

# Non-parametric seismic data recovery with curvelet frames

Felix J. Herrmann\* and Gilles Hennenfent\*

(January 2, 2007)

## 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 multidimensional expansions with respect to recently developed curvelet frames. The frame elements of these transforms locally resemble wavefronts present in the data and this leads to a compressible signal representation. This compression enables us to formulate a new seismic data recovery algorithm through sparsity-promoting inversion. The concept of sparsity-promoting inversion is in itself not new to the geosciences. However, the recent insights from the field of ‘compressed sensing’ are new since they identify the conditions that determine successful recovery. These conditions are carefully examined by means of examples geared towards the seismic recovery problem for data with large percentages ( $>70\%$ ) of traces missing. We show that as long as there is sufficient ‘randomness’ in the acquisition pattern, recovery to within an acceptable error is possible. We also show that our approach compares favor-

---

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

ably with another method and the examples underline the importance of using a transform that exploits the multidimensional geometry of the wavefronts present in seismic data. The presented work is an extension of this random subsampling concept towards a nonlinear recovery from randomly undersampled seismic wavefields. Our results are, in principle, extendible to measurements on irregular grids and to other areas such as global seismology.

## INTRODUCTION

Seismic data regularization is an important step in the seismic processing flow that aims at limiting imprints from incomplete acquisition on seismic images. These acquisition imprints are in many cases caused by missing data due to economical and physical constraints, e.g. cable feathering in marine surveys, dead or severely contaminated traces. This incompleteness of the acquired data volume may adversely affect subsequent steps in processing and imaging. For instance, convolution-based multiple prediction, as part of surface related multiple elimination (SRME, Verschuur and Berkhout, 1997), and most migration schemes assume uniform sampling of the data and a violation of this assumption may lead to serious degradation of the image quality. This paper deals with the seismic data recovery problem for missing data on otherwise regularly-sampled grids.

### Motivation

Recovery from incomplete data remains one of the most important tasks faced by the geophysical practitioner whether he is or she is working in the ‘data-rich’ field of exploration geophysics or in the traditionally ‘data-prone’ field of global seismology. We introduce a new *non-parametric* seismic data regularization method for missing traces (station and shot locations). The method is nonlinear, transform-based and does not require information on the locations and dips of the arriving wavefronts. The method is motivated by recent developments in the field of stable signal recovery from noisy and incomplete data (Candès et al., 2006b; Donoho et al., 2006; Elad et al., 2005; Starck et al., 2004; Donoho, 2006). The main idea behind these methods, also known as ‘compressed sensing’, is that compressible signals can under certain specific conditions stably be recovered from noisy and

incomplete measurements. Generally speaking recovery is possible as long as two conditions are met. Firstly, the to-be-recovered signal must admit a strictly sparse or compressible representation, which means that the signal can be seen as a parsimonious superposition of a relatively small subset of columns of a certain matrix. Secondly, during 'acquisition' the signal must be mixed sufficiently. This mixing depends on the coherency between the basis in which acquisition takes place and the representation in which the to-be-recovered signal is parsimonious. This mixing also depends on the 'randomness' within the acquisition pattern. These factors determine the mixing that translates, as Fig. 1 illustrates for the case of acquisition in the Dirac/spike basis and a 'sparse' Fourier domain signal representation, into a reduction of aliasing. Instead of generating the familiar periodic distortion, due to regular undersampling, undersampling on a random grid gives rise to a noisy spectrum. The better the mixing the more Gaussian the noise and the less harmful the imprint of the undersampling. The observation that random sampling reduces aliasing has been made by many others (see e.g. Masry, 1978; Wisecup, 1998; Malcolm, 2000). What is new is that this observation motivates us to follow recent developments from 'compressed sensing' by combining this mixing property with signal representations that compress. Our recovery is based on a nonlinear optimization procedure during which the compression of seismic data volumes in the curvelet domain is exploited. The success of this recovery depends on above described factors that control the mixing and on the ability of the curvelet transform to compress seismic data (Candès et al., 2006a).

### **An illustrative example**

To illustrate the basic ideas behind the nonlinear recovery, let us consider the following example after Candès et al. (2006b). In this example a spike train, i.e., a signal with a

limited number of non-zero entries, is recovered from incomplete measurements. Define a measurement as an inner product with the row vectors of some matrix that represents the measurement basis. For this example consider measuring in the spike/Dirac basis that consists of integer shifts of the discrete Dirac. Since a spike train with a small number of entries is sparse this choice seems the obvious thing to do. However, there is a caveat because it takes  $N - 1$  inner products to single out a single non-zero entry in the length- $N$  spike train and hence this choice is far from optimal. Alternatively, one could argue that measuring in the Fourier basis could be beneficial but again that also seems futile since the Dirac contains all discrete frequencies up to the Nyquist frequency and the *linear* recovery depends on the cancellation of the Fourier modes at all positions except at the location of the spike. Any departure from complete sampling in this linear context leads to an imperfect reconstruction and that explains the challenges we are faced with when sampling multidimensional seismic wavefields.

The main result from 'compressed sensing' states that a length  $N$  vector with  $k \ll N$  arbitrary non-zero entries can exactly be recovered from  $n \ll N$  measurements roughly proportional to the number of non zeros, i.e.,  $n \sim k$  ( $\sim$  means proportional to within  $\log N$ ). For the above example, this result means that the spike train can be recovered *exactly* from  $n$  random Fourier measurements. The measurement in this case corresponds to taking inner products with the rows of the Fourier matrix for a random subset of frequencies. With these choices for the Fourier *measurement* and Dirac *sparsity* basis, the unknown sparsity vector can, following the techniques from stable signal recovery, exactly be recovered by solving a norm-one *nonlinear* sparsity promoting optimization problem. The solution of this optimization problem seeks the sparsest vector whose Fourier transform restricted to the random set of frequencies equals the Fourier measurements at this subset. Mathematically,

the recovery of the spike train corresponds to inverting a rectangular matrix, given by the row-restricted  $n \times N$  Fourier matrix.

One of the goals is to provide insights in the conditions under which heavily under-sampled signals can successfully be recovered nonlinearly. In particular, we will stress the role played by the incoherence between the measurement and sparsity representations, in conjunction with the randomness of the measurement process. Remember that this randomness, as illustrated in Fig. 1, was responsible for the drastic difference between the Fourier spectrum of regularly undersampled data *versus* the spectrum of data with random traces missing.

## Related work in the seismic community

**The forward model:** The recovery of data undersampled in the source and/or receiver coordinates (cf. Fig. 1) is known as the seismic regularization problem. Solutions of this problem are typically derived from a linear forward model

$$\mathbf{f} = \mathbf{K}\mathbf{x}_0 + \mathbf{n}, \tag{1}$$

where the fully-sampled data,  $\mathbf{f} \in \mathbb{R}^M$ , is considered as a linear superposition of the columns of  $\mathbf{K} \in \mathbb{R}^{M \times N}$  with  $\mathbf{K}$  known as the modeling matrix. Possible measurement errors are accounted for by including a Gaussian noise term,  $\mathbf{n} \in \mathbb{R}^M$ . The modeling matrix with the prototype waveforms as its columns may be rectangular ( $N > M$ ). Throughout this paper bold-faced lower-case symbols are reserved for discrete vector quantities and upper-case bold symbols represent finite-dimensional matrices. The unknown model vector is represented by  $\mathbf{x}_0 \in \mathbb{R}^N$ . Seismic data is represented in terms of a vector that contains the seismic data volumes lexicographically sorted. The length of the  $M$ -vector corresponds to the total

number of samples in the seismic data volume. The recovery consists of estimating this model vector from incomplete, noisy and undersampled data, i.e., from

$$\mathbf{y} = \mathbf{P}(\mathbf{K}\mathbf{x}_0 + \mathbf{n}) \quad (2)$$

or

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}. \quad (3)$$

In these expressions,  $\mathbf{P} \in \mathbb{R}^{M \times M}$  represents the diagonal picking matrix and  $\mathbf{y}$  the measured data and  $\mathbf{n}$  the noise. The picking matrix has  $n$  ones at entries where there is data and  $M - n$  zeros where there is not. To simplify notation, we took the liberty to overload the symbols  $\mathbf{y}$  and  $\mathbf{n}$ . The vector quantities either refer to  $M$ -vectors with  $n$  nonzeros or to  $n$ -vectors with only the nonzero entries. The corresponding matrix quantities are defined accordingly, i.e.,  $\mathbf{A} \in \mathbb{R}^{n \times N}$ . Since the rows that correspond to zeros in the picking operator do not participate in the inversion, the recovery problem concerns the inversion of an underdetermined system of equations with  $N \gg n$  unknowns versus  $n$  known measurements.

**The inverse problem:** To fill in the null space associated with the inversion of the rectangular modeling matrix, solutions to data regularization problems are typically cast into an unconstrained optimization problem

$$\begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda J(\mathbf{x}) \\ \tilde{\mathbf{f}} = \mathbf{K}\tilde{\mathbf{x}}. \end{cases} \quad (4)$$

In this formulation, the functional  $J(\mathbf{x})$  represents a penalty term that contains *prior* information on the model. The importance of this *prior* information with respect to the quadratic data misfit is determined by the parameter  $\lambda$ , known as the Lagrange multiplier (see e.g. Vogel, 2002). The symbol  $\tilde{\phantom{x}}$  is reserved for estimated quantities obtained by solving

an optimization problem that jointly minimizes the quadratic misfit between observed and modeled data and a penalty term. The estimate for the regularized data is obtained by applying the modeling operator to the estimate for the model, i.e.,  $\tilde{\mathbf{f}} = \mathbf{K}\tilde{\mathbf{x}}$ .

The success of the above recovery depends on the picking matrix and choices for the modeling matrix, the penalty functional and the Lagrange multiplier. In seismic regularization, the measurement basis is fixed – shot-receiver positions are missing on a grid that is otherwise assumed to be regular while geo-/hydrophone can be considered to take spatial ‘point’ measurements. Successful seismic data recovery depends on the appropriate choices for the modeling matrix,  $\mathbf{K}$ , the picking matrix and the penalty functional  $J(\mathbf{x})$ . The different regularization methods can roughly be divided into *data-dependent* approaches, assuming *prior* (velocity) information on the wave arrivals, and *non-parametric* approaches that do not make such assumptions. Examples of parametric methods are the so-called data mappings (Bleistein et al., 2001), based on approximate solutions of the wave equation. These methods require information on the seismic velocity. Parabolic, apex-shifted Radon or migration-like transforms such as 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) also fall in this category. Other examples of data-adaptive methods are predictive, dip filtering techniques and plane-wave destructors (see e.g. Spitz, 1999; Fomel et al., 2002) that require a preprocessing step. Examples of non-parametric approaches range from minimal size and minimal structure norms for the penalty term (Claerbout and Muir, 1973; Rudin et al., 1992) to transform-based sparse inversion methods based on the Fourier or other transforms (Sacchi and Ulrych, 1996; Duijndan and Schoneville, 1999; Zwartjes and Gisolf, 2006; Abma and Kabir, 2006).

With its distinct wavefronts, seismic data does not lend itself to be easily captured. To



avoid intricate preprocessing steps to locate the wavefronts, a transform-based technique that borrows from the ideas of 'compressed sensing' will be followed. Through exploitation of mixing, during acquisition, and the redundancy and compression of the curvelet transform, our approach differs from existing norm-one recovery techniques. To understand why mixing and compression are important, we first briefly discuss early findings from the signal/image processing literature, such as inpainting (Starck et al., 2004; Elad et al., 2005), that formed the basis for the field now known as 'compressed sensing' (Candès et al., 2006b; Donoho et al., 2006; Donoho, 2006; Tropp, 2006).

## Related work in the signal/image processing community

**Method of frames:** Redundant signal representations, known as frames, occur frequently in signal processing problems such as analog-to-digital conversion. According to frame theory, the coefficient vector  $\mathbf{x}_0$  can be recovered from samples  $\mathbf{y} = \mathbf{A}\mathbf{x}_0$  with  $\mathbf{A} \in \mathbb{R}^{M \times N}$  for  $N \gg M$  by the so-called method of frames (MOF),

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \quad (5)$$

This method corresponds to finding the set of coefficients with the smallest energy that satisfies the data. This  $\ell_2$  optimization problem permits an explicit solution that reads

$$\tilde{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{y} := \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{y} \quad (6)$$

with  $\mathbf{A}^\dagger$  the *dual* frame (see e.g. Daubechies, 1992; Mallat, 1997), given by the pseudo-inverse of  $\mathbf{A}$ . The symbol  $^T$  is reserved for the matrix transpose. The MOF forms the cornerstone of many results obtained in sampling theory (Duffin and Schaeffer, 1952) and seismic data regularization (Liu and Sacchi, 2004). Despite its success, the MOF is subop-

timal when the signal admits a sparse representation, i.e.,  $\mathbf{x}_0$  is strictly sparse or has a fast decay for the amplitude-sorted coefficients.

**Basis pursuit and recent extensions:** For sparse vectors  $\mathbf{x}_0$ , the recovery can be greatly improved by replacing the quadratic penalty term by a sparsity promoting  $\ell_1$ - norm. As such the idea of using the  $\ell_1$  norm is not exactly new to the geosciences. For instance, since the seminal work of Claerbout and Muir (1973), there exists a large body of applied and theoretical work (Santosa and Symes, 1986) on sparse inversion. Applications range from deconvolution (Oldenburg et al., 1981; Ulrych and Walker, 1982; Levy et al., 1988; Sacchi et al., 1994), seismic data regularization with Fourier and Radon transforms (Sacchi and Ulrych, 1996; Duijndan and Schoneville, 1999; Trad et al., 2003), adaptive subtraction for multiple removal (Guitton, 2004; Herrmann et al., 2006) to compressed extrapolation (Herrmann and Lin, 2006) and Bayesian approaches with long-tailed Cauchy distributions.

With the exception of the work by Santosa and Symes (1986), relatively little is known on the theoretical conditions that lead to a successful sparse inversion. The theory of ‘compressed sensing’ from which this paper derives can be seen as extensions of earlier work on Basis Pursuit (BP) and Basis Pursuit Denoising (BPDN, see e.g. Chen et al., 2001; Starck et al., 2004; Elad et al., 2005) that allow for

- a proof of *equivalence* between the zero-norm optimization problem

$$\mathbf{P}_0 : \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \mathbf{Ax} = \mathbf{y}, \quad (7)$$

seeking a solution with the least number of non-zero entries, and the norm-one optimization problem

$$\mathbf{P}_1 : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{Ax} = \mathbf{y}. \quad (8)$$

While  $\mathbf{P}_0$  is combinatorial and NP-hard, i.e., it has a computational complexity that grows exponentially with the length of the sparsity vector,  $\mathbf{P}_1$  is convex and feasible for (very) large sparsity vectors.

- *unique* recovery, when the number of non-zero entries in  $\mathbf{x}_0$  is sufficiently small compared to the number of observations in  $\mathbf{y}$ ;
- *stable* recovery under additive noise, i.e.,  $\mathbf{x}_0$  can be recovered to within the noise level (the recovery error exceeds the noise level by a moderate constant) via

$$\mathbf{BPDN} : \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (9)$$

with  $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$  the noisy measurements.

## Our approach

Both BP and BPDN form the main motivation for this paper. Following the inpainting approach by Elad et al. (2005) based on redundant transforms, we formulate seismic data regularization as a curvelet-based norm-one recovery problem. We call our approach curvelet recovery by sparsity-promoting inversion (CRSI), which is based on the discrete curvelet transform (Candès and Donoho, 2000a; Candès et al., 2006a; Ying et al., 2005) that is known to compress seismic data, facilitating a stable seismic data recovery (Herrmann and Hennenfent, 2005; Herrmann, 2005; Hennenfent and Herrmann, 2006a,b; Thomson et al., 2006). The recovery takes the form of a constrained nonlinear optimization problem:

$$\mathbf{P}_\epsilon : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 & \text{s.t. } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{C}^T \tilde{\mathbf{x}} \end{cases} \quad (10)$$

with  $\mathbf{A}$  the restricted curvelet synthesis matrix,  $\mathbf{C}^T$  the curvelet synthesis matrix and  $\epsilon$  a control parameter proportional to the noise level.

Our contribution mainly lies in the adaptation of the above recovery problem to the very large-scale seismic sampling problem. We also aim at partially answering what we consider as two important questions that govern the seismic sampling problem, namely

*How many measurements does one need to recover the seismic wavefield to within a prescribed accuracy?*

*To what accuracy can one reconstruct the seismic wavefield given a certain acquisition grid and noise level?*

## **Outline**

First, a series of stylized examples is presented to identify the major factors that contribute to successful recovery. The series starts with the recovery of strictly sparse functions and then works its way up to the recovery of compressible signals. Our experimental findings are used as guidelines for our formulation of the large-scale seismic data recovery problem. This formulation includes a multidimensional transform that exploits continuity along wavefronts and the introduction of a large-scale solver for the recovery. We conclude by discussing the recovery of synthetic and real seismic data volumes, which includes a comparison with another method.

## **A PRIMER: SPARSE RECOVERY OF SINUSOIDALS**

A series of stylized experiments on sinusoidal functions is conducted. These functions are computed applying the transpose of the discrete Fourier transform to realizations of the sparsity vector. Thus, the discrete Fourier transform defines the sparsity basis. The experiments are designed to gain insight in the recovery conditions as a function of the

parsimoniousness, i.e., strict sparsity *versus* compression. Our measurement basis is the Dirac basis, which is the same as the basis in which seismic wavefields are measured. The samples are taken at a subset of random discrete (at integer values) entries from the signal vector. Without loss of generality, we replace the fast Fourier transform by the discrete cosine transform. These choices for the measurement and sparsity bases are equivalent to the choices made as part of seismic data regularization based on sparse Fourier inversion (Sacchi and Ulrych, 1996; Duijndan and Schoneville, 1999; Zwartjes and Gisolf, 2006). Both the strong and the more applicable weak recovery conditions will be discussed for strictly sparse, compressible and noisy samplings.

### Exact recovery of strictly sparse signals

Without loss of generality, consider the recovery of fully sampled sinusoidal functions,  $\mathbf{f} \in \mathbb{R}^N$ , that consist of a superposition of a limited number of sinusoids, i.e.  $\mathbf{f} = \mathbf{F}^T \mathbf{x}_0$  with  $\mathbf{F}^T$  the inverse Fourier transform matrix and  $\mathbf{x}_0 \in \mathbb{R}^N$  the sparsity vector with  $k \ll N$  non-zero entries. We are interested in recovering the function  $\mathbf{f}$ , which is sparse in the Fourier domain, from a small random subset of samples in the physical domain  $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ . We call this vector  $\mathbf{y} \in \mathbb{R}^n$ , with the subset of random samples, the measurement vector. Since  $n \ll N$ , the measurements are incomplete and the recovery of  $\mathbf{f}$  entails the inversion of a rectangular matrix. Recent results from 'compressed sensing' proof that the solution of following optimization problem

$$\mathbf{P}_1 : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 & \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{y} \\ \tilde{\mathbf{f}} = \mathbf{F}^T \tilde{\mathbf{x}} \end{cases} \quad (11)$$

*exactly* recovers the original function (Candès et al., 2006c). This proof depends on certain conditions for the measurement  $\mathbf{y}$ , the synthesis matrix  $\mathbf{A} \in \mathbb{R}^{n \times N}$  and the sparsity  $\mathbf{x}_0$ .

To better understand this nonlinear recovery, let us first identify the different matrices that make up the synthesis matrix,  $\mathbf{A}$ . This matrix decomposes into  $\mathbf{A} := \mathbf{RMS}^T$ , where  $\mathbf{S} := \mathbf{F}$  is the sparsity basis, defined by the Fourier transform;  $\mathbf{M} := \mathbf{I}$  the Dirac measurement basis with  $\mathbf{I}$  the identity matrix and  $\mathbf{R}$  the restriction matrix. This restriction matrix extracts  $k$  rows from the  $N \times N$  Fourier matrix and applies a normalization such that the columns of  $\mathbf{A}$  have unit norm. The symbol  $:=$  denotes 'defined as'.

Without going into many technical details, the conditions for a successful recovery of individual realizations for the sparsity and measurement vector with  $\mathbf{P}_1$  are typically sharp. For instance, the recovery of the sinusoidal function of length  $N = 100$ , plotted in Fig. 2(a), with a  $k = 1$  non-zero entry in  $\mathbf{x}_0$  (see Fig. 2(b)) is successful for a measurement vector  $\mathbf{y}$  consisting of  $n = 5$  elements and fails for  $n = 4$  measurements (see Fig. 2(d)). This sort of sharp transition has also been observed by Santosa and Symes (1986) for spiky deconvolution and suggests the existence of certain bounds, predicting the success of recovery with  $\mathbf{P}_1$  for arbitrary strictly sparse vectors  $\mathbf{x}_0$ .

**Strong recovery conditions:** One of the important results for the exact recovery from incomplete samplings, states that exact recovery is possible as long as the synthesis matrix  $\mathbf{A}$  obeys what is called a *restricted isometry*. An isometry is a function which preserves distances. For example, rotation or translation are isometries in a plane. In that case, every set of columns of size  $\leq k$  (with  $k$  the number of non-zero entries in  $\mathbf{x}_0$ ) approximately behaves as an orthonormal matrix (Candès et al., 2006b). As to be expected, the conditions for the matrix  $\mathbf{A}$  to obey a restricted isometry depend on the choices for the measurement

and sparsity matrices and on the size of the sparsity vector, the number of observations  $n$  and the number of nonzeros  $k$  in the sparsity vector. An important result by Candès et al. (2006b) states that exact recovery of the  $k$  non-zero entries in  $\mathbf{x}_0$  is possible, as long as the number of measurements is roughly proportional to the number of non-zero entries in  $\mathbf{x}_0$ , i.e.,

$$n \propto \mu^2 \cdot k. \quad (12)$$

This condition is strong and holds for arbitrary restrictions,  $\mathbf{x}_0$ , and  $\mathbf{y}$ . The parameter  $\mu \geq 1$  in this expression denotes the mutual coherence between the measurement and sparsity matrices and is defined by

$$\mu(\mathbf{M}, \mathbf{S}) = \sqrt{N} \max_{(i,j) \in [1 \dots N] \times [1 \dots N]} |\langle \mathbf{m}_i, \mathbf{s}_j \rangle| \quad (13)$$

with  $\mathbf{m}_i$  and  $\mathbf{s}_j$  the rows of  $\mathbf{M}$  and  $\mathbf{S}$ , respectively. The mutual coherence between the Dirac and Fourier bases is minimal ( $\mu = 1$ ) because their basis vectors are either maximally concentrated in the space or in the frequency domain. According to Eq. (12), the smaller the mutual coherence the fewer observations are required for successful recovery (Candès et al., 2006b). Intuitively, this phenomenon can be explained from the observation that the Fourier transform of a vector in the Dirac basis corresponds to a sinusoidal function that will always intersect with a (single) spike train in the Fourier domain. Because all sinusoidals intersect with Diracs at arbitrary frequencies, a limited number of inner products contains sufficient information to recover the non-zero entries and hence the signal.

**Weak recovery conditions:** Of course, the above results that hold for arbitrary vectors and restrictions are powerful and mathematically beautiful. Strong recovery conditions, however, tend to be pessimistic requiring many measurements or sparsity vectors with very few non zeros. In addition, the strong conditions, so far, have been derived from matrices

defined by orthonormal bases and not by redundant frame representations such as the curvelet transform. To address some of these issues, we settle for *weak* recovery conditions instead. These weak recovery conditions were recently proposed by Donoho et al. (2006) and provide conditions for which recovery by  $\mathbf{P}_1$  is very likely for *typical* realizations of the restriction and the sparsity vector. Typical realizations in this context refer to measurements that are close to uniformly random distributed and sparsity vectors that do not display too much structure, e.g., an atypical 'chess board' pattern of  $\pm 1$ 's or a regular subsampling.

From a practical viewpoint, these conditions are more easily met and it comes not really to a surprise that the weak recovery conditions derive from the observation that random subsampling and incoherence promote mixing that reduces aliasing. The same observation was made by Donoho et al. (2006), who coined the slogan "noiseless underdetermined problems behave like noisy well-determined problems". Indeed, when the difference between the matched filter and the original sparsity vector are considered,

$$\mathbf{z} = \mathbf{A}^T \mathbf{A} \mathbf{x}_0 - \mathbf{x}_0, \tag{14}$$

one finds a behavior that is close to Gaussian as illustrated in Fig. 3. In that figure, the difference for the Fourier spectrum of the shot record with the randomly missing traces is arguably close to Gaussian, while the uniformly missing data is clearly aliased even though both spectra are derived from the same number of traces.

By virtue of the mixing, recovery from incomplete measurements turns into some sort of a 'denoising' problem for which fast numerical solvers can be derived that approximate the expensive solution of  $\ell_1$  problems by linear programming. In the next section, we will introduce phase diagrams to study the weak recovery conditions in a probabilistic framework by conducting suits of experiments under different settings.



**Phase diagrams:** Interchanging the pessimistic and often difficult to calculate strong recovery conditions, for the more realistic weaker conditions, leads to the additional complication of entering into a probabilistic framework. We can no longer study the recovery of individual realizations for the restrictions and sparsity vectors. Instead, the statistical behavior of many stylized recovery experiments need to be examined jointly, under varying numbers of measurements,  $n$ , and numbers of non-zero entries,  $k$ , in the sparsity vector. For this purpose, the computational efficient method of stage-wise orthonormal matching pursuit (StOMP, Donoho et al., 2006) is used. The drastic computational speedup of StOMP goes, however, at the expense of a slight drop in the recovery performance since StOMP solves  $\mathbf{P}_1$  only approximately.

The results for these experiments are summarized in a diagram that records the number of successful recoveries. These so-called phase diagrams measure the performance of sparse recovery as a function of two ratios, namely the ratio of the number of measurements over the length of the sparsity vector,  $\delta = n/N$ , and the ratio of the number of non-zero entries in the sparsity vector over the number of measurements,  $\rho = k/n$ . The first ratio expresses the aspect ratio of the synthesis matrix  $\mathbf{A}$ . The second ratio expresses the 'skewness' between the number of measurements and the number of unknown non-zero entries in the sparsity vector. For a large enough system size,  $N$ , and number of experiments for each pair  $(\delta, \rho)$ , these diagrams contain reproducible transitions from  $(\delta, \rho)$  combinations that lead to successful or failing recovery. This sort of abrupt transitional behavior is a well-known phenomenon in statistical physics that describes the behavior of large systems with a random component, the random restriction in our case.

The phase diagrams themselves are calculated by counting the number of successful recoveries with StOMP for each parameter pair  $(\delta, \rho) \in (0, 1] \times (0, 1]$  and for a fixed number

of different realizations of the synthesis matrix and sparsity vector. Success or failure of the recovery is measured by counting the number of entries in the recovered vector that differ by more than a small tolerance parameter (typically  $10^{-4}$ ). In this way, issues related to machine imprecision are circumvented. The length of the signals is  $N = 800$  and 25 experiments were conducted for each of the  $25 \times 25$  parameter pairs  $(\delta, \rho)$ . The resulting diagram is plotted in Fig. 4. The bright areas correspond to parameter combinations for which recovery is likely to be successful. Recovery is likely to fail in the darker regions. For large  $\rho$ , recovery is unlikely by lack of relative sparsity compared to the number of measurements. For each  $n$ , there is a different critical number of non-zeros in  $\mathbf{x}_0$  for which recovery becomes possible. The larger the number of measurements, the larger number of non zeros that are recoverable. On the far left, the sparsity vector contains a single spike and the recovery starts for  $\rho \approx 0.2$ , which corresponds to approximately 5 measurements. As one moves to the right, the sparsity vector becomes denser and more measurements are required for the recovery. On the far right, the vector is full and as expected recovery at that point is still possible because the orthonormal discrete cosine transform is used.

For synthesis matrices made out of orthonormal matrices, a theory has been developed by Donoho et al. (2006) that predicts the location of the transition. This theory also predicts that as the length  $N$  of the sparsity vector increases, the transition from recoverable to non recoverable becomes sharper, a behavior also observed for systems in statistical physics. Unfortunately, as the system size increases, the computational complexity of computing the complete phase diagram becomes prohibitively expensive. This observation is important because seismic data recovery is large-scale and does not fit in the idealized setting of orthonormal sparsity representations for which theoretical results exist predicting the location of the transition.

However, when available, phase diagrams are extremely useful because they allow one to derive specific conditions under which one can expect the recovery to be successful. As long as the pair  $(\delta, \rho)$  is sufficiently far away from the phase transition in the bright colored region, the percentage of successful recoveries will be very high for typical realizations of the restrictions and sparsity vectors. Of course, several issues remain, namely how does the recovery behave (i) with respect to compressible rather than strictly sparse signals, (ii) under the presence of noise and (iii) under less than optimal (seismic) acquisition geometries? In the next section, we will shed some light on these important issues.

### **Stable recovery of (compressible) sinusoidals**

In real-life applications of signal recovery solution methods, one typically has to contend with noise-contaminated measurements and signals for which strictly sparse representation remains elusive. Real data simply do not permit strictly sparse representations, while robustness of geophysical methods to noise is another prerequisite. By means of a series of stylized examples, we will show that the sparse recovery method is stable under noise and extendible to compressible signals. For more theoretical background on these extensions refer to Candès et al. (2006b); Donoho et al. (2006).

**Recovery of strictly sparse signals from noisy data:** Suppose the observed data is given by  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$  with  $\mathbf{n} \in \mathbb{R}^n$  zero-centered with standard deviation  $\sigma$  white Gaussian noise. In that case, exact recovery according  $\mathbf{P}_1$  is no longer relevant because of the equality constraint. By replacing this constraint in  $\mathbf{P}_1$  by a constraint on the  $\ell_2$  difference between the synthesized and observed data, we arrive at the following formulation for the sparse

recovery problem

$$\mathbf{P}_\epsilon : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 & \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{F}^T \tilde{\mathbf{x}}. \end{cases} \quad (15)$$

There are numerous papers that study the behavior of this optimization problem (see e.g. Candès et al., 2006b). Exact recovery is no longer possible because of the noise. However, the solution of  $\mathbf{P}_\epsilon$  recovers a strictly sparse signal to within the noise level. The conditions on the synthesis matrix and sparsity vectors are similar to the noise-free case. The recovered signal has been shown to have an error of

$$\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2 \leq C \cdot \epsilon \quad (16)$$

with  $C$  a moderately sized constant (Candès et al., 2006b). This result is accomplished by setting the tolerance such that one remains to within the mean of the  $\ell_2$ -norm of the noise plus or minus two standard deviations. Since  $n_{1\dots n} \in \mathcal{N}(0, \sigma^2)$ , the probability of  $\|\mathbf{n}\|_2^2$  exceeding its mean by plus or minus two standard deviations is small. The  $\|\mathbf{n}\|_2^2$  is distributed according to the  $\chi^2$ -distribution with mean  $n \cdot \sigma^2$  and variance  $\sqrt{2n} \cdot \sigma^2$ . By choosing  $\epsilon^2 = \sigma^2(n + \nu\sqrt{2n})$  with  $\nu = 2$ , we remain well within the mean plus or minus two standard deviations (Candès et al., 2006b).

Fig. 5 contains an example where a strictly sparse signal is recovered from incomplete and noisy measurements. The recovered signal is not the same as the original but the recovery is clearly within the noise level. As in the noise-free case, phase diagrams can be calculated for noisy experiments. The phase transitions in this case are less sharp and they depict the normalized relative  $\ell_2$  error instead of the number of non-recovered entries. See Donoho et al. (2006) for more details.

**Stable recovery of compressible signals:** Besides measurements being noisy, natural occurring signals typically can not be considered as a sparse superposition of columns of the sparsity matrix. However, when sorted, the coefficients,  $\mathbf{x} = \mathbf{S}\mathbf{f}$ , in the transformed domain tend to decay. This behavior can be quantified by a power-law decay rate, i.e.,

$$|x_{i \in I}| \leq C_r \cdot i^{-r} \quad \text{with } r > 1/2, \quad (17)$$

with  $I$  the indices such that  $x_{I(1)} \geq x_{I(2)} \geq \dots \geq x_{I(N)}$  and  $r$  the compression rate with  $C_r$  a constant that depends on the signal's energy. The faster the decay the larger  $r$ . For orthonormal sparsity matrices, this decay is related to the decay rate for the nonlinear approximation error. This error is defined as the  $\ell_2$ -difference between the original signal vector  $\mathbf{f}$  and its nonlinear approximation from the  $k$ -largest coefficients, i.e.,  $\mathbf{f}^k := \mathbf{S}^T \mathbf{x}_{I(1 \dots k)}$  and reads

$$\|\mathbf{f} - \mathbf{f}^k\|_2 \leq C_2 \cdot k^{-r+1/2}. \quad (18)$$

The larger the decay rate, the more the signal's energy is concentrated in a small percentage of large coefficients, and the faster the approximation error decays as a function of  $k$ . Because compressible signals are no longer strictly sparse, their recovery from (noisy) data includes an additional source of error related to the  $k$ -term truncation of the sparsity vector. This truncation stems from the fact that only the  $k$  largest entries of the sparsity vector are recovered. The number  $k$  depends on the properties of  $\mathbf{A}$  and is proportional to the number of measurements  $n$ .

For synthesis matrices  $\mathbf{A}$  that obey a restricted isometry, Candès et al. (2006b) showed that the  $\ell_2$ -error for a recovery based on  $\mathbf{P}_\epsilon$  reads

$$\|\tilde{\mathbf{x}} - \mathbf{x}_0\|_2 \leq C_3 \cdot \epsilon + C_4 \cdot k^{-r+1/2} \quad (19)$$

for compressible signals with a nonlinear approximation rate according to Eq. 18. This error

contains contributions from the noise and from the recovery of only the  $k$ -largest entries of  $\mathbf{x}_0$ . According to the strong condition of Eq. 12, the number of resolvable entries in  $\mathbf{x}_0$  increases with the number of observations. Above result shows that the recovery is stable with a recovery error that depends on a combination of the noise level and the number of observations  $n$ . The larger  $n$ , the larger the number of recoverable entries  $k$  and the smaller the recovery error by virtue of Eq. 19.

To illustrate the recovery of compressible signals, we included Fig. 6. The original signal of Fig. 6(a) is no longer strictly sparse in the transformed domain but as Fig. 6(b) suggests, the signal is compressible. In fact, the entries of the sparsity vector  $\mathbf{x}_0$  were generated by applying random permutations and signs to a sequence of the form  $u_i = \exp(-i/2)$  for  $i = 1 \cdots N$ . The results for the recovery from 20 noise-free measurements is also plotted in Fig. 6. Despite the lack of strict sparsity, the signal is accurately recovered and this performance can be attributed to a successful recovery of the largest entries in the sparsity vector by  $\mathbf{P}_\epsilon$ .

**Recovery diagram:** As with the recovery of strictly sparse signals, weak recovery conditions exist for compressible signals. Again, useful insights can be gained by conducting suits of experiments. In this situation, the experiments are carried out over different numbers of measurements and increasing compression rates, i.e.,  $(\delta, r) \in (0, 1] \times (1/2, 2]$ . The squared relative  $\ell_2$  error given by  $\text{err}_2 = \|\tilde{\mathbf{x}} - \mathbf{x}_0\|^2 / \|\mathbf{x}_0\|^2$  encodes the greyscale. The number of experiments and length of the vector are kept the same as with the phase diagram experiments. The bright region in the recovery diagram (see Fig. 7) corresponds to parameter combinations that favor accurate recovery. As expected, the recovery error decays with  $\delta$  (or the number of measurements for  $N = 800$  fixed). As expected the error decays

rapidly as a function of the compression rate. This observation suggests that finding signal representations that compress is crucial for the recovery.

## Observations from the stylized examples

The recovery examples discussed so far show that strictly sparse and compressible signals can be recovered by solving nonlinear optimization problems:  $\mathbf{P}_1$  for exact recovery of strictly sparse signals and  $\mathbf{P}_\epsilon$  for the stable recovery of compressible signals, possibly in the presence of noise. Both nonlinear programs promote sparsity through norm-one minimization. The recovery diagram show that the recovery hinges on the compression rate, leading to substantial improvements in the recovery error.

We also discussed that strong recovery conditions, valid for all possible restrictions, can be replaced by weaker conditions that apply to typical restrictions and sparsity vectors. These latter conditions lead to a probabilistic framework, yielding a recovery with a certain probability. Phase diagrams proved to be a useful tool in measuring the recoverability of strictly sparse signals as a function of the aspect ratio of the synthesis matrix  $\mathbf{A}$  and the ratio of the number of measurements over the number of unknown entries to be recovered in the sparsity vector.

Extensions to noisy and compressible signals were also discussed showing that (i) the recovery is stable under noise and that (ii) compressible signals can approximately be recovered. This latter feature of the recovery is especially important for the recovery of naturally occurring data sets that typically do not permit a strictly sparse representation. The introduction of the recovery diagram, where the relative recovery error is plotted as a function of  $\delta$  and the nonlinear approximation rate, clearly shows the importance of signal

representations that accomplish high compression rates.

The recovery diagram, including its iso-recovery-error contours, contains another piece of important information. For instance, the measurement size, given a particular permissible recovery error and empirical decay rate, can be calculated from the intercept of the appropriate contour with a line of constant approximation rate. Similarly, given the number of the measurements and the empirical decay rate, one can find the recovery error from the grey value at the specified parameter combination for  $(\delta, r)$ .

The stylized examples presented up to this point are small scale and hence somewhat removed from large-scale seismic recovery problems. The examples did, however, show us the importance of mixing and compression rates. The first property hinges on the randomness in the acquisition and the mutual coherence between the measurement and sparsity representations. The former is well known amongst experts in seismic sampling, who know that random sampling reduces aliasing. The latter property, in conjunction with the observation on the compression rate, shed new light on the recovery problem in particular since there is no control over the measurement basis. In the next sections, we shift gears towards the recovery of seismic data in multiple dimensions using the curvelet transform.

## SELECTION OF THE SYNTHESIS MATRIX

The success of recovery from seismic data with missing traces hinges on the existence of a compressible signal representation for data with multidimensional wave fronts. The recovery also depends on the mixing, largely determined by the incoherence amongst the vectors spanning the measurement and sparsity matrices, and the randomness of the restriction. In



this section, we motivate our choices for the definition of the synthesis matrix  $\mathbf{A} := \mathbf{RMS}^T$ , given the properties of seismic data and constraints on seismic data acquisition.

## **The restriction and measurement matrices**

Assuming point sources and ignoring the directional and temporal frequency response of the geo/hydrophones, it is safe to define the measurement matrix by the identity/Dirac basis,  $\mathbf{M} := \mathbf{I}$ . This matrix represents the sampling along the spatial source/receiver and time coordinates. At a given source/receiver position, the time is regularly sampled. The source/receiver positions themselves are irregularly sampled, leading to seismic data volumes with missing traces (see e.g. Fig. 1). We also assume that the traces are recorded at a regularly-spaced grid along the surface, yielding a dataset with missing traces on an otherwise regular acquisition grid.

## **The sparsity matrix**

The most striking feature of seismic data is the presence of bandwidth-limited wavefronts. These wavefronts come in many shapes, vary in frequency content, direction and may contain conflicting dips and caustics. For these reasons, it has been a challenge to find alternative transforms with coefficients that decay faster than the Fourier transform for functions with curved wavefronts. In 1-D, wavelet coefficients rapidly decay away from singularities and that explains why the wavelet transform yields a better approximation rate for functions with singularities. Unfortunately, this behavior does not extend to higher dimensions, where the singularities may no longer be confined to points. Instead, singularities may lie on curved sheets or wavefronts. The decay of the wavelet coefficients will not be

optimal in this case because fast decays will only occur in the direction perpendicular to the wavefronts. These directions are known as the 'wavefront set' and since wavelets lack directionality they can not not resolve this wavefront set. This incapability explains why wavelets do not improve the compression rate for data with wavefronts as much as recently developed directional curvelet frames do (Candès and Donoho, 2000a, 2004; Candès and Donoho, 2005a,b).

**Curvelet frames:** The key point of this paper is to recover seismic data by exploiting the compression of seismic data volumes in a sparse non-parametric transformed domain. During this recovery minimal assumptions are made regarding the shape, direction and frequency content of the arriving wavefronts. For this purpose, a discretization is needed of a transform that is capable of detecting wavefronts (Candès and Donoho, 2005a,b). The recently introduced fast discrete curvelet transforms (FDCT, Candès et al., 2006a; Ying et al., 2005) offer such a discretization, expanding regularly gridded data with respect to a collection of localized discrete multiscale and multidirectional prototype waveforms that are anisotropically shaped. Without *prior* information, the location and direction of wavefronts are found through the 'principle of alignment' that leads to large inner products between curvelets (rows of the curvelet transform matrix) and wavefronts that locally have the same direction and frequency content. This principle of alignment is illustrated in Fig. 8. Only few curvelet coefficients will be large, the other coefficients will decay rapidly away from the wavefronts. This property translates in a rapid decay for the magnitude-sorted coefficients. By comparison, the wavelet transform decays more slowly for functions with curved singularities (Candès et al., 2006a; Hennenfent and Herrmann, 2006b).

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  $\mathbf{x} \in \mathbb{R}^M$ . 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  $\mathbf{f}$ .

**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. 9 and 10). 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. 10. 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. 9). 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.

**Compression properties of curvelet frames:** In two dimensions and ignoring log-like factors in this discussion, the Fourier transform only attains an asymptotic decay of  $\mathcal{O}(k^{-1/2})$  for functions that are twice-differentiable and that contain singularities along piece-wise twice differentiable curves. 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  $\mathcal{O}(M \log M)$  and is not an issue. 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. 11 of a typical seismic shot 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. 11 seems to be also supported by comparisons for the decay of the normalized amplitude-sorted Fourier, wavelet and curvelet coefficients, included in Fig. 12. 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 define  $\mathbf{S} := \mathbf{C}$ .

### The synthesis matrix

With all the definitions for the restriction, measurement and sparsity matrices in place, we are now ready to define the representation for the noisy seismic data matrix as

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \quad \text{with} \quad \mathbf{A} := \mathbf{R}\mathbf{I}\mathbf{C}^T. \quad (20)$$

Again, the sampled seismic wavefield is contained in the vector  $\mathbf{y} \in \mathbb{R}^n$ . These observations are related to the sparsity vector  $\mathbf{x}_0$  through the synthesis matrix  $\mathbf{A}$ . This synthesis matrix is defined in terms of the restriction,  $\mathbf{R} \in \mathbb{R}^{n \times M}$ , measurement,  $\mathbf{I} \in \mathbb{R}^{M \times M}$ , and sparsity,  $\mathbf{C}^T \in \mathbb{R}^{M \times N}$ , matrices, yielding the following sizes for the synthesis matrix,  $\mathbf{A} \in \mathbb{R}^{n \times N}$  and the sparsity vector  $\mathbf{x}_0 \in \mathbb{R}^N$ .

The seismic recovery problem is extremely large scale. There are two factors contributing to this large scale. Firstly, seismic data is collected along 3 to 5 dimensions, yielding Tera bytes of data. Secondly, the benefit from the angular decomposition in the curvelet domain starts for shot records with at least 256 traces and 256 time samples, i.e.,  $M = 2^{16}$  in 2-D and  $M = 2^{24}$  in 3-D. These sizes prohibit explicit implementation of the matrices and column normalization for the synthesis matrix. The matrices themselves are implemented

in a matrix-free form. The sampled wavefields typically have large percentages ( $> 70\%$ ) of missing traces, yielding aspect ratios for the synthesis matrix that range from  $\delta = n/N \approx 0.04$  in 2-D to  $\delta \approx 0.01$  in 3-D.

### Conditioning of the synthesis matrix

Inverting an underdetermined system of equations of the size mentioned in the previous sections clearly is a daunting task. Not only is the size almost prohibitively large, but the specific properties of seismic data acquisition, where complete traces are missing, also causes problems. These missing traces introduce a certain anisotropy in the problem and are in clear violation of a 'random' subsampling. The missing traces are harmful because they introduce a large mutual coherence with (near) vertically-oriented curvelets (see Fig. 13). To reduce the impact of the seismic sampling, an angular weighting in the curvelet domain is applied. This weighting, as illustrated in Fig. 13, reduces the increased mutual coherence and is implemented by defining the synthesis matrix as follows

$$\mathbf{A} := \mathbf{R}\mathbf{I}\mathbf{C}^T\mathbf{W} \quad \text{with} \quad \mathbf{W} = \text{diag}\{\mathbf{w}\}. \quad (21)$$

The weighting vectors contains zeros at positions that correspond to wedges that contain near to vertical curvelets and ones otherwise.

## SEISMIC DATA RECOVERY WITH CURVELET FRAMES

We are now at the point where the optimization problem  $\mathbf{P}_\epsilon$  needs to be solved for a very large system. To arrive at a practical solution, we first rewrite this constrained optimization problem into a series of simpler unconstrained optimization problems. Each of these subproblems is solved with an iterative soft thresholding method with the threshold playing

the role of a Lagrange multiplier. By carefully lowering this threshold, a solution is obtained where the missing traces are recovered. The method is tested on synthetic data, including a discussion on the potential uplift of 3-D curvelets over 2-D curvelets. We conclude this section, by comparing our recovery method to the method of plane-wave destruction for real data.

### Curvelet Recovery by Sparsity-promoting Inversion (CRSI)

**Definition of the unconstrained subproblems:** For the synthesis matrix  $\mathbf{A}$  defined in Eq.'s 20-21, the recovery of a compressible seismic data volume from incomplete measurements corresponds to inverting an underdetermined system that involves the solution of the sparsity promoting nonlinear program  $\mathbf{P}_\epsilon$ . Following Elad et al. (2005), we replace this constrained optimization problem by a series of simpler unconstrained optimization problems

$$\mathbf{P}_\lambda : \begin{cases} \tilde{\mathbf{x}}_\lambda = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ \tilde{\mathbf{f}}_\lambda = \mathbf{C}^T \tilde{\mathbf{x}}_\lambda. \end{cases} \quad (22)$$

These subproblems depend on the Lagrange multiplier  $\lambda$ . For the noise-free case, the solution of  $\mathbf{P}_\lambda$  for  $\lambda \rightarrow 0$  converges to  $\mathbf{P}_1$ , while the solution in the noisy case is reached by solving  $\mathbf{P}_\lambda$  for  $\lambda \rightarrow \lambda_\epsilon$  with  $\lambda_\epsilon = \sup_\lambda \{\lambda : \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}_\lambda\|_2 \leq \epsilon\}$ . During the nonlinear optimization, the rectangular matrix  $\mathbf{A}$  is inverted by imposing the sparsity promoting  $\ell_1$ -norm. This norm regularizes the inverse problem of finding the unknown coefficient vector (see also Daubechies et al., 2005).

**Solution of each subproblem by iterative thresholding:** Following Daubechies et al. (2005), Elad et al. (2005); Candés and Romberg (2004) and ideas dating back to Figueiredo

and Nowak (2003), subproblems  $\mathbf{P}_\lambda$  can be solved by an iterative thresholding technique that derives from the Landweber descent method (Vogel, 2002). For  $\lambda$  fixed, looping over

$$\mathbf{x} \leftarrow T_\lambda(\mathbf{x} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x})), \quad (23)$$

with

$$T_\lambda(x) := \text{sgn}(x) \cdot \max(0, |x| - |\lambda|) \quad (24)$$

the soft thresholding operator, is shown by Daubechies et al. (2005) to converge to the solution of  $\mathbf{P}_\lambda$  for a large enough number of iterations and as long as the largest singular value of  $\mathbf{A}$  is smaller than 1, i.e.  $\|\mathbf{A}\| < 1$ . The cost of each iteration is a synthesis and subsequent analysis.

**Solution by the cooling method:** Because of the large number of unknowns and the size of the matrices involved, it is impractical to solve each subproblem  $\mathbf{P}_\lambda$  to convergence. Instead a common strategy is to solve  $\mathbf{P}_\lambda$  approximately, by iterating Eq. (23)  $L$  times. If after these iterations the condition  $\|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}_\lambda\|_2 \leq \epsilon$  is not met, the Lagrange multiplier is lowered and the previous solution is used as a first guess for the new subproblem (Starck et al., 2004; Elad et al., 2005). Sparsity is imposed from the beginning by setting  $\lambda_1$  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). The details of the cooling method are presented in Table. 1.

## Seismic data recovery with CSRI



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**

$\mathbf{x}^{i+1} = T_{\lambda_l}^s(\mathbf{x}^i + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^i))$

**end for**

$i = i + 1;$

**end while**

$\tilde{\mathbf{f}} = \mathbf{C}^T \mathbf{x}^i.$

Table 1: The cooling method with iterative thresholding.

**2-D synthetic for a layered earth model:** The synthetic survey was modeled with a 50-foot (15.24-m) receiver interval, 4-ms sampling interval, and 25-Hz central-frequency Ricker wavelet as a source wavelet. Fig. 14(a) shows the resulting common-midpoint (CMP) gather. The dataset contains 256 traces of 500 time samples each. Fig. 14(b) shows more closely an area in the data where there are conflicting dips. The simulated acquired data were obtained by randomly removing about 60% of the traces from the complete data, which corresponds to an average spatial sampling of 125 feet (38.1 m).

Based on the maximum expected dip of the reflection events in the data, we used a minimum velocity constraint of 5000 ft/s (1524 m/s) and no negative dips to limit the number of unknowns. Figs. 14(e) and 14(f) show the results for the CMP reconstruction with the CRSI algorithm for 100 iterations (5 inner- and 20 outer-loops). Figs. 14(g) and 14(h) plot the difference between the recovered and 'ground-truth' complete data. The SNR for the recovery is about 29.8 dB, which corroborates the observation that there is almost no energy on the difference plots. Curvelet reconstruction clearly benefits from continuity along wavefronts in the data and has no issue with conflicting dips thanks to the multidirectional nature of curvelets.

**Sliced versus volumetric interpolation:** A synthetic seismic line is generated using a subsurface velocity model with two-dimensional inhomogeneities. This velocity model consists of a high-velocity layer, which 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 complete prestack dataset can be represented as a three-dimensional volume along the shot, receiver and time coordinates

(Fig. 17(a)). The simulated acquired data (Fig. 17(b)) were obtained by randomly removing 80 % of the receiver positions for each shot, which corresponds to an average spatial sampling of 75 m.

Since the dataset is three-dimensional, the question arises whether it would be beneficial to perform the recovery with the 3-D curvelet transform on the whole 3-D volume instead of recovering individual slices (Fig. 15). The advantage of 3-D curvelets is that they exploit the continuity in two directions along the wavefronts. According to the same principle of alignment, the large entries in the curvelet vector correspond to localized events in the data with the same frequency content and now two dips. Not only will the 3-D curvelet representation be more compressible, but these higher dimensional curvelets will also be more likely to overlap with areas where data is present.

Fig. 16(b) shows one shot from the simulated acquired data, Figs. 16(c) and 16(e) its reconstruction with 2-D and 3-D CRSI, respectively, for 250 iterations and with no minimum velocity constraint applied. Figs. 16(d) and 16(f) show the differences between the complete and the reconstructed data. The SNR's for the 2-D and 3-D results are 3.9 dB and 9.3 dB, respectively. From the plots for the recovered shot records, it is clear that 3-D CRSI benefits from 3-D information that greatly improves the reconstruction over the 2-D CRSI. This improvement is particularly visible for the shallow near zero-offset events. Figs. 17(c) and 17(d) show the 3-D curvelet reconstruction and the difference for the selected time, common-source, and common-receiver slices of the data volume. The overall SNR for this reconstruction is 16.92 dB.

**2-D real data:** Fig. 18(a) shows the first 1.7 s sampled at 4 ms of the first 200 receivers of a shot-record of a seismic survey from offshore Gippsland basin Australia. The group interval

is 12.5 m. Fig. 18(b) shows the same data, but randomly omitting 60 % of the original traces (corresponding average spatial sampling is 31.25 m). CRSI result and difference between reconstruction and original data are presented in Figs. 18(c) and 18(d), respectively. For the comparison (see Hennenfent and Herrmann, 2007, for more details), we also include the result of another state-of-the-art interpolation method using plane-wave destruction (PWD, Fomel et al., 2002) and the corresponding difference in Figs. 18(e) and 18(f), respectively. The SNR for 2-D CRSI is 18.8 dB compared to 5.5 dB for PWD, which clearly confirms that curvelet reconstruction captures better the seismic energy than PWD and thus provides a better interpolation result.

## CONCLUSIONS

In this paper, a new non-parametric seismic data regularization method was proposed that combines existing ideas of sparsity-promoting penalty terms with multiscale and multidirectional signal expansions that compress seismic data. The compression by curvelet frames, in conjunction with a sufficiently random acquisition, led to a scheme that we called curvelet recovery by sparsity-promoting inversion (CRSI). This scheme was able to recover data with up to 80 % of its traces missing.

### Initial findings

The success of non-parametric seismic data recovery for data with large percentages of its traces missing hinges on the interplay of four key factors, namely (i) the size of the measurement (percentage of present traces); (ii) the 'randomness' in the shot/receiver locations of the acquisition grid; (iii) the compression rate attained by the sparsity transform and (iv)

the coherence between the vectors in the measurement and sparsity matrices. Depending on each of these factors, recovery to within an acceptable error will either be feasible or infeasible.

**Stylized examples:** By means of a set of carefully chosen experiments, we were able to numerically confirm the importance of the above key factors. In this experimental part of the paper, we also discussed the merits of weak recovery conditions over the often too pessimistic strong recovery conditions. These weak conditions hold for typical recovery problems, excluding the more difficult situations such as the recovery from regularly sub-sampled data. A phase diagram was computed to assess the probabilistic nature of the weak recovery conditions for strictly sparse signals. This diagram confirmed that there is a sharp transition between regions of failed and successful recovery as a function of ratios expressing the strict sparsity over the number of measurements and the aspect ratio of the restricted synthesis matrix. We also demonstrated that the recovery is stable under noise and can be extended to compressible rather than strictly sparse signals. We introduced a recovery diagram that summarizes the results of experiments where both the aspect ratio of the synthesis matrix and the compression rate were varied. This graph showed a strong dependence of the recovery error on the compression rate. For recovery problems where the compression rate is known *a priori*, the recovery diagram proved to be particularly useful because it allows one, in principle, to either design an acquisition grid from which data can be recovered with a prescribed accuracy or given a certain acquisition it can determine to what accuracy the data can be recovered.

**Selection of the sparsity transform for seismic data:** The stylized examples and in particular the recovery diagram demonstrated the importance of high nonlinear approximation rates. Extending this observation to the 'real-life' (extremely) large-scale seismic recovery problem was also discussed. We argued that curvelet frames are a good choice for the sparsity transform. The reconstruction from the largest 1% of the curvelet coefficients was distinctively better than the results from equal percentages of Fourier and wavelet coefficients. This improved performance can be partially explained on theoretical grounds and from the percentage-wise concentration of energy in the coefficients. Linking these observations with the stylized examples is somewhat of a challenge because of the redundancy of the curvelet transform. This redundancy leads to an increase in the number of unknowns. The partial recovery examples, on the other hand, point to the tentative conclusion that the approximation rate as a function of relative numbers of coefficients is important. Since there is not yet a theory for nonlinear recovery with redundant transforms, some important questions remain to be answered.

**Application to the large-scale seismic setting:** Of all contributing factors, there is in seismic data acquisition no control over the basis in which seismic measurements are taken. There is some control over the 'randomness' of the acquisition and the selection of the sparsity matrix. We argued that a decomposition of seismic data volumes with respect to multiscale and multidirectional curvelets leads to a favorable compression and hence an improved recovery. The examples on synthetic and real data presented in this paper support this statement. Not only was the recovery successful for data with large percentages of traces missing, but the recovery also improved significantly with the three-dimensional curvelet transform. In 3-D, continuity along two directions is exploited and led to an improvement of

several dB's for the recovery. This observation underlines the importance of exploiting the multidimensional geometry of wavefronts in seismic data. Conflicting dips and less-than-ideal real data were also correctly handled. Because of the large-scale nature of the seismic recovery problem, it remains a challenge to put more definitive quantitative numbers on the performance of CRSI.

**Comparison with another method:** The advantage of a non-adaptive transform-based approach is that minimal *prior* information is necessary for the recovery. Our 2-D examples, for instance, only use a minimum-velocity constraint limiting the maximum dip. Similarly, Fourier-based techniques do not require much *prior* information. Of course, the success of these methods depend on the compression rate of the transform, which in turn depends on the properties of the signal. Non-stationary signals, such as seismic data, violate the stationarity assumption of the Fourier transform. Introducing carefully chosen window functions, may provide a remedy but limits the frequency and adds an extra parameter. Data-adaptive methods are an alternative, but they also depend on certain assumptions and on a preprocessing step. The method of plane-wave destruction is an example of a data-adaptive method. Comparing the performance of this method on real data shows that curvelet-based recovery is still successful even though some of the smoothness conditions along the wavefronts may be violated. A multiscale transform such as the curvelet transform is arguably robust and performs well on real data.

## Relation to existing approaches

The ideas presented in this paper do not stand by themselves. They constitute an interesting mix of existing approaches in seismic sampling and data processing with recent ideas

from the theoretical field of information theory. For instance, it was relatively well-known in the seismic community that random sampling reduces aliasing and it is interesting to see that these ideas are seemingly independently extended to the more general setting of 'compressed sensing'. Promotion of sparsity and incompleteness of measurements are central in both approaches. The approaches differ in emphasis. Seismic data recovery is driven by a practical force to make the recovery work for very large-scale problems while the 'compressed sensing' community has and still is developing a theory. This theory not only provides a deep insight, it also lists the conditions that make recovery possible. Another difference is that the seismologist is satisfied when the recovery result 'looks good', irrespective of whether the  $\ell_1$ -norm solver ran to convergence. This means that certain theoretical questions remain on the performance of a nonlinear sampling theory for large-scale seismic wavefields.

### **The large-scale challenge**

**The  $\ell_1$  solver:** The success of seismic data recovery depends for a large part on the solution of the nonlinear  $\ell_1$ -norm optimization problem. A relatively simple threshold-based cooling method was presented that performs well on very large systems. Sparked by the results of 'compressed sensing', there is a current surge of interest in developing large-scale  $\ell_1$ -norm solvers. Our algorithm can only benefit from these developments.

**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. The redundancy of the curvelet transform prohibits extension to higher dimensions. By



applying a domain decomposition in three dimensions, the first problem has successfully been addressed (Thomson et al., 2006). The second problem is still open and may require combination with other transforms.

## **Extensions of CRSI to unstructured data**

In this paper, we limited ourselves to data with missing data on an otherwise regular grid. This assumption limits the applicability of CRSI. Fourier-based sparsity promoting methods (see e.g. Zwartjes and Gisolf, 2006) are designed to work with data on irregular grids. With the non-uniform fast discrete curvelet transform developed by the authors (Hennenfent and Herrmann, 2006b), CRSI can be extended to irregular data. This extension would also bring CRSI in the traditionally 'data-prone' field of global seismology, where irregular sampling and spherical coordinate systems prevail.

## **ACKNOWLEDGMENTS**

The authors would like to thank the authors of CurveLab and the authors of SparseLab for making their codes available at [www.curvelet.org](http://www.curvelet.org) and [sparselab.stanford.edu](http://sparselab.stanford.edu), respectively. This paper was prepared with Madagascar, a reproducible research package ([rsf.sourceforge.net](http://rsf.sourceforge.net)). 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 Felix J. Herrmann 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 Up-

stream 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.

## 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., L. Demanet, D. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: *SIAM Multiscale Model. Simul.*, **5**, 861–899.
- Candès, E. and D. Donoho, 2005a, Continuous Curvelet Transform I: Resolution of the Wavefront Set: *Appl. Comput. Harmon. Anal.*, **19**, 162–197.
- , 2005b, Continuous Curvelet Transform II: Discretization and Frames: *Appl. Comput. Harmon. Anal.*, **19**, 198–222.
- Candès, E., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: *Comm. Pure Appl. Math.*, **59**, 1207–1223.
- Candès, E. J. and D. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities: *Comm. Pure Appl. Math.*, **57**, 219–266.
- 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.
- Candès, E. J. and J. Romberg, 2004, Practical signal recovery from random projections: Presented at the Wavelet Applications in Signal and Image Processing XI.

- Candès, E. J., J. Romberg, and T. Tao, 2006c, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: *IEEE Trans. Inform. Theory*, **52**, 489–509.
- Canning, A. and G. H. F. Gardner, 1996, Regularizing 3-D data sets with DMO: *Geophysics*, **61**, 1103–1114.
- Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: *SIAM Journal on Scientific Computing*, **43**, 129–159.
- Claerbout, J. and F. Muir, 1973, Robust modeling with erratic data: *Geophysics*, **38**, 826–844.
- Daubechies, I., 1992, *Ten lectures on wavelets*: SIAM.
- Daubechies, I., M. Defrise, and C. de Mol, 2005, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints: *CPAM*, 1413–1457.
- Donoho, D., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of underdetermined linear equations by stagewise orthonormal matching pursuit. Preprint.
- Donoho, D. L., 2006, Compressed sensing: *IEEE Trans. Inform. Theory*, **52**, 1289–1306.
- Duffin, R. J. and A. C. Schaeffer, 1952, A class of nonharmonic Fourier series: *Trans. Amer. Math. Soc.*, **72**.
- Duijndan, A. and M. Schoneville, 1999, Nonuniform fast Fourier transform: *Geophysics*, **64**, 539–551.
- Elad, M., J. Starck, P. Querre, and D. Donoho, 2005, Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA): *Appl. Comput. Harmon. Anal.*, **19**, 340–358.
- Figueiredo, M. and R. Nowak, 2003, An EM algorithm for wavelet-based image restoration: *IEEE Trans. Image Processing*.

- 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 Kirchhoff migration: *Geophys. Pros.*, **50**, 577–588.
- Guittou, A., 2004, Adaptive subtraction of multiples using the  $\ell_1$ -norm: *Geophys Prospect*, **52**, 27–27.
- 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. Herrmann, 2006a, Application of stable signal recovery to seismic interpolation: Presented at the SEG International Exposition and 76th Annual Meeting.
- Hennenfent, G. and F. J. Herrmann, 2006b, Seismic denoising with non-uniformly sampled curvelets: *IEEE Comp. in Sci. and Eng.*, **8**, 16–25.
- , 2007, Seismic data regularization: a comparative study. In preparation.
- Herrmann, F. and G. Hennenfent, 2005, Non-linear data continuation with redundant frames: Presented at the CSEG National Convention.
- 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, 2006, Nonlinear primary-multiple separation with directional curvelet frames. In revision.
- Herrmann, F. J. and T. Lin, 2006, Compressed extrapolation. Submitted for publication.
- Levy, S., D. Oldenburg, and J. Wang, 1988, Subsurface imaging using magnetotelluric data: *Geophysics*, **53**, 104–117.
- Liu, B. and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**, 1560–1568.

- Malcolm, A., 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.
- Mallat, S. G., 1997, *A wavelet tour of signal processing*: Academic Press.
- Masry, E., 1978, Alias-free sampling : an alternative conceptualization and its applications: *IEEE Trans. Inform. Theory*, **IT-22**, 298–312.
- Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: *Geophysics*, **46**, 1528–1542.
- Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: *Proceedings of the eleventh annual international conference of the Center for Nonlinear Studies on Experimental mathematics : computational issues in nonlinear science*, 259–268, Elsevier North-Holland, Inc.
- Sacchi, M. and T. Ulrych, 1996, Estimation of the discrete Fourier transform, a linear inversion approach: *Geophysics*, **61**, 1128–1136.
- 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. 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. Donoho, 2004, Redundant multiscale transforms and their application to morphological component separation: *Advances in Imaging and Electron Physics*, **132**.

- Thomson, D., G. Hennenfent, H. Modzelewski, and F. Herrmann, 2006, A parallel windowed fast discrete curvelet transform applied to seismic processing: Presented at the SEG International Exposition and 76th Annual Meeting.
- Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse radon transform: *Geophysics*, **68**, 386–399.
- Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: *Geophysics*, **68**, 2043–2054.
- Tropp, T., 2006, Just relax: convex programming methods for identifying sparse signals in noise: *IEEE Trans. Inform. Theory*, **52**, 1030–1051.
- Ulrych, T. J. and C. Walker, 1982, Analytic minimum entropy deconvolution: *Geophysics*, **47**, 1295–1302.
- 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**.
- Ying, L., L. Demanet, and E. Candés, 2005, 3d discrete curvelet transform: , 591413, SPIE.
- Zwartjes, P. and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: *Geophysics*, **71**, V171–V186.

## LIST OF FIGURES

1 Fourier spectra for incomplete subsampled data. **(a)** Regularly missing data that lead to a strongly aliased spectrum plotted in **(b)**. **(c)** Undersampled data with data missing on an uniform random grid that gives rise to a noisy Fourier spectrum plotted in **(d)**. Observe that the Fourier spectrum for the random subsampled data looks noisy while the regular undersampled data displays the well-known and harmful periodic imprint of aliasing.

2 Recovery of a strictly sparse signal in the discrete cosine transform (DCT) domain from 5 sample points of the original signal. **(a)** Original (plain line), recovered (+) and the measured signals (line with the 5 non-zero measurements). **(b)** DCT representations of original (plain line) and recovered signals (+). **(c-d)** the same as **(a-b)** but now for 4 measurements for which recovery fails. When successful the recovery is perfect and the transition from success to failure is sharp for a given experiment.

3 Differences between the matched filter and the sparsity vector (cf. Eq. 14) from Fig. 1. **(a)** the aliased case for the regular subsampling. **(b)** the 'noisy' case for the random subsampling. Observe that the residual for the random subsampled data looks like 'Gaussian noise', while the regular undersampled data contains harmful the periodic imprint of aliasing.

4 Example of a phase diagram for strictly sparse length  $N = 800$  signals and noise-free measurements. The number of independent experiments for each parameter pair  $(\delta, \rho) \in (0, 1] \times (0, 1]$  is 25. The grey-scale of each of the  $25 \times 25$  pixels represents the number of entries in the sparsity vector that deviate by more than  $10^{-4}$ . The darker the pixel, the less likely the recovery for that specific parameter pair  $(\delta, \rho)$ . Observe that there is a relatively sharp transition between the regions where recovery is successful and where



it fails. Starting for the very sparse single spike vector on the left, the recovery starts to be successful for approximately 5 measurements ( $\rho \approx 0.2$ ) and works its way gradually up as the vector becomes less sparse. Recovery on the far right, for non-sparse vectors, is still possible because the cosine transform is orthonormal.

5 Recovery of a strictly sparse signal in the DCT domain from 15 noisy samples (with signal-to-noise ratio (SNR) of 6 dB). **(a)** Original (plain line), recovered (+) and signals (line with 15 non-zero noisy measurements). **(b)** DCT representations of the original (plain line) and recovered (with SNR 9 dB) signals (+). Observe that the recovery is not exact since the algorithm can only recover to within the noise level.

6 Recovery of a compressible signal in the DCT domain from 20 sample points of the original signal. **(a)** Original (plain line) and recovered (+) signals. **(b)** DCT representations of original (plain line) and recovered signals (+). Note that the large DCT coefficients are recovered i.e. most of the signal's energy is captured.

7 Example of a recovery diagram for parameter combinations  $(\delta, r) \in (0, 1] \times (1/2, 2]$  on a regular grid of  $25 \times 25$ . Notice that the relative  $\ell_2$  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.

8 Example of the alignment of curvelets with curved events.

9 Discrete curvelet partitioning of the 2-D Fourier plane into second dyadic coronae and sub-partitioning of the coronae into angular wedges.

10 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  $90^\circ$  rotation is a property of the Fourier transform.

11 Example of the reconstruction from 1 % of the largest Fourier, wavelet and curvelet coefficients. **(a)** A shot record from a real marine dataset. Reconstruction from 1 % of the largest **(b)** Fourier; **(c)** wavelet and **(d)** curvelet coefficients. The curvelet reconstruction clearly performs best.

12 Decay of the transform coefficients for a typical synthetic (the fully sampled data set that corresponds to Fig. 1(a)) and real data set (Fig. 11(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.

13 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.

14 Curvelet reconstruction using 2-D synthetic with a layered earth model. **(a)** Complete CMP gather and **(b)** a zoom of an area in the CMP where there are conflicting dips. **(c)** Simulated acquired data with about 60 % randomly missing traces and **(d)** a zoom of

the same area as **(b)**. **(e)** The curvelet reconstruction and **(f)** the zoom. **(g)** The difference between reconstruction and complete data and **(h)** the zoom. Virtually all the initial seismic energy is recovered without error as illustrated by the difference plots (SNR = 29.8 dB).

15 Illustration of sliced versus volumetric interpolation.

16 Comparison between sliced and volumetric CRSI. **(a)** One complete shot from a 2-D synthetic prestack dataset and **(b)** the corresponding simulated acquired data with 80% randomly missing traces. **(c)** 2-D CRSI reconstruction and **(d)** difference between **(c)** and **(a)**. **(e)** 3-D CRSI reconstruction and **(f)** difference between **(e)** and **(a)**. 3-D CRSI clearly benefits from 3-D information that greatly improves the reconstruction over 2-D CRSI.

17 2-D prestack data interpolation using 3-D CRSI. **(a)** Complete synthetic prestack dataset, **(b)** simulated acquired data with 80% randomly missing traces, **(c)** 3-D CRSI reconstruction, and **(d)** difference between **(c)** and **(a)**.

18 2-D real data interpolation using CRSI and PWD. **(a)** Shot-record of a seismic survey from offshore Gippsland basin Australia. Group interval is 12.5 m. **(b)** Same data as **(a)**, but randomly omitting 60% of the original traces (corresponding average spatial sampling is 31.25 m). **(c)** and **(d)** are CRSI result and difference with **(a)**, respectively. **(e)** and **(f)** are PWD result and difference with **(a)**, respectively. SNR's are 18.8 dB for CRSI and 5.5 dB for PWD.

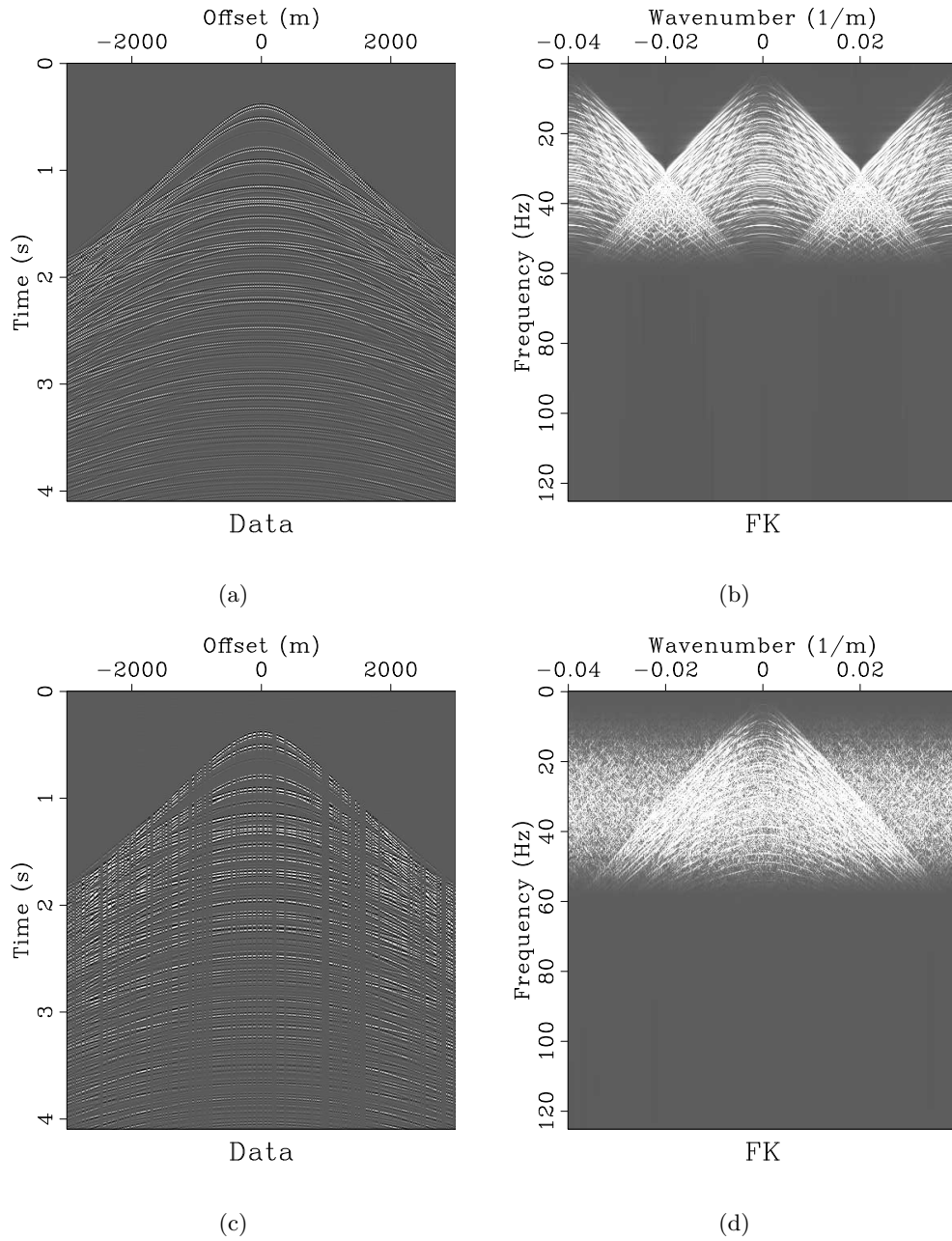


Figure 1: Fourier spectra for incomplete subsampled data. **(a)** Regularly missing data that lead to a strongly aliased spectrum plotted in **(b)**. **(c)** Undersampled data with data missing on an uniform random grid that gives rise to a noisy Fourier spectrum plotted in **(d)**. Observe that the Fourier spectrum for the random subsampled data looks noisy while the regular undersampled data displays the well-known and harmful periodic imprint of aliasing.

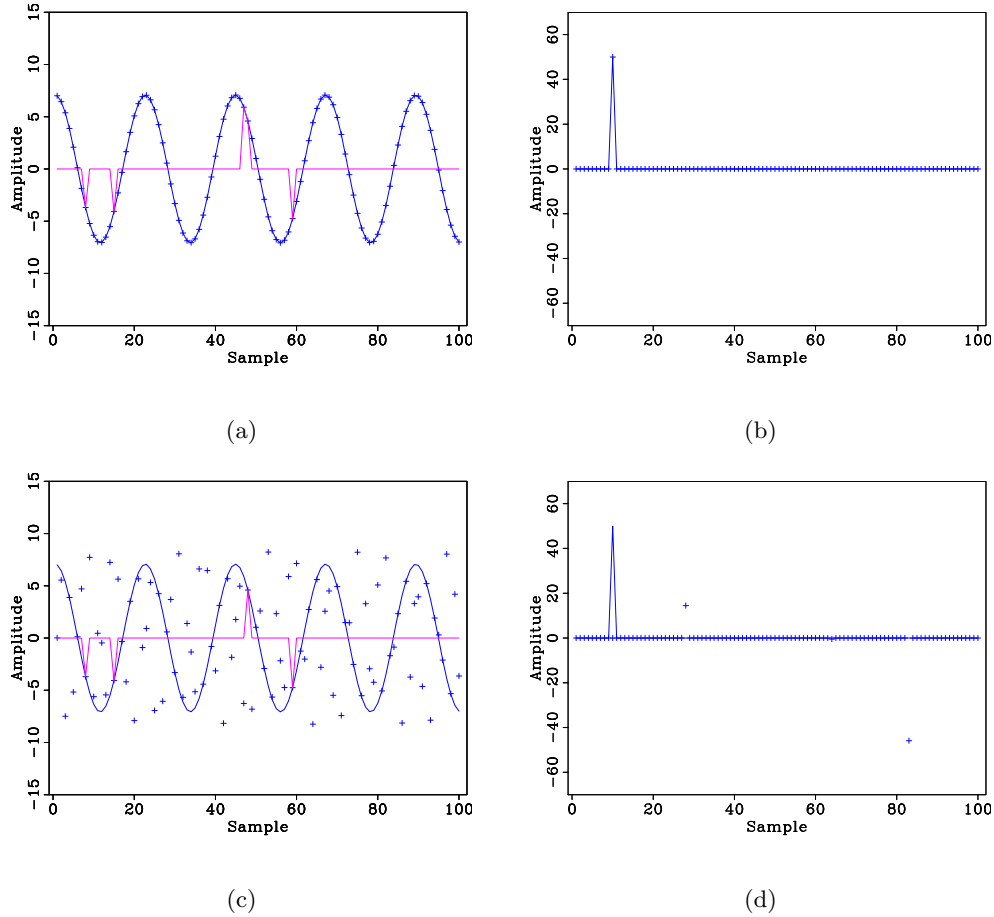


Figure 2: Recovery of a strictly sparse signal in the discrete cosine transform (DCT) domain from 5 sample points of the original signal. **(a)** Original (plain line), recovered (+) and the measured signals (line with the 5 non-zero measurements). **(b)** DCT representations of original (plain line) and recovered signals (+). **(c-d)** the same as **(a-b)** but now for 4 measurements for which recovery fails. When successful the recovery is perfect and the transition from success to failure is sharp for a given experiment.

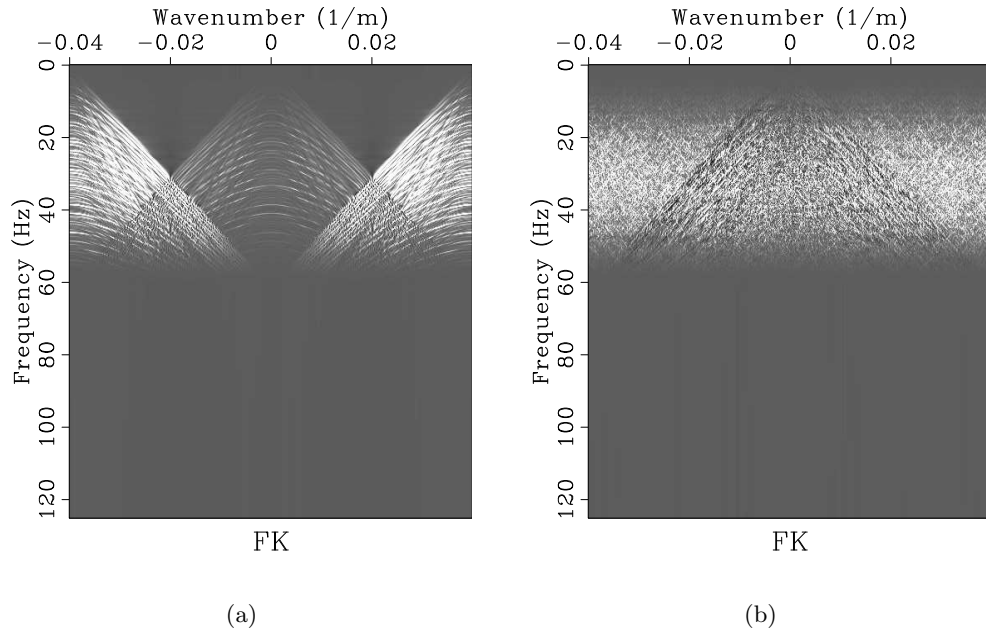


Figure 3: Differences between the matched filter and the sparsity vector (cf. Eq. 14) from Fig. 1. **(a)** the aliased case for the regular subsampling. **(b)** the 'noisy' case for the random subsampling. Observe that the residual for the random subsampled data looks like 'Gaussian noise', while the regular undersampled data contains harmful the periodic imprint of aliasing.

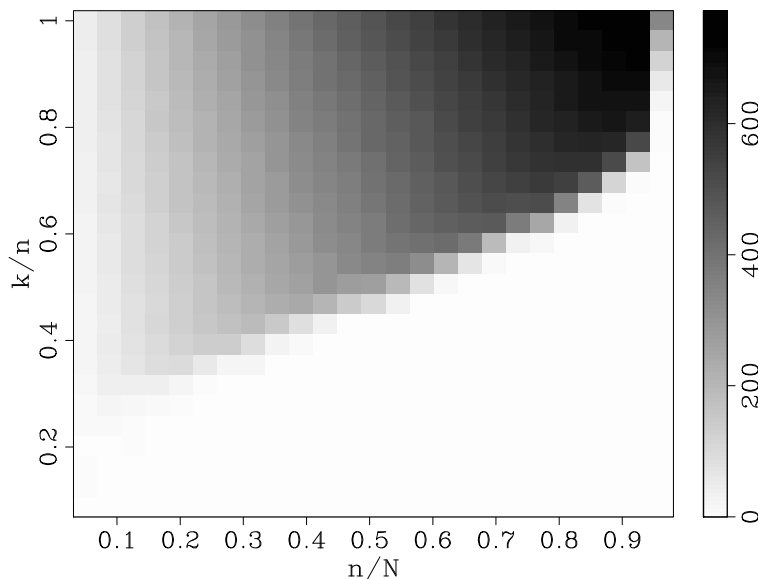


Figure 4: Example of a phase diagram for strictly sparse length  $N = 800$  signals and noise-free measurements. The number of independent experiments for each parameter pair  $(\delta, \rho) \in (0, 1] \times (0, 1]$  is 25. The grey-scale of each of the  $25 \times 25$  pixels represents the number of entries in the sparsity vector that deviate by more than  $10^{-4}$ . The darker the pixel, the less likely the recovery for that specific parameter pair  $(\delta, \rho)$ . Observe that there is a relatively sharp transition between the regions where recovery is successful and where it fails. Starting for the very sparse single spike vector on the left, the recovery starts to be successful for approximately 5 measurements ( $\rho \approx 0.2$ ) and works its way gradually up as the vector becomes less sparse. Recovery on the far right, for non-sparse vectors, is still possible because the cosine transform is orthonormal.

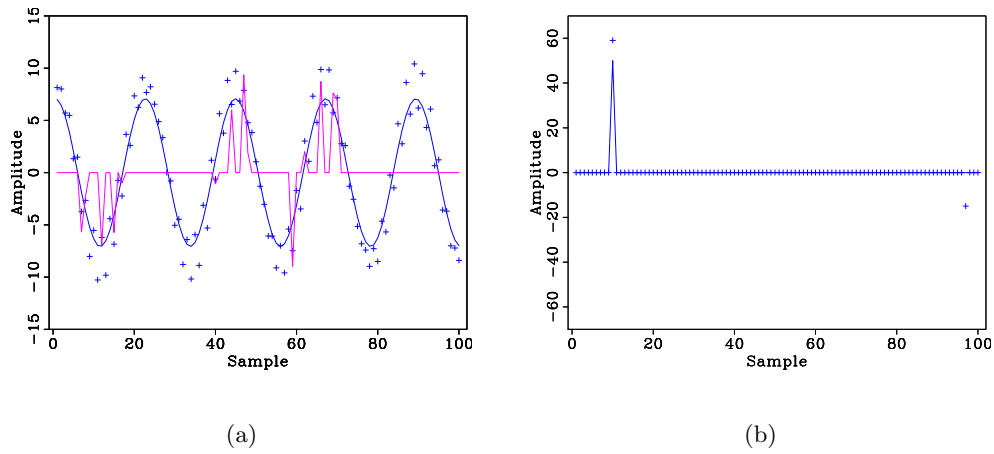


Figure 5: Recovery of a strictly sparse signal in the DCT domain from 15 noisy samples (with signal-to-noise ratio (SNR) of 6 dB). **(a)** Original (plain line), recovered (+) and signals (line with 15 non-zero noisy measurements). **(b)** DCT representations of the original (plain line) and recovered (with SNR 9 dB) signals (+). Observe that the recovery is not exact since the algorithm can only recover to within the noise level.



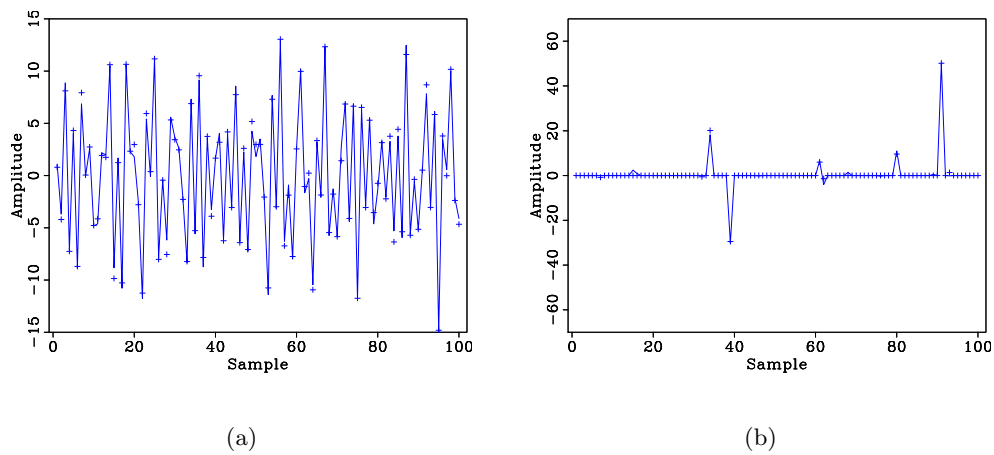


Figure 6: Recovery of a compressible signal in the DCT domain from 20 sample points of the original signal. **(a)** Original (plain line) and recovered (+) signals. **(b)** DCT representations of original (plain line) and recovered signals (+). Note that the large DCT coefficients are recovered i.e. most of the signal's energy is captured.

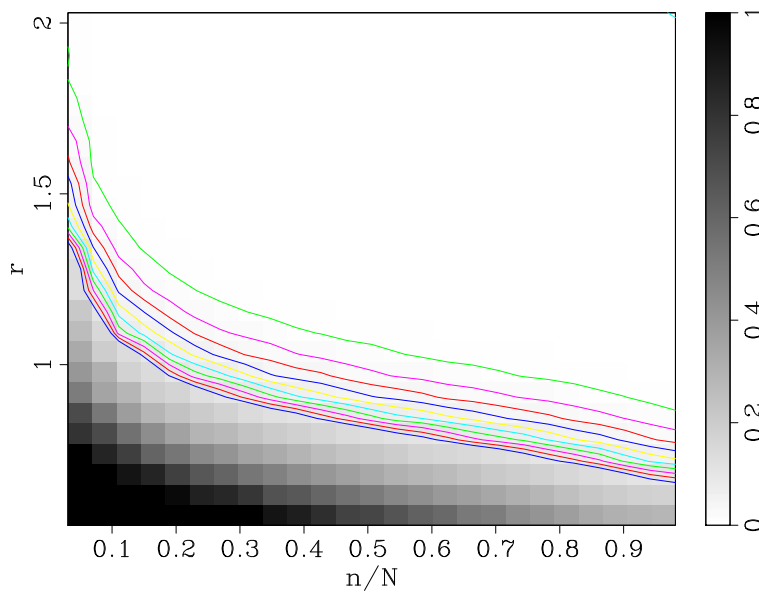


Figure 7: Example of a recovery diagram for parameter combinations  $(\delta, r) \in (0, 1] \times (1/2, 2]$  on a regular grid of  $25 \times 25$ . Notice that the relative  $\ell_2$  error decays the most rapidly with  $r$ . The contour lines represent 1% decrements in the recovery error starting at 10% on the lower-left corner and decaying to 1% in the direction of the upper-right corner.

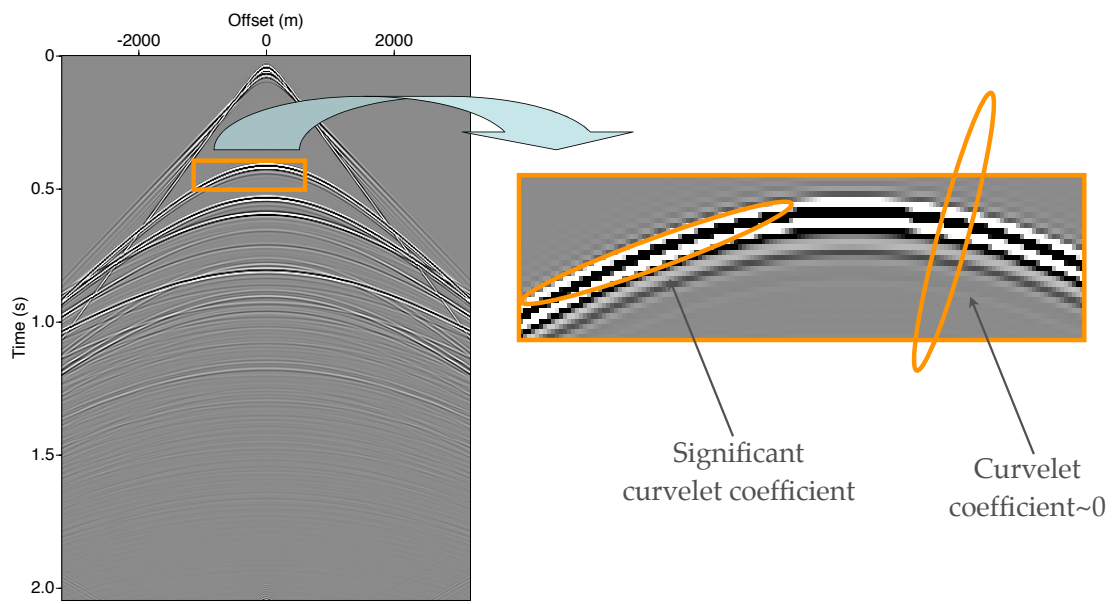


Figure 8: Example of the alignment of curvelets with curved events.

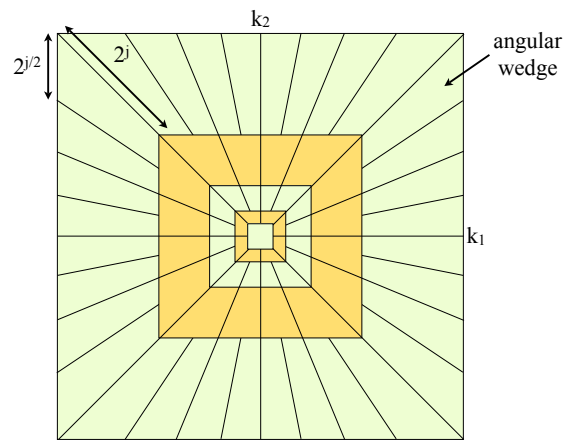


Figure 9: Discrete curvelet partitioning of the 2-D Fourier plane into second dyadic coronae and sub-partitioning of the coronae into angular wedges.

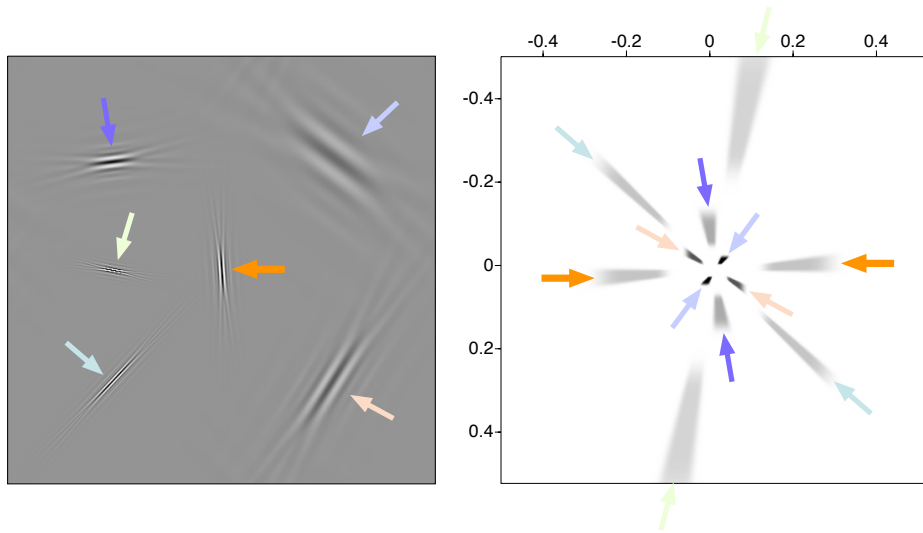


Figure 10: Spatial and frequency representation of curvelets. **(a)** Six different curvelets in the spatial domain at five different scales. **(b)** Dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. Each pair of opposing 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  $90^\circ$  rotation is a property of the Fourier transform.

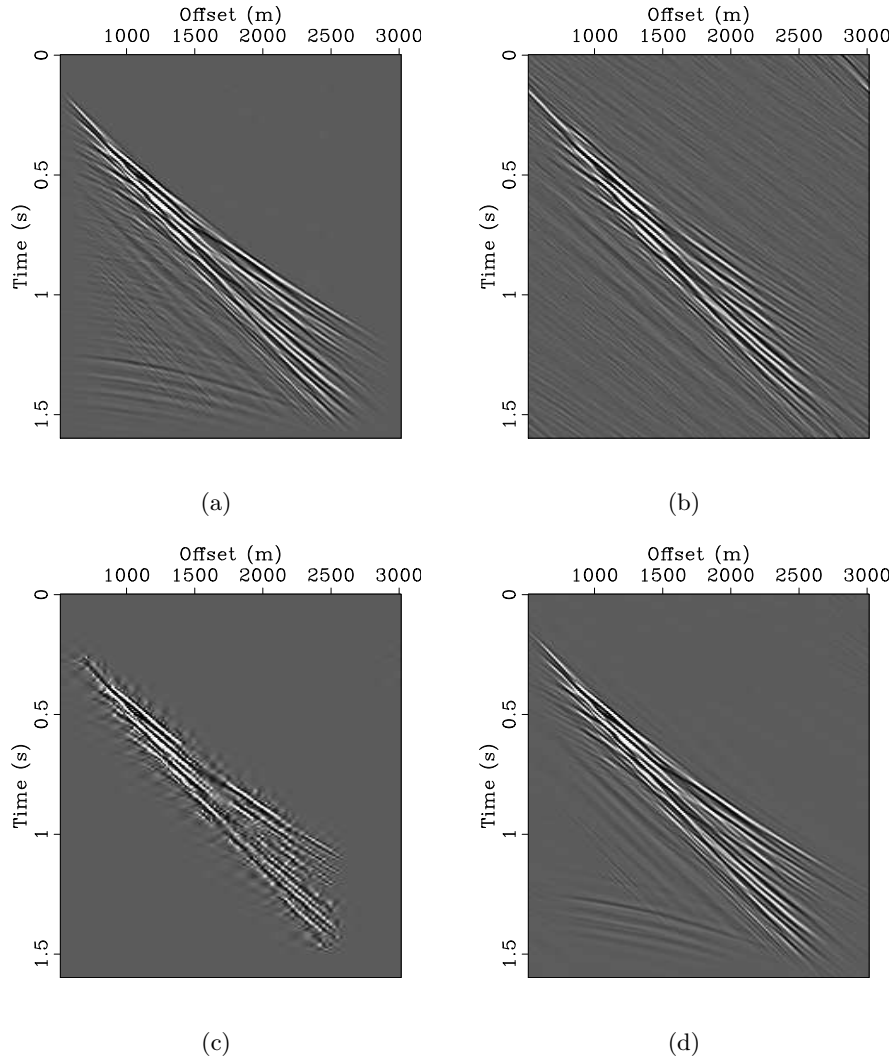
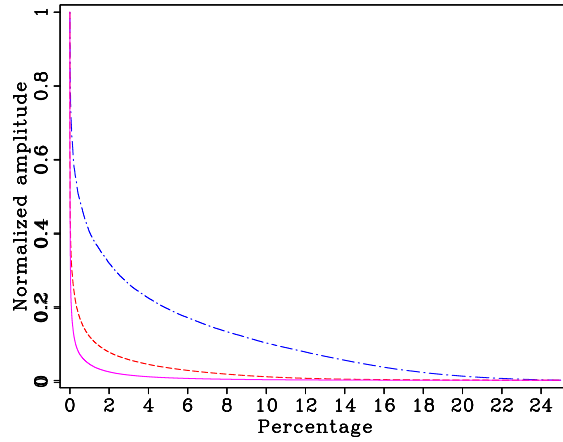
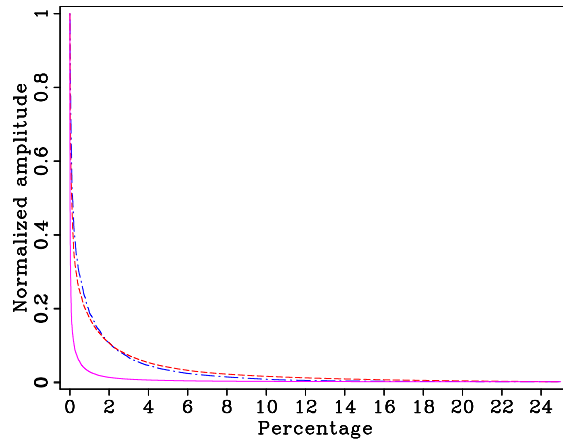


Figure 11: Example of the reconstruction from 1% of the largest Fourier, wavelet and curvelet coefficients. **(a)** A shot record from a real marine dataset. Reconstruction from 1% of the largest **(b)** Fourier; **(c)** wavelet and **(d)** curvelet coefficients. The curvelet reconstruction clearly performs best.



(a)



(b)

Figure 12: Decay of the transform coefficients for a typical synthetic (the fully sampled data set that corresponds to Fig. 1(a)) and real data set (Fig. 11(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.

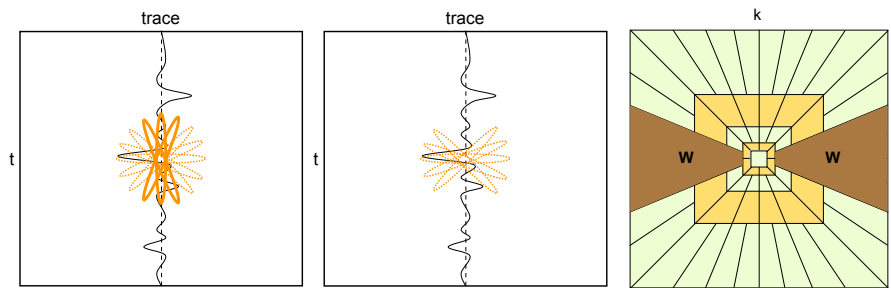


Figure 13: Illustration of the angular weighting designed to reduce the adverse effects of seismic sampling. On the left, the increased mutual coherence between near vertical-oriented curvelets and a missing trace. In the middle, a schematic of the curvelets that survive the angular weighting illustrated on the right.



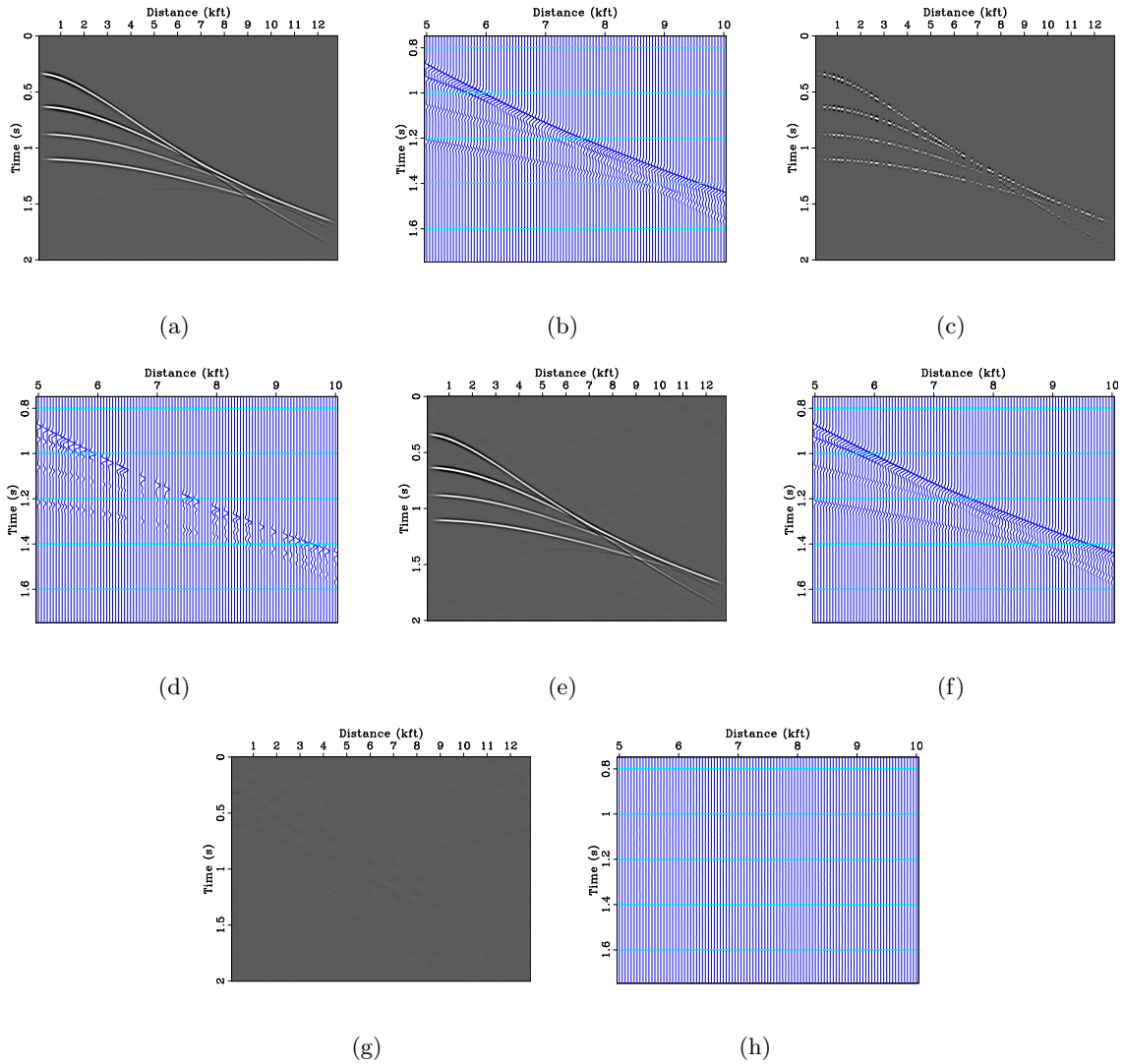


Figure 14: Curvelet reconstruction using 2-D synthetic with a layered earth model. (a) Complete CMP gather and (b) a zoom of an area in the CMP where there are conflicting dips. (c) Simulated acquired data with about 60% randomly missing traces and (d) a zoom of the same area as (b). (e) The curvelet reconstruction and (f) the zoom. (g) The difference between reconstruction and complete data and (h) the zoom. Virtually all the initial seismic energy is recovered without error as illustrated by the difference plots (SNR = 29.8 dB).

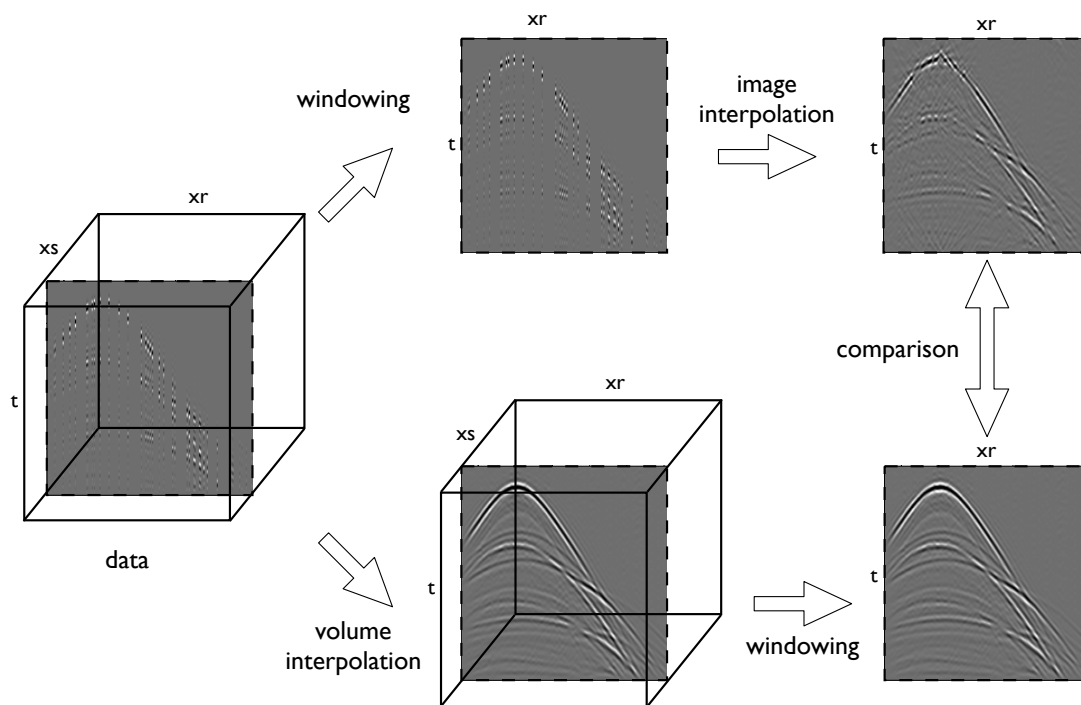


Figure 15: Illustration of sliced versus volumetric interpolation.

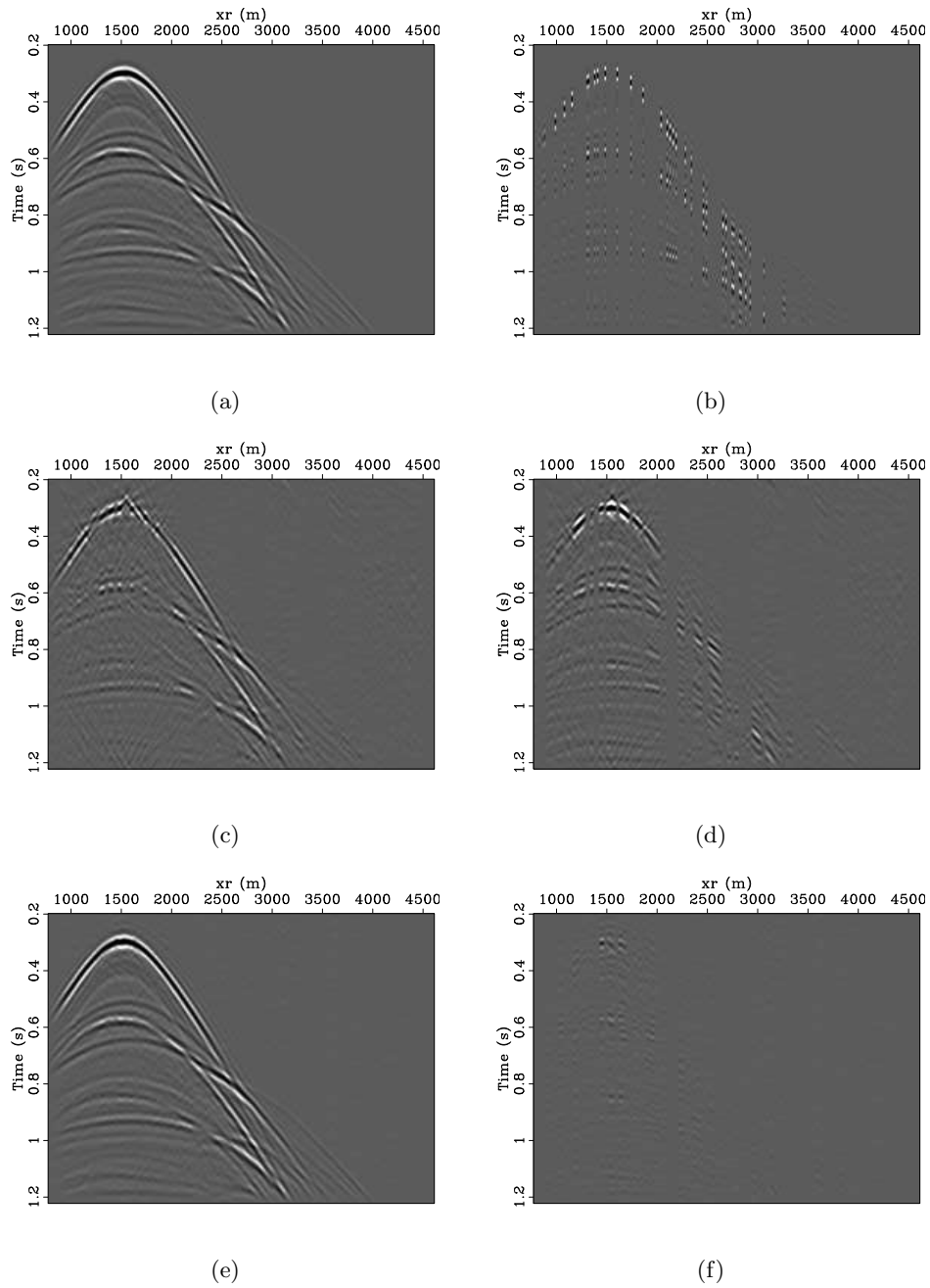


Figure 16: Comparison between sliced and volumetric CRSI. **(a)** One complete shot from a 2-D synthetic prestack dataset and **(b)** the corresponding simulated acquired data with 80% randomly missing traces. **(c)** 2-D CRSI reconstruction and **(d)** difference between **(c)** and **(a)**. **(e)** 3-D CRSI reconstruction and **(f)** difference between **(e)** and **(a)**. 3-D CRSI clearly benefits from 3-D information that greatly improves the reconstruction over 2-D CRSI.

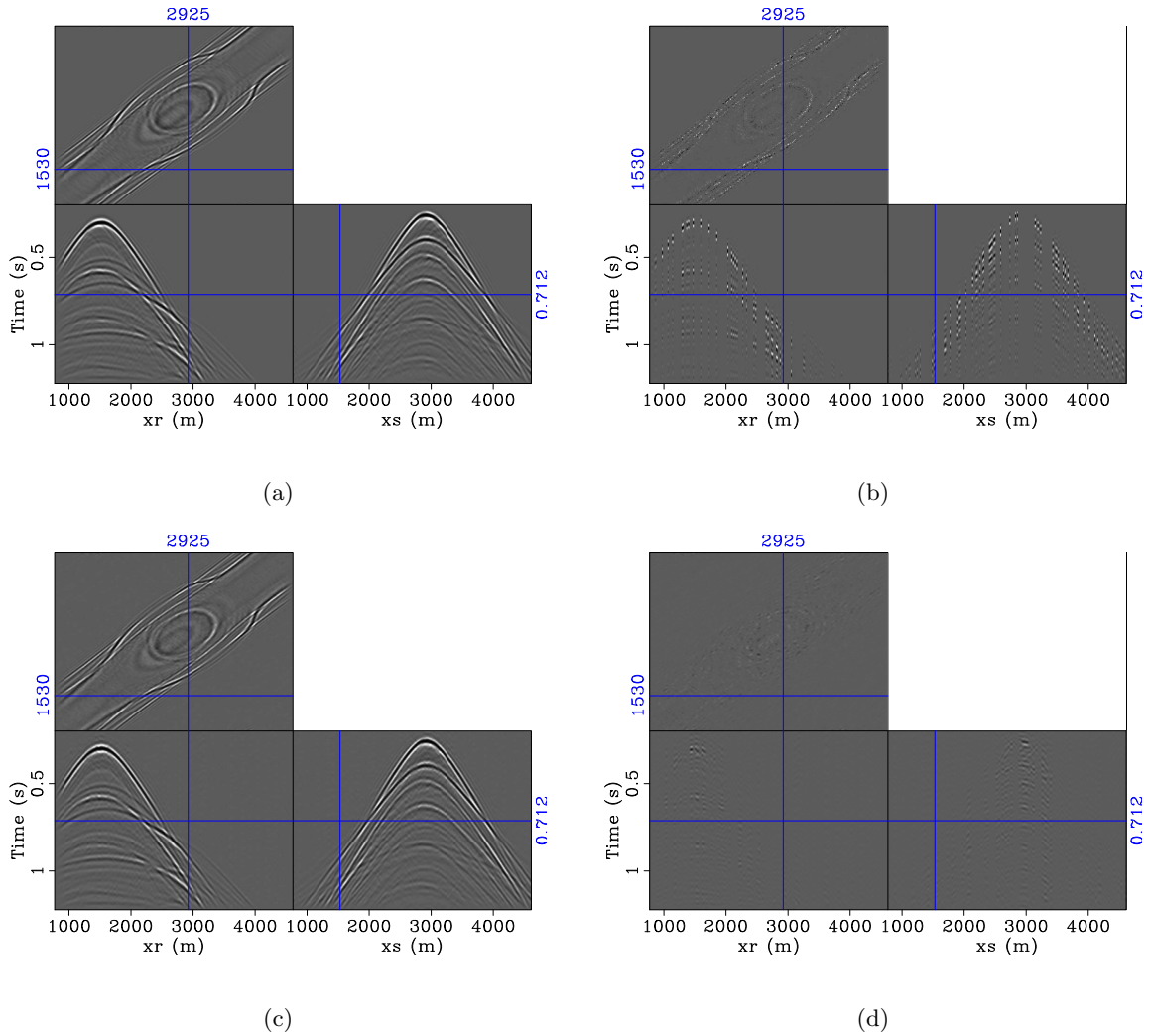


Figure 17: 2-D prestack data interpolation using 3-D CRSI. **(a)** Complete synthetic prestack dataset, **(b)** simulated acquired data with 80% randomly missing traces, **(c)** 3-D CRSI reconstruction, and **(d)** difference between **(c)** and **(a)**.

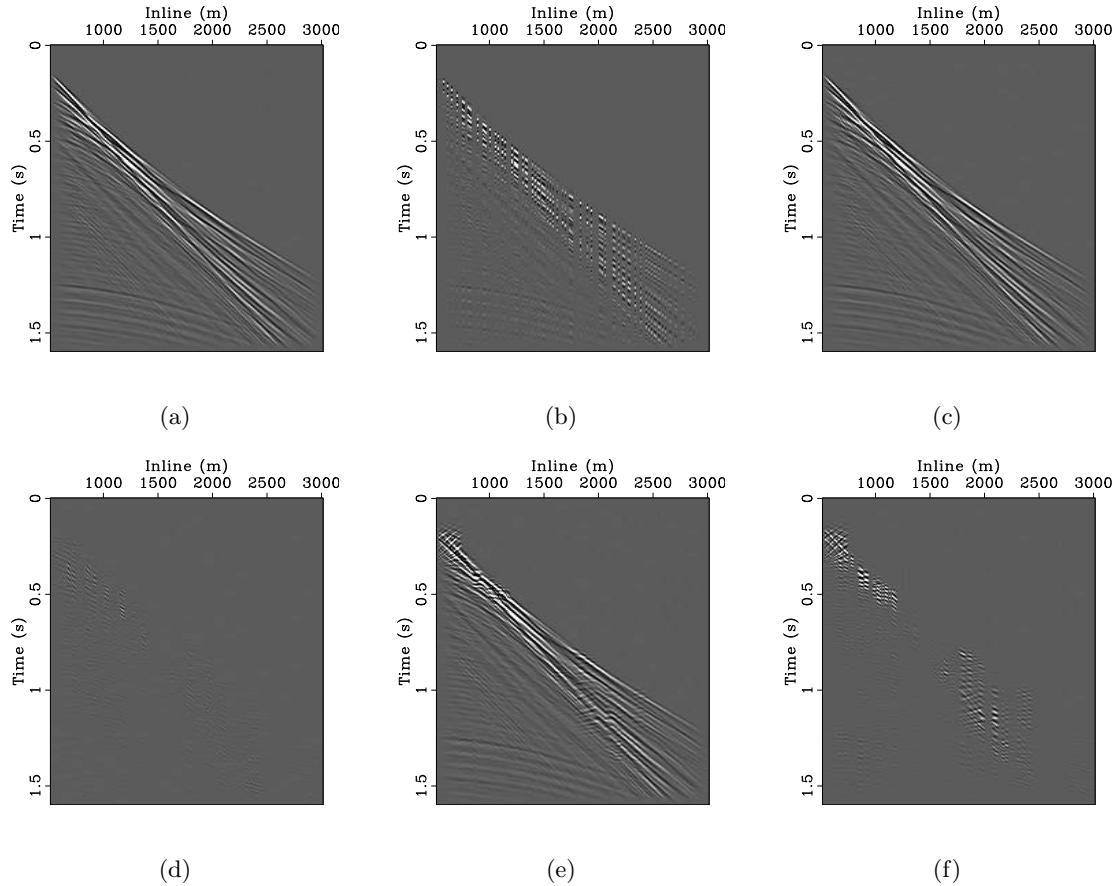


Figure 18: 2-D real data interpolation using CRSI and PWD. **(a)** Shot-record of a seismic survey from offshore Gippsland basin Australia. Group interval is 12.5 m. **(b)** Same data as **(a)**, but randomly omitting 60% of the original traces (corresponding average spatial sampling is 31.25 m). **(c)** and **(d)** are CRSI result and difference with **(a)**, respectively. **(e)** and **(f)** are PWD result and difference with **(a)**, respectively. SNR's are 18.8 dB for CRSI and 5.5 dB for PWD.