Randomized sampling and sparsity: getting more information from fewer samples

Felix J. Herrmann[1]

ABSTRACT

Many seismic exploration techniques rely on the collection of massive data volumes that are subsequently mined for information during processing. While this approach has been extremely successful in the past, current efforts toward higher-resolution images in increasingly complicated regions of the Earth continue to reveal fundamental shortcomings in our workflows. Chiefly amongst these is the so-called "curse of dimensionality" exemplified by Nyquist's sampling criterion, which disproportionately strains current acquisition and processing systems as the size and desired resolution of our survey areas continues to increase. In this paper, we offer an alternative sampling method leveraging recent insights from compressive sensing towards seismic acquisition and processing for data that are traditionally considered to be undersampled. The main outcome of this approach is a new technology where acquisition and processing related costs are no longer determined by overly stringent sampling criteria, such as Nyquist. At the heart of our approach lies randomized incoherent sampling that breaks subsampling related interferences by turning them into harmless noise, which we subsequently remove by promoting transform-domain sparsity. Now, costs no longer grow significantly with resolution and dimensionality of the survey area, but instead depend on transform-domain sparsity only. Our contribution is twofold. First, we demonstrate by means of carefully designed numerical experiments that compressive sensing can successfully be adapted to seismic exploration. Second, we show that accurate recovery can be accomplished for compressively sampled data volumes sizes that exceed the size of conventional transform-domain data volumes by only a small factor. Because compressive sensing combines transformation and encoding by a single linear encoding step, this technology is directly applicable to acquisition and to dimensionality reduction during processing. In either case, sampling, storage, and processing costs scale with transform-domain sparsity. We illustrate this principle by means of number of case studies.

[1]Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, the University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada. Email: fherrmann@eos.ubc.ca

The University of British Columbia Technical Report. TR-2010-01, 2010-05-19

INSPIRATION

Nyquist sampling and the curse of dimensionality

The livelihood of exploration seismology depends on our ability to collect, process, and image extremely large seismic data volumes. The recent push towards full-waveform only exacerbates this reliance, and we, much like researchers in many other fields in science and engineering, are constantly faced with the challenge to come up with new and innovative ways to mine this overwhelming barrage of data for information.

In exploration seismology, this challenge is especially daunting because our data volumes sample wavefields that exhibit structure in up to five dimensions (two coordinates for the sources, two for the receivers, and one for time). To resolve this high-dimensional structure, we are not only confronted with Nyquist's sampling criterion but we also face the so-called "curse of dimensionality", which refers to the exponential increase in volume when adding extra dimensions to our data collection.

These two challenges are amongst the largest impediments to progress in the application of the seismic method to oil and gas exploration. In this paper, we introduce a new methodology adapted from the field of "compressive sensing" or "compressive sampling" (CS—in short throughout the paper Candès et al., 2006; Donoho, 2006a; Mallat, 2009), which is aimed at removing these impediments via dimensionality reduction techniques based on randomized subsampling. With this dimensionality reduction, we arrive at a sampling framework where the sampling rates are no longer dominated by the gridsize but by transform-domain compression; more compressible data requires less sampling.

Dimensionality reduction by compressive sensing

Current nonlinear data-compression techniques are based on high-resolution linear sampling (e.g., sampling by a CCD chip in a digital camera) followed by a nonlinear encoding technique that consists of transforming the samples to some transformed domain, where the signal's energy is encoded by a relatively small number of significant transform-domain coefficients (Mallat, 2009). Compression is accomplished by keeping only the largest transform-domain coefficients. Because this compression is lossy, there is an error after decompression. A compression ratio expresses the compressed-signal size as a fraction of the size of the original signal. The better the transform captures the energy in the sampled data, the larger the attainable compression ratio for a fixed loss.

Even though this technique underlies the digital revolution of many consumer devices, including digital cameras, iPods, etc., this approach is not well suited for exploration seismology because it is wasteful in two important ways. First, high-resolution data has to be collected during the linear sampling step, which is prohibitively expensive. Second, the encoding phase is nonlinear. This means that if

we select a compression ratio that is too high, the decompressed signal may have an unacceptable error, in the worst case making it necessary to repeat collection of the high-resolution samples, which is infeasible in cases where storage of high-resolution data is a concern or where the cost of acquisition forms the main impediment such as in exploration seismology.

By replacing the combination of high-resolution sampling and nonlinear compression by a single randomized subsampling technique that combines sampling and encoding in one single linear step, CS addresses many of the above shortcomings. First of all, randomized subsampling has the distinct advantage that the encoding is linear and does not require access to high-resolution data during encoding. This opens possibilities to sample incrementally and to process data in the compressed domain. Second, encoding through randomized sampling suppresses subsampling related artifacts. Coherent subsampling related artifacts—whether these are caused by periodic missing traces or by cross-talk between coherent simultaneous-sources—are turned into relatively harmless incoherent Gaussian noise by randomized subsampling (see e.g. Herrmann and Hennenfent, 2008; Hennenfent and Herrmann, 2008; Herrmann et al., 2009b, for seismic applications of this idea).

By solving a sparsity-promoting problem (Candès et al., 2006; Donoho, 2006a; Herrmann et al., 2007; Mallat, 2009), we reconstruct high-resolution data volumes from the randomized samples at the moderate cost of a minor oversampling factor compared to data volumes obtained after conventional compression (see e.g. Donoho et al., 1999, for wavelet-based compression). With sufficient sampling, this nonlinear recovery outputs a set of largest transform-domain coefficients that produces a reconstruction with a recovery error comparable with the error incurred during conventional compression. As in conventional compression this error is controllable, but in case of CS this recovery error depends on the sampling ratio—i.e., the ratio between the number of samples taken and the number of samples of the high-resolution data. Because compressively sampled data volumes are much smaller than high-resolution data volumes, we reduce the dimensionality and hence the costs of acquisition, storage, and possibly of data-driven processing.

In this paper, we mainly consider recovery methods that derive from compressive sampling. Therefore our method differs from interpolation methods based on pattern recognition (Spitz, 1999), plane-wave destruction (Fomel et al., 2002) and data mapping (Bleistein et al., 2001), including parabolic, apex-shifted Radon and DMO-NMO/AMO (Trad, 2003; Trad et al., 2003; Harlan et al., 1984; Hale, 1995; Canning and Gardner, 1996; Bleistein et al., 2001; Fomel, 2003; Malcolm et al., 2005). To benefit fully from this new sampling paradigm, we will translate and adapt its ideas to exploration seismology while evaluating their performance. Here lies our main contribution. Before we embark on this mission we first share some basic insights from compressive sensing in the context of two well-known problems in geophysics, namely recovery from time-harmonic signals, which is relevant for missing-trace interpolation, and spiky deconvolution.

Compressive sensing is based on three key elements: randomized sampling, spar-

sifying transforms, and sparsity-promotion recovery by convex optimization . By themselves, these elements are not new to geophysics. Spiky deconvolution and high-resolution transforms are all based on sparsity-promotion (Taylor et al., 1979; Oldenburg et al., 1981; Ulrych and Walker, 1982; Levy et al., 1988; Sacchi et al., 1994) and analyzed by mathematicians (Santosa and Symes, 1986; Donoho and Logan, 1992); wavelet transforms are used for seismic data compression (Donoho et al., 1999); randomized samples have been shown to benefit Fourier-based recovery from missing traces (Trad et al., 2003; Xu et al., 2005; Abma and Kabir, 2006; Zwartjes and Sacchi, 2007b). The contribution of CS lies in the combination of these concepts into a comprehensive theoretical framework that provides design principles and performance guarantees.

Examples

Periodic versus uniformly random subsampling

Because Nyquist's sampling criterion guarantees perfect reconstruction of arbitrary bandwidth-limited signals, it has been the leading design principle for seismic data acquisition and processing. This explains why acquisition crews go at length to place sources and receivers as finely and as regularly as possible. Although this approach spearheaded progress in our field, CS proves that periodic sampling at Nyquist rates may be far from optimal when the signal of interest exhibits some sort of structure, such as when the signal permits a transform-domain representation with few significant and many zero or insignificant coefficients. For this class of signals (which includes nearly all real-world signals) it suffices to sample randomly at a rate that is lower than Nyquist.

As an example consider recovery of harmonic signals that are made of superpositions of a limited number of Fourier modes (e.g. cosines). According to Nyquist, this type of signal can be fully recovered from periodic measurements, sampled at a rate at least twice the rate of the highest mode. However, according to compressive sensing, we can recover this type of signals from far fewer randomly placed samples. In the seismic situation, this corresponds to using seismic arrays with fewer geophones selected uniformly random from regular sampling grids with spacings defined by Nyquist. By taking these samples randomly instead of periodically, we turn coherent aliases into Gaussian white noise (Donoho et al., 2006; Hennenfent and Herrmann, 2008; Donoho et al., 2009) as illustrated in Figure 1. We observe that for the same number of samples the subsampling artifacts can behave very differently.

In the geophysical community, subsampling-related artifacts are commonly known as "spectral leakage" (Xu et al., 2005), where energy from each frequency is leaked to other frequencies. Understandably, the amount of spectral leakage depends on the degree of subsampling: the higher this degree the more leakage. However, the characteristics of the artifacts themselves depend on the irregularity of the sampling.

The more uniformly random our sampling is, the more the leakage behaves as zero-centered Gaussian noise spread over the entire frequency spectrum.

Compressive sensing schemes aim to design acquisition that specifically create Gaussian-noise like subsampling artifacts (Donoho et al., 2006, 2009). As opposed to coherent subsampling related artifacts (Figure 1(f)), these noisy artifacts (Figure 1(d)) can subsequently be removed by a sparse recovery procedure, during which the artifacts are separated from the signal and amplitudes are restored. Of course, the success of this method also hinges on the degree of subsampling, which determines the noise level, and the sparsity level of the signal.

By carrying out a random ensemble of experiments, where random realizations of harmonic signals are recovered from randomized samplings with decreasing sampling ratios, we confirm this behavior empirically. Our findings are summarized in Figure 2. The estimated spectra are obtained by solving a sparsifying program with the Spectral Projected Gradient for ℓ_1 solver (SPGL1 - Berg and Friedlander, 2008) for signals with k non-zero entries in the Fourier domain. We define these spectra by randomly selecting k entries from vectors of length 600 and populating these with values drawn from a Gaussian distribution with unit standard deviation. As we will show below, the solution of each of these problems corresponds to the inversion of a matrix whose aspect ratio, the ratio of the number of columns over the number of rows, increases as the number of samples decreases.

To get reasonable estimates, each experiment is repeated 100 times for the different subsampling schemes and for varying sampling ratios ranging from 1/2 to 1/6. The reconstruction error is the number of vector entries at which the estimated spectrum and the true spectrum disagree by more than 10^{-4}. This error counts both false positives and false negatives. The averaged results for the different experiments are summarized in Figures 2(a) and 2(b), which correspond to regular and random subsampling, respectively. The horizontal axes in these plots represent the relative underdeterminedness of the system, i.e., the ratio of the number k of nonzero entries in the spectrum to the number of acquired data points n. The vertical axes denote the percentage of erroneous entries. The different curves represents the different subsampling factors. In each plot, the curves from top to bottom correspond to sampling ratios of 1/2 to 1/6.

Figure 2(a) shows that, regardless of the subsampling factor, there is no range of relative underdeterminedness for which the spectrum and hence the signal can accurately be recovered from regular subsamplings. Sparsity is not enough to discriminate the signal components from the spectral leakage. The situation is completely different in Figure 2(b) for the random sampling. In this case, one can observe that for a subsampling ratio of 1/2 exact recovery is possible for $0 < k/n \lesssim 1/4$. The main purpose of these plots is to qualitatively show the transition from successful to failed recovery. The quantitative interpretation for these diagrams of the transition is less well understood but also observed in phase diagrams in the literature (Donoho et al., 2006; Donoho and Tanner, 2009; Donoho et al., 2009). A possible explanation for the observed behavior of the error lies in the nonlinear behavior of the solvers and on an

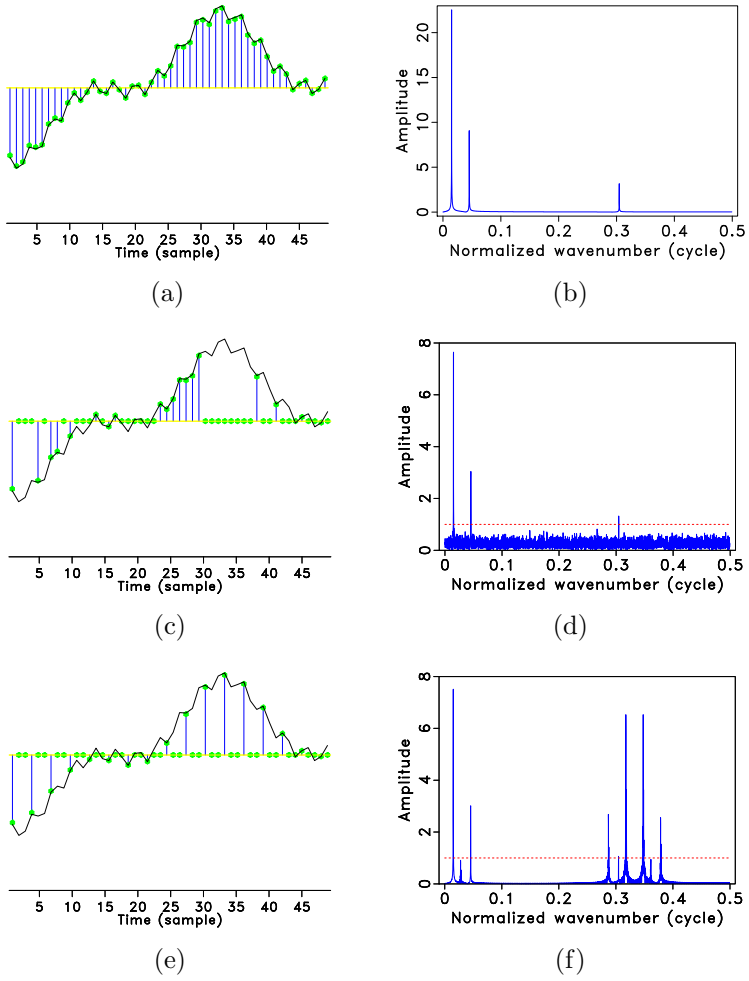
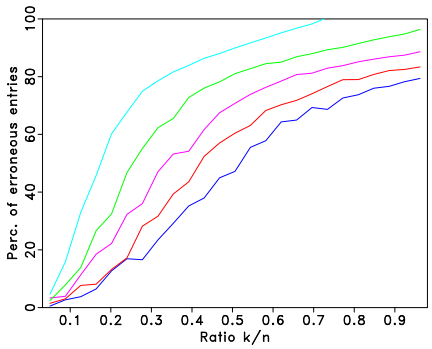
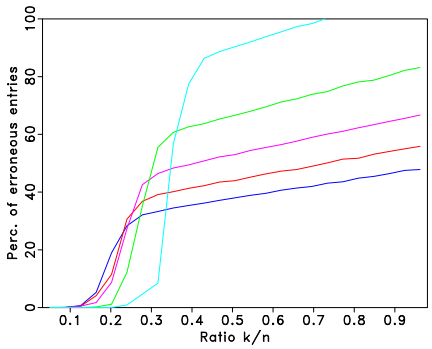


Figure 1: Different (sub)sampling schemes and their imprint in the Fourier domain for a signal that is the superposition of three cosine functions. Signal **(a)** regularly sampled above Nyquist rate, **(c)** randomly three-fold undersampled according to a discrete uniform distribution, and **(e)** regularly three-fold undersampled. The respective amplitude spectra are plotted in **(b)**, **(d)** and **(f)**. Unlike aliases, the subsampling artifacts due to random subsampling can easily be removed using a standard denoising technique, e.g., nonlinear thresholding (dashed line), effectively recovering the original signal. (adapted from (Hennenfent and Herrmann, 2008))

error not measured in the ℓ_2 sense.



(a)



(b)

Figure 2: Averaged recovery error percentages for a k-sparse Fourier vector reconstructed from n time samples taken **(a)** regularly and **(b)** uniformly randomly. In each plot, the curves from top to bottom correspond to a subsampling factor ranging from two to six. (adapted from Hennenfent and Herrmann (2008))

Impulsive versus randomized incoherent sources

To underline the importance of randomization in sampling, let us also study the quintessential seismic inversion problem known as "spiky deconvolution" (see Santosa and Symes, 1986, for a mathematical treatment of this topic). The objective of spiky deconvolution is to estimate a high-resolution impulse train from data that is given by the convolution of a spike train with a known source function. Unfortunately, this method relies on accurate information on the source wavelet, which is generally unavailable, and therefore requires the solution of a blind deconvolution problem involving estimation of the source and the reflectivity. For new developments on this topic, we refer to recent work by Lin and Herrmann (2009b), who estimate the surface-free Green's function and source function by solving a bi-convex optimization problem.

Again compressive sensing offers interesting insights, an observation also made by

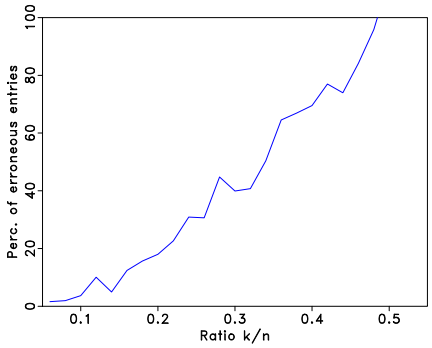
Mallat (2009). In this deconvolution problem, the roles of the sampling and the sparsifying transform are reversed, and the deconvolved reflectivity is found by calculating the vector with the smallest one-norm that, after convolution with the known source function, explains the data. Following Mallat (2009), let us consider two scenarios where we take roughly the same number of Fourier-domain measurements—i.e., we use source functions that have the same number of non-zero Fourier coefficients.

In the first scenario, we consider a seismic experiment where the observed reflectivity is given by the convolution of a spike train with a bandwidth limited impulsive source, say a Ricker wavelet. The Fourier spectrum of a 20 Hz central-frequency Ricker wavelet approximately corresponds to a "sampling" of 54 frequencies contiguous over the band of the Ricker wavelet (this corresponds to a sampling ratio of 0.09 for a signal length of 600). According to CS, this type of sampling is unfavorable. In the second scenario, we replace the impulsive Ricker wavelet with the same number of samples, but now taken uniformly random across the entire spectrum. In the time domain, this second now random sampling corresponds to a source function that behaves as a realization of white Gaussian noise. Thus, the source is no longer impulsive (see also Lin and Herrmann, 2009a, for a discussion of the design of random phase-encoded sweeps).

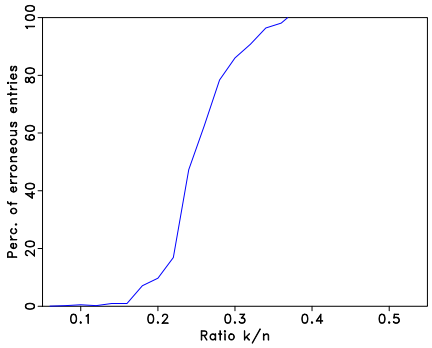
CS predicts improved recovery for the random source. We confirm this by carrying out a second series of experiments where we compare spiky deconvolution for the Ricker and random source functions as a function of the sparsity level (the number of spikes) with the source function and hence the number of measurements fixed. The results of this exercise are summarized in Figure 3 and the results confirm the prediction that the recovery from the Ricker wavelet are problematic and only successful for very small numbers of spikes. Recovery from the randomized sampling, on the other hand, is successful as long as the number of spikes does not exceed 11 ($k/n \approx 0.2$ from Figure 3(b), yielding $k \approx 0.2 \times n \approx 11$ with $n = 54$).

Main contributions

We propose and analyze randomized sampling schemes, termed compressive seismic acquisition. Under specific conditions, these schemes create favorable recovery conditions for seismic wavefield reconstructions that impose transform-domain sparsity in Fourier or Fourier-related domains (see e.g. Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007a; Herrmann et al., 2007; Hennenfent and Herrmann, 2008; Tang et al., 2009). Our contribution is twofold. First, we demonstrate by means of carefully designed numerical experiments on synthetic and real data that compressive sensing can successfully be adapted to seismic acquisition, leading to a new generation of randomized acquisition and processing methodologies where high-resolution wavefields can be sampled and reconstructed with a controllable error. We introduce a number of performance measures that allow us to compare wavefield recoveries based on different sampling schemes and sparsifying transforms. Second, we show that accurate recovery can be accomplished for compressively sampled data volumes sizes



(a)



(b)

Figure 3: Averaged recovery error percentages for $k \in [3, 30]$-sparse physical-domain vectors reconstructed from 54 "Fourier" samples taken **(a)** by convolution with a Ricker wavelet and **(b)** by a wavelet defined by taking random samples across the whole spectrum.

that exceed the size of conventional transform-domain compressed data volumes by a small factor. Because compressive sensing combines transformation and encoding by a single linear encoding step, this technology is directly applicable to seismic acquisition and to dimensionality reduction during processing. We verify this claim by a series of experiments on real data. We also show that the linearity of CS allows us to extend this technology to seismic data processing. In either case, sampling, storage, and processing costs scale with transform-domain sparsity.

Outline

First, we briefly present the key principles of CS, followed by a discussion on how to adapt these principles to the seismic situation. For this purpose, we introduce measures that quantify reconstruction and recovery errors and expresses the overhead that CS imposes. We use these measures to compare the performance of different transform domains and sampling strategies during reconstruction. We use this information to evaluate and apply this new sampling technology towards acquisition and processing of a 2-D seismic line.

BASICS OF COMPRESSIVE SENSING

In this section, we give a brief overview of CS and concise recovery criteria. CS relies on specific properties of the compressive-sensing matrix and the sparsity of the to-be-recovered signal.

Recovery by sparsity-promoting inversion

Consider the following linear forward model for sampling

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^n$ represents the compressively sampled data consisting of n measurements. Suppose that the high-resolution data $\mathbf{f}_0 \in \mathbb{R}^N$, with N the ambient dimension, has a sparse representation $\mathbf{x}_0 \in \mathbb{R}^N$ in some known transform domain. For now, we assume that this representation is the identity basis—i.e., $\mathbf{f}_0 = \mathbf{x}_0$. We will also assume that the data is noise free. According to this model, measurements are defined as inner products between rows of $\mathbf{A}$ and high-resolution data.

The sparse recovery problem involves the reconstruction of the vector $\mathbf{x}_0 \in \mathbb{R}^N$ given incomplete measurements $\mathbf{b} \in \mathbb{R}^n$ with $n \ll N$. This involves the inversion of an underdetermined system of equations defined by the matrix $\mathbf{A} \in \mathbb{R}^{n \times N}$, which represents the sampling operator that collects the acquired samples from the model, $\mathbf{f}_0$.

The main contribution of CS is to come up with conditions on the compressive-sampling matrix $\mathbf{A}$ and the sparse representation $\mathbf{x}_0$ that guarantee recovery by solving a convex sparsity-promoting optimization problem. This sparsity-promoting program leverages sparsity of $\mathbf{x}_0$ and hence overcomes the singular nature of $\mathbf{A}$ when estimating $\mathbf{x}_0$ from $\mathbf{b}$. After sparsity-promoting inversion, the recovered representation for the signal is given by

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{b} = \mathbf{A}\mathbf{x}. \tag{2}$$

In this expression, the symbol $\tilde{}$ represents estimated quantities and the ℓ_1 norm $\|\mathbf{x}\|_1$ is defined as $\|\mathbf{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{N} |x[i]|$, where $x[i]$ is the i^{th} entry of the vector $\mathbf{x}$.

Minimizing the ℓ_1 norm in equation 2 promotes sparsity in $\mathbf{x}$ and the equality constraint ensures that the solution honors the acquired data. Among all possible solutions of the (severely) underdetermined system of linear equations ($n \ll N$) in equation 1, the optimization problem in equation 2 finds a sparse or, under certain conditions, the sparsest (i.e., smallest ℓ_0 norm (Donoho and Huo, 2001)) possible solution that exactly explains the data.

Recovery conditions

The basic idea behind CS (see e.g. Candès et al., 2006; Mallat, 2009) is that recovery is possible and stable as long as any subset S of k columns of the $n \times N$ matrix $\mathbf{A}$—with $k \leq N$ the number of nonzeros in $\mathbf{x}$—behave approximately as an orthogonal basis. In that case, we can find a constant $\hat{\delta}_k$ for which we can bound the energy of the signal from above and below —i.e.,

$$(1 - \hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2 \leq \|\mathbf{A}_S\mathbf{x}_S\|_{\ell_2}^2 \leq (1 + \hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2, \tag{3}$$

where S runs over sets of all possible combinations of columns with the number of columns $|S| \leq k$ (with $|S|$ the cardinality of S). The smaller $\hat{\delta}_k$, the more energy is captured and the more stable the inversion of $\mathbf{A}$ becomes for signals $\mathbf{x}$ with maximally k nonzero entries.

The key factor that bounds the restricted-isometry constants $\hat{\delta}_k > 0$ from above is the mutual coherence amongst the columns of $\mathbf{A}$—i.e.,

$$\hat{\delta}_k \leq (k-1)\mu \tag{4}$$

with

$$\mu = \max_{1 \leq i \neq j \leq N} |\mathbf{a}_i^H \mathbf{a}_j|, \tag{5}$$

where $\mathbf{a}_i$ is the i^{th} column of $\mathbf{A}$ and ^{H} denotes the Hermitian transpose.

Matrices for which $\hat{\delta}_k$ is small contain subsets of k columns that are incoherent. Random matrices, with Gaussian *i.i.d.* entries with variance n^{-1} have this property, whereas deterministic constructions almost always have structure.

For these random Gaussian matrices (there are other possibilities such as Bernouilli or restricted Fourier matrices that accomplish approximately the same behavior, see e.g. Candès et al., 2006; Mallat, 2009), the mutual coherence is small. For this type of CS matrices, it can be proven that Equation 3 holds and Equation 2 recovers $\mathbf{x}_0$'s exactly with high probability as long as this vector is maximally k sparse with

$$k \leq C \cdot \frac{n}{\log_2(N/n)}, \tag{6}$$

where C is a moderately sized constant. This result proves that for large N, recovery of k nonzeros only requires an oversampling ratio of $n/k \approx C \cdot \log_2 N$, as opposed to taking all N measurements.

The above result is profound because it entails an oversampling with a factor $C \cdot \log_2 N$ compared to the number of nonzeros k. Hence, the number of measurements that are required to recover these nonzeros is much smaller than the ambient dimension ($n \ll N$ for large N) of high-resolution data. Similar results hold for compressible instead of strictly sparse signals while measurements can be noisy (Candès et al., 2006; Mallat, 2009). In that case, the recovery error depends on the noise level and on the transform-domain compression rate—i.e., the decay of the magnitude-sorted coefficients.

The results also extend to more general compressive-sampling matrices of the type

$$\mathbf{A} \stackrel{\text{def}}{=} \mathbf{RMS}^H, \tag{7}$$

where $\mathbf{R}$ is a restriction matrix, which selects $n \ll N$ rows randomly, $\mathbf{M}$ is an orthogonal measurement matrix, and $\mathbf{S}^H$ is a synthesis matrix defined by an orthonormal sparsifying transform. The resulting compressive sampling matrix can be expected to perform well as long as the coherence

$$\mu = \max_{1 \leq i \neq j \leq N} |\left(\mathbf{RMs}^i\right)^H \mathbf{RMs}^j| \tag{8}$$

is small. In this expression, $\mathbf{s}^i$ refers to the i^{th} row of the sparsifying basis $\mathbf{S}$.

In summary, according to CS (Candès et al., 2006b; Donoho, 2006b), the solution $\tilde{\mathbf{x}}$ of equation 2 and $\mathbf{x}_0$ coincide when two conditions are met, namely 1) $\mathbf{x}_0$ is sufficiently sparse, i.e., $\mathbf{x}_0$ has few nonzero entries, and 2) the subsampling artifacts are incoherent, which is a direct consequence of measurements with a matrix whose action mimics that of a Gaussian matrix. The first condition requires that the energy of $\mathbf{f}_0$ is well concentrated in the sparsifying domain. The second condition of incoherence requires understanding of interactions between the sparsifying transform $\mathbf{S}$, the measurement basis, $\mathbf{M}$, and the restriction operator $\mathbf{R}$.

Intuitively, successful recovery requires that the artifacts introduced by subsampling the original signal $\mathbf{f}_0$ are not sparse in the transformed domain. When this condition is not met, sparsity alone is no longer an effective prior to solve the recovery

problem. Although qualitative in nature, the second condition provides a fundamental insight in choosing subsampling schemes that favor recovery by sparsity-promoting inversion.

Unfortunately, most rigorous results from CS, except for work by Rauhut et al. (2008), are valid for orthonormal measurement and sparsity bases only and the computation of the recovery conditions for realistically sized seismic problems remains computational prohibitive. To overcome these important shortcomings, we will in the next section introduce a number of practical and computable performance measures that allow us to design and compare different compressive-seismic acquisition strategies.

COMPRESSIVE-SENSING DESIGN AND ASSESSMENT

As we have seen, the machinery that supports sparse recovery from incomplete data depends on specific properties of the compressive-sensing matrix. Consequently, CS can not be applied to arbitrary linear inversion problems. To the contrary, the success of CS hinges on the design of acquisition strategies that are (physically and/or) practically feasible and that lead to favorable conditions for sparse recovery. The CS matrix needs to both be realizable and behave as a Gaussian matrix. The following key components need to be in place:

1. **a sparsifying signal representation** that exploits the signal's structure by mapping the energy into a small number of significant transform-domain coefficients. The smaller the number of significant coefficients, the better the recovery;

2. **sparse recovery by transform-domain one-norm minimization** that is able to handle large system sizes. The fewer the number of matrix-vector evaluations, the faster and more practically feasible the wavefield reconstruction;

3. **randomized seismic acquisition** that breaks coherent interferences induced by deterministic subsampling schemes. Randomization renders subsampling related artifacts—including aliases and simultaneous source crosstalk—harmless by turning these artifacts into incoherent Gaussian noise;

Given the complexity of seismic data in high dimensions and field practicalities of seismic acquisition, the mathematical formulation of CS outlined in the previous section does not readily apply to seismic exploration. Therefore, we will focus specifically on the design of source subsampling schemes that favor recovery and on the selection of the appropriate sparsifying transform. Because theoretical results are mostly lacking, we will guide ourselves by numerical experiments that are designed to measure recovery performance.

Sampling and sparse recovery

During seismic data acquisition, data volumes are collected that represent discretizations of analog finite-energy wavefields in two or more dimensions including time. So, we are concerned with the acquisition of an analog spatio-temporal wavefield $\bar{f}(t,x) \in L_2((0,T] \times [-L,L])$ with time T in the order of seconds and length L in the order of kilometers. The sampling intervals are of the order of milliseconds and of meters.

Mathematically, high-resolution sampling of a continuous seismic wavefield can be written as

$$f[q] = \overline{\Phi}_s \bar{f}[q], \quad q = 0 \cdots N-1, \tag{9}$$

where N is the total number of samples, and where $\overline{\Phi}_s$ is a linear high-resolution analog-to-digital converter (Mallat, 2009). After inclusion of the analog-to-digital converter, which in our case models the directivity of geophones or sources, compressively sampled data can be modeled via

$$b[q] = \mathbf{RM}\overline{\Phi}_s \bar{f}[q], \quad q = 0 \cdots n-1. \tag{10}$$

After organizing these high-resolution samples into the vector $\mathbf{f} \stackrel{\text{def}}{=} \left\{f[q]\right\}_{q=0\cdots N-1} \in \mathbb{R}^N$ and absorbing the converter into the definition of the measurement matrix, we arrive at the following expression for compressively-sampled measurements:

$$\mathbf{b} = \mathbf{RMf} \in \mathbb{R}^n \quad \text{with} \quad n \ll N. \tag{11}$$

In this formulation, incomplete seismic acquisition is modeled by taking inner products between the discrete vector $\mathbf{f}$ and $n \ll N$ randomly selected rows from the measurement matrix $\mathbf{M}$ as dictated by $\mathbf{R}$.

We recover the discretized wavefield $\mathbf{f}$ by inverting the compressive-sampling matrix

$$\mathbf{A} \stackrel{\text{def}}{=} \overbrace{\mathbf{R}}^{\text{restriction}} \underbrace{\mathbf{M}}_{\text{measurement}} \overbrace{\mathbf{S}^H}^{\text{synthesis}} \tag{12}$$

with the sparsity-promoting program:

$$\widetilde{\mathbf{f}} = \mathbf{S}^H \widetilde{\mathbf{x}} \quad \text{with} \quad \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \stackrel{\text{def}}{=} \sum_{p=0}^{P-1} |x[i]| \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}. \tag{13}$$

This formulation differs from standard compressive sensing because we allow for a wavefield representation that is redundant—i.e., $\mathbf{S} \in \mathbb{C}^{P \times N}$ with $P \geq N$. Aside from results reported by Rauhut et al. (2008), which show that recovery with redundant frames is determined by the RIP constant $\hat{\delta}$ of the restricted sampling and sparsifying matrices that is least favorable, there is no practical algorithm to compute these constants. Therefore, our hope is that the above sparsity-promoting optimization program, which finds amongst all possible transform-domain vectors the vector $\widetilde{\mathbf{x}} \in \mathbb{R}^P$ that has the smallest ℓ_1-norm, recovers high-resolution data $\widetilde{\mathbf{f}} \in \mathbb{R}^N$.

Seismic wavefield representations

One of the key ideas of CS is leveraging structure within signals to reduce sampling. Typically, structure translates into transform-domains that concentrate the signal's energy in an as few as possible significant coefficients. The size of seismic data volumes, along with the complexity of its high-dimensional and highly directional wavefront-like features, makes it difficult to find a transform that accomplishes this task.

To meet this challenge, we only consider transforms that are fast (at the most $N \log N$ with N the number of samples), multiscale (splitting the Fourier spectrum into dyadic frequency bands), and multidirectional (splitting Fourier spectrum into second dyadic angular wedges). For reference, we also include separable 2-D wavelets in our study. We define this wavelet transform as the Kronecker product (denoted by the symbol $\otimes$) of two 1D wavelet transforms: $\mathbf{W} = \mathbf{W}_1 \otimes \mathbf{W}_1$ with $\mathbf{W}_1$ the 1D wavelet-transform matrix.

Separable versus non-separable transforms

There exists numerous signal representations that decompose a multi-dimensional signal with respect to directional and localized elements. Because we are concerned with finding the appropriate representation for seismic wavefields, we limit our search to non-separable curvelets (Candès et al., 2006a) and wave atoms (Demanet and Ying, 2007). This choice is motivated by the fact that the elements of these transforms behave approximately as high-frequency asymptotic eigenfunctions of wave equations (see e.g. Smith, 1998; Candès and Demanet, 2005; Candès et al., 2006a; Herrmann et al., 2008b). This property makes these two representations particularly well suited for our task of representing seismic data parsimoniously.

Unlike wavelets, which compose curved wavefronts into a superposition of multiscale "fat dots" with limited directionality, curvelets and wave atoms compose wavefields as a superposition of highly anisotropic localized and multiscale waveforms, which obey a so-called parabolic scaling principle. For curvelets in the physical domain, this principle translates into an support that is anisotropic with length $\sim$ width^2. At the fine scales, this scaling leads to curvelets that become increasingly anisotropic, i.e., needle-like. Each dyadic frequency band is split into a number of overlapping angular wedges that double in every other dyadic scale. This partitioning results in increased directionality at the fine scales. This construction makes curvelets well adapted to data with impulsive wavefront-like features. Wave atoms, on the other hand, are anisotropic because their wavelength depends quadratically on their width—i.e., wavelength $\sim$ width^2. By construction, wave atoms are more appropriate for data with oscillatory patterns. Because seismic data sits somewhere between these two extremes, we include both transforms in our study.

Approximation error

Like many other naturally occurring phenomena, seismic wavefields do not permit strictly sparse representations—i.e., representations where many coefficients are strictly zero. However, for an appropriately chosen representation magnitude-sorted transform-domain coefficients often decay rapidly–i.e., the magnitude of the j^{th} largest coefficient is $\mathcal{O}(j^{-s})$ with $s \geq 1/2$. For orthonormal bases, this decay rate is directly linked to the decay of the nonlinear approximation error (see e.g. Mallat, 2009). This error expresses the difference between the discretized wavefield and its reconstruction from the largest k transform-domain coefficients—expressed by

$$\sigma(k) = \|\mathbf{f} - \mathbf{f}_k\| = \mathcal{O}(k^{1/2-s}), \tag{14}$$

with $\mathbf{f}_k$ the reconstruction from the largest k - coefficients. Notice that this error does not account for discretization errors (cf. Equation 9), which we ignore.

Unfortunately, this relationship between the decay rates of the magnitude-sorted coefficients and the decay rate of the nonlinear approximation error does not hold for redundant transforms. Another complicating factor is that expansions with respect to this type of signal representations are not unique. This means that there are many coefficient sequences that explain the data $\mathbf{f}$ making them less sparse. For instance, analysis by the curvelet transform of a single curvelet does not produce a single non-zero entry in the curvelet coefficient vector.

To address this issue, we use an alternative definition for the nonlinear approximation error, which is based on the solution of a sparsity-promoting program. With this definition, the k-term nonlinear-approximation error is computed by taking the $k-$largest coefficients from the vector that solves

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{S}^H\mathbf{x} = \mathbf{f}. \tag{15}$$

Because this vector is obtained by inverting the synthesis operator $\mathbf{S}^H$ with a sparsity-promoting program, this vector is always sparser than the vector obtained by applying the analysis operator $\mathbf{S}$ directly.

To account for different redundancies in the transforms, we study signal-to-noise ratios (SNRs) as a function of the sparsity ratio $\rho = k/P$ (with $P = N$ for orthonormal bases) defined as

$$\text{SNR}(\rho) = -20 \log \frac{\|\mathbf{f} - \mathbf{f}_\rho\|}{\|\mathbf{f}\|}. \tag{16}$$

The smaller this ratio, the more coefficients we ignore, the sparser the transform-coefficient vector becomes, which in turn leads to a smaller SNR. This latter number expresses the ratio of the energy difference between the original signal and its reconstruction and over the energy of original signal. In our study, we include $\mathbf{f}_\rho$ that are derived from either the analysis coefficients or from the synthesis coefficients. The latter coefficients are solutions of the above sparsity-promoting program (Equation 15).

Empirical approximation errors

The above definition gives us a metric to compare recovery SNRs of seismic data for wavelets, curvelets, and wave atoms. We make this comparison on a common-receiver gather (Figure 4) extracted from a Gulf of Suez dataset. Because the current implementations of wave atoms (Demanet and Ying, 2007) only support data that is square, we padded the 178 traces with zeros to 1024 traces. The temporal and spatial sampling interval of the high-resolution data are 0.004s and 25m, respectively. Because this zero-padding biases the ρ, we apply a correction.

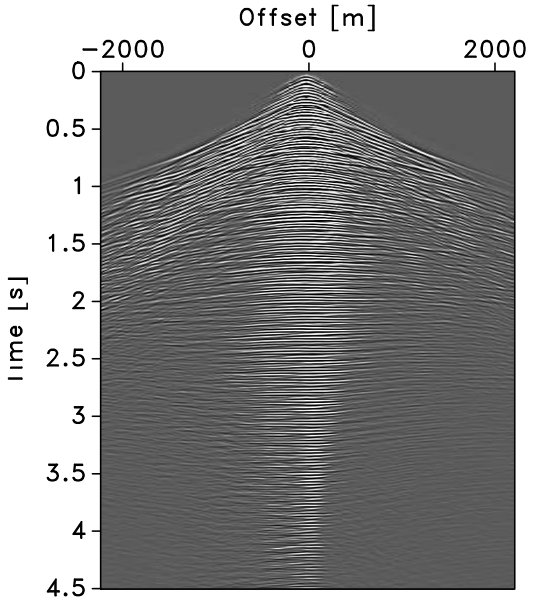


Figure 4: Real common-receiver gather from Gulf of a Suez data set.

Our results are summarized in Figure 5 and they clearly show that curvelets lead to rapid improvements in SNR as the sparsity ratio increases. This effect is most pronounced for synthesis coefficients, benefiting remarkably from sparsity promotion. By comparison, wave atoms benefit not as much, and wavelet even less. This behavior is consistent with the overcompleteness of these transforms, the curvelet transform matrix has the largest redundancy (a factor of about eight in 2-D) and is therefore the tallest. Wave atoms only have a redundancy of two and wavelets are orthogonal. Since our method is based on sparse recovery, this experiment suggests that sparse recovery from subsampling would potentially benefit most from curvelets. However, this is not the only factor that determines the performance of our compressive-sampling scheme.

Subsampling of shots

Aside from obtaining good reconstructions from small compression ratios, breaking the periodicity of coherent sampling is paramount to the success of sparse recovery—whether this involves selection of subsets of sources or the design of incoherent

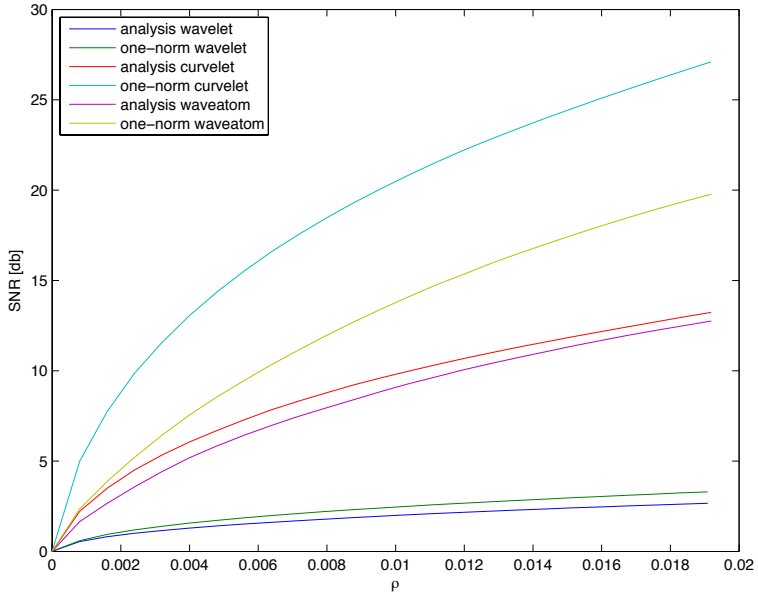


Figure 5: Signal-to-noise ratios (SNRs) for the nonlinear approximation errors of the common-receiver gather plotted in Figure 4. The SNRs are plotted as a function of the sparsity ratio $\rho \in (0,\ 0.02]$. The plots include curves for the errors obtained from the analysis and one-norm minimized synthesis coefficients. Notice the significant improvement in SNRs for the synthesis coefficients obtained by solving Equation 15.

simultaneous-source experiments. To underline the importance of maximizing incoherence in seismic acquisition, we conduct two experiments where common-source gathers are recovered from subsets of sequential and simultaneous-source experiments. To make useful comparisons, we keep for each survey the number of source experiments, and hence the size of the collected data volumes, the same.

Coherent versus incoherent sampling

Mathematically, sequential and simultaneous acquisition only differ in the definition of the measurement basis. For sequential-source acquisition, this sampling matrix is given by the Kronecker product of two identity bases—i.e., $\mathbf{I} \stackrel{\text{def}}{=} \mathbf{I}_{N_s} \otimes \mathbf{I}_{N_t}$, which is a $N \times N$ identity matrix with $N = N_t \times N_s$, the product of the number of time samples N_t and the number of shots N_s. For simultaneous acquisition, where all sources fire simultaneously, this matrix is given by $\mathbf{M} \stackrel{\text{def}}{=} \mathbf{G}_{N_s} \otimes \mathbf{I}_{N_t}$ with $\mathbf{G}_{N_s}$ a $N_s \times N_s$ Gaussian matrix with *i.i.d.* entries. In both cases, we use a restriction operator $\mathbf{R} \stackrel{\text{def}}{=} \mathbf{R}_{n_s} \otimes \mathbf{I}_{N_t}$ to model the collection of incomplete data by reducing the number of shots to $n_s \ll N_s$. This restriction acts on the source coordinate only.

Roughly speaking, CS predicts superior recovery for compressive-sampling matrices with smaller coherence. According to Equation 8, this coherence depends on the

interplay between the restriction, measurement, and synthesis matrices. To make a fair comparison, we keep the restriction matrix the same and study the effect of having measurement matrices that are either given by the identity or by a random Gaussian matrix. Physically, the first CS experiment corresponds to surveys with sequential shots missing. The second CS experiment corresponds to simultaneous-source experiments with simultaneous source experiments missing. Examples of both measurements for the real common-receiver gather of Figure 4 are plotted in Figure 6(a) and 6(b), respectively. Both data sets have 50% of the original size. Remember that the horizontal axes in the simultaneous experiment no longer has a physical meaning. Notice also that there is no observable coherent crosstalk amongst the simultaneous sources.

Multiplication of orthonormal sparsifying bases by random measurement matrices turns into random matrices with a small mutual coherence amongst the columns. This property also holds (but only approximately) for redundant signal representations with synthesis matrices that are wide and have columns that are linearly dependent. This suggests improved performance using random incoherent measurement matrices. To verify this statement empirically, we compare sparse recoveries with Equation 13 from data plotted in Figures 6(a) and 6(b).

Despite the fact that simultaneous acquisition with all sources firing simultaneously may not be easily implementable in practice[1], this approach has been applied successfully to reduce simulation and imaging costs (Herrmann et al., 2009b; Herrmann, 2009; Lin and Herrmann, 2009a,b). In the "eyeball norm", the recovery from the simultaneously data is as expected clearly superior (cf. Figures 6(c) and 6(d)). The difference plots (cf. Figures 6(e) and 6(f)) confirm this observation and show very little coherent signal loss for the recovery from simultaneous data. This is consistent with CS, which predicts improved performance for sampling schemes that are more incoherent. Because this qualitative statement depends on the interplay between the sampling and the sparsifying transform, we conduct an extensive series of experiments to get a better idea on the performance of these two different sampling schemes for different sparsifying transforms. We postpone our analysis of the quantitative behavior of the recovery SNRs to after that discussion.

Sparse recovery errors

The examples of the previous section clearly illustrate that randomized sampling is important, and that randomized simultaneous acquisition leads to better recovery compared to randomized subsampling of sequential sources. To establish this observation more rigorously, we calculate estimates for the recovery error as a function of the sampling ratio $\delta = n/N$ by conducting a series of 25 controlled recovery experiments. For each $\delta \in [0.2, 0.8]$, we generate 25 realizations of the randomized compressive-sampling matrix. Applying these matrices to our common-receiver gather (Figure 4)

[1]Although one can easily imagine a procedure in the field where a "supershot" is created by some stacking procedure.

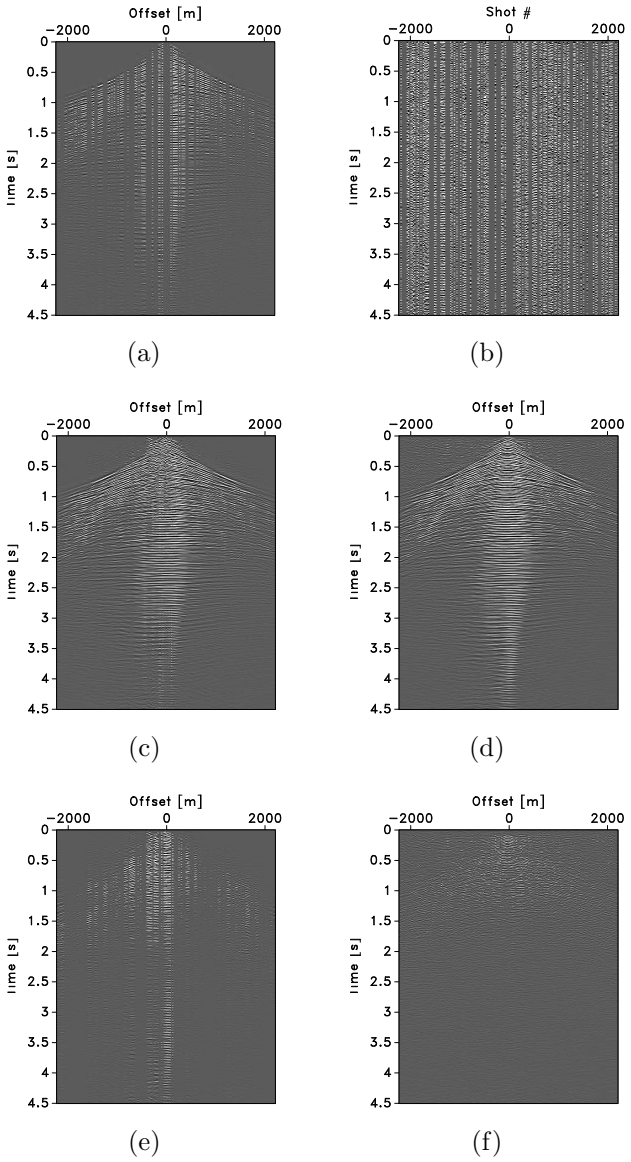


Figure 6: Recovery from a compressively-sampled common receiver gather with 50% ($\delta = 0.5$) of the sources missing. **(a)** Receiver gather with sequential shots selected uniformly at random. **(b)** The same but for random simultaneous shots. **(c)** Recovery from incomplete data in **(a)**. **(d)** The same but now for the data in **(b)**.**(e)** Difference plot between the data in Figure 4 and the recovery in **(c)**. **(f)** The same but now for recovery from simultaneous data in **(b)**. Notice the remarkable improvement in the recovery from simultaneous data.

produces 25 different data sets that are subsequently used as input to sparse recovery with wavelets, curvelets, and wave atoms. For each realization, we calculate the $\mathrm{SNR}(\delta)$ with

$$\mathrm{SNR}(\delta) = -20 \log \frac{\|\mathbf{f} - \widetilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|}, \tag{17}$$

where

$$\widetilde{\mathbf{f}}_\delta = \mathbf{S}^H \widetilde{\mathbf{x}}_\delta \quad \text{and} \quad \widetilde{\mathbf{x}}_\delta = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}_\delta \mathbf{x} = \mathbf{b}.$$

For each experiment, the recovery of $\widetilde{\mathbf{f}}_\delta$ is calculated by solving this optimization problem for 25 different realizations of $\mathbf{A}_\delta$ with $\mathbf{A}_\delta \stackrel{\text{def}}{=} \mathbf{R}_\delta \mathbf{M}_\delta \mathbf{S}^H$, where $\mathbf{R}_\delta \stackrel{\text{def}}{=} \mathbf{R}_{n_s} \otimes \mathbf{I}_{N_t}$ with $\delta = n_s/N_s$. For each simultaneous experiment, we also generate different realizations of the measurement matrix $\mathbf{M} \stackrel{\text{def}}{=} \mathbf{G}_{N_s} \otimes \mathbf{I}_{N_t}$.

From these randomly selected experiments, we calculate the average SNRs for the recovery error, $\overline{\mathrm{SNR}}(\delta)$, including its standard deviation. By selecting δ evenly on the interval $\delta \in [0.2, 0.8]$, we obtain reasonable reliable estimates with error bars. Results of this exercise are summarized in Figure 7. From these plots it becomes immediately clear that simultaneous acquisition greatly improve recovery for all three transforms. Not only are the SNRs better, but the spread in SNRs amongst the different reconstructions is also much less, which is important for quality assurance. The plots validate CS, which predicts improved recovery for increased sampling ratios. Although somewhat less pronounced as for the approximation SNRs in Figure 5, our results again show superior performance for curvelets compared to wave atoms and wavelets. This observation is consistent with our earlier empirical findings.

Empirical oversampling ratios

The main advantage of CS is that it provides access to the largest, and hence most significant, transform-domain coefficients without the necessity of conducting a complete high-resolution survey followed by the computation of the k-term nonlinear approximation. Conversely, sparse recovery from incoherent samples requires significantly less samples, but this reduction goes at the expense of conducting incoherent sampling in conjunction with the solution of a computationally intensive large-scale sparse recovery problem. Therefore, the key factor that establishes CS is the sparsity ratio ρ that is required to recover wavefields with errors that do not exceed a predetermined nonlinear approximation error (cf. Equation 16). The latter sets the fraction of largest coefficients that needs to be recovered to meet a preset minimal SNR for reconstruction.

Motivated by Mallat (2009), we introduce the oversampling ratio $\delta/\rho \geq 1$. For a given δ, we obtain a target SNR from $\overline{\mathrm{SNR}}(\delta)$. Then, we find the smallest ρ for which the nonlinear recovery SNR is greater or equal to $\overline{\mathrm{SNR}}(\delta)$. Thus, the oversampling ratio $\delta/\rho \geq 1$ expresses the sampling overhead required by compressive sensing. This measure helps us to determine the performance of our CS scheme numerically. The

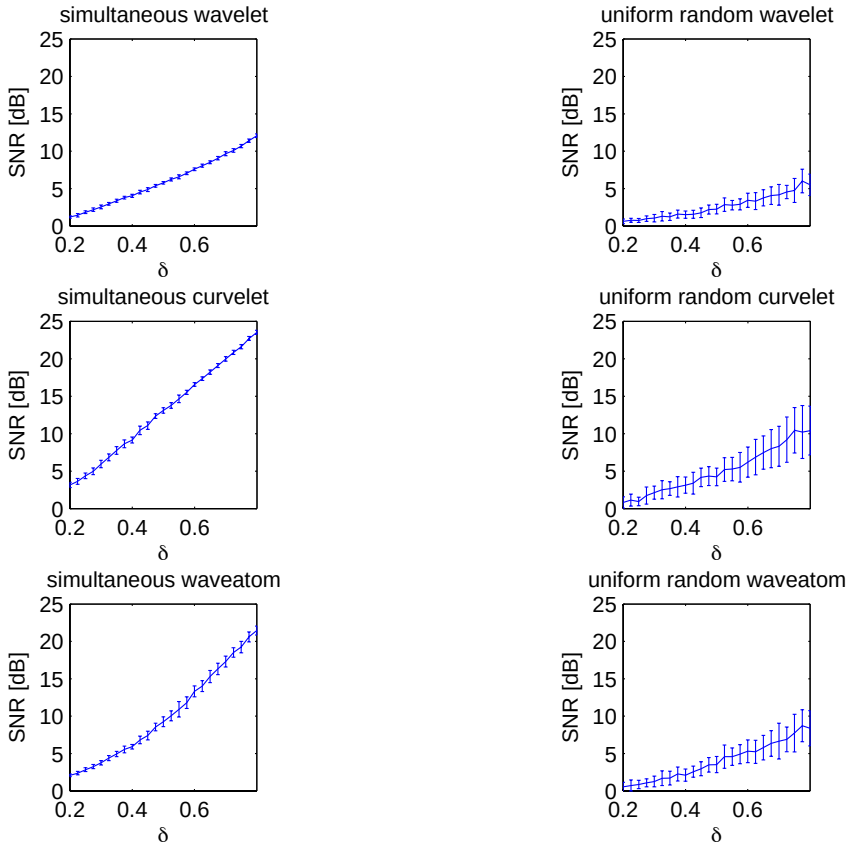


Figure 7: SNRs (cf. Equation 17) for nonlinear sparsity-promoting recovery from compressively sampled data with $20\% - 80\%$ of the sources missing ($\delta \in [0.2, 0.8]$). The results summarize 25 experiments for 25 different values of $\delta \in [0.2, 0.8]$. The plots include estimates for the standard deviations. From these results, it is clear that simultaneous acquisition (results in the left column) is more conducive to sparsity-promoting recovery. Curvelet-based recovery seems to work best, especially towards high percentages of data missing.

smaller this ratio, the smaller the overhead and the more economically favorable this technology becomes compared to conventional sampling schemes.

We can compute this oversampling ratio for each δ by finding the sparsity ratio ρ for which the recovery $\overline{\text{SNR}}(\delta)$ is smaller or equal to the nonlinear approximation $\text{SNR}(\rho)$, i.e., we calculate for each $\delta \in [0.2, 0.8]$

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\text{SNR}}(\delta) \leq \text{SNR}(\tilde{\rho})\}. \tag{18}$$

When the sampling ratio approaches one from below ($\delta \to 1$), the data becomes more and more sampled leading to smaller and smaller recovery errors. To match this decreasing error, the sparsity ratio $\rho \to 1$ and consequently we can expect this oversampling ratio to go to one, $\delta/\rho \to 1$.

Remember that in the CS paradigm, acquisition costs grow with the permissible recovery SNR that determines the sparsity ratio. Conversely, the costs of conventional sampling grow with the size of the sampling grid irrespective of the transform-domain compressibility of the wavefield, which in higher dimensions proves to be a major difficulty.

The numerical results of our experiments are summarized in Figure 8. Our calculations use empirical SNRs for both the approximation errors of the synthesis coefficients as a function of ρ and the recovery errors as a function of δ. The estimated curves lead to the following observations. First, as the sampling ratio increases the oversampling ratio decreases, which can be understood because the recovery becomes easier and more accurate. Second, recoveries from simultaneous data have significantly less overhead and curvelets outperform wave atoms, which in turn perform significantly better than wavelets. All curves converge to the lower limit (depicted by the dashed line) as $\delta \to 1$. Because of the large errorbars in the recovery SNRs (cf. Figure 7), the results for the recovery from missing sequential sources are less clear. However, general trends predicted by CS are also observable for this type of acquisition, but the performance is significantly worse than for recovery with simultaneous sources. Finally, the observed oversampling ratios are reasonable for both curvelet and wave atoms.

SYNTHETIC CASE STUDIES

We have seen that coherent subsampling-related interferences antagonize wavefield recovery by sparsity promotion. Well-designed randomized subsampling schemes—through jittered sampling (Hennenfent and Herrmann, 2008) of e.g., source positions or through randomly phase-encoded simultaneous sources (Beasley et al., 1998; Beasley, 2008; Berkhout, 2008; Neelamani et al., 2008; Herrmann et al., 2009b)—lead to manageable subsampling artifacts that are incoherent with a noise level that depends on the degree of subsampling: the more subsampled the higher the noise level. As we have seen, herein lies an unique opportunity as long as we are able to separate subsampling interferences from the desired signal. This is where transform-domain

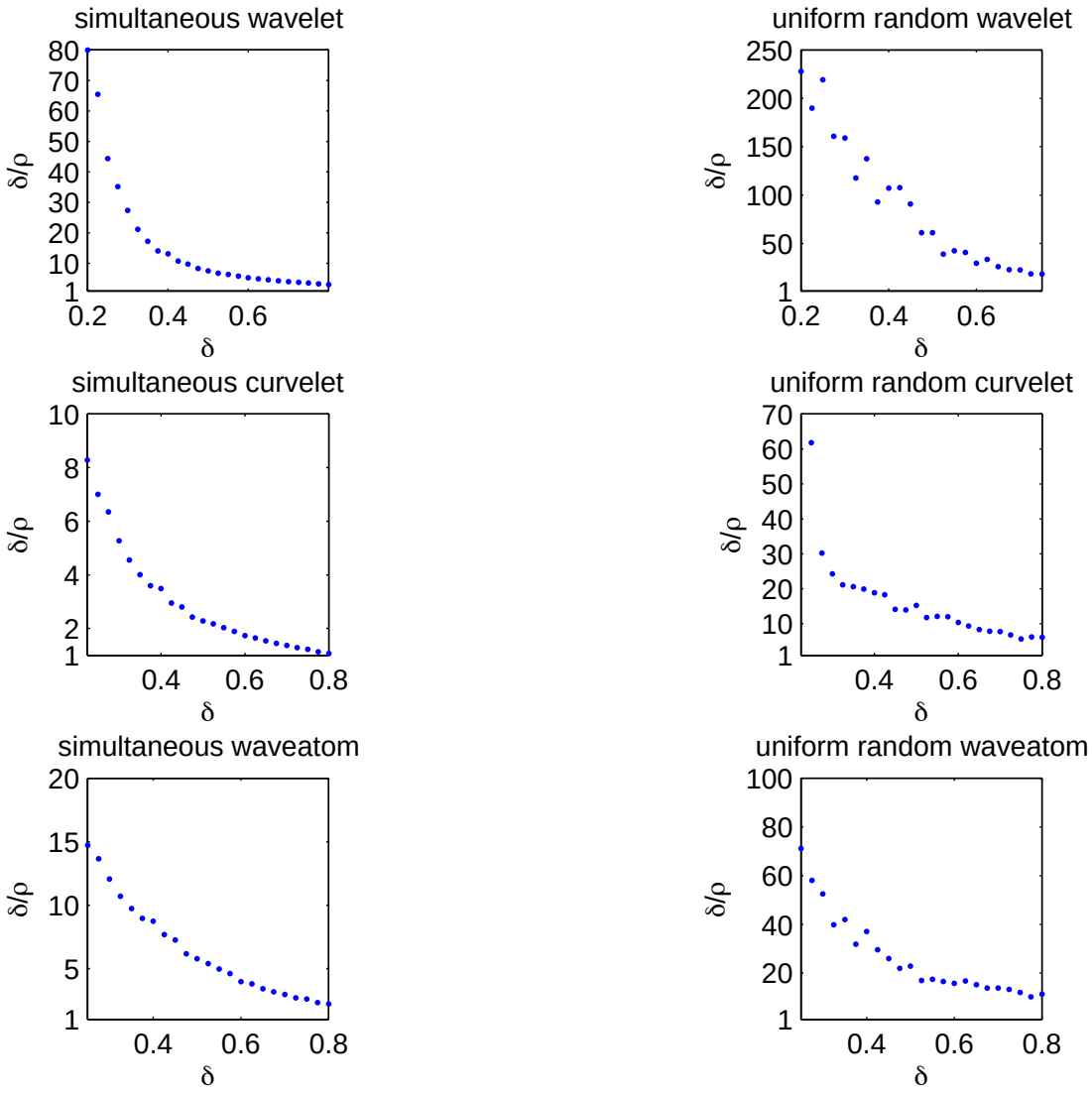


Figure 8: Oversampling ratio δ/ρ as a function of the sampling ratio δ (cf. Equation 18) for sequential- and simultaneous-source experiments. As expected, the overhead is smallest for simultaneous acquisition and curvelet-based recovery.

sparsity entered into the equation, because the sparser we represent our desired wavefield, the better we can remove the noise and correct the amplitudes.

Transform-domain sparsity, however, is not the only technique at our disposal to recover severely subsampled data. Recent developments in curvelet-based wavefield inversion (Herrmann et al., 2008a; Herrmann and Wang, 2008) have shown that transform-based sparsity-promoting recovery benefits from the inclusion of operators that model wave propagation. Inverting these operators, as part of a compressive sensing procedure, focuses the wavefield's energy and this potentially enhances transform-domain sparsity. The inclusion of a multiple-generating surface also improves "the mixing", which make the measurement matrix act more like a Gaussian matrix.

To shed more light on the above statements in the context of practical wavefield recovery let us formulate two concrete questions:

> *For a given sampling ratio, is it better to use periodic or jittered subsamplings of sequential shots or is it better to use a subset of simultaneous-shot experiments?*

and a question related to seismic processing

> *Is it better to first reconstruct a wavefield from simultaneous data and then estimate the surface-free impulse response or is it better to estimate this impulse response directly from simultaneous data?*

In the next two sections we aim to answer these questions empirically on synthetic data obtained by finite differences on an acoustic model that includes salt, surrounded by sedimentary layers and a water bottom that is not completely flat. To answer above questions, we conduct two recovery experiments of high-resolution 2-D seismic lines from shot subsamplings. To exploit the multi-dimensionality of this problem, we use the 2-D discrete curvelet transform along the source and receiver coordinates, and the discrete wavelet transform along the remaining time coordinate:

$$\mathbf{S} \stackrel{\text{def}}{=} \mathbf{C}_2 \otimes \mathbf{W}. \tag{19}$$

The first example is a straight-forward extension of our work on common-receiver gathers and is aimed to firmly establish the findings of the previous sections for the more realistic setting of a seismic 2-D line. The second example takes CS a step further by leveraging recent work on the estimation of primaries by sparse inversion (EPSI-van Groenestijn and Verschuur, 2009; Lin and Herrmann, 2009b), which involves the inversion of the following monochromatic relationship:

$$\underbrace{\widehat{\mathbf{G}}}_{\text{surface-free impulse response}} \overbrace{\left[\widehat{\mathbf{Q}} - \widehat{\mathbf{P}}\right]}^{\text{downgoing wavefield}} \approx \overbrace{\widehat{\mathbf{P}}}^{\text{upgoing wavefield}} \tag{20}$$

between the known source function $\widehat{\mathbf{Q}}$, the known upgoing wavefield $\widehat{\mathbf{P}}$, and the unknown surface-free data $\widehat{\mathbf{G}}$ (for details see e.g. Lin and Herrmann, 2009b). Again, to make a fair comparison between these sparse recovery schemes, we keep the number of measurements the same in all experiments.

Sparse recovery

To verify the prediction of CS, which states that wavefield reconstructions improve for CS matrices that behave more like Gaussian matrices, we compare curvelet-based recoveries from surveys that have 50 % shots missing. We compare wavefield reconstructions from deterministic periodic and randomized jittered subsampled sequential source locations and subsampled randomly phase-encoded simultaneous source experiments (Hennenfent and Herrmann, 2008; Herrmann et al., 2009b). The first two acquisition scenarios differ in the definition of the restriction matrix while measurements with phase-encoded sources are obtained by replacing the identity matrix with the measurement matrix defined by (Romberg, 2009)

$$\mathbf{M} \stackrel{\text{def}}{=} \left[\mathbf{I} \otimes \mathcal{F}_s^* \,\text{diag}\left(e^{\hat{i}\boldsymbol{\theta}}\right) \mathcal{F}_s \otimes \mathbf{I}\right], \tag{21}$$

with $\mathcal{F}_s$ the Fourier transform along the source coordinate and $\theta \in [0, \pi]$ random phase rotations uniformly drawn on the interval. This measurement matrix turns sequential sources into simultaneous sources. As before, the restriction operator selects the subset of simultaneous sources.

Results from sparsity-promoting recovery for 50 % of the sources missing are included in Figure 9. Comparison between the results for regular, jittered, and simultaneous shots shows again drastic improvements in recovery quality as we move from regular subsampled (8.9 dB), to randomly jittered sequential sources (10.9 dB) to randomized simultaneous sources (16.1 dB). These findings clearly confirm the importance of randomization in the collection of seismic data. This example also nicely illustrates that randomization of the source locations by itself is not necessarily optimal and that a lot is to be gained by designing randomized simultaneous-source acquisitions.

Estimation of primaries

We conclude our discussion by showing an example where we include more information on the physics—e.g., information on the surface-free boundary condition and the source function—in the definition of the CS matrix. Consider an experiment where we estimate the primaries from simultaneously acquired data with 50 % of the simultaneous shots missing. A superficial juxtaposition of the recovery (Figure 10(b)) with the fully sampled wavefield (Figure 10(a)) suggest a satisfactory recovery. However, if we use this recovery as input to primary estimation the result (Figure 10(c)) is

inadequate because lots of surface-related multiple energy remains present after the inversion. If, on the other hand, we invert Equation 20 for simultaneous data (see e.g. van Groenestijn and Verschuur, 2009; Lin and Herrmann, 2009b), we find an estimate for the primaries (Figure 10(d)) that is greatly improved. In this case, there is little surface-related multiple energy remaining and the imprint of the source function has mostly been removed. This result is important because it not only shows that primaries can be estimated directly from simultaneously acquired data, but this result also shows that data processing can be carried out in the compressively sampled domain, which reduces the cost of processing.

The explanation for these improvements is threefold. First, Equation 20 encodes the surface-free impulse response redundantly in the total data. This means that as long as the surface-related multiples are within the available time window of our input data, sparse inversion maps these multiples back to the surface-free impulse response. The presence of these multiple copies helps recovery. Second, the above relation mixes the primaries and this improves the incoherence. Third, because surface-related multiples are mapped to primaries, the number of events is reduced and this enhances transform-domain sparsity. It is clear that all these factors contribute to improvement of our estimate of surface-free data. This example also shows how delicate relations of the type in Equation 20 are, even in cases where the free-surface boundary condition is known exactly. This sensitivity can be used to estimate the source-function using a bi-convex optimization problem (Lin and Herrmann, 2009b) on which we will report elsewhere in more detail.

DISCUSSION

The presented results illustrate that we are at the cusp of exciting new developments where acquisition and processing workflows are no longer impeded by a fear of creating coherent subsampling related artifacts. Instead, we arrive at a formulation with control over these artifacts. We accomplish this achievement by applying the following three new design principles:

1. **randomize**—break coherent aliases by introducing randomness, e.g. by designing randomly perturbed acquisition grids, or by designing randomized simultaneous sources.

2. **sparsify**—utilize sparsifying transforms in conjunction with sparsity-promoting programs that separate signal and subsampling artifacts and that restore amplitudes.

3. **focus**–leverage wave physics to focus seismic energy, enhancing transform-domain sparsity and sparse recovery.

EXTENSIONS

The implications of randomized incoherent sampling go far beyond the examples presented in this paper. For instance, our approach is applicable to land acquisition for physically realizable sources (Krohn and Neelamani, 2008; Romberg, 2009) and can be used to compute solutions to wavefield simulations (Herrmann et al., 2009b) and to compute full waveform inversion (Herrmann et al., 2009a) faster. Because randomized sampling is linear (Bobin et al., 2008), wavefield reconstructions and processing can be carried out incrementally as more compressive data becomes available.

CONCLUSIONS

Following ideas from compressive sensing, we made the case that seismic wavefields can be reconstructed with a controllable error from randomized subsamplings. By means of carefully designed numerical experiments on synthetic and real data, we established that compressive sensing can indeed successfully be adapted to seismic data acquisition, leading to a new generation of randomized acquisition and processing methodologies.

With carefully designed experiments and the introduction of performance measures for nonlinear approximation and recovery errors, we established that curvelets perform best in recovery, closely followed by wave atoms, and with wavelets coming in as a distant third, which is consistent with the directional nature of seismic wavefronts. This finding is remarkable for the following reasons: (i) it underlines the importance of sparsity promotion, which offsets the "costs" of redundancy and (ii) is shows that the relative sparsity ratio effectively determines the recovery performance rather than the absolute number of significant coefficients. Our observation of significantly improved recovery for simultaneous-source acquisition also confirms predictions of compressive sensing. Finally, our analysis showed that accurate recoveries are possible from compressively sampled data volumes that exceed the size of conventionally compressed data volumes by only a small factor.

The fact that compressive sensing combines sampling and compression in a single linear encoding step has profound implications for exploration seismology that include: a new randomized sampling paradigm, where the cost of acquisition are no longer dominated by resolution and size of the acquisition area, but by the desired reconstruction error and transform domain sparsity of the wavefield; and a new paradigm for randomized processing and inversion, where dimensionality reductions will allow us to mine high-dimensional data volumes for information in ways, which previously, would have been computationally infeasible.

ACKNOWLEDGMENTS

This paper was written as a follow up to the author's presentation during the "Recent advances and Road Ahead" session at the 79$^{\text{th}}$ annual meeting of the Society of Exploration Geophysicists. I would like to thank the organizers of this session Dave Wilkinson and Masoud Nikravesh for their invitation. I also would like to thank Dries Gisolf and Eric Verschuur of the Delft University of Technology for their hospitality during my sabbatical and Tim Lin and Gilles Hennenfent for preparing some of the figures. This publication was prepared using CurveLab (`curvelet.org`), a toolbox implementing the Fast Discrete Curvelet Transform, WaveAtom a toolbox implementing the Wave Atom transform (`http://www.waveatom.org/`), Madagascar (`rsf.sf.net`), a package for reproducible computational experiments, SPGℓ_1 (`cs.ubc.ca/labs/scl/spgl1`), and SPARCO (`cs.ubc.ca/labs/scl/sparco`), a suite of linear operators and problems for testing algorithms for sparse signal reconstruction. FJH was in part financially supported by NSERC Discovery Grant (22R81254) and by CRD Grant DNOISE 334810-05. The industrial sponsors of the Seismic Laboratory for Imaging and Modelling (SLIM) BG Group, BP, Chevron, Petrobras, and WesternGeco are also gratefully acknowledged.

REFERENCES

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

Beasley, C. J., 2008, A new look at marine simultaneous sources: The Leading Edge, **27**, 914–917.

Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: SEG Technical Program Expanded Abstracts, **17**, 133–135.

Berg, E. v., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Technical Report 2, Department of Computer Science, University of British Columbia.

Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

Bleistein, N., J. Cohen, and J. Stockwell, 2001, Mathematics of Multidimensional Seismic Imaging, Migration and Inversion: Springer.

Bobin, J., J. Starck, and R. Ottensamer, 2008, Compressed sensing in astronomy: IEEE J. Selected Topics in Signal Process, **2**, 718–726.

Candès, E., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Comm. Pure Appl. Math., **59**, 1207–1223.

Candès, E. J., and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Comm. Pure Appl. Math, **58**, 1472–1528.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, 861–899.

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

Canning, A., and G. H. F. Gardner, 1996, Regularizing 3-d data sets with dmo: **61**, 1103–1114.

Demanet, L., and L. Ying, 2007, Wave atoms and sparsity of oscillatory patterns: Applied and Computational Harmonic Analysis, **23**, 368–387.

Donoho, D., A. Maleki, and A. Montanari, 2009, Message passing algorithms for compressed sensing: Proceedings of the National Academy of Sciences.

Donoho, D., and J. Tanner, 2009, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, **367**, 4273.

Donoho, D. L., 2006a, Compressed sensing: IEEE Trans. Inform. Theory, **52**, 1289–1306.

———, 2006b, Compressed sensing: IEEE Transactions on Information Theory, **52**, 1289 – 1306.

Donoho, D. L., and X. Huo, 2001, Uncertainty principles and ideal atomic decomposition: IEEE Transactions on Information Theory, **47**, 2845–2862.

Donoho, D. L., and B. F. Logan, 1992, Signal Recovery and the Large Sieve: SIAM Journal on Applied Mathematics, **52**, 577–591.

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of under-

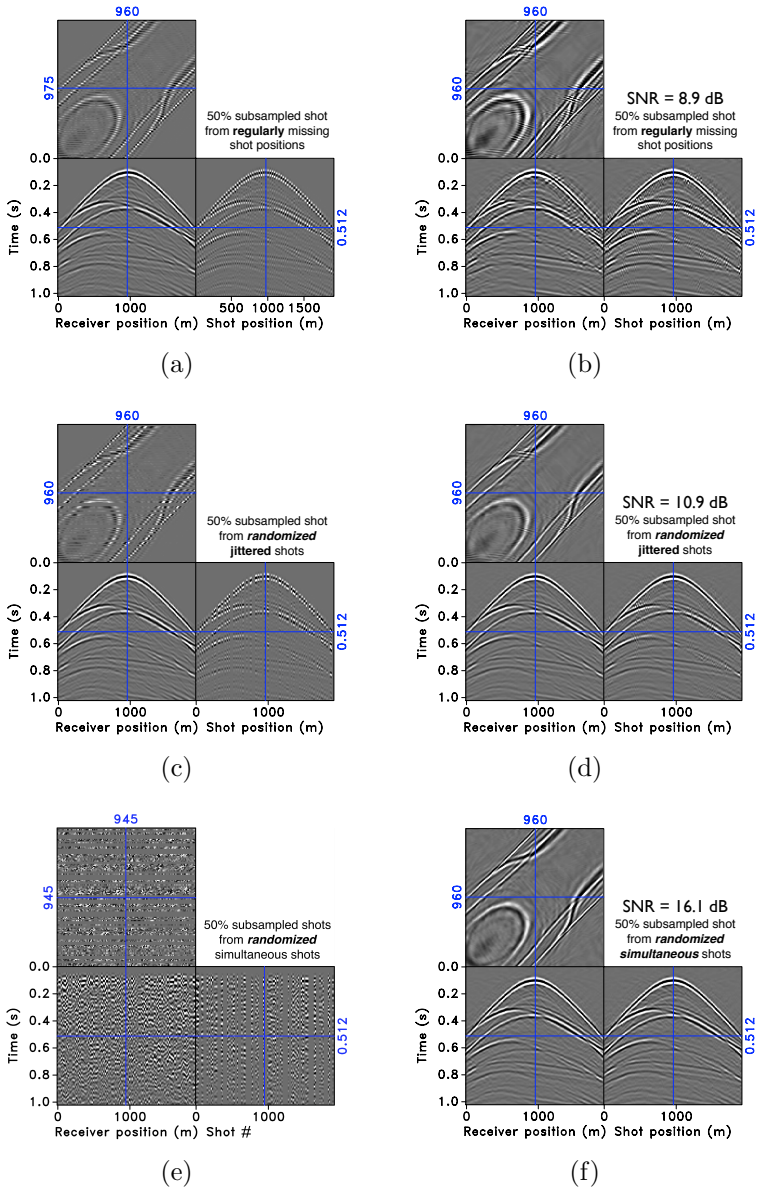


Figure 9: Sparsity-promoting recovery from 50% of the shots missing. **(a)** Regularly subsampled shots. **(b)** Recovery from regularly subsampled shots (8.9 dB). **(c)** Jittered subsampled shots (10.9 dB) **(d)** Recovery from jittered subsampled shots (16.1 dB). **(e)** Subsampled randomized simultaneous shots. **(f)** Recovery from randomized simultaneous shots. Notice the remarkable improvement in recovery from the simultaneously acquired data.

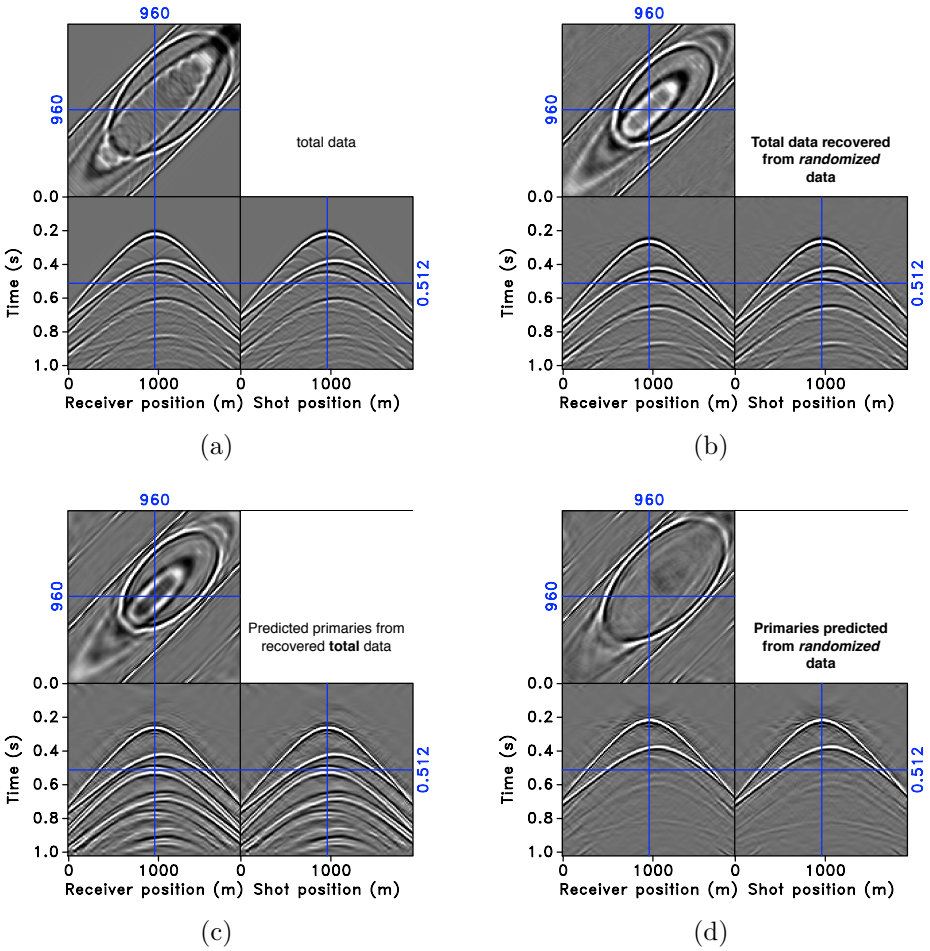


Figure 10: Sparsity-promoting recovery from simultaneous-source data with 50 % of the shots missing. **(a)** Original data. **(b)** Estimation of the total data, including surface-related multiples, by sparsity promotion. **(c)** Estimation of primaries from recovered total data under **(b)**. **(d)** Estimation of primaries directly from the simultaneously acquired data. Notice the remarkable improvement in the estimation of the primaries directly from the simultaneously acquired data.

determined linear equations by stagewise orthogonal matching pursuit: Technical report, Stanford Statistics Department. (TR-2006-2. `http://stat.stanford.edu/~idrori/StOMP.pdf`).

Donoho, P., R. Ergas, and R. Polzer, 1999, Development of seismic data compression methods for reliable, low-noise performance: SEG International Exposition and 69th Annual Meeting, 1903–1906.

Fomel, S., 2003, Theory of differenctial offset continuation: **68**, 718–732.

Fomel, S., J. G. Berryman, R. G. Clapp, and M. Prucha, 2002, Iterative resolution estimation in least-squares Kirchoff migration: Geophys. Pros., **50**, 577–588.

Hale, D., 1995, DMO processing: Geophysics Reprint Series.

Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.

Hennenfent, G., and F. J. Herrmann, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**.

Herrmann, F., D. Wang, G. Hennenfent, and P. Moghaddam, 2008a, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5. (doi:10.1190/1.2799517).

Herrmann, F. J., 2009, Compressive imaging by wavefield inversion with group sparsity: SEG Technical Program Expanded Abstracts, SEG, SEG, 2337–2341.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Non-linear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799.

Herrmann, F. J., Y. A. Erlangga, and T. Lin, 2009a, Compressive sensing applied to full-waveform inversion.

Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009b, Compressive simultaneous full-waveform simulation. TR-2008-09. to appear as a letter in geophysics, august, 2009.

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.

Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008b, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, 150–173. (doi:10.1016/j.acha.2007.06.007).

Herrmann, F. J., and D. Wang, 2008, Seismic wavefield inversion with curvelet-domain sparsity promotion: SEG Technical Program Expanded Abstracts, **27**, 2497–2501.

Krohn, C., and R. Neelamani, 2008, Simultaneous sourcing without compromise: Rome 2008, 70th EAGE Conference & Exhibition, B008.

Levy, S., D. Oldenburg, and J. Wang, 1988, Subsurface imaging using magnetotelluric data: Geophysics, **53**, 104–117.

Lin, T., and F. J. Herrmann, 2009a, Designing simultaneous acquisitions with compressive sensing: Presented at the EAGE Technical Program Expanded Abstracts, EAGE.

———, 2009b, Unified compressive sensing framework for simultaneous acquisition with primary estimation: SEG Technical Program Expanded Abstracts, SEG, 3113–3117.

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., 2009, A Wavelet Tour of Signal Processing: the Sparse Way: Academic Press.

Neelamani, N., C. Krohn, J. Krebs, M. Deffenbaugh, and J. Romberg, 2008, Efficient seismic forward modeling using simultaneous random sources and sparsity: SEG International Exposition and 78th Annual Meeting, 2107–2110.

Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: Geophysics, **46**, 1528–1542.

Rauhut, H., K. Schnass, and P. Vandergheynst, 2008, Compressed sensing and redundant dictionaries: IEEE Trans. Inform. Theory, **54**, 2210–2219.

Romberg, J., 2009, Compressive sensing by random convolution: SIAM Journal on Imaging Sciences, **2**, 1098–1128.

Sacchi, M. D., T. J. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Trans. Signal Process., **46**, 31–38.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, 938–945.

Santosa, F., and W. Symes, 1986, Linear inversion of band-limited reflection seismogram: SIAM J. of Sci. Comput., **7**.

Smith, H. F., 1998, A Hardy space for Fourier integral operators.: J. Geom. Anal., **8**, 629–653.

Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58.

Tang, G., J. Ma, and F. Herrmann, 2009, Design of two-dimensional randomized sampling schemes for curvelet-based sparsity-promoting seismic data recovery.: Technical Report TR-2009-03, the University of British Columbia.

Taylor, H. L., S. Banks, and J. McCoy, 1979, Deconvolution with the ℓ_1 norm: Geophysics, **44**, 39.

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.

Ulrych, T. J., and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, 1295–1302.

van Groenestijn, G. J. A., and D. J. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87 – V95.

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

———, 2007b, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, V21–V32.