

Non-parametric seismic data recovery with curvelet frames

Felix J. Herrmann¹ and Gilles Hennenfent¹

ABSTRACT

Seismic data recovery from data with missing traces on otherwise regular acquisition grids forms a crucial step in the seismic processing flow. For instance, unsuccessful recovery leads to imaging artifacts and to erroneous predictions for the multiples, adversely affecting the performance of multiple elimination. A non-parametric transform-based recovery method is presented that exploits the compression of seismic data volumes by recently developed curvelet frames. The elements of this transform are multidimensional and directional and locally resemble wavefronts present in the data, which leads to a compressible representation for seismic data. This compression enables us to formulate a new curvelet-based seismic data recovery algorithm through sparsity-promoting inversion. The concept of sparsity-promoting inversion is in itself not new to geophysics. However, the recent insights from the field of ‘compressed sensing’ are new since they clearly identify the three main ingredients that go into a successful formulation of a recovery problem, namely a sparsifying transform, a sampling strategy that subdues coherent aliases and a sparsity-promoting program that recovers the largest entries of the curvelet-domain vector while explaining the measurements. These concepts are illustrated with a stylized experiment that stresses the importance of the degree of compression by the sparsifying transform. With these findings, a curvelet-based recovery algorithm is developed, which recovers seismic wavefields from seismic data volumes with large percentages of traces missing. During this construction, we benefit from the main three ingredients of compressive sampling, namely the curvelet compression of seismic data, the existence of a favorable sampling scheme and the formulation of a large-scale sparsity-promoting solver based on a cooling method. The recovery performs well on synthetic as well as real data and performs better by virtue of the sparsifying property of curvelets. Our results are applicable to other areas such as global seismology.

INTRODUCTION

The methodology presented in this paper addresses two important issues in seismic data acquisition, namely the mediation of imaging artifacts caused by physical con-

¹Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada

¹**e-mail:** fherrmann,ghennenfent@eos.ubc.ca

straints encountered during acquisition, and the design of a more economic acquisition, limiting the number of source and receiver positions within the survey. In either case, the data is incomplete and it is our task to recover a fully-sampled seismic data volume as required by wave-equation based multiple elimination (SRME, Verschuur and Berkhout, 1997) and imaging (Symes, 2006). This paper deals with the specific case of seismic data recovery from a regularly-sampled grid with traces missing. As a consequence, the data is undersampled and the Nyquist sampling criterion is violated, giving rise to a Fourier spectrum that may contain harmful aliases.

A multitude of solutions have been proposed to mitigate the impact of coherent aliases on seismic imaging. Our approach derives from three key ingredients, namely a sparsifying transform, a sampling strategy that limits the occurrence of harmful aliases and a nonlinear recovery scheme that promotes transform-domain sparsity and consistency with the acquired data. These three key ingredients form the basis of the emerging field of “compressive sampling” (Candès et al., 2006b; Donoho et al., 2006b) with several applications that include MRI-imaging (Lustig et al., 2007) and A/D conversion (Tropp et al., 2006). Compressive sampling can be seen as a theoretically rigorous justification of empirical ideas on sparsity-promoting inversion that existed in the geophysical literature with applications that include “spiky deconvolution” (Taylor et al., 1979; Oldenburg et al., 1981; Ulrych and Walker, 1982; Levy et al., 1988; Sacchi et al., 1994) analyzed by mathematicians (Santosa and Symes, 1986; Donoho and Logan, 1992) to Fourier and Radon transform-based seismic data recovery, an approach initially proposed by Sacchi et al. (1998) and extended by numerous authors (Trad et al., 2003; Xu et al., 2005; Abma and Kabir, 2006; Zwartjes and Sacchi, 2007). Amongst all these methods, it was observed that a successful solution of these problems depends critically on the number of measurements (or the frequency passband for deconvolution) and the signal’s sparsity in some transformed domain, e.g. spikes for deconvolution and Fourier for sparse recovery.

Compressive sampling provides insights in the conditions that determine successful recovery from incomplete data. We leverage these new insights towards a formulation of the large-scale seismic data regularization problem, where a sparsifying transform, anti-alias sampling and a sparsity-promoting solver are used to solve this problem for acquisitions with large percentages of traces missing. These theoretical developments are important since they provide a better intuition of the overriding principles that go into the design of a recovery method and into explicit construction of a sparsifying transform, the sampling strategy and the sparsity-promoting solver.

In this paper, we consider a recovery method that derives from this intuition by using a generic sparsifying transform that requires minimal prior information (although our method benefits like Fourier-based interpolation (Zwartjes and Sacchi, 2007) from dip discrimination by means of specifying a minimum apparent velocity). In that respect our method differs from interpolation methods based on pattern recognition (Spitz, 1999), plane-wave destruction (Fomel et al., 2002) and data mapping (Bleistein et al., 2001), including parabolic, apex-shifted Radon and DMO-NMO/AMO (Trad, 2003; Trad et al., 2003; Harlan et al., 1984; Hale, 1995; Canning and Gardner,

1996; Bleistein et al., 2001; Fomel, 2003; Malcolm et al., 2005), which require, respectively, the omission of surface waves, specific knowledge on the dominant dips and a velocity model.

Our main contribution

The success of our recovery method for seismic data, named curvelet-based recovery by sparsity-promoting inversion (CRSI), derives from a sparsifying transform in conjunction with a sampling scheme that favors recovery. With their well-documented sparsity for seismic data with wavefronts and Fourier-domain localization property (Candès et al., 2006a; Hennenfent and Herrmann, 2006; Herrmann et al., 2007a), curvelets render sparsity-promoting inversion into a powerful constraint for the recovery of seismic data. Our contribution, first reported in Herrmann (2005), lies in the application of this transform (see e.g. Candès et al., 2006a; Ying et al., 2005, for details on the definition and implementation of the discrete curvelet transform) to the seismic recovery problem. Our work includes the adaptation towards a geophysically feasible sampling scheme that eliminates harmful aliases and allows for a dip discrimination by means of a minimum apparent velocity. This combination of sparsity-promotion and sampling permits a solution of a very large-scale ℓ_1 -minimization problem at a computational cost comparable to iterative-re-weighted least-squares (IRLS Gersztenkorn et al., 1986).

Our formulation for the solution of the seismic data recovery problem reads

$$\mathbf{P}_\epsilon : \quad \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i| & \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{S}^T \tilde{\mathbf{x}} \end{cases} \quad (1)$$

and is reminiscent of the solution of the ‘‘inpainting problem’’, the problem of infilling missing data, reported by Elad et al. (2005). In this expression, \mathbf{y} is the vector with the incomplete data and \mathbf{x} the unknown coefficient vector that generates the decimated data through the modeling matrix, \mathbf{A} . The solution of the recovery problem corresponds to finding the sparsity vector, \mathbf{x} with minimal ℓ_1 norm subject to fitting the data to within a noise-dependent ℓ_2 error ϵ . The estimate for the recovered data vector, $\tilde{\mathbf{f}}$, is obtained by applying the inverse transform, \mathbf{S}^T , to the recovered sparsity vector, $\tilde{\mathbf{x}}$, that solves \mathbf{P}_ϵ . Above formulation for the recovery problem is known to be stable and extends to (seismic) signals that are not strictly sparse but compressible (Candès et al., 2006b). In that case, the recovery error becomes smaller for transforms that concentrate the signal’s energy amongst a smaller fraction of the coefficients.

At this point, the well established ability of curvelets (Candès et al., 2006a; Hennenfent and Herrmann, 2006; Herrmann et al., 2007a) enters into the equation. Compared to discrete wavelets, used for digital storage of multidimensional seismic data volumes (Donoho et al., 1999), curvelets truly honor the behavior of seismic wavefields. They correspond to localized ‘little plane waves’ that are oscillatory in one direction and smooth in the other direction(s) (Candès and Donoho, 2000a, 2004).

Like directional isotropic wavelets, they are multiscale and multi-directional, but unlike wavelets, they have an anisotropic shape – they obey the so-called parabolic scaling relationship, yielding a width $\propto \text{length}^2$ for the support of curvelets in the physical domain. Curvelets are also strictly localized in the Fourier domain and quasi localized in the space domain, i.e., they decay rapidly away from the crest where they are maximal. The anisotropic scaling is necessary to detect wavefronts (Candès and Donoho, 2005b,a) and explains their high compression rates on seismic data (Candès et al., 2006a; Herrmann et al., 2007a,b).

Outline

To maximally leverage the recent insights gained from compressive sampling, we tie the important aspects of this theory into the formulation of the seismic recovery problem. After presenting a brief overview of this theory, including an intuitive explanation, we emphasize the importance of compression rates on the quality of the recovery by means of a series of stylized experiments. Based on this experience, the appropriate sparsifying transform, sampling strategy and minimal velocity constraint that controls the mutual coherence are reviewed, followed by the formulation of our sparsity-promoting inversion method. We conclude by applying this method to various datasets with a focus on improvements of curvelet-based recovery over recovery with plane-wave destruction and the additional benefits from shot-receiver interpolation with 3-D curvelets over recovery from shot records with 2-D curvelets.

COMPRESSIVE SAMPLING

The basics

Compressive sampling states that a signal with a sparse Fourier spectrum can be recovered exactly from sub-Nyquist sampling by solving a sparsity-promoting program that seeks, amongst all possible solutions, a spectrum with the smallest ℓ_1 norm whose inverse Fourier transform equals the sampled data. During the recovery, the rectangular modeling matrix, \mathbf{A} , linking the unknown sparsity N -vector, \mathbf{x} , to the incomplete n -data vector, \mathbf{y} , is inverted. The recovered data is calculated by taking the inverse Fourier transform of the recovered sparsity vector that solves (denoted by the tilde symbol $\tilde{}$) the sparsity promoting program. Compressive sampling provides the conditions under which this underdetermined system of equations ($n \ll N$) can be inverted. This theory also applies to more general situations, including the presence of noise, compressible instead of strictly sparse signals and more general measurement and sparsity bases, replacing the Fourier basis.

To be specific, compressive sampling theory states that \mathbf{P}_ϵ (cf. Eq. 1) recovers in the noise-free case (for $\epsilon \rightarrow 0$) the k non-zero entries of the Fourier N -vector exactly from $n \sim k \times \log N$ samples in the vector, $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ (Candès et al., 2006b). For

random sampling, this condition was recently improved to $n = k \times 2 \log(N/k)$ by Donoho and Tanner (2007) in the regime $N \gg k$.

So, what is the rationale behind these sampling criteria for k -sparse Fourier vectors? Intuitively, one may argue that taking a single time sample corresponds to probing the data by an inner product with a complex exponential in the Fourier domain. This sinusoidal function intersects with any non-zero entry of the unknown Fourier spectrum. One can argue that two intersections from two arbitrary samples should suffice to determine the amplitude and phase for each non-zero entry of the spectrum. Extending this argument to a k -sparse spectrum turns this into a combinatorial problem, seeking the smallest number of nonzero entries in the sparsity vector with an inverse Fourier transform that fits the data. The theory of compressive sampling provides conditions under which the above combinatorial problem can be replaced by \mathbf{P}_ϵ for which practical solvers exist. This theory also provides guidelines for sampling strategies that limit the imprint of interference that leads to coherent aliases. After illustrating the importance of compression for the recovery on a series of stylized experiments, we discuss the design of a compressive sampling procedure that is favorable for the recovery of seismic data with traces missing.

A stylized experiment

Sparsifying transforms form the key component of compressive sampling. As we will show below, the accuracy of the recovery depends on the degree of compression achieved by the sparsifying transform. For signals that are not strictly sparse but compressible, their sparsity properties can be measured by the compression rate, r , defined by the exponent for the powerlaw decay of the magnitude-sorted coefficients. The larger r , the faster the decay of the reconstruction error, measuring the energy difference between the original signal and its approximation from the k largest coefficients. Because \mathbf{P}_ϵ (cf. Eq. 1) recovers the largest k coefficients, the recovery of compressible signals improves in a transformed domain with a large compression rate. The challenge is to find a sparsifying transform that also permits a favorable sampling condition.

A series of experiments is conducted that measures the performance of the recovery as a function of the compression rate and the aspect ratio of the modeling matrix, $\delta = n/N$. This aspect ratio is related to the undersampling rate. As before, a modeling matrix defined in terms of the decimated Fourier matrix is used. The experiments are carried out for varying numbers of measurements, n , and for increasing compression rates, i.e., $(\delta, r) \in (0, 1] \times (1/2, 2]$. For each parameter combination, twenty different pseudo-random realizations are generated defining the random sampling and the entries in the sparsity vector, \mathbf{x}_0 . For each r , this vector is calculated by applying random permutations and signs flips to a sequence that decays with i^{-r} for $i = 1 \cdots N$ with $N = 800$. The incomplete data is generated for each realization with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ and is used as input to StOMP (Donoho et al., 2006a), a solver that solves \mathbf{P}_ϵ approximately, for $\epsilon = 0$. As a performance metric, the squared relative ℓ_2

error, $\text{err}_2 = \|\tilde{\mathbf{x}} - \mathbf{x}_0\|^2 / \|\mathbf{x}_0\|^2$, is calculated and averaged amongst the realizations for fixed $(\delta, r) \in (0, 1] \times (1/2, 2]$. This error is encoded in the greyscale of the recovery diagram, which is included in Fig. 1. Bright regions correspond to parameter combinations that favor accurate recovery. For r fixed, the relative error decays as the number of measurements increases. For each undersampling ratio, $\delta = n/N$, the error decays rapidly as a function of the compression rate, r . This example underlines the importance of finding a representation that has a high compression rate.

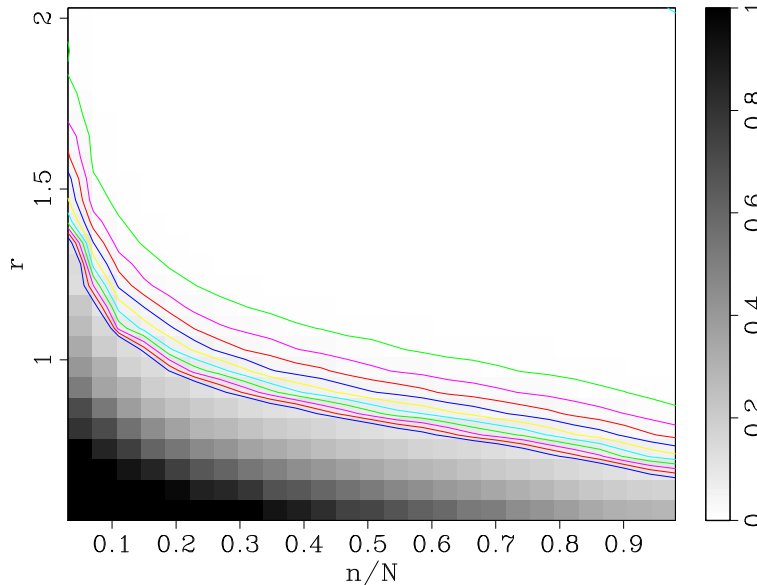


Figure 1: Example of a recovery diagram for parameter combinations $(\delta, r) \in (0, 1) \times (1/2, 2)$ on a regular grid of 25×25 . Notice that the relative ℓ_2 error decays the most rapidly with r . The contour lines represent 1% decrements in the recovery error starting at 10% on the lower-left corner and decaying to 1% in the direction of the upper-right corner.

The recovery diagram contains another piece of important information. For a user-defined recovery error and empirical decay rate, the degree of undersampling can be calculated from the intercept of the appropriate contour with a line of constant approximation rate. Conversely, for a given degree of undersampling, the relative recovery error can be determined by looking at the grey value at the specified parameter combination for (δ, r) .

Approximately a decade ago Sacchi et al. (1998) showed that a sparse Fourier spectrum can be recovered from sub-Nyquist sampling by a Bayesian argument that amounted to the solution of an optimization problem close in spirit to \mathbf{P}_ϵ . While this work has recently been expanded to large-scale problems in higher dimensions by Trad et al. (2006) and Zwartjes and Sacchi (2007), compressive sampling and the presented recovery diagram provide new insights regarding the abruptness of the recovery as a function of the undersampling and the sparsity, and the importance of the compression rate on the quality of the recovery. Unfortunately, the large number

of experiments required to compute the recovery diagram preclude a straightforward extension of these experiments to the seismic situation, where problem sizes exceed ($N = \mathcal{O}(2^{30})$). However, this does not mean that abstract concepts of compressive sampling are not useful in the design of a compressive sampling scheme for seismic data.

COMPRESSIVE SAMPLING OF SEISMIC DATA

Application of the seismic recovery problem according to the principles of compressive sampling requires a number of generalizations. To make these extensions explicit, the modeling matrix is factored into $\mathbf{A} := \mathbf{RMS}^T$, where \mathbf{S}^T (cf. Eq.1) represents the synthesis matrix of the sparsifying transform, \mathbf{M} the measurement matrix and \mathbf{R} the restriction or sampling matrix. The measurement matrix represents the basis in which the measurements are taken and corresponds to the Dirac (identity) basis in seismology and to the Fourier basis in MRI imaging (Lustig et al., 2007). The sampling matrix models missing data by removing zero traces at locations (rows) where data is missing, passing the remaining rows unchanged. The above definition for the modeling matrix is commensurate with the formulation of compressive sampling. As predicted by compressive-sampling theory, the recovery depends quadratically on a new quantity that measures the mutual coherence, $\mu \geq 1$, between the vectors of the measurement and sparsity bases. This mutual coherence is defined as

$$\mu(\mathbf{M}, \mathbf{S}) = \sqrt{M} \max_{(i,j) \in [1 \dots M] \times [1 \dots N]} |\langle \mathbf{m}_i, \mathbf{s}_j \rangle| \quad (2)$$

with \mathbf{m}_i and \mathbf{s}_j the rows of \mathbf{M} and \mathbf{S} , respectively. For the Dirac-Fourier pair, where measurements are taken in Euclidean space of a signal that is sparse in Fourier space, this quantity attains its minimum at $\mu = 1$. Because this property quantifies the spread of the vectors from the measurement basis in the sparsifying domain, it explains successful recovery of signals that are sparse in the Fourier domain from a limited number of Euclidean samples. Compressive-sampling theory extends this idea to different measurement and sparsity matrix pairs and this incoherence quantity proves, aside from the compressibility of the to-be-recovered signal, to be one of the important factors that determines the recovery performance.

Choice for the sparsifying transform

Despite the presence of curved wavefronts with conflicting dips, caustics and a frequency content that spans at least three decades, the curvelet transform attains high compression on synthetic as well as on real seismic data. An intuitive explanation for this behavior lies in the 'principle of alignment', predicting large correlations between curvelets and wavefronts that locally have the same direction and frequency content. This principle is illustrated in Fig. 2 and explains that only a limited number of curvelet coefficients interact with the wavefront while the other coefficients decay

rapidly away from a wavefront. Remark that curvelets require no knowledge on the location of the wavefronts and do not rely on a NMO correction to reduce the spatial bandwidth. However, additional steps such as focusing (see Herrmann et al., 2008) or spatial-frequency content reduction by NMO will improve the recovery but these require extra prior information.

This compression property of curvelets leads, as shown in Fig. 3, to a reconstruction from the largest 1% coefficients that is far superior compared to Fourier- or wavelet-based reconstructions from the same percentage of coefficients. The curvelet result in Fig. 3(d) is artifact free while the Fourier (Fig. 3(b)) and wavelet (Fig. 3(c)) reconstructions both suffer from unacceptable artifacts. Both for synthetic and real data the observed decays of the magnitude-sorted coefficients, as plotted in Fig. C-1 of Appendix C, support the superior performance of curvelets. By virtue of this property, the curvelet transform is the appropriate choice for our sparsifying transform and we set, $\mathbf{S} := \mathbf{C}$ with $\mathbf{C} \in \mathbb{R}^{N \times M}$ the discrete curvelet transform (Candès et al., 2006a; Ying et al., 2005) with $N > M$ the number of curvelet coefficients and M the size of the fully-sampled data volume, $\mathbf{f}_0 \in \mathbb{R}^M$. See the appendices for further detail on the curvelet transform and its performance on seismic data.

Unlike the Fourier and wavelet bases, curvelets form a frame with a moderate redundancy. Frames share many properties with bases but their redundancy requires care in computing the curvelet coefficients, which are no longer unique. Despite the loss of orthogonality, a technical condition required by compressive sampling, curvelets lead to excellent recovery results, which can be understood intuitively.

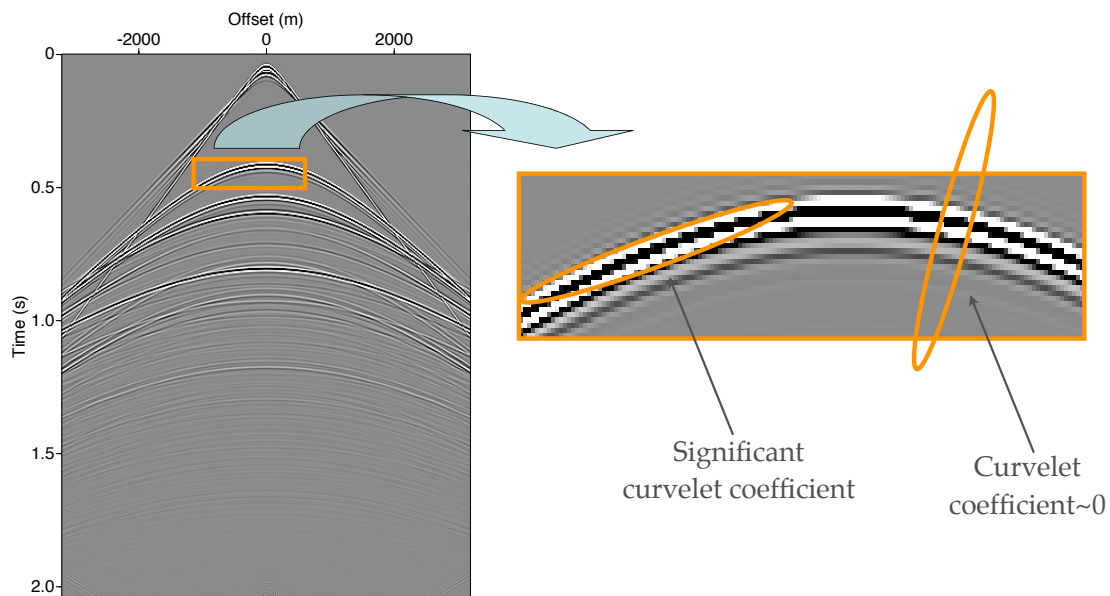


Figure 2: Example of the alignment of curvelets with curved events.

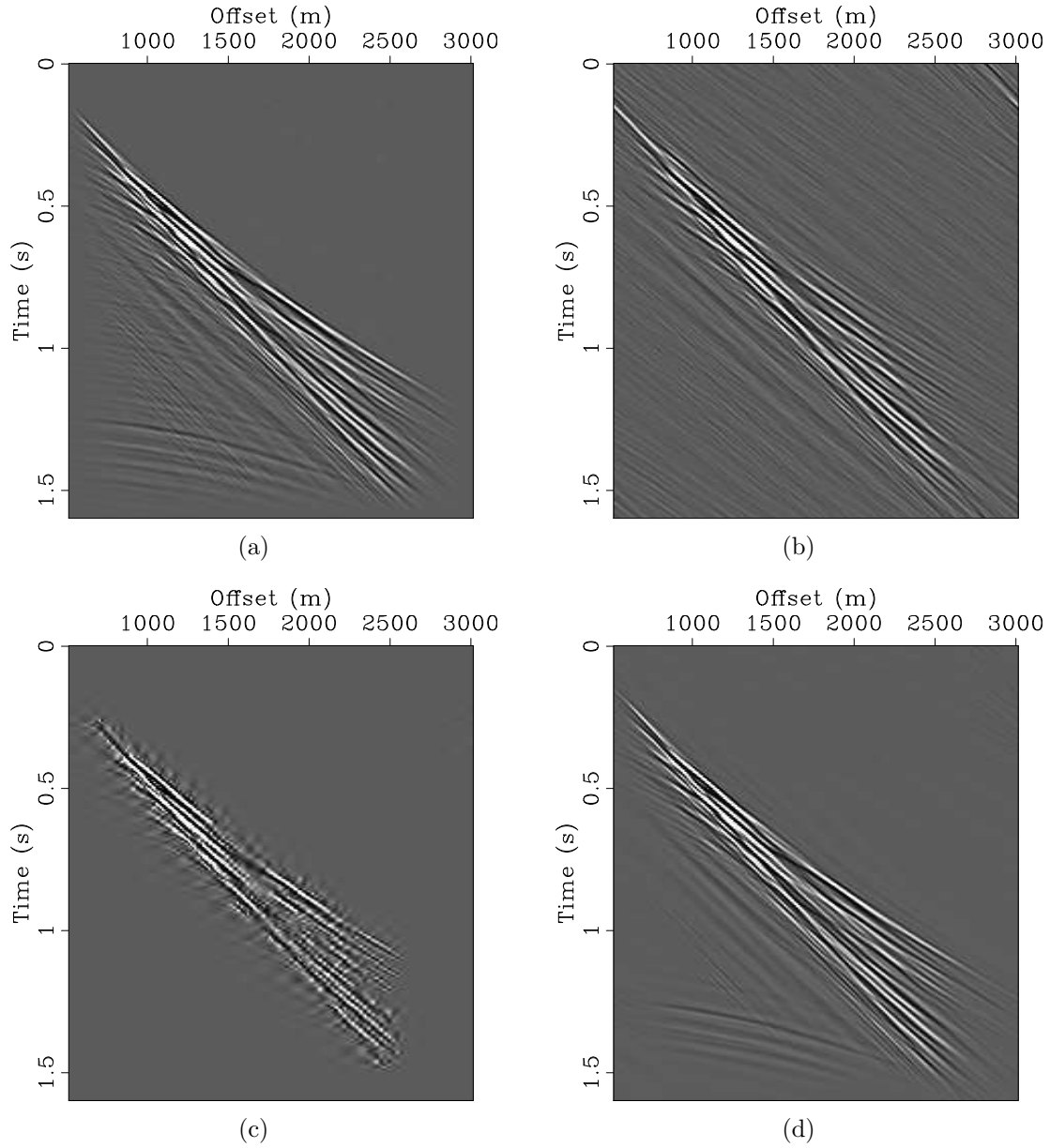


Figure 3: Partial reconstruction in different transform domains. **(a)** Original shot record reconstructed from its 1% amplitude-largest **(b)** Fourier, **(c)** wavelet and **(d)** curvelet coefficients. The curvelet reconstruction is clearly the most accurate approximation.

The measurement matrix

Sampling of seismic wavefields during acquisition can be considered as taking measurements in the Dirac basis, i.e., $\mathbf{M} := \mathbf{I}$ with \mathbf{I} the identity matrix. This is a good approximation for omnidirectional point sources that are impulsive and for receivers with no directivity and a flat frequency response. For this “choice” of measurement basis – the physics of seismic wavefield acquisition limits this choice to this specific type of measurement basis – the recovery conditions are reasonably favorable according to compressive sampling because the Dirac basis is arguably reasonably incoherent with curvelets, whose Fourier spectrum is confined to localized angular wedges (see Fig. 4). We argue that this loss of mutual coherence with respect to the Dirac-Fourier pair is offset by the improved sparsity attained by curvelets (see also our discussion on the role of compression in the stylized examples section). In 3-D this argument gains more strength by virtue of improved sparsity and reduced mutual coherence, i.e., fewer 3-D curvelets are required to capture sheet-like wavefronts while more 3-D curvelets are necessary to cancel each other to approximate a discrete delta Dirac.

Aside from this argument, most if not all practical compressive sampling schemes use sparsifying transforms that are not ideal. For instance, in MRI people use Fourier measurement bases and wavelets as the sparsity basis (Lustig et al., 2007; Candès et al., 2007). At the coarse scales, wavelets become more Fourier-like and hence would adversely affect the recovery. In practice, these less-than-ideal circumstances do not necessarily translate into unfavorable recovery.

Another complication is related to the fact that seismic data is sampled regularly in time and at a subset of source/receiver positions that belong to the acquisition grid. This means that data is fully sampled in time and irregularly along the source/receiver coordinates. This asymmetric trace-by-trace sampling is unfavorable for the recovery because it introduces correlations between vertically-oriented curvelets and vertically-oriented traces along which the data is collected. Fig. 4 illustrates this problem schematically.

To incorporate this additional complication in our formalism, we extend the formal definition of mutual coherence (cf. Eq. 2) by studying the pseudo mutual coherence between the rows of the acquisition matrix, \mathbf{RM} , and the columns of the curvelet synthesis matrix. From this perspective, enforcing a dip discrimination by means of specifying a minimum apparent velocity (see e.g. Zwartjes and Sacchi, 2007), has a natural interpretation in the context of compressive sampling because this discrimination removes steeply dipping curvelets and hence reduces the “mutual coherence” (see Fig. 4). This dip discrimination corresponds to Fourier-domain dip filtering and is equivalent to replacing the Dirac measurement basis with a Toeplitz matrix derived from a dip-filtered discrete delta Dirac. In this case, the mutual coherence will also be reduced, yielding a more favorable recovery condition. This observation is consistent with reports in the geophysical literature, where maximal dip limitation for the recovered wavefields are known to improve recovery (Zwartjes and Sacchi, 2007).

Because curvelets are angular selective, it is straightforward to implement the dip

discrimination as a diagonal weighting matrix in the curvelet domain. This choice not only avoids having to put infinities in a weighting for the ℓ_1 -norm in Eq. 1 but it also allows us to redefine the synthesis matrix as

$$\mathbf{S}^T := \mathbf{C}^T \mathbf{W} \quad \text{with} \quad \mathbf{W} = \text{diag}\{\mathbf{w}\} \quad (3)$$

with $\mathbf{C}^T \in \mathbb{R}^{M \times N}$ the inverse discrete curvelet transform. The weighting vector, \mathbf{w} , contains zeros at positions that correspond to wedges that contain near vertical curvelets and ones otherwise (see Fig. 4). However, this redefinition does not impact the actual wavefield because near vertical events can not occur and leads to a reduced mutual coherence between the rows of the acquisition matrix and the columns of the now restricted curvelet synthesis matrix. This restriction removes the curvelets that correlate with traces in the acquisition and therefore leads to a reduction of the mutual coherence, i.e., the sum in Eq. 2 no longer runs over the vertically oriented curvelets. The observation that reduced coherence leads to favorable recovery conditions is consistent with the theory of compressive sampling.

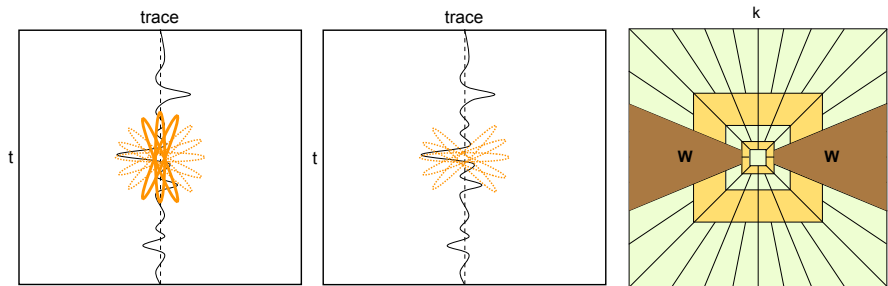


Figure 4: Illustration of the angular weighting designed to reduce the adverse effects of seismic sampling. On the left, the increased mutual coherence between near vertical-oriented curvelets and a missing trace. In the middle, a schematic of the curvelets that survive the angular weighting illustrated on the right.

The restriction/sampling matrix

Curvelet-based recovery performs less well in the presence of strong coherent aliases caused by regular undersampling. These coherent aliases are harmful because they lead to artifacts that have large inner products with curvelets, which may lead to falsely recovered curvelets. The performance of transform-based recovery methods depends on a reduction of these aliases that are caused by constructive interference induced by a regular decimation of the data.

Random subsampling according to a discrete uniform distribution – each discrete grid point is equally probable to be sampled – is known to break aliases. For the restricted Fourier matrix, which consists of the Fast Fourier transform (FFT) applied to a vector with zeros inserted at locations where samples are missing, this random sampling turns aliases into a relatively harmless random noise (according to the slogan

“noiseless underdetermined problems behave like noisy well-determined problems” by Donoho et al., 2006b), allowing for a separation of signal from incoherent interference by a denoising procedure that exploits the sparsifying property of curvelets on seismic data (Hennenfent and Herrmann, 2007a,c). Roughly speaking, this can be understood by arguing that random subsampling according to a discrete uniform distribution corresponds to some sort of a perturbation of the regularly decimated grid that is known to create coherent aliases. As shown in Hennenfent and Herrmann (2007c), this type of sampling, and our extension to jitter sampling, creates a noisy spectrum, where for all wave numbers aliased energy is distributed over the seismic temporal frequency band.

The observation that irregular sampling favors recovery is well known amongst scientists and engineers (Sun et al., 1997; Wisecup, 1998; Malcolm, 2000). Albeit not strictly necessary, we will, for the remainder of this paper, assume that the data is sampled according to a discrete uniform distribution. In practice, there is no need to insist on this condition as long as there is some control on the clustering of the measurements and the size of the largest gaps in the acquisition. Details on this important topic are beyond the scope of this paper and the reader is referred to Donoho and Logan (1992) and to recent applied work by the authors Hennenfent and Herrmann (2007b,c) on jitter sampling.

The modeling matrix

With the sampling and sparsifying matrices in place, the representation for noisy seismic data can now be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \quad \text{with} \quad \mathbf{A} := \mathbf{RIS}^T, \quad (4)$$

$\mathbf{y} \in \mathbb{R}^n$ the noisy measurements and $\mathbf{n} \in \mathbb{R}^n$ a zero-centered pseudo-white Gaussian noise. According to this model, the measurements are related to the sparsity vector \mathbf{x}_0 through the modeling matrix $\mathbf{A} \in \mathbb{R}^{n \times N}$. This modeling matrix is defined by compounding the restriction, $\mathbf{R} \in \mathbb{R}^{n \times M}$; measurement, $\mathbf{I} \in \mathbb{R}^{M \times M}$; and inverse transform, $\mathbf{S}^T \in \mathbb{R}^{M \times N}$ matrices. The noisy measurements themselves are given by $\mathbf{y} = \mathbf{R}\mathbf{f}_0 + \mathbf{n}$ with $\mathbf{R} \in \mathbb{R}^{n \times M}$ the restriction matrix taking $n \ll M$ random samples from the full data vector, $\mathbf{f}_0 \in \mathbb{R}^M$. Because the curvelets transform is redundant, the length of the curvelet vector exceeds the length of the full data vector ($N > M > n$). Therefore, our main task is to invert the modeling matrix \mathbf{A} for situations where $\delta = n/N \approx 0.04$ in 2-D and $\delta \approx 0.01$ in 3-D.

CURVELET RECOVERY BY SPARSITY-PROMOTING INVERSION (CRSI)

The seismic data regularization problem is solved with matrix-free implementations for the fast discrete curvelet transform (defined by the fast discrete curvelet transform,

FDCT, with wrapping, a type of periodic extension, see Candès et al., 2006a; Ying et al., 2005) and the restriction operator. The solution of \mathbf{P}_ϵ (cf. Eq. 1) is cast into a series of simpler unconstrained subproblems. Each subproblem is solved with an iterative soft-thresholding method with thresholds that are carefully lowered. For (extremely) large problems, this cooling leads to the solution of \mathbf{P}_ϵ with a relatively small number ($\mathcal{O}(100)$) of matrix-vector multiplications.

The unconstrained subproblems

The inversion of the underdetermined system of equations in Eq. 4 lies at the heart of compressive sampling. The large system size of seismic data and the redundancy of the curvelet transform exacerbate this problem. Our main thesis is that the matrix, \mathbf{A} , can be successfully inverted with an iterative solution of the sparsity-promoting program \mathbf{P}_ϵ (cf. Eq. 1) by means of a descent method supplemented by thresholding.

Following Elad et al. (2005), the constrained optimization problem, \mathbf{P}_ϵ , is replaced by a series of simpler unconstrained optimization problems

$$\mathbf{P}^\lambda : \begin{cases} \tilde{\mathbf{x}}_\lambda = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ \tilde{\mathbf{f}}_\lambda = \mathbf{S}^T \tilde{\mathbf{x}}_\lambda. \end{cases} \quad (5)$$

These subproblems depend on the Lagrange multiplier λ , determining the emphasis of the ℓ_1 -norm over the ℓ_2 data misfit. The solution of \mathbf{P}_ϵ is reached by solving \mathbf{P}^λ for $\lambda \downarrow \lambda_\epsilon$ with $\lambda_\epsilon = \sup_\lambda \{\lambda : \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}_\lambda\|_2 \leq \epsilon\}$. During the solution of the nonlinear optimization problem \mathbf{P}^λ , the rectangular matrix \mathbf{A} is inverted by first emphasizing the sparsity-promoting ℓ_1 -norm, yielding sparse approximate solutions, followed by a relaxation as λ decreases, increasing the energy captured from the data.

Solution by iterative thresholding

Following Daubechies et al. (2004), Elad et al. (2005); Candès and Romberg (2004) and ideas dating back to Figueiredo and Nowak (2003), the subproblems \mathbf{P}^λ are solved by an iterative thresholding technique that derives from the Landweber descent method (Vogel, 2002). According to Daubechies et al. (2004) looping over

$$\mathbf{x} \leftarrow T_\lambda (\mathbf{x} + \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{x})), \quad (6)$$

converges to the solution of \mathbf{P}^λ with

$$T_\lambda(x) := \text{sgn}(x) \cdot \max(0, |x| - |\lambda|) \quad (7)$$

the soft-thresholding operator. This convergence requires a large enough number of iterations and a largest singular value of \mathbf{A} that is smaller than 1, i.e. $\|\mathbf{A}\| < 1$. Each iteration requires two matrix-vector multiplications.

The descent update, $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x})$, minimizes the quadratic part of Eq. 5. This update is subsequently projected onto the ℓ_1 ball by the soft thresholding. Even though this procedure provably converges to the solution of \mathbf{P}^λ , the large scale of the seismic regularization problem precludes running these iterations to convergence within a reasonable number of matrix-vector multiplications.

Final solution by cooling

Cooling is a common strategy to solve large to extremely large-scale problems. During this cooling process, the subproblems \mathbf{P}^λ are solved approximately for λ decreasing. Because of its simplicity, the iterative-thresholding technique, presented in Eq. 6, lends itself particularly well for this approach since it offers a warm start, typically given by the previous outer loop, and control over the accuracy. This accuracy is related to the number of iterations, L , of the inner loop. The higher L the more accurate the solutions of the subproblems become.

The convergence of the overall problem is improved by using the approximate solution of the previous subproblem, the warm start, as input to the next problem for which λ is slightly decreased (Starck et al., 2004; Elad et al., 2005). Sparsity is imposed from the beginning by setting λ_1 close to the largest curvelet coefficient, i.e. $\lambda_1 < \|\mathbf{A}^T\mathbf{y}\|_\infty$. As the Lagrange multiplier is lowered, more coefficients are allowed to enter the solution, leading to a reduction of the data misfit. A similar approach, derived from POCS (Bregman, 1965), was used by Candés and Romberg (2004) and Abma and Kabir (2006). The details of the cooling method are presented in Table. 1.

In practice, five inner loops and 10-50 outer loops suffice to solve for \mathbf{x} with the series of subproblems \mathbf{P}^λ . When the cooling is appropriately chosen, the solution of the subproblems converges to the solution of \mathbf{P}_ϵ . The final solution to the seismic data regularization problem, $\tilde{\mathbf{f}}$, is obtained by applying the weighted-inverse curvelet transform to $\tilde{\mathbf{x}}$, i.e., $\tilde{\mathbf{f}} = \mathbf{S}^T\tilde{\mathbf{x}}$. The total number of matrix-vector multiplications required by this method is similar to those required by iterative-re-weighted least-squares (Gersztenkorn et al., 1986).

SEISMIC DATA RECOVERY WITH CRSI

The performance of our recovery algorithm is evaluated on synthetic as well as on real data. The first synthetic example is designed to highlight our ability to handle conflicting dips. Next, a synthetic seismic line is used to study the potential uplift for a recovery with 3-D curvelets over a recovery with 2-D curvelets. Finally, our method is tested on real data and compared to a regularization method based on plane-wave destruction (Fomel et al., 2002).

```

Initialize:
i = 0;  $\mathbf{x}^0 = \mathbf{0}$ ;
Choose: L,  $\|\mathbf{A}^T \mathbf{y}\|_\infty > \lambda_1 > \lambda_2 > \dots$ 
while  $\|\mathbf{y} - \mathbf{A}\mathbf{x}^i\|_2 > \epsilon$  do
  for l = 1 to L do
     $\mathbf{x}^{i+1} = T_{\lambda_i}(\mathbf{x}^i + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^i))$ 
  end for
  i = i + 1;
end while
 $\tilde{\mathbf{f}} = \mathbf{S}^T \mathbf{x}^i$ .

```

Table 1: The cooling method with iterative thresholding.

2-D synthetic for a layered earth model

Consider the reflection response of a medium with four plane layers, modeled with a 50-foot (15.24-m) receiver interval, 4-ms sampling interval and a source function given by a Ricker wavelet with a central-frequency of 25-Hz. The dataset contains 256 traces of 500 time samples each. The resulting common-midpoint (CMP) gather after incomplete acquisition is shown in Fig. 5(a) together with a close-up in Fig. 5(b) of an area with conflicting dips. The incomplete acquisition was simulated by randomly removing 60% of the traces. This undersampling corresponds to a sub-Nyquist average spatial sampling of 125 feet (38.1 m).

Based on the maximum expected dip of the reflection events in the data, a minimum velocity constraint of 5000 ft/s (1524 m/s) was used. To limit the number of unknowns, the negative dips were excluded. Figs. 5(c) and 5(d) show the results for the CMP reconstruction with the CRSI algorithm for 100 iterations (5 inner- and 20 outer-loops). The starting Lagrange multiplier was chosen such that 99.5% of the coefficients do not survive the first threshold. Since the data is noise free, the Lagrange multiplier is lowered such that 99% of the coefficients survives the final threshold. This corresponds to the situation where \mathbf{P}_ϵ is solved with a constraint that is close to an equality constraint, i.e., nearly all energy of the incomplete data is captured.

Figs. 5(e) and 5(f) plot the difference between the recovered and 'ground-truth' complete data. The SNR for the recovery, defined as $SNR = 20 \log \|\tilde{\mathbf{f}} - \mathbf{f}_0\| / \|\mathbf{f}_0\|$, is about 29.8 dB, which corroborates the observation that there is almost no energy in the difference plots. Curvelet reconstruction clearly benefits from continuity along the wavefronts in the data and has no issue with conflicting dips thanks to the multidirectional properties of curvelets.

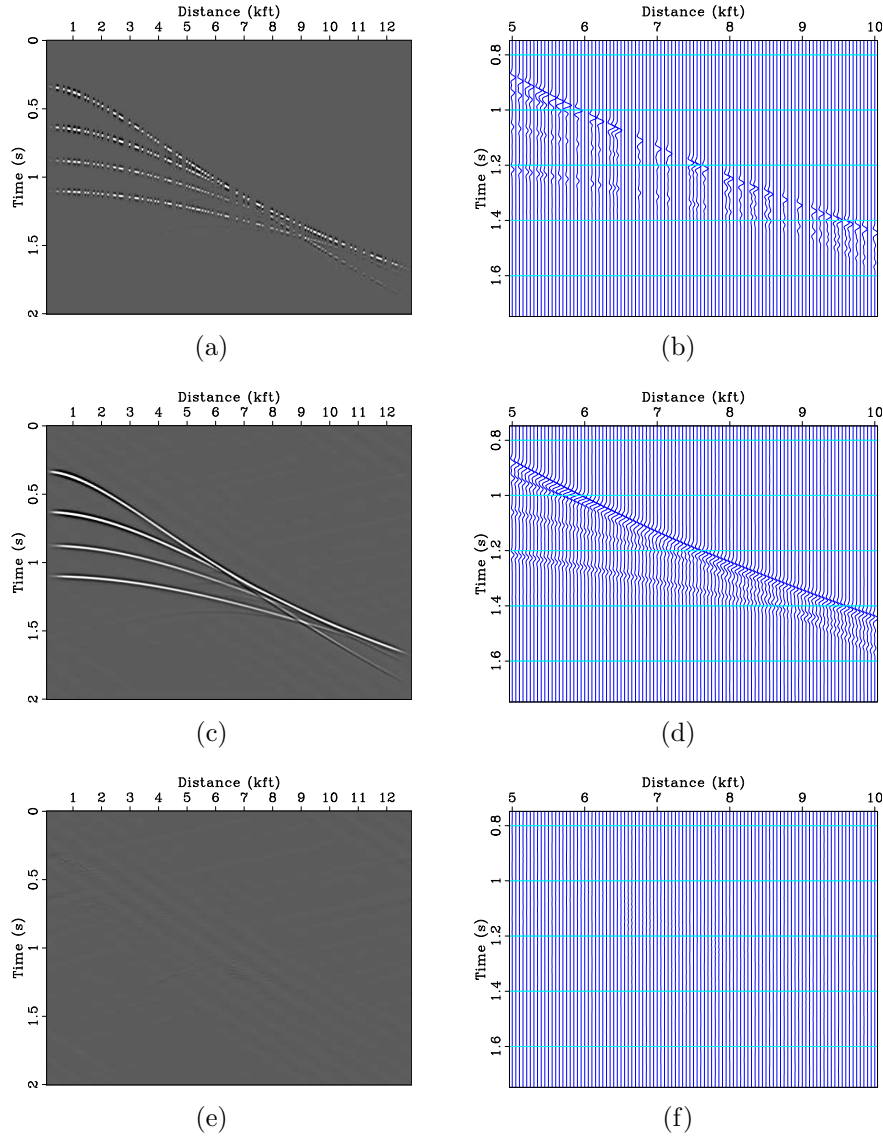


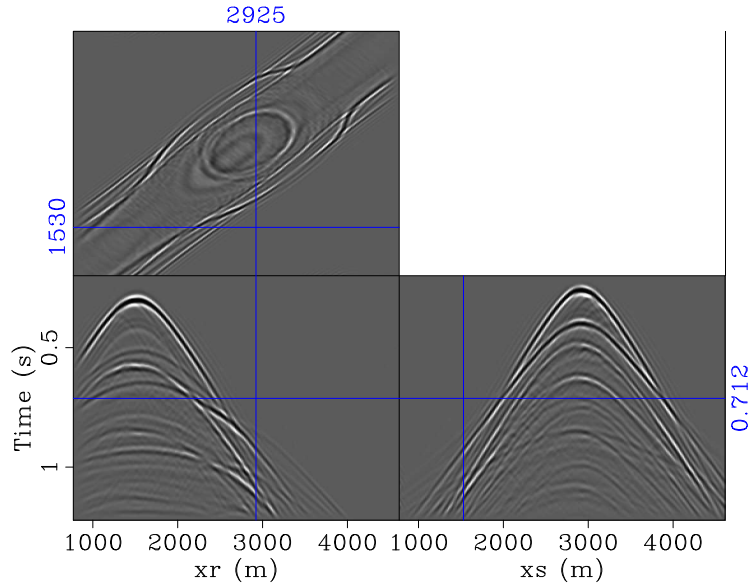
Figure 5: Synthetic example of curvelet 2-D reconstruction. (a) Simulated acquired data with about 60% randomly missing traces and (b) zoom in a complex area of the CMP gather. (c) Curvelet reconstruction and (d) same zoom as (c). (e) Difference between reconstruction and complete data (not shown here) and (f) zoom. Virtually all the initial seismic energy is recovered without error as illustrated by the difference plots (SNR = 29.8 dB).

Common-shot/receiver versus shot-receiver interpolation

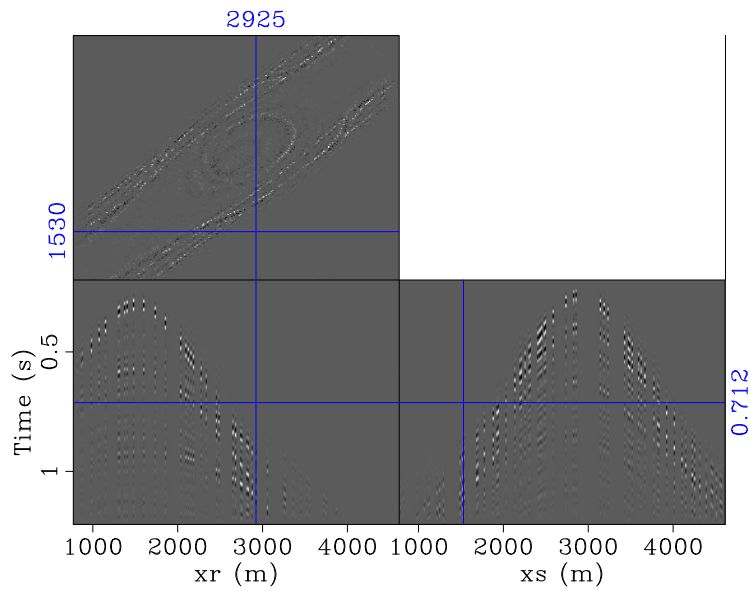
Curvelets derive their success in seismology from honoring the multi-dimensional geometry of wavefronts in seismic data. To illustrate the potential benefit from exploiting this high-dimensional geometry, a comparison is made between common-shot interpolation with 2-D curvelets and shot-receiver interpolation with 3-D curvelets. For this purpose, a synthetic seismic line is simulated with a finite-difference code for a subsurface velocity model with two-dimensional inhomogeneities. This velocity model consists of a high-velocity layer that represents salt, surrounded by sedimentary layers and a water bottom that is not completely flat. Using an acoustic finite-difference modeling algorithm, 256 shots with 256 receivers are simulated on a fixed receiver spread with receivers located from 780 to 4620 m with steps of 15 m. The temporal sample interval is 4 ms. The data generated by these simulations can be organized in a 3-D data volume (shot-receiver volume) along the shot, x_s , receiver, x_r and time, t coordinates. The full data and the incomplete acquisition are depicted in Fig. 6(a) and 6(b). The incomplete acquisition is simulated by randomly removing 80% of the receiver positions for each shot, which corresponds to an average spatial sampling interval of 75 m. Again the full data serves as the ground truth.

To make the comparison, we either solve a series of 2-D dimensional problems on individual shot gathers or we solve the full 3-D interpolation problem. This procedure is outlined in Fig. 7 with results for one selected shot record summarized in Fig. 8. These results show a clear improvement for the interpolation with the 3-D curvelet transform over the recovery from individual shot records with 2-D curvelets. For both cases results were obtained with 250 iterations and without imposing a minimal velocity constraint. We omitted this constraint because we want to study the uplift without interference from this velocity constraint. Contrasting the results in Figs. 8(c) and 8(e) confirms the improved recovery by exploiting the 3-D structure, an observation corroborated by the difference plots. The improvement in continuity is particularly visible for the shallow near zero-offset traces where the events have a large curvature. The SNR's for the 2- and 3-D curvelet-based recovery are 3.9 dB and 9.3 dB, respectively, which confirms the visual improvement.

As a possible explanation for the observed performance gain for 3-D curvelets, we argue that 3-D curvelets make up for the increased redundancy (a factor of 24 for 3-D compared to only a factor of 8 in 2-D) by exploiting continuity of wavefronts along an extra tangential direction. This extra direction leads to an improved concentration of the energy amongst relatively fewer curvelet coefficients. The increased dimensionality of 3-D curvelets also makes intersections with areas where data is present more likely. Finally, the theory of compressive sampling tells us that the recovery performance is proportional to the mutual coherence. In 2-D, curvelets are locally line like while 3-D curvelets are locally plate like. Consequently, the mutual coherence between a vertical-oriented 3-D curvelet and a trace is smaller than its 2-D counterpart and this also explains the improved recovery. The result plotted in Fig. 9(a) and the difference plot in Fig. 9(b) confirm the expected improvement and the recovered



(a)



(b)

Figure 6: Synthetic data volume. **(a)** Complete dataset consisting of $256 \times 256 \times 256$ samples along the source, x_s , receiver, x_r and time coordinates. **(b)** Simulated acquired data with 80% randomly missing traces.

data displays a nice continuity along the reconstructed wavefronts. Moreover, there is only minor residual energy in the difference plots for a time slice, common-shot and common-receiver panels. The positions of these slices are indicated by the vertical and horizontal in the different panels. The SNR for the 3-D recovery with the 3-D curvelets is 16.92 dB, which is by all means acceptable.

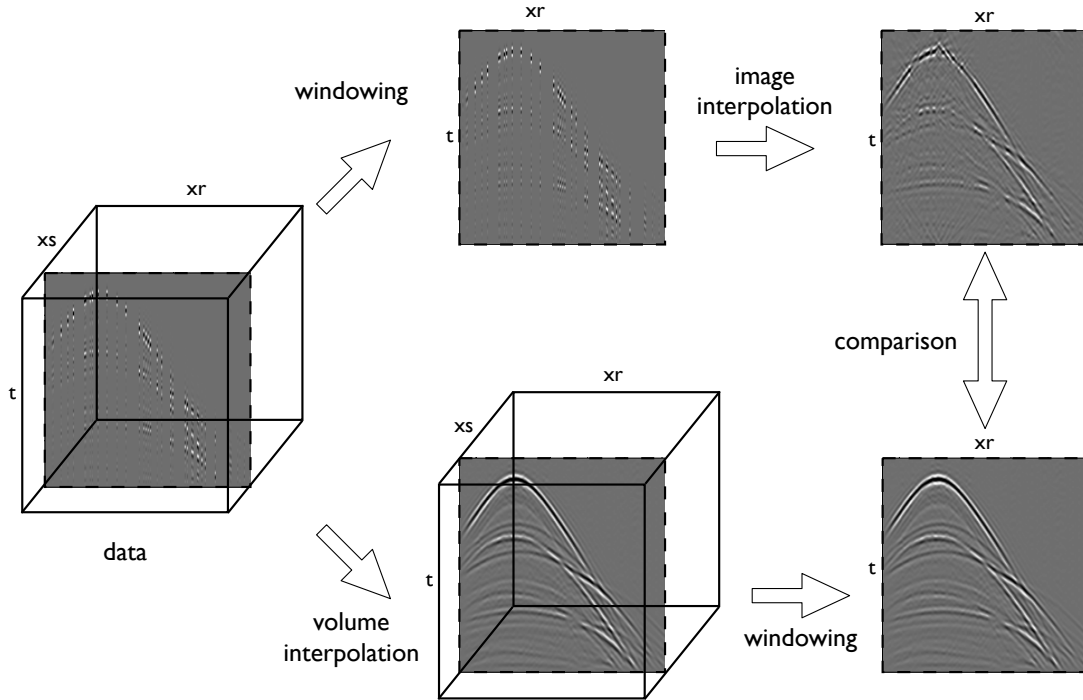


Figure 7: Illustration of common shot versus shot-receiver interpolation on the complete data volume.

Comparison between CRSI and plane-wave destruction on 2-D real data

To conclude the discussion, our method is contrasted with an interpolation method based on plane-wave destruction (Fomel et al., 2002). Fig. 10(a) displays a real shot record that is used for the comparison. This record is taken from a seismic survey, collected at the offshore Gippsland basin in Australia, and contains traces with the first 1.7 s of data received at 200 hydrophones. The data is sampled at 4 ms with a receiver spacing of 12.5 m. The data is decimated by randomly removing 60% of the traces, which corresponds to an average spatial sampling interval of 31.25 m. The results obtained with CRSI and the plane-wave destruction method are included in Fig. 10. The CRSI result shows a nice recovery with a small residual error. The interpolation result and difference plot for the plane-wave destruction method are included in Figs. 10(e) and 10(f). These results clearly indicate the challenges imposed by real data, with the recovery performing well for regions with

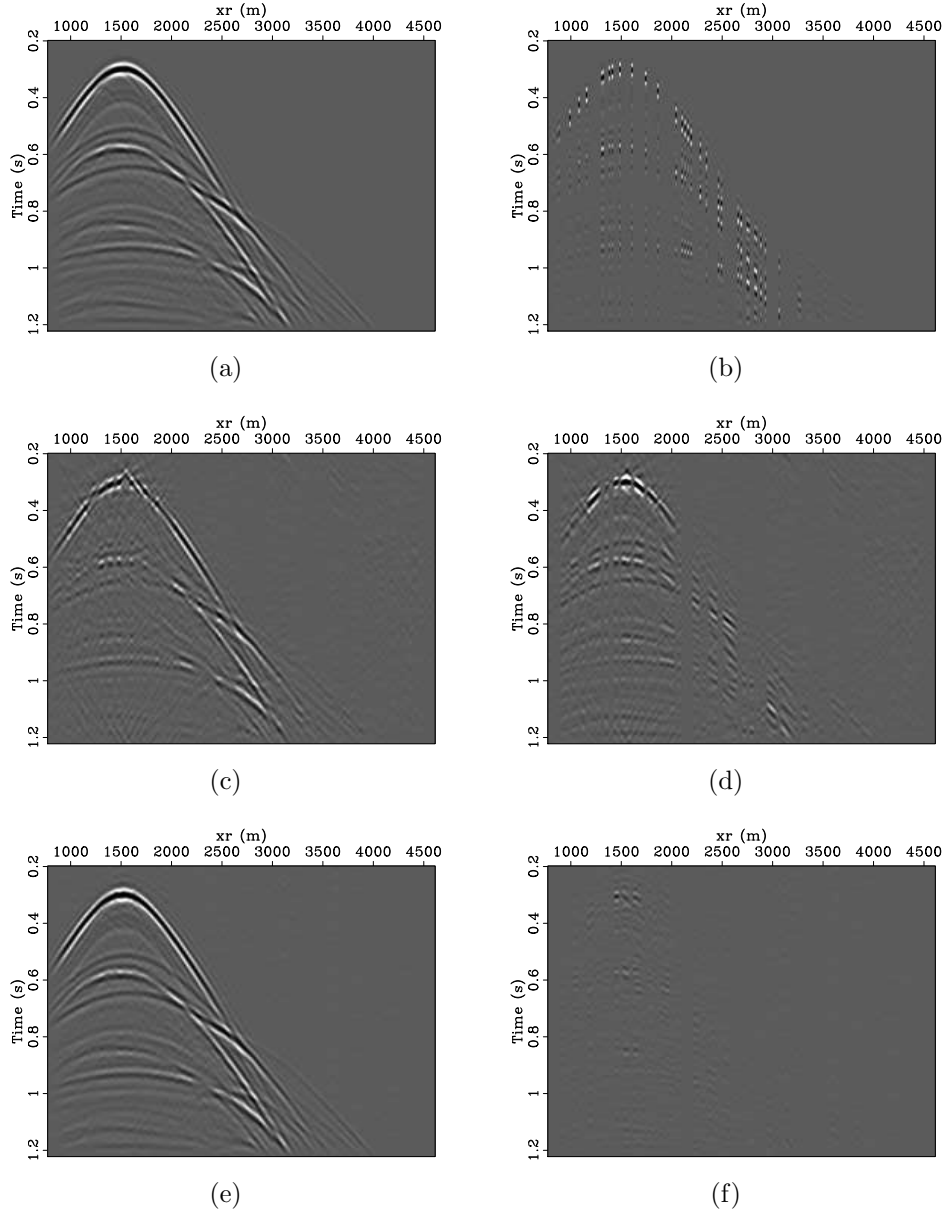
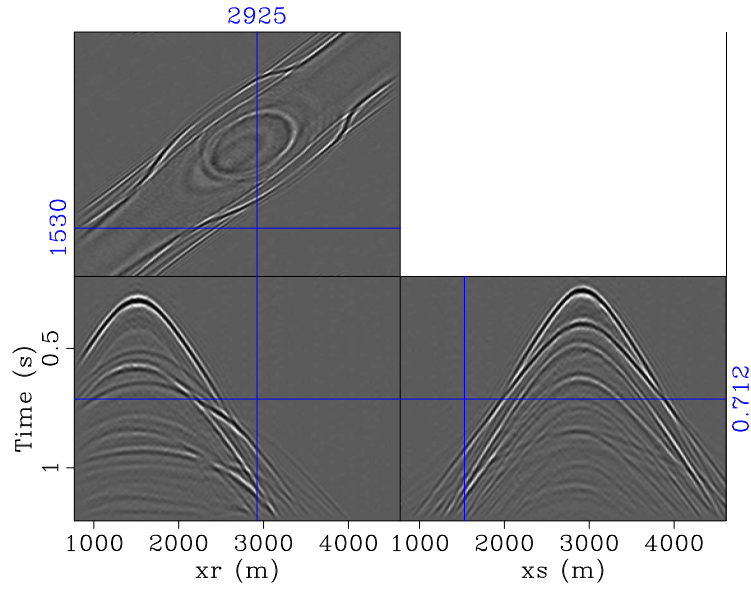
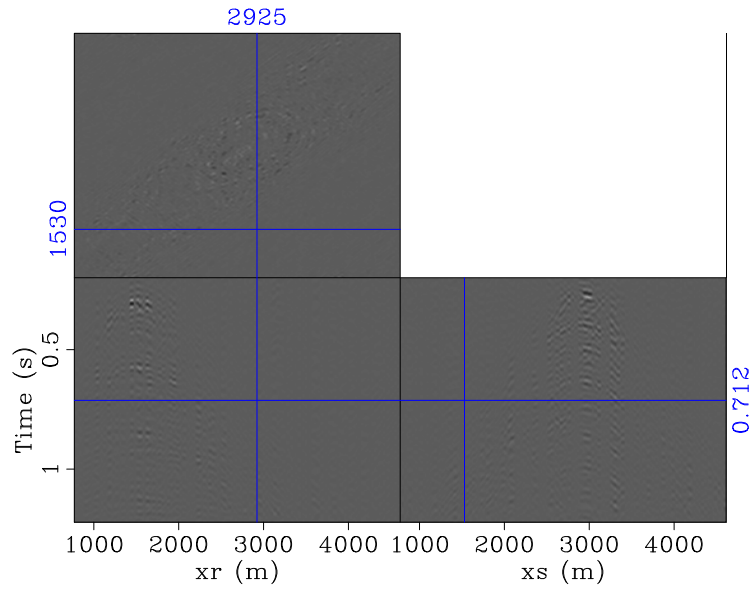


Figure 8: Comparison between common-shot (2-D) and shot-receiver (3-D) CRSI. **(a)** Shot from the original data volume. **(b)** Corresponding simulated incomplete data with 80% randomly missing traces. **(c)** 2-D CRSI result. **(d)** Difference between **(c)** and **(a)**. **(e)** Shot extracted from 3-D CRSI result. **(f)** Difference between **(e)** and **(a)**. 3-D CRSI clearly benefits from 3-D information that greatly improves the reconstruction over 2-D CRSI.



(a)



(b)

Figure 9: Synthetic example of curvelet volume interpolation. **(a)** 3-D CRSI result based on the simulated acquired data of Fig. 8(b). **(d)** Difference between Fig. 8(a) and **(a)**. Notice the continuity and the small difference in the common-shot, common-receiver and time slice. The positions in the cube are indicated by the numbered lines.

low complexity. However, the plane-wave destruction method does not perform so well for regions where there is more complexity and in particular in regions with conflicting dips. In those areas our curvelet-based method maintains its performance while the plane-wave destruction creates events with erroneous dips. This problem can be related to the inability to assign unique slopes to the reflection events. Curvelets do not experience these difficulties because they can handle multiple dips at the same location. Again, the improved performance is reflected in the SNR's, which is 18.8 dB for 2-D CRSI compared to 5.5 dB for the plane-wave destruction.

DISCUSSION

Initial findings

Compressive sampling: We showed that the concepts of compressive sampling apply to the seismic recovery problem. Indeed, some of the ideas of compressive sampling are not exactly new to (exploration) seismology, where Fourier, Radon and even migration-based high-resolution approaches have been used to solve the seismic regularization problem. However, compressive sampling offers a clear and concise framework that gives insights in the workings of a successful recovery. These insights offered guidance while making specific choices to exploit the inherent geometry within the seismic wavefield and to eliminate aliases and correlations due to trace-by-trace sampling. Most importantly, compressive sampling tells us that the largest entries of the sparsity vector are recovered thereby underlining the importance of sparsifying transform for seismic data.

Sparsifying transform: An important factor contributing to the performance of our method is the ability of curvelets to parsimoniously capture the essential characteristics of seismic wavefields. This property explains the rapid decay for the magnitude-sorted coefficients and the relative artifact-free reconstruction from a relatively small percentage of largest coefficients. The moderate coherence between the seismic measurement basis and curvelets and the inclusion of the minimal-velocity constraint all contribute to the success of our method. Finally, the results from shot-receiver interpolation showed significant improvement over interpolation on shot records. This behavior is consistent with findings in the literature on Fourier-based recovery (Zwartjes and Sacchi, 2007).

The cooling method: Despite its large scale, the seismic recovery problem lends itself particularly well for a solution by iterative thresholding with cooling. As the threshold is lowered, additional components enter into the solution, which leads to an improved data misfit and controlled loss of sparsity. We find it quite remarkable that this relatively simple threshold-based solver performs so well on the solution of ℓ_1 problems that can be considered as large to extremely large. In a future paper, we

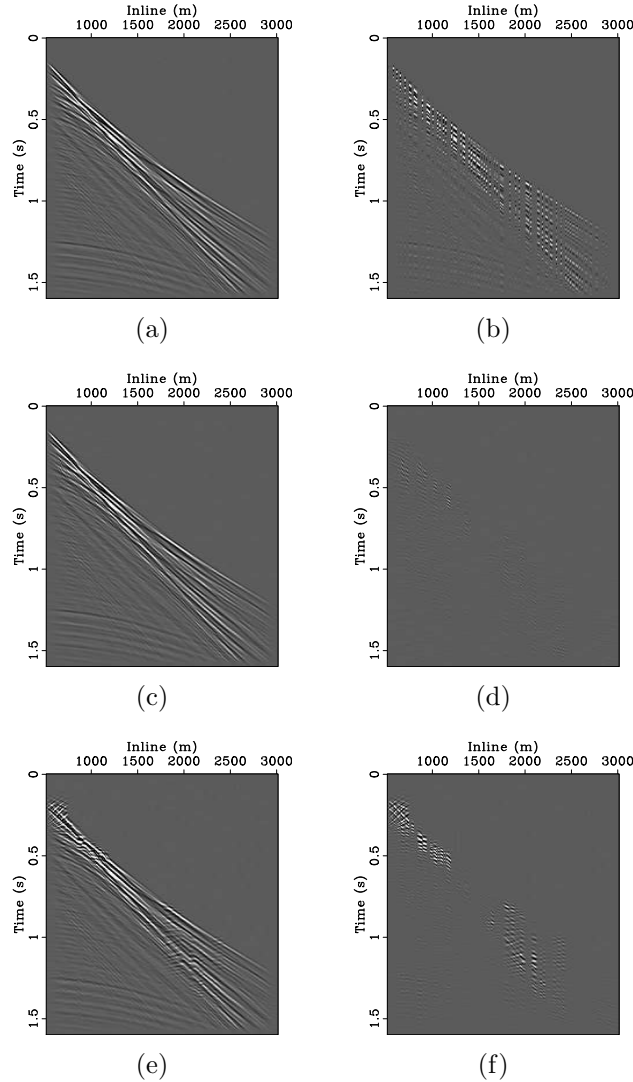


Figure 10: Comparison of plane-wave destruction and curvelet-based 2-D recovery real data. **(a)** Shot-record of a seismic survey from offshore Gippsland basin Australia. Group interval is 12.5 m. **(b)** Incomplete data derived from **(a)** by randomly removing 60% of the traces (corresponding to average spatial sampling is 31.25 m). **(c)** Result obtained with CRSI. **(d)** Difference between CRSI result and ground truth. **(e)** and **(f)** the same as **(c)** and **(d)** but now obtained with plane-wave destruction. The improvement of the curvelet-based method over the plane-wave destructions is corroborated by the SNR's which are 18.8 dB 5.5 dB, respectively.

plan to report on the properties of this solver compared to other recent developments in solver technology, emerging within the field of compressive sampling (Tibshirani, 1996; Candès and Romberg, 2005; Donoho et al., 2005; Figueiredo et al., 2007; Koh et al., 2007; van den Berg and Friedlander, 2007).

Extensions

Focused CRSI: Our recovery method can be improved when additional information on the wavefield is present. For instance, as part of SRME estimates for the primaries in the data are available. These estimates can be used to focus the energy by compounding the modeling matrix of CRSI with an operator defined by the estimate for the major primaries. As shown by Herrmann et al. (2007c, 2008), this inclusion leads to a better recovery that can be attributed to an improved compression due to focusing with the primaries.

The parallel curvelet transform: Aside from the large number of unknowns within the recovery, seismic datasets typically exceed the memory size of compute nodes in a cluster. The fact that seismic data is acquired in as many as five dimensions adds to this problem. Unfortunately, the redundancy of the curvelet transform makes it difficult to extend this transform to higher dimensions. By applying a domain decomposition in three dimensions, some progress has been made (Thomson et al., 2006). The second problem is still open and may require combination with other transforms.

Jitter sampling: During random sampling there is no precise control over the size of the gaps. This lack of control may lead to an occasional failed recovery. Recently, Hennenfent and Herrmann (2007b) have shown that this problem can be avoided by jitter sampling. During this jitter sampling, the size of the gaps and the occurrence of coherent aliases are both controlled. We report on this recent development elsewhere (Hennenfent and Herrmann, 2007c).

CRSI for unstructured data: The presented interpolation method assumed data to be missing on otherwise regular grids. With the non-uniform fast discrete curvelet transform developed by the authors (Hennenfent and Herrmann, 2006), CRSI no longer requires data to be collected on some underlying grid. This extension makes CRSI applicable in other fields such as global seismology, where irregular sampling and spherical coordinate systems prevail.

Fast (reweighted) ℓ_1 solvers: The success of compressed sensing depends on the ability to solve large-scale ℓ_1 optimization problems. As a result, there has been a surge in research activity addressing this important issue (Tibshirani, 1996; Candès

and Romberg, 2005; Donoho et al., 2005; Figueiredo et al., 2007; Koh et al., 2007). One development is particularly relevant and that is the discussion (see Candès et al., 2007, for further detail) whether to solve the recovery problem according to Eq. 1, known as the synthesis problem or, according to

$$\mathbf{P}_\epsilon^a : \quad \tilde{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{Cf}\|_1 \quad \text{s.t.} \quad \|\mathbf{RMf} - \mathbf{y}\|_2 \leq \epsilon, \quad (8)$$

which is known as the analysis problem. Even though there are reports in the literature (Candès et al., 2007) that state that the analysis form (cf. Eq. 8) leads to improved recovery results, our experience with (extremely) large problems in CRSI have shown better recovery with the synthesis formulation (cf. Eq. 1). Because current hardware affords only $\mathcal{O}(100)$ matrix-vector multiplies, the future challenge will be the inclusion of more sophisticated ℓ_1 -norm solvers and the investigation of potential benefits from a possible reweighting and a formulation in the analysis form. The latter corresponds to an approximate solution for the ℓ_0 problem for which encouraging results have been reported (Candès et al., 2007). In a future paper, we plan to report on these issues.

CONCLUSIONS

A new non-parametric seismic data regularization technique was proposed that combines existing ideas from sparsity-promoting inversion with parsimonious transforms that expand seismic data with respect to elements that are multiscale and multidirectional. The compression attained by these elements, which form the redundant curvelet frame, in conjunction with an acquisition that is not too far from random, led to a compressive sampling scheme that recovers seismic wavefields from data with large percentages of traces missing.

Treating the seismic data regularization problem in terms of a compressive sampling problem enabled us to design a scheme that favored recovery. The success of this scheme can be attributed to three main factors, namely the compression seismic wavefields by curvelets, the control of aliases by (close to) random sampling and the solution of (extremely) large-scale ℓ_1 problems by a cooled iterative thresholding. This combination allowed us to reconstruct seismic wavefields from data with up to 80% of its traces missing at a cost comparable to other sparsifying transform-based methods. Our method was successfully applied to synthetic and real data. A significant improvement was witnessed for shot-receiver interpolation during which the 3-D geometry of seismic wavefields is fully exploited by 3-D curvelets. Our results also showed a significant improvement on real data with conflicting dips amongst the wave arrivals.

Unfortunately, compressive sampling does not offer explicit sampling criteria for a curvelet-based recovery of seismic wavefields. However, this theory has given us insights that justified the design of our recovery method, where the seismic data regularization problem is solved by sparsity promotion in the curvelet domain.

ACKNOWLEDGMENTS

The authors would like to thank the authors of CurveLab, SparseLab and the Rice Wavelet Toolbox for making their codes available. In particular, we would like to thank Laurent Demanet for providing us with further insights into the curvelet transform. This paper was prepared with Madagascar (rsf.sourceforge.net) supplemented by SLIMpy operator overloading, developed by S. Ross Ross. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and Collaborative Research and Development Grant DNOISE (334810-05) of F.J.H. and was carried out as part of the SINBAD project with support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and Shell. The authors would also like to thank ExxonMobil Upstream Research Company for providing us with the real dataset and the Institute of Pure and Applied Mathematics at UCLA supported by the NSF under grant DMS-9810282. Finally, the authors would like to thank the anonymous reviewers whose constructive comments helped improve this paper.

REFERENCES

- Abma, R. and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: *Geophysics*, **71**, E91–E97.
- Bleistein, N., J. Cohen, and J. Stockwell, 2001, *Mathematics of Multidimensional Seismic Imaging, Migration and Inversion*: Springer.
- Bregman, L., 1965, The method of successive projection for finding a common point of convex sets: *Soviet Math. Dokl.*, **6**, 688–692.
- Candès, E. and D. L. Donoho, 2005a, Continuous Curvelet Transform II: Discretization and Frames: *Appl. Comput. Harmon. Anal.*, **19**, 198–222.
- Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: *SIAM Multiscale Model. Simul.*, **5**, 861–899.
- Candès, E. J. and D. L. Donoho, 2000a, Curvelets – a surprisingly effective nonadaptive representation for objects with edges: Presented at the *Curves and Surfaces*, Vanderbilt University Press.
- , 2000b, Recovering Edges in Ill-posed Problems: Optimality of Curvelet Frames: *Ann. Statist.*, **30**, 784–842.
- , 2004, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities: *Comm. Pure Appl. Math.*, **57**, 219–266.
- , 2005b, Continuous Curvelet Transform I: Resolution of the Wavefront Set: *Appl. Comput. Harmon. Anal.*, **19**, 162–197.
- Candès, E. J. and J. Romberg, 2004, Practical signal recovery from random projections: Presented at the *Wavelet Applications in Signal and Image Processing XI*.
- , 2005, ℓ_1 -magic. (Software, <http://www.acm.caltech.edu/l1magic/>).
- Candès, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: *Comm. Pure Appl. Math.*, **59**, 1207–1223.
- Candès, E. J., M. B. Wakin, and S. P. Boyd, 2007, Enhancing sparsity by reweighted ℓ_1 minimization.
- Canning, A. and G. H. F. Gardner, 1996, Regularizing 3-D data sets with DMO: *Geophysics*, **61**, 1103–1114.
- Daubechies, I., M. Defrise, and C. de Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints: *CPAM*, 1413–1457.
- Donoho, D. L., I. Drori, V. Stodden, and Y. Tsaig, 2005, SparseLab. (Software, <http://sparselab.stanford.edu/>).
- , 2006a, SparseLab.
- Donoho, D. L. and B. F. Logan, 1992, Signal Recovery and the Large Sieve: *SIAM Journal on Applied Mathematics*, **52**, 577–591.
- Donoho, D. L. and J. Tanner, 2007, How many random projections does one need to recover a k -sparse vector. (submitted for publication).
- Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006b, Sparse solution of underdetermined linear equations by stagewise orthonormal matching pursuit. (Preprint).
- Donoho, P., R. Ergas, and R. Polzer, 1999, Development of seismic data compression methods for reliable, low-noise performance: *SEG International Exposition and 69th Annual Meeting*, 1903–1906.

- Elad, M., J.-L. Starck, P. Querre, and D. L. Donoho, 2005, Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA): *Appl. Comput. Harmon. Anal.*, **19**, 340–358.
- Figueiredo, M. and R. D. Nowak, 2003, An EM algorithm for wavelet-based image restoration: *IEEE Trans. Image Processing*.
- Figueiredo, M., R. D. Nowak, and S. J. Wright, 2007, Gradient Projection for Sparse Reconstruction. (Software, <http://www.lx.it.pt/~mtf/GPSR/>).
- Fomel, S., 2003, Theory of differential offset continuation: *Geophysics*, **68**, 718–732.
- Fomel, S., J. G. Berryman, R. G. Clapp, and M. Prucha, 2002, Iterative resolution estimation in least-squares Kirchoff migration: *Geophys. Pros.*, **50**, 577–588.
- Gersztenkorn, A., J. B. Bednar, and L. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: *Geophysics*, **51**, 357–369.
- Hale, D., 1995, DMO processing: *Geophysics Reprint Series*.
- Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: *Geophysics*, **49**, 1869–1880.
- Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: *IEEE Comp. in Sci. and Eng.*, **8**, 16–25.
- , 2007a, Irregular sampling: from aliasing to noise: Presented at the EAGE 69th Conference & Exhibition.
- , 2007b, Random sampling: new insights into the reconstruction of coarsely-sampled wavefields: Presented at the SEG International Exposition and 77th Annual Meeting.
- , 2007c, Simply denoise: wavefield reconstruction via coarse nonuniform sampling: *Geophysics*. (To appear).
- Herrmann, F. J., 2005, Robust curvelet-domain data continuation with sparseness constraints: Presented at the EAGE 67th Conference & Exhibition Proceedings.
- Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007a, Nonlinear primary-multiple separation with directional curvelet frames: *Geoph. J. Int.*, **170**, 781–799.
- Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2007b, Sparsity- and continuity-promoting seismic imaging with curvelet frames: *Appl. Comput. Harmon. Anal.* (To appear, doi:10.1016/j.acha.2007.06.007).
- Herrmann, F. J., D. Wang, and G. Hennenfent, 2007c, Multiple prediction from incomplete data with the focused curvelet transform: Presented at the SEG International Exposition and 77th Annual Meeting.
- Herrmann, F. J., D. Wang, G. Hennenfent, and P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: *Geophysics*, **73**.
- Koh, K., S. J. Kim, and S. Boyd, 2007, Simple matlab solver for l1-regularized least squares problems. (Software, <http://www-stat.stanford.edu/~tibs/lasso.html>).
- Levy, S., D. Oldenburg, and J. Wang, 1988, Subsurface imaging using magnetotelluric data: *Geophysics*, **53**, 104–117.
- Lustig, M., D. L. Donoho, and J. M. Pauly, 2007, Sparse mri: The application of compressed sensing for rapid mr imaging: *Magn. Reson. Med.* (to appear).
- Malcolm, A. E., 2000, Unequally spaced fast Fourier transforms with applications to seismic and sediment core data.: Master’s thesis, University of British Columbia.

- Malcolm, A. E., M. V. de Hoop, and J. A. Rousseau, 2005, The applicability of DMO/AMO in the presence of caustics: *Geophysics*, **70**, S1–S17.
- Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: *Geophysics*, **46**, 1528–1542.
- Sacchi, M. D., T. J. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: *IEEE Trans. Signal Process.*, **46**, 31–38.
- Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: *Geophysics*, **59**, 938–945.
- Santosa, F. and W. W. Symes, 1986, Linear inversion of band-limited reflection seismogram: *SIAM J. of Sci. Comput.*, **7**.
- Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: *The Leading Edge*, **18**, 55–58.
- Starck, J.-L., M. Elad, and D. L. Donoho, 2004, Redundant multiscale transforms and their application to morphological component separation: *Advances in Imaging and Electron Physics*, **132**.
- Sun, Y., G. T. Schuster, and K. Sikorski, 1997, A quasi-Monte Carlo approach to 3-D migration: *Theory: Geophysics*, **62**, 918–928.
- Symes, W. W., 2006, Reverse time migration with optimal checkpointing: Technical Report 06-18, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.
- Taylor, H. L., S. Banks, and J. McCoy, 1979, Deconvolution with the ℓ_1 norm: *Geophysics*, **44**, 39.
- Thomson, D., G. Hennenfent, H. Modzelewski, and F. J. Herrmann, 2006, A parallel windowed fast discrete curvelet transform applied to seismic processing: Presented at the SEG International Exposition and 76th Annual Meeting.
- Tibshirani, R., 1996, Least Absolute Shrinkage and Selection Operator. (Software, <http://www-stat.stanford.edu/~tibs/lasso.html>).
- Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: *Geophysics*, **68**, 2043–2054.
- Trad, D. O., J. Deere, and S. Cheadle, 2006, Wide azimuth interpolation: Presented at the 2006 Annual Meeting of the Can. Soc. Expl. Geophys.
- Trad, D. O., T. Ulrych, and M. D. Sacchi, 2003, Latest views of the sparse radon transform: *Geophysics*, **68**, 386–399.
- Tropp, J., M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, 2006, Random filters for compressive sampling and reconstruction: Presented at the Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP).
- Ulrych, T. J. and C. Walker, 1982, Analytic minimum entropy deconvolution: *Geophysics*, **47**, 1295–1302.
- van den Berg, E. and M. Friedlander, 2007, In pursuit of a root: Technical report, UBC Computer Science. (TR-2007-16).
- Verschuur, D. J. and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, part II: practical aspects and examples: *Geophysics*, **62**, 1596–1611.
- Vogel, C., 2002, *Computational Methods for Inverse Problems*: SIAM.

- Wisecup, R., 1998, Unambiguous signal recovery above the nyquist using random-sample-interval imaging: *Geophysics*, **63**.
- Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: *Geophysics*, **70**, V87 – V95.
- Ying, L., L. Demanet, and E. J. Candès, 2005, 3-D discrete curvelet transform: , 591413, SPIE.
- Zwartjes, P. M. and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: *Geophysics*, **72**, V21–V32.

APPENDIX A

THE DISCRETE CURVELET TRANSFORM

The FDCT by wrapping (see e.g. Candès et al., 2006a; Ying et al., 2005) perfectly reconstructs data after decomposition by applying the transpose of the curvelet transform, i.e., we have $\mathbf{f} = \mathbf{C}^T \mathbf{C} \mathbf{f}$ for an arbitrary finite-energy vector \mathbf{f} . In this expression, $\mathbf{C} \in \mathbb{R}^{N \times M}$ represents the curvelet decomposition matrix. The curvelet coefficients are given by $\mathbf{x} = \mathbf{C} \mathbf{f}$ with $\mathbf{x} \in \mathbb{R}^N$. The curvelet transform is an overcomplete signal representation. The number of curvelets, i.e, the number of rows in \mathbf{C} . exceeds the number of data ($M \ll N$). The redundancy is moderate (approximately 8 in two dimensions and 24 in three dimensions). This redundancy implies that \mathbf{C} is not a basis but rather a tight frame for our choice of curvelet transform. This transform preserves energy, $\|\mathbf{f}\|^2 = \|\mathbf{C} \mathbf{f}\|^2$. Because $\mathbf{C} \mathbf{C}^T$ is a projection, not every curvelet vector is the forward transform of some function \mathbf{f} . Therefore, the vector \mathbf{x}_0 can not readily be calculated from $\mathbf{f} = \mathbf{C}^T \mathbf{x}_0$, because there exist infinitely many coefficient vectors whose inverse transform equals \mathbf{f} .

APPENDIX B

CURVELET PROPERTIES

Curvelets are directional frame elements that represents a tiling of the two-/three-dimensional frequency domain into multiscale and multi-angular wedges (see Fig. B-1 and B-2). Because the directional sampling increases every-other scale, curvelets become more and more anisotropic for finer and finer scales. They become 'needle-like' as illustrated in Fig. B-2. Curvelets are strictly localized in the Fourier domain and of rapid decay in the physical domain with oscillations in one direction and smoothness in the other direction(s). Their effective support in the physical domain is given by ellipsoids. These ellipsoids are parameterized by a width $\propto 2^{j/2}$, a length $\propto 2^j$ and an angle $\theta = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale, $j = 1 \cdots J$ and l the angular index with the number of angles doubling every other scale doubling (see Fig. B-1). Curvelets are indexed by the multi-index $\gamma := (j, l, \mathbf{k}) \in \mathcal{M}$ with \mathcal{M} the multi-index set running

over all scales, j , angles, l , and positions \mathbf{k} (see for more details Candès et al., 2006a; Ying et al., 2005). Therefore, conflicting angles are possible.

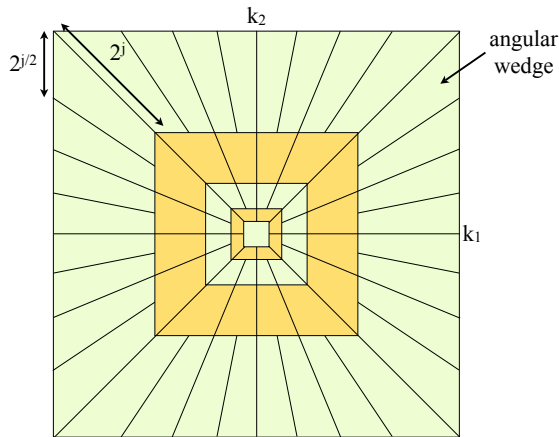


Figure B-1: Discrete curvelet partitioning of the 2-D Fourier plane into second dyadic coronae and sub-partitioning of the coronae into angular wedges.

APPENDIX C

COMPRESSION PROPERTIES OF CURVELET FRAMES

For 2-D functions that are twice-differentiable and that contain singularities along piece-wise twice differentiable curves, the Fourier transform (ignoring log-like factors in this discussion) only attains an asymptotic decay of the nonlinear k -term nonlinear approximation error of $\mathcal{O}(k^{-1/2})$. For this class of functions, this decay is far from the optimal decay rate $\mathcal{O}(k^{-2})$ (Candès and Donoho, 2000b). Wavelets improve upon Fourier, but their decay $\mathcal{O}(k^{-1})$ is suboptimal. Curvelets, on the other hand, attain the optimal rate $\mathcal{O}(k^{-2})$. In three dimensions, similar (unpublished) results hold and this is not surprising because curvelets can in that case explore continuity along two directions.

Continuous-limit arguments underly these theoretical estimates, somewhat limiting their practical relevance. Additional facts, such as the computational overhead, the redundancy and the nonlinear approximation performance on real data, need to be taken into consideration. The computational complexity of the curvelet transform is $\mathcal{O}(M \log M)$. The redundancy of the curvelet transform, however, maybe of concern. Strictly speaking wavelets yield the best SNR for the least *absolute* number of coefficients, suggesting wavelets as the appropriate choice. Experience in seismic data recovery, backed by the evaluation of the reconstruction and recovery performance in the 'eye-ball norm', suggest otherwise. Performance measures in terms of the decay rate as a function of the *relative* percentages of coefficients are more informative. For instance, when the reconstruction in Fig. 3 of a typical seismic shot

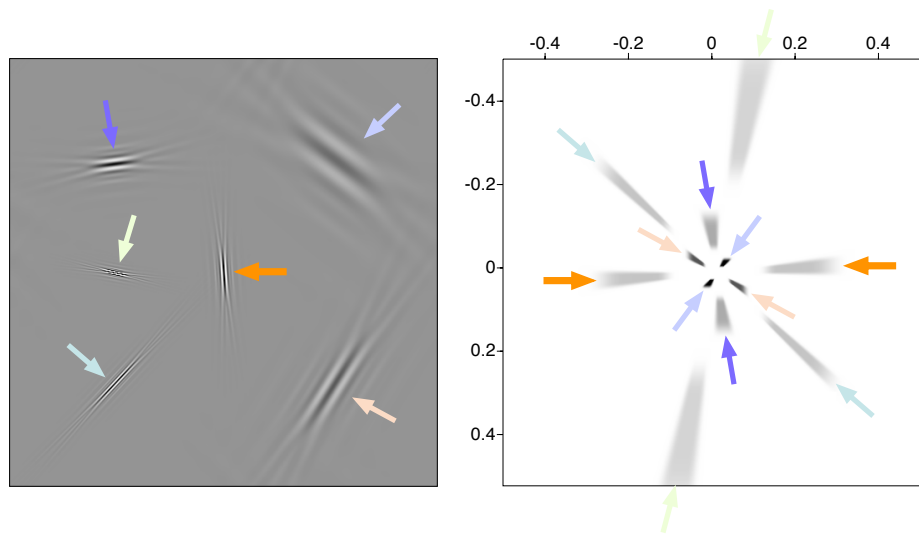
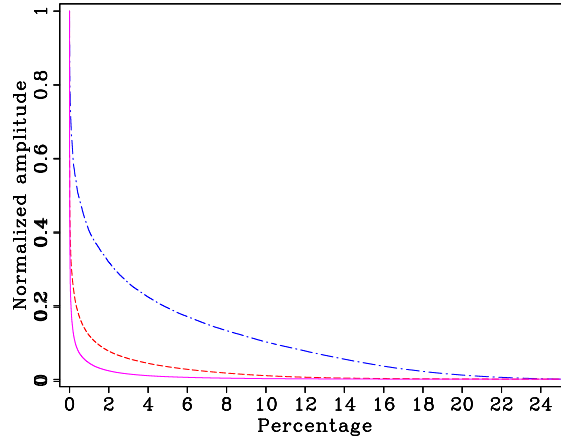
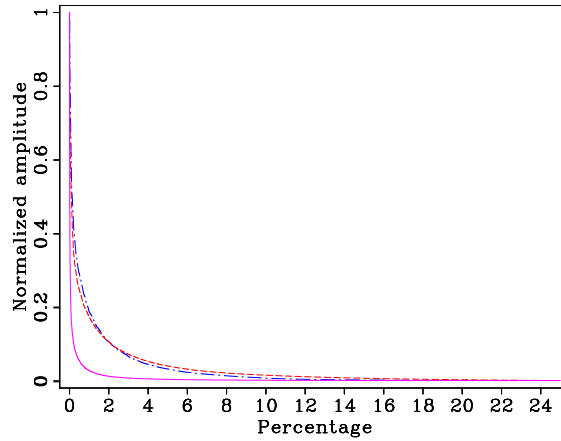


Figure B-2: Spatial and frequency representation of curvelets. **(a)** Six different curvelets in the spatial domain at five different scales. **(b)** Dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. Each pair of opposing wedges represents a real curvelet. The variable j is the curvelet scale. Each scale is represented at a number of angles that double at every other scale. This figure illustrates the correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientation of curvelets in the two domains. The 90° rotation is a property of the Fourier transform.

record from only 1% of the coefficients is considered, it is clear that curvelets give the best result. The corresponding reconstructions from Fourier and wavelets coefficients clearly suffer from major artifacts. These artifacts are related to the fact that seismic data does not lend itself to be effectively approximated by superpositions of monochromatic plane waves or 'fat' wavelet 'point scatterers'. This superior performance of the curvelet reconstruction in Fig. 3 is also supported by comparisons for the decay of the normalized amplitude-sorted Fourier, wavelet and curvelet coefficients, included in Fig. C-1. In three dimensions, we expect a similar perhaps even more favorable behavior by virtue of the higher dimensional smoothness along the wavefronts. These observations, suggest that curvelets are the appropriate choice for the sparsity representation so we set $\mathbf{S} := \mathbf{C}$.



(a)



(b)

Figure C-1: Decay of the transform coefficients for a typical synthetic (the fully sampled data set that corresponds to Fig. 2) and real data set (Fig. 3(a)). Comparison is made between the Fourier, wavelet and curvelet coefficients. **(a)** The normalized coefficients for a typical 2-D synthetic seismic shot record. **(b)** The same for a real shot record. Coefficients in the Fourier are plotted with the blue – dashed and dotted line, the wavelet coefficients with the red – dashed line, and the curvelet with the pink – solid line. The seismic energy is proportionally much better concentrated in the curvelet domain thus providing a sparser representation of seismic data than Fourier and wavelets.