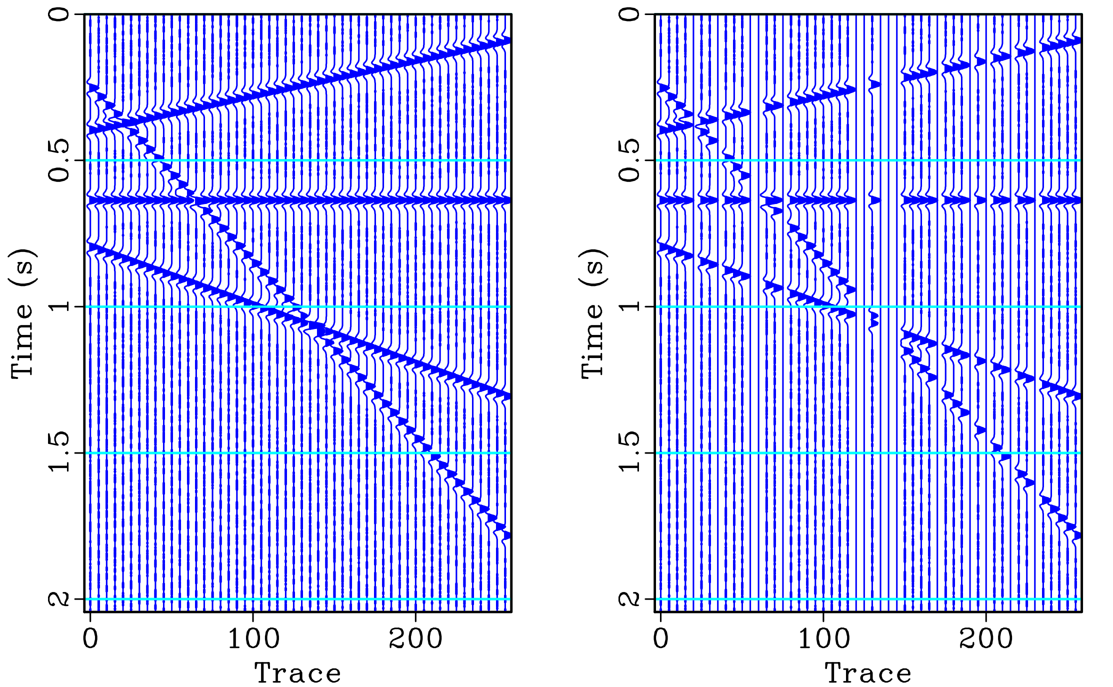
We use two synthetic examples and one field data set to compare the performance of the aforementioned percentile-half-thresholding approach with the conventional thresholding approaches. For simplicity, we only compare half thresholding with soft thresholding in two cases: percentile and constant-value threshold.

The first example is a synthetic example that is composed of four linear events, one of which crosses the other three events. The original data is shown in Figure

1a. After randomly removing 30% traces, we create the decimated section, as shown in Figure 1b. Figure 2 show the reconstruction results for the first synthetic example. From the result using percentile half thresholding (Figure 2d) and constant-value half thresholding (Figure 2b), and their corresponding estimation error sections (the difference between the true data and estimated data), we can observe that half thresholding nearly do a perfect job while the percentile thresholding strategy helps obtain a even better result. Both conventional constant-value and percentile soft-thresholding results seem to have significant estimation error. We can also observe that the percentile soft thresholding outperforms the constant-value soft thresholding by obtaining slightly less estimation error. In order to better compare the error sections using different approaches, we amplified the magnitude of each error section shown in Figures 2e, 2f, 2g and 2h by multiplying the magnitude by 10 times. The magnitude amplified error sections are shown in Figures 2i, 2j, 2k and 2l. It's much clearer that the percentile-half-thresholding strategy obtains a much better performance than any other approaches. The diagrams shown in Figure 3 further prove the observations in that the converged result for percentile half thresholding has the largest SNR, which is followed by constant-value half thresholding, and the constant-value soft thresholding obtains the smallest SNR.

The second example is a hyperbolic-events synthetic example. The original data and decimated data with 30 % traces randomly removed traces are shown in Figure 4. We apply four different thresholding approaches as mentioned above to this example and get the reconstructed results and their corresponding reconstruction error sections, which are shown in Figure 5. Similarly, the reconstructed result for the proposed percentile half thresholding (shown in Figure 5d) obtains the best results, causing negligible reconstruction error (shown in Figure 5h). The constant-value half thresholding also performs well because the error section (shown in Figure 5b) contains small amount of coherent signals. The reconstruction results for soft thresholding using either percentile or constant-value strategy (shown in Figures 5a and 5c)are not pleasant. Because they still contain some zero or weak-amplitude traces and cause a significant energy loss in the error sections as shown in Figures 5e and 5g, which indicates a failure in interpolating the missing traces. From the convergence diagram as shown in Figure 6, we can conclude that the half-thresholding approaches outperforms the conventional soft-thresholding approaches, while the percentile-thresholding strategy can helps obtain better results that the constant-value strategy.

The third example is a field marine data, which was investigated widely in the literature (Fomel, 2002; Liu and Fomel, 2011). We can obtain a similar conclusion as the previous two examples. However, from the reconstructed data and the estimation error sections, the conventional soft-thresholding (both constant-value and percentile) results (shown in Figures 8a and 8c) suffer from a heavy useful-energy loss as shown in Figures 8e and 8g and the reconstructed data as shown in Figures 8a and 8c still contain some obvious zero or weak-amplitude traces, which is no longer acceptable. From the convergence diagrams shown in Figure 9, the final SNRs of conventional soft-thresholding approaches are lower than the half-thresholding approaches.. Although the constant-value half thresholding can get slightly higher SNR than percentile half thresholding during the first several iterations, the percentile half thresholding converges to a higher final SNR than constant-value half thresholding.



(a) (b)

Figure 1: (a) Synthetic data (linear events). (b) Decimated synthetic data (created by randomly removing 30% traces).

CONCLUSIONS

We have introduced a general thresholding framework for IST algorithm. The only difference between half thresholding and conventional soft or hard thresholding is owing to the thresholding operator, so the percentile-thresholding strategy can be easily inserted into the half-thresholding algorithm. We compared four different types of thresholding strategies by using two synthetic examples and one field data example. The results show that apart from the threshold-setting convenience, the percentile thresholding also has the possibility for improving the recovery performance. Compared with soft thresholding, half thresholding tends to have a more precise reconstructed result.

ACKNOWLEDGMENTS

We thank Sanyi Yuan, Shan Qu, and two anonymous reviewers for their constructive suggestions, which helped improve this paper. We are grateful to developers of

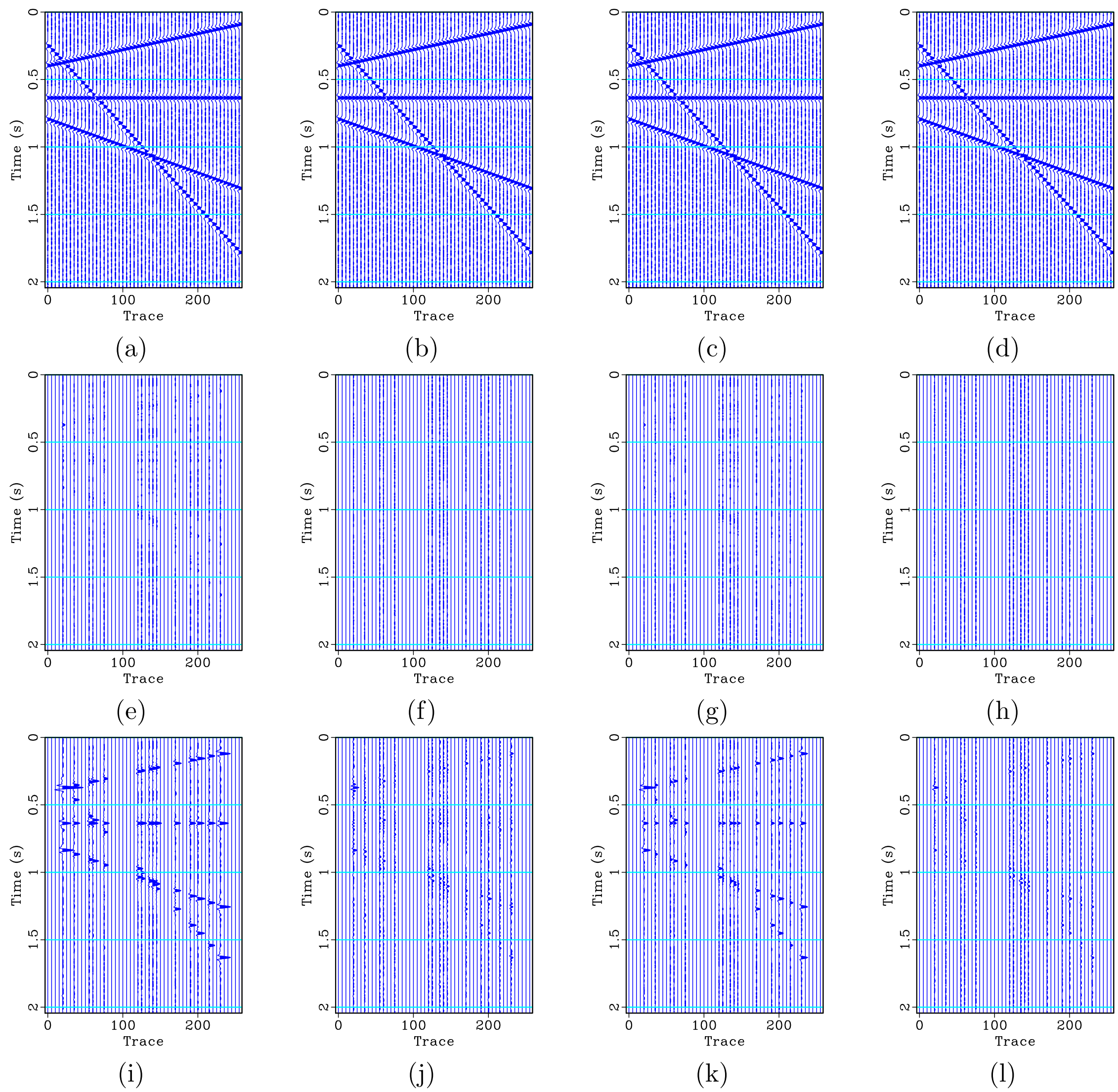


Figure 2: Reconstructed results for synthetic data (linear events). (a) Using constant-value soft thresholding. (b) Using constant-value half thresholding. (c) Using percentile soft thresholding. (d) Using percentile half thresholding. (e)-(h) Reconstruction error sections corresponding to (a)-(d), respectively. (i)-(l) Amplitude amplified error sections by $\times 10$, corresponding to (e)-(h), respectively.

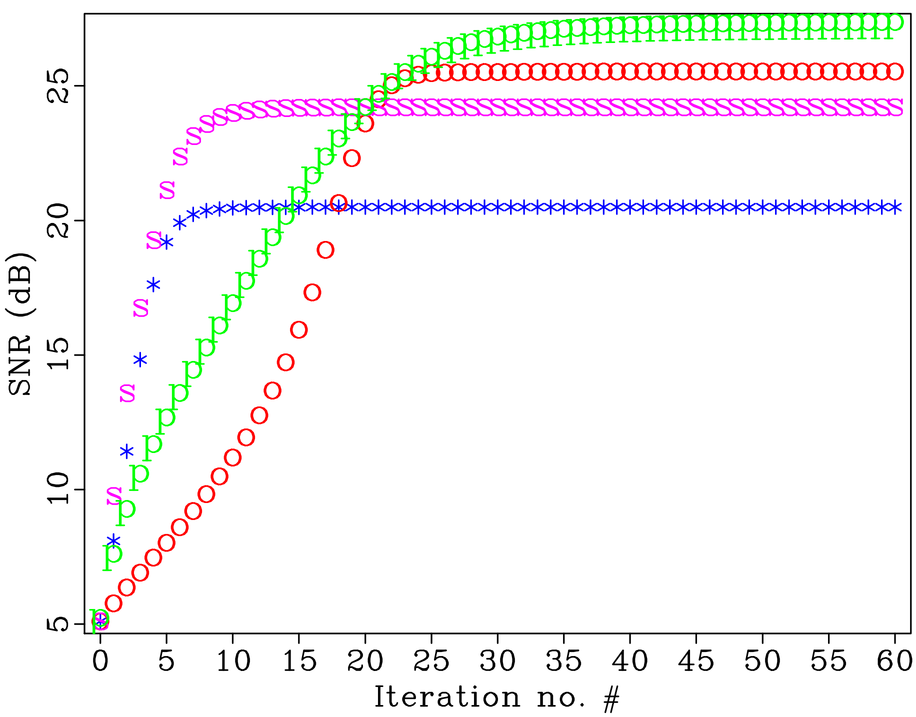


Figure 3: Convergence diagrams for synthetic data (linear events). "p" corresponds to percentile half thresholding. "o" corresponds to constant-value half thresholding. "s" corresponds to percentile soft thresholding. "*" corresponds to constant-value soft thresholding.

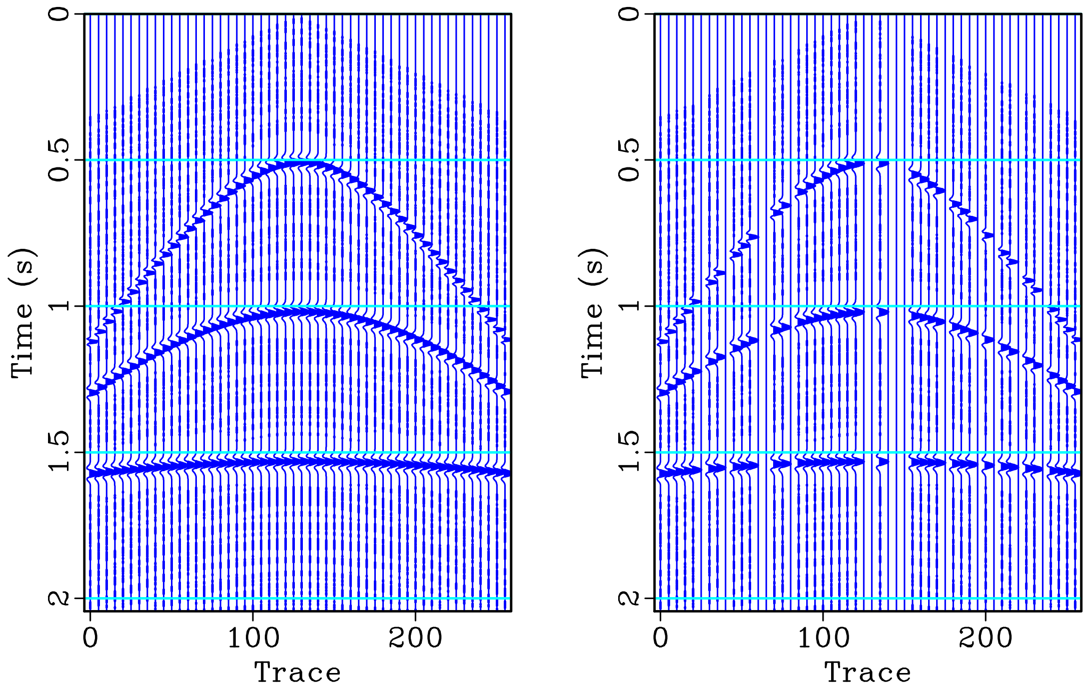


Figure 4: (a) Synthetic data (hyperbolic events). (b) Decimated synthetic data (created by randomly removing 30% traces).

the Madagascar software package for providing corresponding codes for testing the algorithms and preparing the figures.

REFERENCES

Abma, R., and N. Kabir, 2006, 3d interpolation of irregular data with a pocs algorithm: Geophysics, **71**, E91–E97.

Bleistein, N., and H. Jaramillo, 2000, A platform for kirchhoff data mapping in scalar models of data acquisition: Geophysical Prospecting, **48**, 135–161.

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, 489–509.

Canning, A., and G. H. F. Gardner, 1996, Reducing 3d acquisition footprint for 3d dmo and 3d prestack migration: Geophysics, **63**, 1177–1183.

Donoho, D. L., 2006, Compressed sensing: IEEE transactions on Information Theory, **52**, 1289–1306.

Fomel, S., 2002, Application of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

———, 2003, Seismic reflection data interpolation with differential offset and shot continuation: Geophysics, **68**, 733–744.

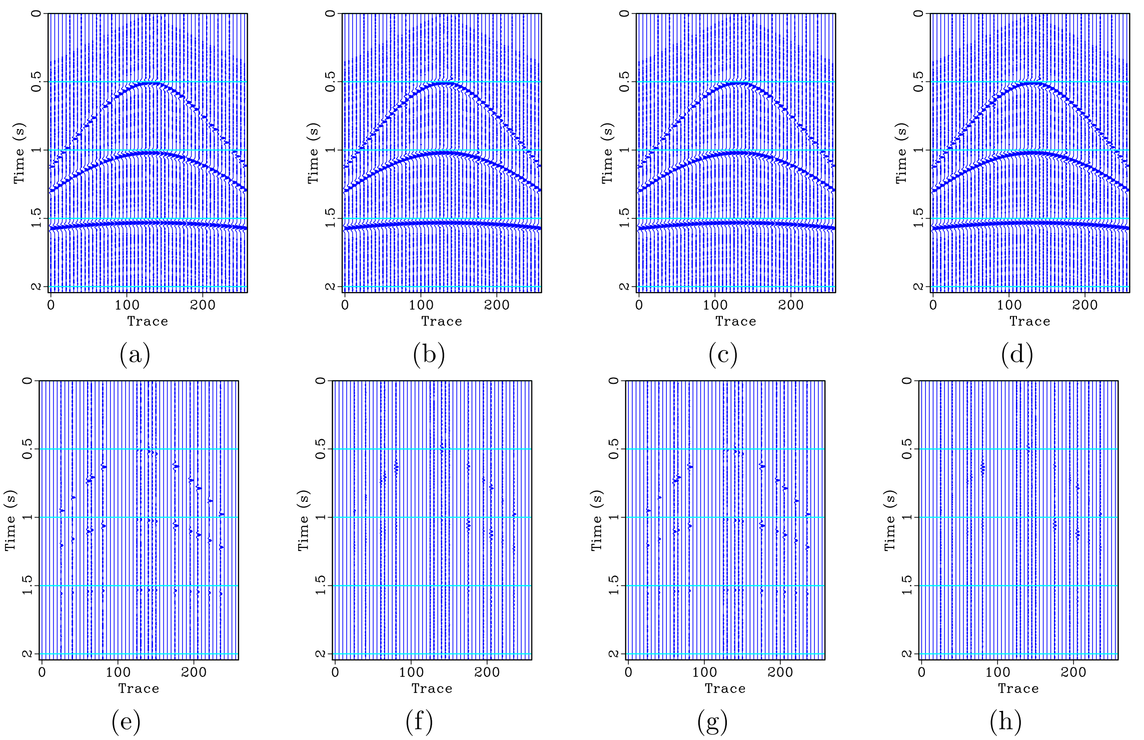


Figure 5: Reconstructed results for synthetic data (hyperbolic events). (a) Using constant-value soft thresholding. (b) Using constant-value half thresholding. (c) Using percentile soft thresholding. (d) Using percentile half thresholding. (e)-(h) Reconstruction error sections corresponding to (a)-(d), respectively.

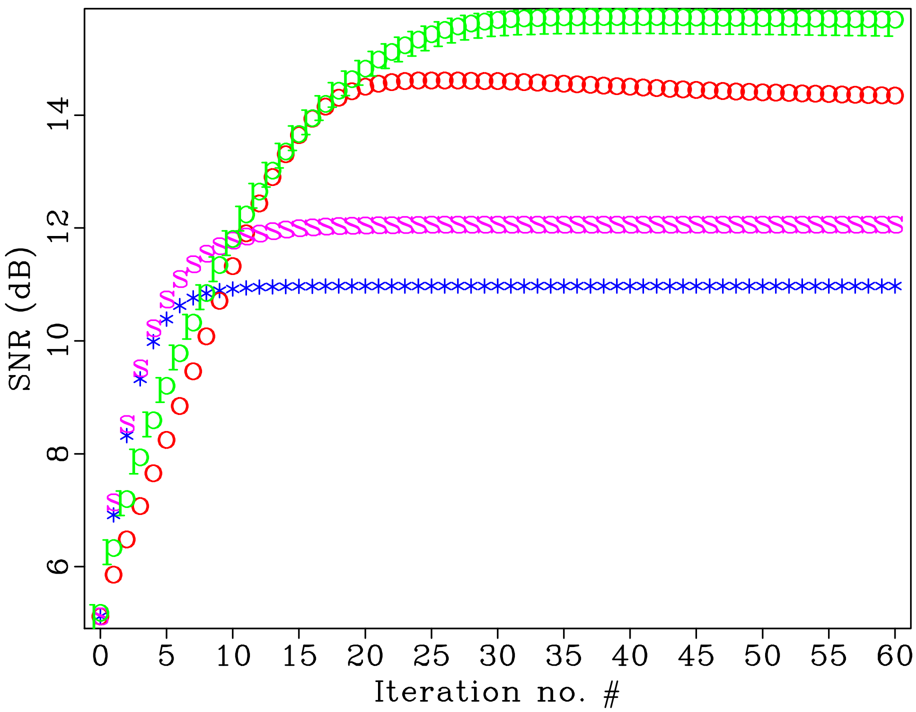
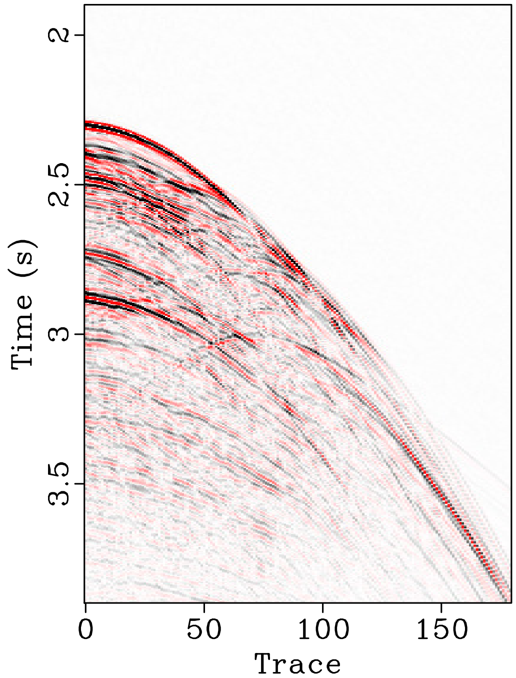
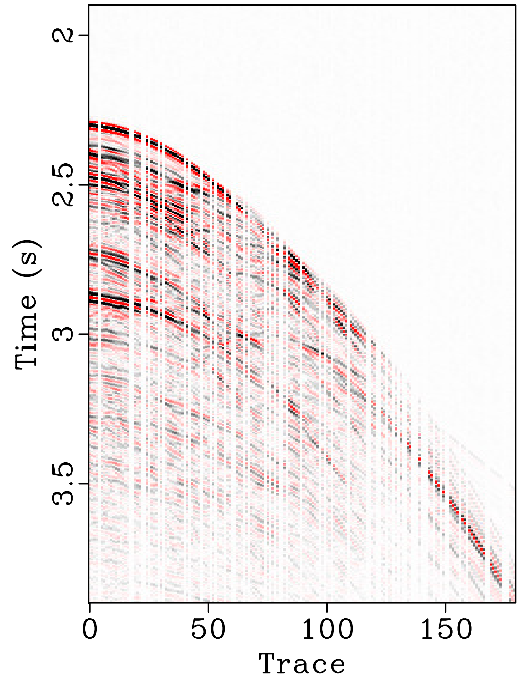


Figure 6: Convergence diagrams for synthetic data (hyperbolic events). "p" corresponds to percentile half thresholding. "o" corresponds to constant-value half thresholding. "s" corresponds to percentile soft thresholding. "*" corresponds to constant-value soft thresholding.



(a)



(b)

Figure 7: (a) Field data. (b) Decimated field data (created by randomly removing 30% traces).

Gao, J., X. Chen, G. Liu, and J. Ma, 2010, Irregular seismic data reconstruction based on exponentional threshold model of pocs method: Applied Geophysics, 229–238.

Hennenfent, G., and F. Herrmann, 2006, Seismic denoising with nonunformly sampled curvelets: Computing in Science & Engineering, **8**, 16–25.

Herrmann, F. J., 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, WB173–WB187.

Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Stacking seismic data using local correlation: Geophysics, **74**, V43–V48.

Liu, Y., and S. Fomel, 2011, Seismic data interpolation beyond aliasing using regularized nonstationary autoregression: Geophysics, **76**, V69–V77.

Mahdad, A., P. Doulgeris, and G. Blacquiere, 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, Q9–Q17.

Mallat, S. G., 2009, A wavelet tour of signal processing: The sparse way: Academic Press.

Naghizadeh, M., and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: Geophysics, **72**, V111–V118.

Porsani, M. J., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, 1461–1467.

Sacchi, M., T. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: IEEE Transactions on Signal Processing,

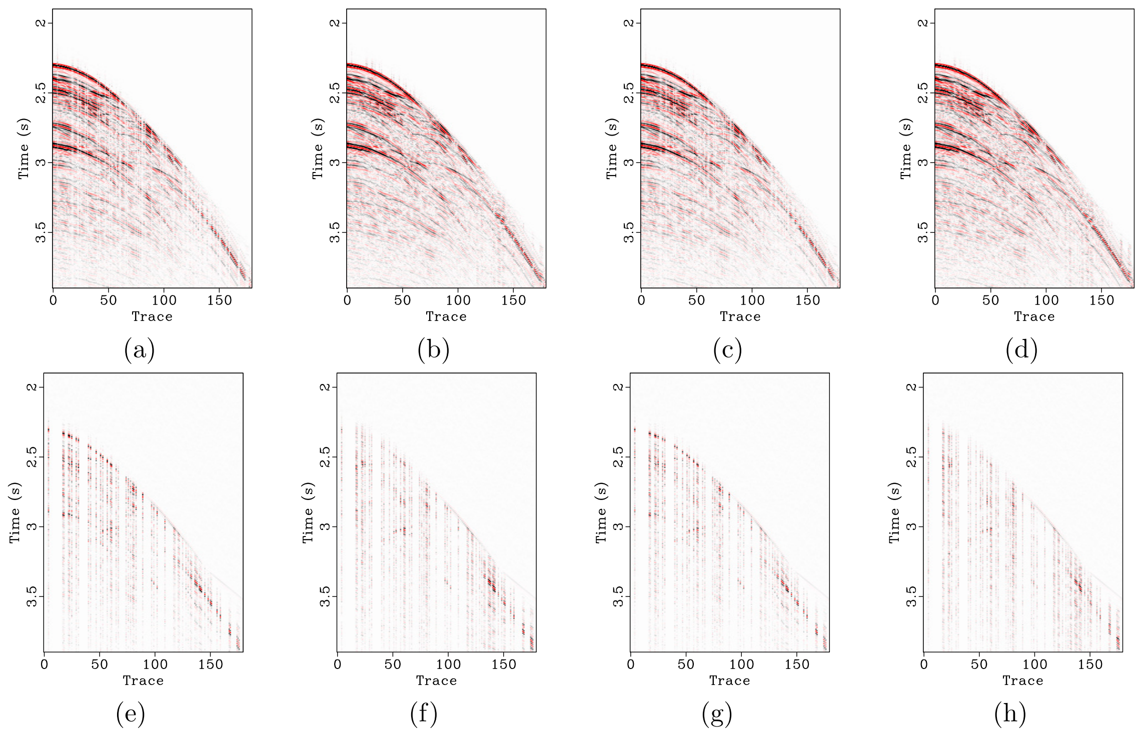


Figure 8: Reconstructed results for field data. (a) Using constant-value soft thresholding. (b) Using constant-value half thresholding. (c) Using percentile soft thresholding. (d) Using percentile half thresholding. (e)-(h) Reconstruction error sections corresponding to (a)-(d), respectively.

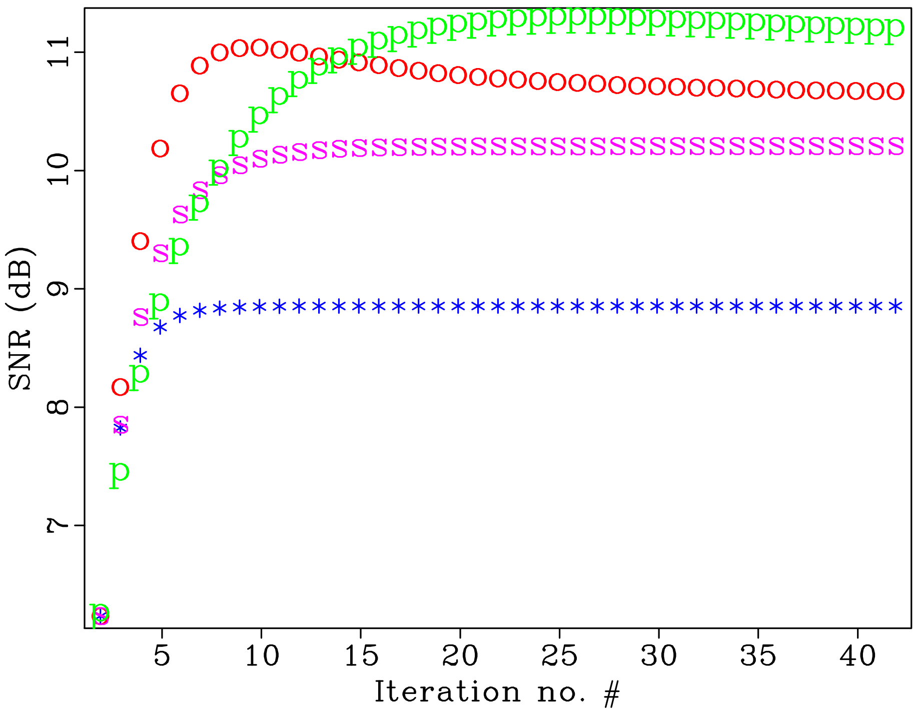


Figure 9: Convergence diagrams for field data. ”p” corresponds to percentile half thresholding. ”o” corresponds to constant-value half thresholding. ”s” corresponds to percentile soft thresholding. ”*” corresponds to constant-value soft thresholding.

46, 31–38.

Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Geophysics, **56**, 785–794.

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, 890–908.

Wang, D., C. Liu, Y. Liu, and G. Liu, 2008, Application of wavelet transform based on lifting scheme and percentiles soft-threshold to elimination of seismic random noise: Progress in Geophysics (in Chinese), **23**, 1124–1130.

Wang, Y., 2002, Seismic trace interpolation in the fxy domain: Geophysics, **67**, 1232–1239.

———, 2003, Sparseness-constrained least-squares inversion: application to seismic wave propagation: Geophysics, **68**, 1633–1638.

Xu, S., Y. Zhang, and G. Lambare, 2010, Antileakage fourier transform for seismic data regularization in higher dimensions: Geophysics, **75**, WB113–WB120.

Xu, Z., X. Chang, F. Xu, and H. Zhang, 2012, L1/2 regularization: A thresholding representation theory and a fast solver: IEEE Transactions on neural networks and learning systems, **23**, 1013–1027.

Yang, P., J. Gao, and W. Chen, 2012, Curvelet-based pocs interpolation of nonuniformly sampled seismic records: Journal of Applied Geophysics, **79**, 90–99.

———, 2013a, An iterative half thresholding method for seismic data interpolation:

83rd Annual International Meeting, SEG, Expanded Abstracts, 3579–3584.

———, 2013b, On analysis-based two-step interpolation methods for randomly sampled seismic data: Computers & Geosciences, **51**, 449–461.

Zwartjes, P., and A. Gisolf, 2007, Fourier construction with sparse inversion: Geophysical Prospecting, **55**, 199–221.

Zwartjes, P., and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, V21–V32.