

Application of stable signal recovery to seismic data interpolation

Gilles Hennenfent* and Felix J. Herrmann

Earth & Ocean Sciences Department, University of British Columbia

SUMMARY

We propose a method for seismic data interpolation based on 1) the reformulation of the problem as a stable signal recovery problem and 2) the fact that seismic data is sparsely represented by curvelets. This method does not require information on the seismic velocities. Most importantly, this formulation potentially leads to an explicit recovery condition. We also propose a large-scale problem solver for the ℓ_1 -regularization minimization involved in the recovery and successfully illustrate the performance of our algorithm on 2D synthetic and real examples.

INTRODUCTION

Seismic data is often irregularly (and sparsely) sampled along spatial coordinates due to practical and economical constraints. This problem of irregularly sampled and missing data mainly occurs in 3D seismic settings although it does occasionally happen in 2D. On land, it can be due to the presence of a lake, dead or severely contaminated traces, etc. In marine surveys, it can be caused, for example, by the proximity of a platform or by the cable feathering. As most of the multi-trace processing algorithms do not handle irregular sampled (and aliased) data, interpolation to a regular grid is necessary in order to process and subsequently interpret the data. Indeed, data irregularities often transform into image artifacts and poor spatial resolution in the migrated image. Relatively simple approaches can be used to handle missing traces in otherwise regularly sampled data. For instance, one can linearly interpolate neighboring traces. In the more general case of missing traces in irregularly sampled seismic data, one can collect the traces in bins and stack them. Unfortunately, these approaches do not account for the presence of wavefronts in the data and consequently these types of solutions to the seismic data interpolation problem do not necessarily give optimal results. In the practice of seismic exploration, several methods published in the geophysical literature are applied. These methods can be divided in three main groups. *Data mapping* methods are based on migration-like operators – e.g. offset continuation (see e.g. Stovas and Fomel, 1993), shot continuation (see e.g. Schwab, 1993). Claerbout (1992) and Spitz (1991) describe a second type of methods based on the formulation of the data interpolation problem as an iterative optimization problem with a convolution operator. Finally, there are methods based on particular transforms – e.g. the (non)uniform discrete Fourier transform (see e.g. Sacchi and Ulrych, 1996; Duijndam and Schonewille, 1999; Zwartjes, 2005), the linear Radon transform (see e.g. Thorson and Claerbout, 1985) and the parabolic Radon transform (see e.g. Hampson, 1986).

In this paper, we limit ourselves to the case of missing traces in otherwise regularly sampled data and improve upon Hennenfent and Herrmann (2005) by extending the idea of stable signal recovery (SSR) from few and noise-contaminated observations, as introduced in Candès et al. (2005b), to seismic data interpolation. This transform-based method explores the sparsity of seismic data in the curvelet domain and does not require information on the seismic velocities. Most importantly, it potentially leads to an explicit recovery condition.

THEORY

Seismic data interpolation

Seismic data interpolation is an underdetermined inverse problem. Interpolation approaches typically involve some sort of a minimization

problem that aims at jointly minimizing a quadratic distance to the acquired data penalized by a functional on the unknown (regularization term). The general form for these minimization problems is as follows

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda J(\mathbf{x}). \quad (1)$$

In this expression, the n -vector \mathbf{y} represents the acquired data, λ the regularization parameter balancing the emphasis of data misfit *versus a priori* information residing in the penalty term $J(\mathbf{x})$, and \mathbf{A} the operator – e.g. restriction operator combined with a generic transform – that relates the m -vector \mathbf{x} – e.g. representation of the interpolated seismic data in the transform domain – to the acquired data \mathbf{y} . The success of data interpolation depends upon choices for 1) the operator \mathbf{A} and the penalty term $J(\mathbf{x})$, and 2) the regularization parameter λ . High-resolution methods explore for example the sparsity (minimum structure) of seismic data in a transform domain. Sacchi and Ulrych (1996) followed later by Zwartjes (2005) use for instance the sparsity of seismic data in the Fourier domain. Although these methods are successful, it is unclear how to assess their performance.

Stable signal recovery

Consider a vastly underdetermined system of linear equations. The problem is then to recover an unknown m -vector \mathbf{x}_0 from $n \ll m$ contaminated observations $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}$, where \mathbf{A} is an $n \times m$ matrix and \mathbf{n} is an n -vector Gaussian error term of variance σ^2 . Such systems have infinitely many solutions and one may wonder whether it is possible to recover \mathbf{x}_0 accurately. Recent results from information theory (Candès et al., 2004, 2005b) show that, if \mathbf{A} obeys some specific (weak) condition and, if \mathbf{x}_0 has few nonzero entries (i.e. \mathbf{x}_0 is sparse), then the solution $\tilde{\mathbf{x}}$ of the following minimization problem

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{y}\|_2 \leq \varepsilon \quad (2)$$

is within the noise level $\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2 \leq C_1 \cdot \varepsilon$. In this expression, ε is the size of the error term \mathbf{n} and C_1 a positive well behaved (i.e. small) constant. Note that, in the noise-free case, \mathbf{x}_0 can be recovered *exactly* (Candès et al., 2005b). Roughly speaking, the condition on \mathbf{A} requires that there exists a positive constant $S < m$ such that every subset of S or less columns of \mathbf{A} approximately behaves like an orthonormal system. In this case, solving (2) recovers *any* sparse signal \mathbf{x}_0 with the number of nonzero entries of about S or less.

Similar results were obtained in a more realistic setting where the signal of interest \mathbf{f}_0 is assumed to be approximately sparse or compressible – i.e. sorted absolute values of entries decay like a power law – in some orthonormal sparsity basis \mathbf{S} and incompletely measured in a second orthonormal basis \mathbf{M} . The measurement – also called observation – y of the vector \mathbf{f}_0 against the vector \mathbf{m} is defined as the projection of \mathbf{f}_0 onto \mathbf{m} – i.e. $y = \mathbf{m}^H \mathbf{f}_0$ (the symbol H denotes the Hermitian transpose). In that case, $\mathbf{A} := \mathbf{RMS}^H$, where \mathbf{R} is a restriction matrix which extracts n rows from the $m \times m$ orthonormal matrix \mathbf{MS}^H , while normalizing the columns to unit norm. Then the solution $\tilde{\mathbf{x}}$ to

$$\begin{cases} \min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}\|_1 & \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2 \leq \varepsilon \\ \tilde{\mathbf{f}} = \mathbf{S}^H \tilde{\mathbf{x}} \end{cases} \quad (3)$$

verifies

$$\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2 \leq C_2 \cdot \varepsilon + C_3 \cdot \frac{\|\mathbf{x}_0 - \mathbf{x}_{0,S}\|_1}{\sqrt{S}} \quad (4)$$

provided

$$S \leq C_4 \cdot \frac{1}{\mu^2} \cdot n \quad (5)$$

ignoring log-like factors. Eq. (5) is called the recovery condition. In these expressions, $\mathbf{x}_{0,S}$ is the truncated vector corresponding to the S largest entries (in absolute value) of $\mathbf{x}_0 = \mathbf{S}\mathbf{f}_0$, $(C_i)_{i=2,3,4}$ positive well behaved constants, and μ the *mutual coherence* given by

$$\mu(\mathbf{M}, \mathbf{S}) = \sqrt{m} \cdot \max_{k,l} |\mathbf{m}_k(\mathbf{s}_l)^H| \quad (6)$$

with \mathbf{m}_k the k^{th} row of \mathbf{M} and \mathbf{s}_l the l^{th} row of \mathbf{S} .

This result is significant for several reasons. First of all, it essentially states that recovering the S largest entries (in absolute value) of the decomposition \mathbf{x}_0 in the sparsity basis \mathbf{S} of \mathbf{f}_0 requires only $n \propto S \cdot \mu^2$ observations in the measurement basis \mathbf{M} . It thus provides a robust recovery criterion. Secondly, this criterion directly links the required number of measurements n to the sparsity S of the to-be-recovered function and the mutual coherence μ (dependent on the choice of \mathbf{S} and \mathbf{M}). Note that the more similar \mathbf{m}_k and \mathbf{s}_l , the larger the mutual coherence and consequently the larger the number of observations required for the recovery. Thirdly, this result gives an upper bound of the approximation error which shows that the estimate is almost as good as if we knew the S -largest entries of \mathbf{x}_0 . In the seismic context, this framework can thus address important questions such as 1) how well one can expect to recover seismic data given an acquisition geometry, and 2) what the 'optimal' acquisition geometry is in order to recover seismic data within a given accuracy.

Fig. 1 illustrates this recovery result with a simple example. The original 1024-sample signal \mathbf{f}_0 (Fig. 1(a)) is defined as a single nonzero entry in the Discrete Cosine Transform (DCT) domain. The Dirac/spike basis is chosen as the measurement basis \mathbf{M} because it has a low mutual coherence with the sparsity basis DCT. Indeed, there is no vector from \mathbf{S} that has a large inner product with any vector of \mathbf{M} . Experimentally, we measure $\mu(\text{DCT}, \text{Dirac}) = \sqrt{2}$. In order to recover \mathbf{x}_0 and consequently \mathbf{f}_0 by solving Eq. (3), 5 random observations of \mathbf{f}_0 are typically required (Tsaig and Donoho, 2005). The recovered signal $\hat{\mathbf{f}}$ (Fig. 1(b)) virtually overlays \mathbf{f}_0 .

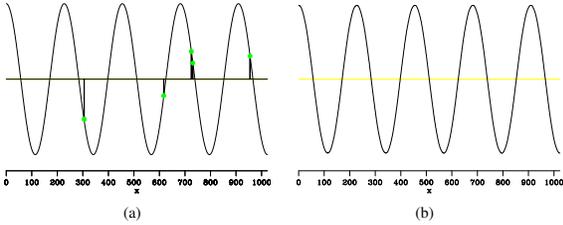


Figure 1: Perfect recovery of a 1024-sample signal from 5 random observations in the Dirac basis. (a) The original signal \mathbf{f}_0 is randomly measured 5 times in the Dirac basis; (b) the recovered signal $\hat{\mathbf{f}}$ virtually overlays \mathbf{f}_0 .

Application of SSR to seismic data interpolation

In the case of missing traces in otherwise regularly sampled data, the acquired data \mathbf{y} is a subgroup of traces of the ideal data \mathbf{f}_0 . Mathematically it translates as the action of a restriction operator (or so-called picking operator) \mathbf{R} on the ideal data. The seismic experiment can thus be written as

$$\mathbf{y} = \mathbf{R}\mathbf{M}\mathbf{f}_0 + \mathbf{n}. \quad (7)$$

\mathbf{M} is the identity/Dirac basis \mathbf{I} . Each of its row contains a discrete Dirac function that singles out a unique time sample among all possible time samples in the vector \mathbf{f}_0 . \mathbf{R} extracts those rows from \mathbf{M} that

represent time samples actually acquired. \mathbf{R} and \mathbf{M} are thus fixed by the acquisition.

The recovery condition is a tradeoff between sparsity/compressibility and mutual coherence. For a given number of observations, an increase of mutual coherence by a factor of 2 requires the solution to be 4 times sparser in order to still recover it for example. Herrmann and Hennenfent (2006) show that the emphasis should be put on compressibility rather than mutual coherence and propose to solve for \mathbf{f}_0 as a sparse superposition of curvelet coefficients (i.e. \mathbf{S} defined as the curvelet transform analysis operator \mathbf{C}). Curvelets achieve indeed a much better compressibility of seismic data than Fourier (most incoherent with seismic spatial sampling) at the expense of a moderate loss of incoherency with the seismic acquisition (see e.g. Candès et al., 2005a; Herrmann and Hennenfent, 2006, for more details about the curvelet transform and its applications to seismic data processing). The interpolated data $\hat{\mathbf{f}}$ is then obtained by $\hat{\mathbf{f}} = \mathbf{C}^H \bar{\mathbf{x}}$, where

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \underbrace{\mathbf{R}\mathbf{I}\mathbf{C}^H}_{\mathbf{A}} \mathbf{x}\|_2 \leq \varepsilon \quad (8)$$

with ε the size of the noise present in the acquired data \mathbf{y} . This minimization problem has the significant advantage over Eq. (1) to relieve the user of testing several regularization parameters λ .

Large-scale problem solver for ℓ_1 -regularization minimization

We propose a parallelizable (see Thomson et al., 2006) algorithm to solve Eq. (8). The solver is based on cooling method optimization and an iterative thresholding algorithm (Daubechies et al., 2004). The cooling method aims at finding the optimal multiplier λ^* for $\mathcal{L}(\mathbf{x}, \lambda) := \lambda \|\mathbf{x}\|_1 + \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - \varepsilon^2$, the Lagrangian function of (8), such that the residual $\mathbf{r} := \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon$. The algorithm is as follows

```

 $\mathbf{x}_0 :=$  initial guess
 $\lambda_0 :=$  initial Lagrange multiplier
while  $\mathbf{r} > \varepsilon$ 
     $\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda_k)$ 
     $\lambda_{k+1} = \alpha_k \lambda_k$  with  $0 < \alpha_k < 1$ 
end while.
```

The critical part of this algorithm is the minimization of $\mathcal{L}(\mathbf{x}, \lambda_k)$ done by the iterative thresholding algorithm presented in Daubechies et al. (2004). At each sub-iteration, evaluation of

$$\mathbf{x}_{i+1} = \mathcal{S}_{\lambda_k}(\mathbf{x}_i + \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_i)) \quad (9)$$

with

$$\mathcal{S}_{\lambda_k}(x) := \text{sign}(x) \cdot \max(|x| - \lambda_k, 0) \quad (10)$$

yields an approximate estimate for \mathbf{x} which converges to the solution of the subproblem for a large enough number of iterations. In practice, one only needs to approximately solve each subproblem, which significantly accelerates the overall procedure.

RESULTS

Synthetic examples

For the first example, we limit ourselves to 150 iterations (30 updates of the Lagrange multiplier and 5 sub-iterations). Fig. 2(a) shows the synthetic dataset with two events, one linear and the other hyperbolic, with 40% randomly-distributed missing traces. The events show an amplitude-versus-offset (AVO) effect. Fig. 2(b) shows the recovered events (signal-to-noise ratio (SNR) = 39.2 dB) and Fig. 2(c) the difference between the original and recovered dataset plotted using the same color scale as Fig. 2(a) and Fig. 2(b). There is hardly any recovery error beside a light dimming in the largest gap around trace number 310, which is more visible in the recovered linear event AVO effect compared to the original one (Fig. 2(d)). Note that this small amplitude

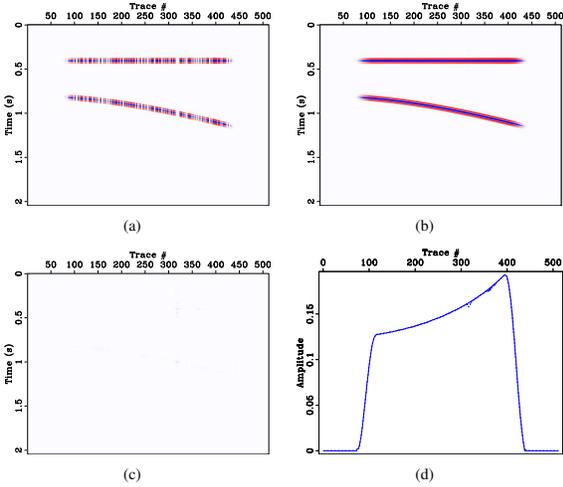


Figure 2: Reconstruction of linear and hyperbolic events from synthetic data with 40% randomly-distributed missing traces. The events show an AVO effect. (a) input data with gaps; (b) recovered events; (c) difference between original and recovered events plotted using the same color scale as (a) and (b); (d) Original AVO effect (plain line) and recovered AVO effect (dash line). Shape and amplitude of events are recovered with high fidelity.

decline can be corrected by doing more Lagrange multiplier updates – i.e. more iterations.

For the second example, we limit ourselves to 400 iterations (80 updates of the Lagrange multiplier and 5 sub-iterations), which is still practical for algorithms based on fast transforms. Fig. 3(a) shows the synthetic dataset, Fig. 3(b) the input data with 20% randomly-distributed missing traces, Fig. 3(c) the recovered seismic signal (SNR = 24.4 dB), and Fig. 3(d) the difference between the original and recovered dataset plotted using the same color scale as Fig. 3(a), Fig. 3(b) and Fig. 3(c). Despite small errors mainly concentrated near the apex, the overall recovery quality is satisfactory. Note especially that amplitudes and continuity are well preserved along wavefronts.

Field data example

In this example, we limit ourselves to 400 iterations (80 updates of the Lagrange multiplier and 5 sub-iterations). Fig. 4(a) shows a real marine dataset, Fig. 4(b) the input data with 40% randomly-distributed missing traces, Fig. 4(c) the recovered seismic signal (SNR = 20.5 dB), and Fig. 4(d) the difference between the original and recovered dataset plotted using the same color scale as Fig. 4(a), Fig. 4(b) and Fig. 4(c). Despite small errors mainly concentrated near the zero offset and in the 150-180 trace range where the local ratio of missing traces is higher, the overall recovery quality is good. Note again that amplitudes and continuity are quite well preserved along wavefronts.

CONCLUSIONS

We have presented a method for seismic data interpolation which is an extension to the stable signal recovery as introduced in Candès et al. (2005b). Instead of resorting to an orthonormal basis, we proposed the redundant curvelet transform as the sparsity representation for seismic data. This fast-transform-based method – the curvelet transform is of same numerical complexity as the FFT – exploits the multiscale, multi-directional and continuity properties of seismic wavefronts which lead to sparsity of seismic data in the curvelet domain without requiring information on the seismic velocities. The recovery is carried out with a large-scale ℓ_1 -regularization minimization problem solver and the ex-

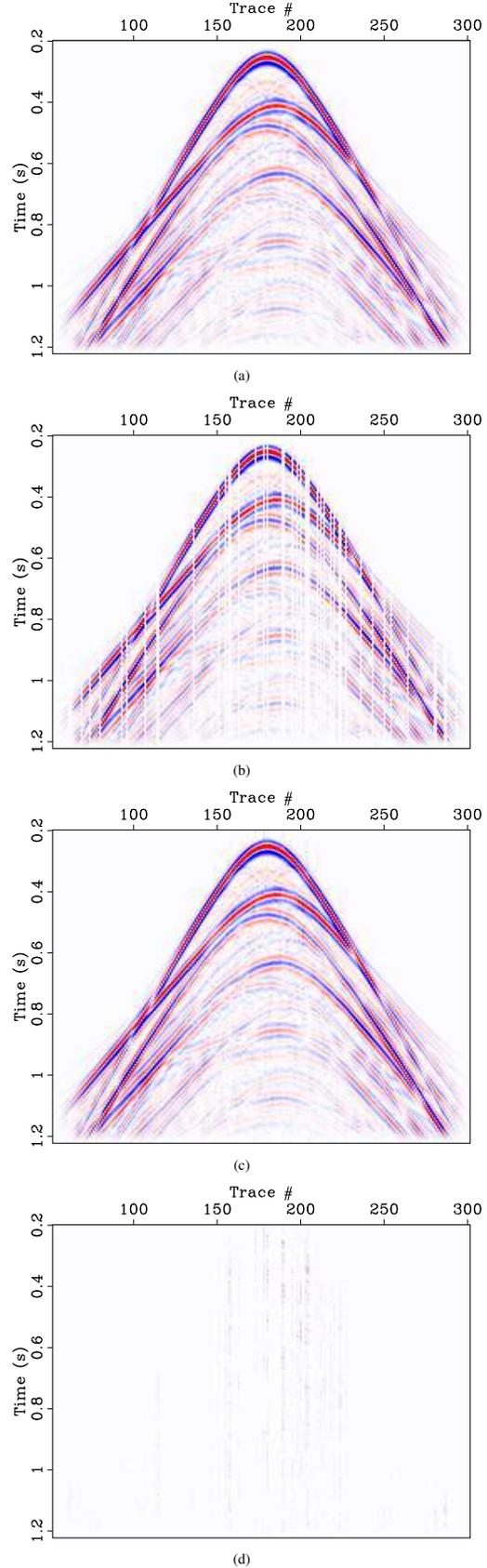


Figure 3: Stable seismic signal recovery. (a) synthetic data; (b) input data with 20% randomly-distributed traces missing; (c) interpolated data; (d) difference between (a) and (c) plotted using the same color scale.

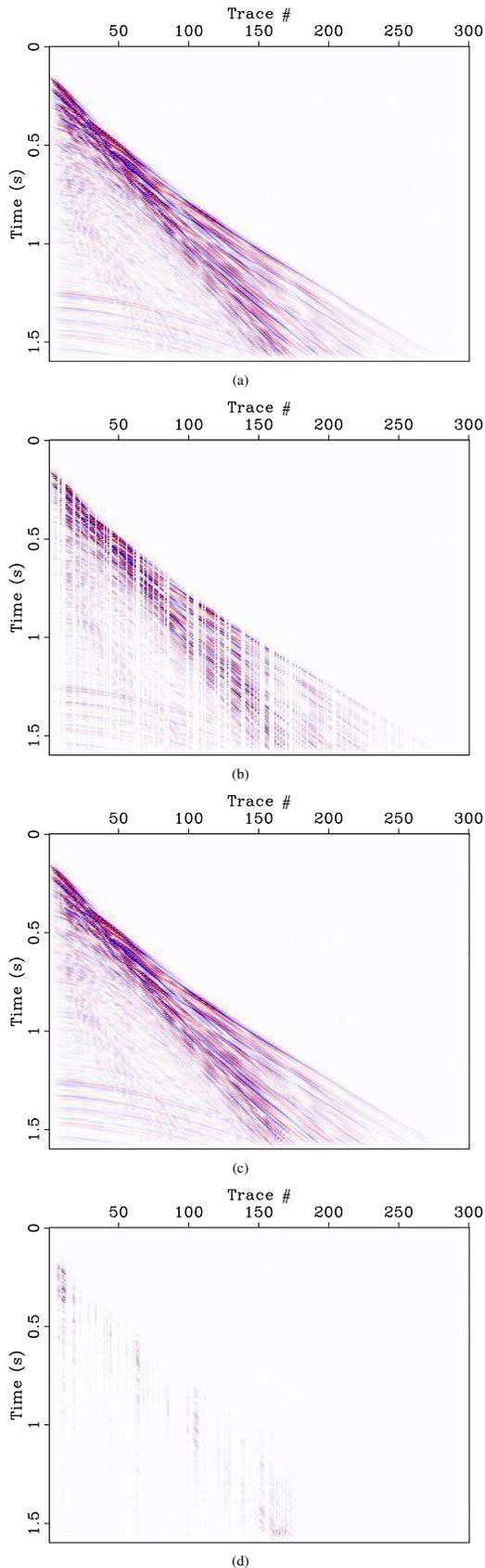


Figure 4: Stable seismic signal recovery. (a) real marine data; (b) input data with 40% randomly-distributed traces missing; (c) interpolated data; (d) difference between (a) and (c) plotted using the same color scale.

amples clearly illustrate the performance of our algorithm in the 2D source-receiver domain on synthetic and real examples. Extension to 3D is straightforward and our methodology may – as long as the continuity along the wavefronts is preserved (the wavefronts may contain caustics) – also be integrated in so-called data mapping approaches.

ACKNOWLEDGMENTS

The authors would like to thank S. Fomel for his new seismic processing software RSF, the authors of CurveLab, D.J. (Eric) Verschuur for the synthetic dataset and ExxonMobil for the real dataset. This work was carried out in part when G.H. was hosted at the Universidade Federal do Rio Grande do Norte (UFRN - Natal, Brazil) by S.J. Bezerra and is part of the SINBAD project with financial support, secured through the Industry Technology Facilitator (ITF), from the following organizations: BG, BP, Chevron, ExxonMobil and SHELL. Additional funding came from the NSERC Discovery Grant 22R81254.

REFERENCES

- Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2005a, CurveLab: Fast Discrete Curvelet Transform.
- Candès, E. J., J. Romberg, and T. Tao, 2004, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
- 2005b, Stable signal recovery from incomplete and inaccurate measurements.
- Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific publishing.
- Daubechies, I., M. Defrise, and C. de Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: *Comm. Pure Appl. Math.*, 1413–1457.
- Duijndam, A. J. W. and M. A. Schonewille, 1999, Reconstruction of band-limited signals, irregularly sampled along spatial direction: *Geophysics*, **64**, 524–538.
- Hampson, D., 1986, Inverse velocity stacking for multiple elimination: *J. Can. Soc. of Expl. Geophys.*, **22**, 44–55.
- Hennenfent, G. and F. J. Herrmann, 2005, Sparseness-constrained data continuation with frames: Applications to missing traces and aliased signals in 2/3D: Presented at the 75th SEG Conference & Exhibition.
- Herrmann, F. J. and G. Hennenfent, 2006, Non-parametric seismic data recovery with curvelet frames. (to be submitted).
- Sacchi, M. D. and T. J. Ulrych, 1996, Estimation of the discrete Fourier transform, a linear inversion approach: *Geophysics*, **61**, 1128–1136.
- Schwab, M., 1993, Shot gather continuation: Technical Report 77, SEP.
- Spitz, S., 1991, Seismic trace interpolation in the f-x domain: *Geophysics*, **56**, 785–794.
- Stovas, A. M. and S. B. Fomel, 1993, Kinematically equivalent DMO operators: *Russian Geology and Geophysics*, **37**, 102–113.
- Thomson, D., G. Hennenfent, H. Modzelewski, and F. J. Herrmann, 2006, A parallel windowed curvelet transform applied to seismic processing. (submitted for presentation to the 76th SEG Conference & Exhibition).
- Thorson, J. R. and J. F. Claerbout, 1985, Velocity-stack and slant-stack stochastic inversion: *Geophysics*, **50**, 2727–2741.
- Tsaig, Y. and D. L. Donoho, 2005, Extensions of compressed sensing: Technical report, Stanford university.
- Zwartjes, P., 2005, Fourier reconstruction with sparse inversion: PhD thesis, Delft university of technology.