Fighting the curse of dimensionality: compressive sensing in exploration seismology

Felix J. Herrmann, Michael P. Friedlander, Özgür Yılmaz
August 9, 2011

***Abstract*—Many seismic exploration techniques rely on the collection of massive data volumes that are mined for information during processing. This approach has been extremely successful, but current efforts toward higher-resolution images in increasingly complicated regions of the Earth continue to reveal fundamental shortcomings in our typical workflows. The "curse of dimensionality" is the main roadblock, and is 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.**

We offer an alternative sampling strategy that leverages 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.

Compressive sensing is a novel nonlinear sampling paradigm, effective for acquiring signals that have a sparse representation in some transform domain. We review basic facts about this new sampling paradigm that revolutionized various areas of signal processing, and illustrate how it can be successfully exploited in various problems in seismic exploration to effectively fight the curse of dimensionality.

***Index Terms*—Compressive sensing, curvelet transform, sparsity promotion, exploration seismology, seismic acqusition, seismic imaging, seismic inversion, and convex optimization**

I. The curse of dimensionality in seismic exploration

Modern-day seismic-data processing, imaging, and inversion rely increasingly on computationally and data-intensive techniques to meet society's continued demand for hydrocarbons. This is problematic because this leads to exponentially increasing costs as the size of the area of interest increases. Motivated by recent findings from compressive sensing (CS) and earlier work in seismic data regularization [55] and phase encoding [52], we confront the challenge of the curse of dimensionality with a randomized dimensionality-reduction approach that decreases the cost of acquisition and subsequent processing significantly. Before we discuss possible solutions to the "curse of dimensionality in exploration seismology", let us first briefly discuss how sampling is conducted in exploration seismology.

Felix J. Herrmann is with the Department of Earth and Ocean Sciences, The University of British Columbia, Vancouver, Canada.

Michael P. Friedlander is with the Department of Computer Science, The University of British Columbia, Vancouver, Canada.

Özgür Yılmaz is with the Department of Mathematics, The University of British Columbia, Vancouver, Canada.

A. Classical approaches

During seismic data acquisition, data volumes are collected that represent dicretizations of analog finite-energy wavefields in up to five dimensions including time. So, we are concerned with the acquisition of an analog spatio-temporal wavefield $\bar{f}(t,x) \in L^2((0,\,T] \times [-X,X])$ with time T in the order of seconds and length X 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

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

where N is the total number of samples, and where $\overline{\Phi}_s$ is the sampling operator and models the characteristics of the devices that are used to collect the measurements, i.e., geophones, including analog-to-digital conversion.

It is convenient to organize these high-resolution samples into a vector $f := \left\{f[q]\right\}_{q=0,\cdots,N-1} \in \mathbb{R}^N$. Note that in practice often we have missing samples, i.e., instead of f, the acquired data is $b = Rf$ where R is a $n \times N$ *restriction matrix* that consists of n rows of the $N \times N$ identity matrix.

B. Bottlenecks and compromises

Unfortunately, pressures for increased resolution (i.e., increasing N) and increasing dimenions (with $\bar{f}(t,x) \in L_2((0,\,T] \times [-X,X]^4)$, two dimensions for the source and two dimensions for the receivers) make complete sampling ($n = N$) economically and physically infeasible and data is sampled at a rate below Nyquist, i.e., $n \ll N$. For the spatial coordinates, this typically corresponds to periodic subsampling of the sources/receivers while the total acquistion time is reduced by reducing the time of the sequential single-source experiments. Unfortunately, these subsamplings can lead to serious artifacts and a lot of research has recently been devoted to come up with improved sampling schemes that randomize spatial locations of sources and receivers or that randomize the sources, e.g., by random dithering of marine sources or by source encodings on land.

C. Dimensionality reduction by Compressive Sensing

While recent proposals to expedite seismic acquisition or computations through simultaneous sourcing have proven successful, the proposed methods miss a rigorous framework that would allow for the design of rigorous workflows. By recognizing these clever new sampling schemes as instances

of CS, we are able to make a start towards sampling and computation strategies that employ structure in seismic data, which translates into transform-domain sparsity. This attribute allows us to come up with sub-Nyquist sampling schemes whose sampling is proportional to the sparsity rather than to the dimensionality of the problem. The success of these techniques hinges on subsamplings that break periodicty of conventional samplings. To demonstrate how this works, we first give a brief introduction to the theory CS, followed by its application to problems in exploration seismology. Because recovery from the subsamplings depends on solving large-scale optimization problems, we reserve a discussion of this important topic towards the end of this paper.

II. Compressive sampling as a dimension reduction method

Various classes of signals such us audio, images, and seismic signals admit *sparse approximations*, i.e., they can be well-approximated by a linear superposition of a few atoms of an appropriate basis, or more generally, a redundant frame. Compressed sensing (or compressive sampling)—following the breakthroughs by Candès, Romberg, and Tao [10] and Donoho [23]—has emerged as a novel paradigm for sensing such signals more efficiently as compared to the classical approach based on Shannon-Nyquist sampling theory. Signals that admit sparse approximations can be acquired from significantly fewer measurements than their ambient dimension by means of nonlinear recovery algorithms, e.g., one-norm minimization or greedy algorithms such as OMP [59] (the anti-leakage Fourier transform, [63] is a special case of OMP). However, because OMP is not suitable for curvelet-based recovery because it only brings a single component into the solution per iteration.

One of the main messages of CS is that the number of samples required to achieve a certain accuracy scales logarithmically with the ambient dimension of the underlying sparse signal, which, in the case of spatial sampling, is the sampling grid size. Thus, in problems where the sheer number of the measurements that need to be obtained (according to the classical Shannon-Nyquist sampling theory) is prohibitively large, the theory of CS is especially invaluable.

The aim of this section is to introduce the mathematical framework behind CS and discuss the particular challenges we face in exploration seismology.

A. Compressive acquisition of sparse signals

The main assumption in the classical sampling theory is that the signals of interest are in a linear subspace (i.e., all functions with a certain bandwidth). However, the main signal model of CS is nonlinear: the signals are *sparse* (only a few of the entries are non-zero) or *compressible* (can be well-approximated by a sparse signal), either in the canonical basis or in some appropriate transform domain.

Formally, consider a high-dimensional signal $x \in \mathbb{R}^N$. We first make the naive assumption that x is k-sparse, i.e., $\|x\|_0 \leq k$, where $\|x\|_0$ denotes the number of non-zero entries of the vector x. (We later relax the sparsity assumption to make way for more realistic signal ensembles including seismic.) Our goal is to obtain x (or an approximation) from non-adaptive linear measurements $y = \Psi x$, where Ψ is an appropriate $n \times N$ *measurement matrix*. Clearly, we can recover all $x \in \mathbb{R}^N$ exactly if $n \geq N$ and Ψ has full rank. Furthermore, setting Ψ to be the $N \times N$ identity matrix corresponds to measuring every entry of x separately.

On the other hand, if $n < N$, i.e., the number of measurements is less than the ambient dimension, the system

$$\Psi z = y$$

has infinitely many solutions, rendering it generally impossible to recover x from y. CS considers this scenario and aims to recover x by utilizing the prior information that x is sparse (or compressible): among all solutions of this equation, find the solution x^* with the smallest number of non-zero entries, i.e., solve the optimization problem

$$\underset{z}{\text{minimize}} \quad \|z\|_0 \quad \text{subject to} \quad \Psi z = y. \tag{1}$$

This is referred to as the *sparse recovery problem*. It can be shown that if every n-by-n submatrix of Ψ is nonsingular, then $x^* = x$, i.e., (1) recovers every k-sparse x *exactly* whenever $k < n/2$, e.g., see [22]. In other words, to acquire a k-sparse signal, we only need to obtain $n > 2k$ measurements regardless of the ambient dimension N and then solve the optimization problem (1). Unfortunately, this observation is not very useful in practice because the program (1) is NP-hard [46] and sensitive to the sparsity assumption and additive noise. The major breakthrough in CS has been to specify explicit conditions under which the minimizer of (1) also solves the convex—and hence computationally tractable—optimization problem

$$\underset{z}{\text{minimize}} \quad \|z\|_1 \quad \text{subject to} \quad \Psi z = y. \tag{2}$$

Specifically, these conditions [10], [23] determine what measurement matrices Ψ can be used so that (2) is guaranteed to recover all k-sparse x in $\mathbb{R}^N$ from n measurements given by $y = \Psi x$. In words, the main requirement is that Ψ nearly preserves the length of all sparse vectors.

Various random matrix ensembles have been shown to be effective compressive measurement matrices, e.g., Gaussian and Bernoulli matrices, and Fourier matrices with randomly selected rows [11], [53]. An important question is how the number of measurements required for exact recovery scales with the sparsity level k, the number of measurements n, and the ambient dimension N. (In the classical sampling theory, k is analogous to "bandwidth", n is analogous to sampling frequency, and N is analogous to the size of the sampling grid.) The following theorem, adapted from [10], summarizes the answer to this question.

Theorem 1. *Let Σ_k^N be the set of all k-sparse vectors in $\mathbb{R}^N$. Suppose Ψ is an $n \times N$ random matrix the entries of which are drawn i.i.d. from a sub-Gaussian distribution. If*

$$n \gtrsim k \log(N/n),$$

with overwhelming probability (on the draw of Ψ) (2) *recovers*

all $x \in \Sigma_k^N$ from the compressive measurements $y = \Psi x$.

In words, if the measurement matrix is chosen appropriately, the number of measurements scales only logarithmically with the ambient dimension N—a tremendous improvement over linear scaling of the classical sampling theory.

Remark 1.1. One-norm minimization via (2) is the most common and well-understood method of solving the sparse recovery problem and it is our main focus. However, it is important to note that various alternative approaches have been proposed and analyzed, e.g., greedy algorithms such as OMP [59], CoSaMP [47], iterative hard thresholding [5], and p-norm minimization with $0 < p < 1$ [13], [29], [54].

B. Compressible signals and robustness to noise

For CS to be practicable, two major modifications to the setting in the previous section need to be considered. First, it is naive to expect signals in practice to be exactly sparse. A more realistic model is that the magnitude-sorted coefficients decay rapidly, leaving us with a vector with only few large entries and many small ones. Such signals can be well approximated by sparse signals and usually are said to be *compressible*. It is crucial that CS is stable in the case of CS. Second, in practical applications typically the measurements are contaminated by noise, and it is again crucial that CS paradigm is robust to noise.

The following result by Candès, Romberg, and Tao [10] shows that this is indeed the case. For $x \in \mathbb{R}^N$, let

$$\sigma_k(x)_{\ell_1} := \min_{v \in \Sigma_k^N} \|x - v\|_1 \tag{3}$$

be the best k-term approximation error of x in ℓ_1. Note that the more compressible x is, the smaller is $\sigma_k(x)_{\ell_1}$.

Theorem 2. *Let $x \in \mathbb{R}^N$ be arbitrary, and let Ψ be an appropriate $n \times N$ measurement matrix (e.g., a Gaussian matrix). Suppose that the noisy measurements are given by $y = \Psi x + e$ where e is additive noise with $\|e\|_2 \leq \epsilon$. Denote by x^* the solution of the following convex program:*

$$\underset{z}{\text{minimize}} \ \|z\|_1 \text{ subject to } \|\Psi z - y\| \leq \epsilon.$$

Then for absolute constants C_1 and C_2,

$$\|x - x^*\|_2 \leq C_1 \epsilon + C_2 k^{-1/2} \sigma_k(x)_{\ell_1},$$

whenever $n = \mathcal{O}(k \log[N/n])$.

In words, the recovered approximation is within the noise level and nearly as accurate as the approximation we would obtain by measuring directly the largest k entries of x.

C. Extensions

As we mentioned before, CS can be used to recover signals that admit a sparse representation with respect to a basis or frame. In sections II-A and II-B we considered the case where the signals of interest were simply sparse in the canonical basis, which, of course, is rarely the case in practice. Images, for example, are sparse with respect to appropriate orthonormal wavelet bases, and seismic signals admit sparse approximations in terms of curvelets [9], [21], [32]. Formally, consider signals $f \in \mathbb{R}^N$ that are sparse with respect to a basis or frame S, i.e.,

$$f = S^H x, \quad x \text{ sparse.}$$

Above S is a $P \times N$ matrix, with $P \geq N$, that admits a left-inverse, and the superscript H denotes the adjoint.

In the next sections we discuss how and to what extent the CS paradigm can be used when S is either an orthonormal basis or a redundant frame. (See, e.g., [43] for a comprehensive review of frame theory.)

1) S is an orthonormal basis: In the case when S is an orthonormal basis (i.e., $S^{-1} = S^H$), CS theory applies essentially unchanged. Specifically, the compressive samples of the signal f is given by $y = \Psi f = \Psi S^H x$ where x is sparse (or compressible). In turn, the effective measurement matrix is ΨS^H and if this matrix satisfies the requirements of Theorems 1 and 2, the conclusions of these theorems remain valid, including the compressibility error bounds of Theorem 2 as S is orthonormal, thus an isometry.

The main challenge is how to choose Ψ for a given S so that ΨS^H is still a good "measurement matrix". Recall that in CS each measurement is an inner product with an appropriate "measurement vector", a row of Ψ. One possible way of constructing a good Ψ tailored to a given sparsity basis S is to first choose an appropriate *measurement basis* M that is *incoherent* with S. The coherence of two bases S and M is reflected by the largest-in-magnitude entry of the matrix MS^H. Once we choose an incoherent M, we discard all but n rows from M and use the resulting $n \times N$ matrix as our measurement matrix. More precisely, we set $\Psi = RM$ where R is an $n \times N$ restriction matrix (consisting of n rows of the $N \times N$ identity matrix). It can be shown that such a Ψ can be used for CS. For example, if the signal u is sparse in Fourier domain, i.e., S is the DFT matrix, then an optimally incoherent measurement basis is given by $M = I_N$, the $N \times N$ identity basis.

In summary, we obtain the compressive samples of the signal f by $y = RMf + e$ (where e is again additive noise) and recover f by solving the convex program

$$\underset{z}{\text{minimize}} \ \|z\|_1 \text{ subject to } \|RMS^H z - y\| \leq \epsilon.$$

We should also note that there is a universal strategy for choosing Ψ that does not require prior knowledge of the sparsity basis S: if we choose Ψ to be an appropriate random measurement matrix that satisfies a certain concentration of measure property [2], e.g., Gaussian and Bernoulli, then ΨS^H is guaranteed to be also a good measurement matrix independent of the orthonormal basis S. The downside of this approach is the computational challenges related to generating, storing, and applying unstructured random matrices.

2) S is a redundant frame: The problem becomes significantly more challenging if the sparsifying dictionary is an overcomplete frame. This means that the signal f can be decomposed as $f = S^H x$ where S is $P \times N$ with $P > N$ and S admits a left inverse. For example, seismic signals are compressible with respect to curvelet frames, which

are overcomplete. Compared to the orthonormal bases, the differences of this set-up are: (i) the expansion coefficients are not unique, i.e., there are (infinitely) many x that explain the same signal f, and (ii) the columns of S^H must be correlated. Accordingly, the approaches used in the orthonormal case do not readily generalize immediately to this case. Empirically, the CS paradigm has been observed to be effective for acquisition of signals that admit sparse approximations in appropriately redundant frames—see Section III-A for an empirical study of compressive seismic-data acquisition using curvelet frames as the sparsifying transform. Finally we note that recent theoretical results, e.g., [12], corraborate with these empirical observations.

D. Challenges in seismic

According to CS, successful dimensionality reduction hinges on an incoherent sampling strategy where coherent aliases are turned into relatively harmless white Gaussian noise. The challenges of adapting this approach to real-life problems in exploration seismology are threefold. First, seismic data acquisition is subject to physical constraints on the placement, type, and number of (possibly simultaneous) sources, and numbers of receivers. These constraints in conjunction with the extreme large size of seismic data call for seismic problem-specific solutions. Second, while CS offers significant opportunities for dimensionality reduction, there remain still challenges in adapting the scientific-computing workflow to this new approach, and again, CS offers an opportunity to make computation more efficient. Third, seismic wavefields are highly multiscale, multidirectional, and are the solution of the wave equation. This calls for the use of directional and anisotropic transforms, e.g., curvelets.

III. Compressive seismic-data acquisition

Perhaps it is too early to claim that CS will constitute a paradigm shift in seismic acquisition. The first breakthrough was the identification of seismic data regularization and simultaneous/continuous acquisition as instances of CS [36]. Further encouraging progress has been made in the selection of the sparsifying transform and the design of randomized sampling schemes that are realizable in the field.

We discuss progress in each of these areas by means of carefully designed examples that include real field data.

A. Selection of the sparsifying transform

CS leverages structure within signals to reduce the required sampling rates. Typically, this structure translates into compressible representations, using an appropriate transform, that concentrate the signal's energy into a small percentage of large 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.

We thus only consider transforms that are fast ($\mathcal{O}(N \log N)$), multiscale (split the Fourier spectrum into dyadic frequency bands), and multidirectional (split the Fourier spectrum into second dyadic angular wedges). For completeness, we also include separable 2-D wavelets in our study. Unlike wavelets, which compose curved wavefronts into a superposition of multiscale "fat dots" with limited directionality, curvelets [9] and wave atoms [21] compose wavefields as a superposition of highly anisotropic, localized, and multiscale waveforms, which obey the so-called parabolic-scaling principle. For curvelets, this principle translates into a support where length is proportional to the square of the width. At fine scales, this leads to needle-like curvelets. Curvelets, with their near invariance under wave propagation [8], are thus highly suitable for compressing seismic data. Wave atoms share with curvelets this invariance, and they are also anisotropic because their wavelength depends quadratically on their width. While curvelets are optimal for data with delta-like wavefronts, wave atoms are more appropriate for compressing data with oscillatory wavefronts. Seismic data sits somewhere between these two extremes, and we include both transforms in our study.

1) Approximation error: For an appropriately chosen representation magnitude-sorted transform-domain coefficients often decay rapidly. For orthonormal bases, the decay rate is directly linked to the decay of the nonlinear approximation error, see e.g. [43]. This error can be expressed by

$$\sigma_k(f)_{\ell_2} := \min_{x \in \Sigma_k^N} \|f - S^H x\|_2,$$

where f_k is optimal argument, which gives the best k-term approximation in the ℓ_2 sense; cf. (3). When S is orthonormal, f_k is uniquely determined by taking the largest-in-magnitude k-entries of Sx. Unfortunately, such a direct way of finding f_k is not available when S is redundant, because redundant expansions are not unique: there are many coefficient sequences that explain the discrete data f, and these different sequences may have varying decay rates.

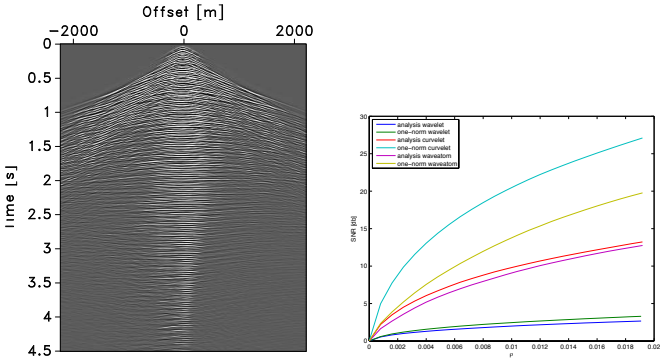
To address this issue, we use an alternative definition for the nonlinear approximation error, which is based on the solution of a sparsity-promoting program. With this definition, the k-term sparse approximation error is computed by taking the k-largest coefficients from the vector that solves

$$\underset{x}{\text{minimize}} \ \|x\|_1 \text{ subject to } S^H x = f, \tag{4}$$

where the P-by-N matrix S is the curvelet analysis operator. As before, N is the ambient dimension, and the transform is redundant if $P > N$.

The solution of (4) is typically sparser than the vector obtained by applying the analysis operator S directly. To be able to compare various redundant transforms with different degrees of redundancy, we study the signal-to-noise ratio $\mathrm{SNR}(\rho) = -20 \log \frac{\|f - f_{\rho P}\|}{\|f\|}$, where $\rho = k/P$ is a compression ratio. A smaller ratio implies a larger fraction of ignored coefficients and sparser transform-coefficient vector, which leads to a smaller SNR. In our study, we include $f_{\rho P}$ that are derived from either the analysis coefficients, i.e., the largest ρP coefficients of Sf, or from the synthesis coefficients that are solutions of the above sparsity-promoting program (4).

2) Empirical approximation errors: Parametrizing the SNR by ρ allows us to compare the recovery quality of seismic data



(a) (b)

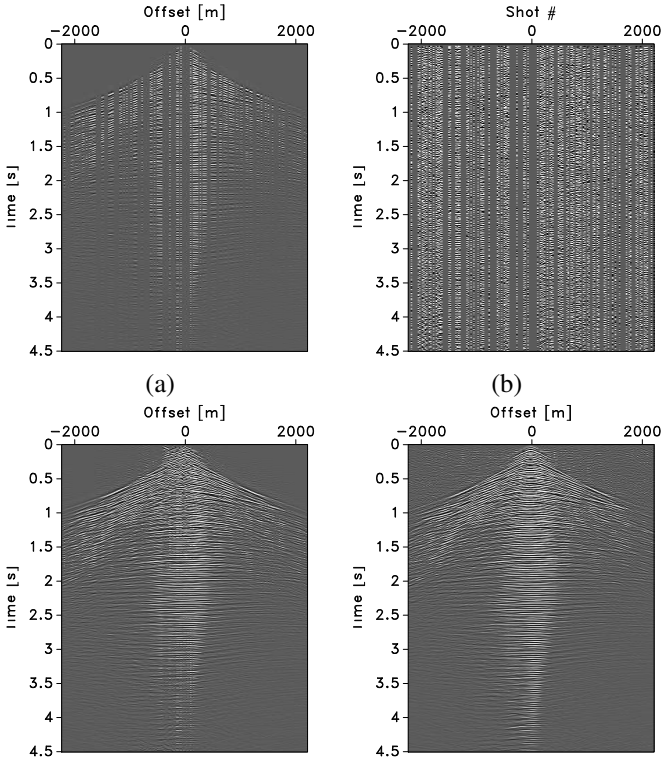
Fig. 1: (Adapted from [38]) Signal-to-noise ratios (SNRs) for the nonlinear approximation errors of a common-receiver gather (a) from a Gulf of Suez data set. The SNRs (b) are plotted as a function of the sparsity ratio $\rho \in (0,\, 0.02]$. The plots include curves for the errors obtained from the analysis and one-norm minimized synthesis coefficients. Notice the significant improvement in SNRs for the synthesis coefficients obtained by solving (4).

using various transforms, such as wavelets, curvelets, and wave atoms. Figure 1 compares the performance of these transforms on a common-receiver gather extracted from a Gulf of Suez dataset. Our results in Figure 1 clearly show that curvelets and wave atoms benefit significantly from sparsity promotion, though wave atoms lag behind curvelets. This effect is most pronounced for synthesis coefficients. Because wavelets are orthogonal, they can not benefit, as expected. Note that the observed behavior is consistent with the degree of redundancy of each transform: the curvelet transform has the largest redundancy (a factor of about eight in 2-D), wave atoms have only a redundancy of two, and wavelets are not redundant. This suggests that sparse recovery from subsampling would potentially benefit most from curvelets. However, this may not be the only factor that determines the performance of CS.

B. Acquisition schemes

Before discussing the application of CS to realistic data examples, we briefly discuss differences between recovery from missing shots, which is an instance of seismic data regularization, and recovery from simultaneous data. The seismic data regularization problem can be considered as the seismic-version of inpainting.

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., $I \stackrel{\text{def}}{=} I_{N_s} \otimes I_{N_t}$, which is the N-by-N identity matrix where $N = N_s N_t$, the product of the number of shots N_s and the number of time samples N_t. For simultaneous acquisition, where all sources fire simultaneously, this matrix is given by $M \stackrel{\text{def}}{=} G_{N_s} \otimes I_{N_t}$ with G_{N_s} a N_s-by-N_s Gaussian matrix with *i.i.d.* entries. In both cases, we use a restriction operator $R \stackrel{\text{def}}{=} R_{n_s} \otimes I_{N_t}$ to model the collection of incomplete data by



(a) (b) (c) (d)

Fig. 2: (Adapted from [38]) Recovery from a compressively-sampled common-receiver gather with 50% 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)**. Notice the remarkable improvement in the recovery from simultaneous data.

reducing the number of shots to $n_s \ll N_s$. This restriction acts on the source coordinate only. For both recovery experiments, we use 2-D curvelets as the sparsifying transform S (cf. II-C).

CS predicts superior recovery for compressive-sampling matrices with smaller coherence. This coherence depends on the interplay between the restriction, measurement, and synthesis matrices. To make a fair comparison, we keep the randomized restriction matrix the same and compare the recoveries for measurement matrices 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 half of the experiments missing. Examples of both measurements for the real common-receiver gather of Figure 1 are plotted in Figure 2. Both CS experiments are using the same amount of data.

Comparing the recovery quality for data for both experiments confirms the insight from CS that states that incoherent measurement matrices favor sparsity-promoting recovery. This opens the possibility of designing efficient acquisition strategies in the field, or of dimensionality reduction in the computer.

Example: coil sampling. The quality of 3-D seismic

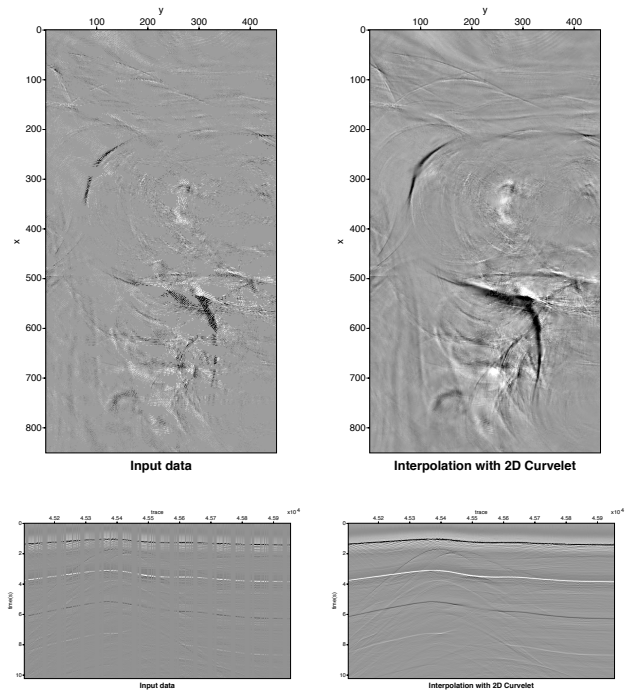


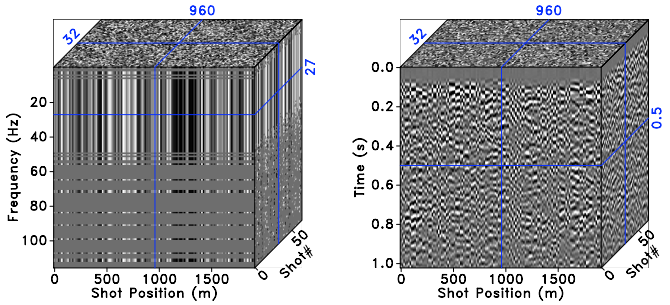
Fig. 3: Synthetic common-azimuth/offset example of coil sampling. The data is simulated with finite-differences on the SEAM model [25]. (Top) Time slice with 66% data missing data and recovery. (Bottom) Cross-section with 66% missing traces and recovered section. Notice the excellent recovery even in regions with strong complexity.

imaging and full-waveform inversion depends largely on azimuthal coverage. While full-azimuth acquisition is getting within reach on land, full-azimuth sampling in marine remains challenging.

Moldoveanu adopted Hennenfent and Herrmann's jittered-sampling approach [33] to devise a coil-sampling strategy that addresses the full-coverage issue, and shoots with several vessels in tandem while navigating along coils whose centers are randomized [45]. To illustrate the performance of sparsity-promoting recovery for this type of sampling, we consider a binned common-azimuth/offset volume simulated from the SEAM model [25] with 66% missing data. The recovery results (where we use curvelets in the lateral directions and wavelets along time to sparsify) are included in Figure 3 for a time slice and a cross section. The results for these slices were obtained with a relatively low number of iterations of the one-norm solver SPGL1 [61] and show excellent recovery from this sampling even in regions of large complexity. This example is a good illustration of the validity of this technology on industry-type data volumes.

IV. Compressive seismic computation

We have so far concentrated on applying CS to seismic acquisition. While the invocation of CS in acquisition potentially reaps major increases in efficiency, CS can also be applied to increase the efficiency of wavefield simulations, imaging, and inversion.



(a) (b)

Fig. 4: (Adapted from [36]) Compressive sampling with simultaneous sources. **(a)** Amplitude spectrum for the source signatures emitted by each source as part of the simultaneous-source experiments. These signatures appear noisy in the shot-receiver coordinates because of the phase encoding (cf. Equation V). Observe that the frequency restrictions are different for each simultaneous source experiment. **(b)** CS-data after applying the inverse Fourier transform. Notice the noisy character of the simultaneous-shot interferences.

A. Compressive simulation

To simulate seismic surveys, one needs for each source experiment to solve a large linear system that discretizes the underlying wave equation. Because seismic surveys consist of many source experiments, we must reduce the number of PDE solves by exploiting linearity of the wave equation in the sources. This linearity allows us to combine sequential sources into a smaller number of "supershots", each consisting of a random superposition of all sequential shots. Neelamani et al. [48] and Herrmann et al. [36] identify this principle, also known as "phase encoding" [52], as an instance of CS, and demonstrate that it can be used to make wave simulations more efficient by reducing the number of sources.

This technique allowed us to significantly speedup simulations with the time-harmonic Helmholtz equation. Used in combination with randomized importance sampling in the temporal frequency band, we achieved speedups proportional to the subsampling ratio. A 4-times reduced data volume simulated for these incoherent sources is plotted in Figure 4(a). As shown in [36], sequential simulations can be recovered from this compressively-sampled simulation by solving a sparsity-promoting program with a cost of $\mathcal{O}(n^3 \log n)$, where n the number of sources, receivers, depth levels, and frequencies. This additional cost is very small compared to the cost of solving the Helmholtz system, which is $\mathcal{O}(n^4)$.

B. Compressive imaging

Even though useful for wavefield simulations, compressive simulation is not particularly suitable for making wave-equation based seismic imaging more efficient because we would have to solve a sparsity-promoting program for each PDE solve. To overcome this problem, [26] proposes to image directly with simultaneous data, making the least-squares

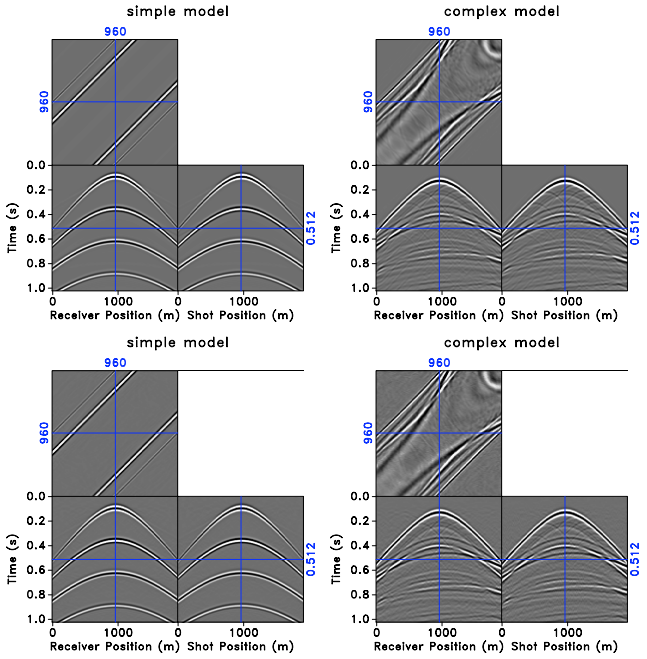


Fig. 5: (Adapted from [36]) Comparison between conventional and compressive simulations for simple and complex velocity models. **(a)** Seismic line for the simple model. **(b)** The same for the complex model. **(c)**. Recovered simulation (with a SNR of 28.1 dB) for the simple model from 25 % of the samples with the ℓ_1-solver running to convergence. **(d)** The same but for the complex model now with a SNR of 18.2 dB.

migration problem

$$\underset{\Delta x}{\text{minimize}} \ \frac{1}{2K}\sum_{i=1}^{K}\|\Delta b_i - A_i\Delta x\|_2^2 = \frac{1}{2K}\|\Delta b - A\Delta x\|_2^2 \quad (5)$$

more efficient. Here, Δb_i are the vectorized monochromatic shot records with linearized data (residue), A_i is the linearized Born scattering matrix, and Δx the unknown seismic image with the medium perturbations. (The quantities Δb and A aggregate the data across experiments.) In certain cases, the residual $\Delta b_i - A_i\Delta x$ can be interpreted as a linearization of the forward map defined by the wave equation; see (6).

This is a difficult problem because each iteration requires a large number of PDE solves, in particular, $4K$ solves, where $K = N_f \cdot N_s$, and N_f and N_s are the number of frequencies and sources. In order to do the inversion, we must be careful to limit the cost of each matrix-vector multiply, which we accomplish by dimensionality reduction. In particular, we use the same supershots as defined in section IV-A.

The optimization problem defined by (5), however, differs fundamentally from the standard CS problem because now the system is overdetermined—there are more equations then unknowns. In addition, the scattering matrix is ill conditioned due to limitations in aperture that may lead to shadow zones. However, for reflectors that are in the range of the scattering operator, the wave-equation Hessian $A^H A$ is near unitary, and curvelets are nearly invariant under the action of the Hessian [39]. This, and the optimality of curvelets on images with reflectors that may include conflicting dips [39], motivates us to replace (5) by a sparsity-promoting program that solves for the curvelet coefficients of the migrated image by replacing A with $\underline{A} := RMAS^H$ and Δb with $\Delta\underline{b} = RM\Delta b$. After applying the sampling, the N_s sequential sources are replaced by $n_s \ll N_s$ supershots with $n_f \ll N_f$ frequencies.

Unfortunately, the degree of randomized dimensionality reduction determines the amount of cross-talk that results from the inversion, and hence we can not reduce the problem size too much. We use a stochastic optimization approach [3], [24], [49], and cast the original imaging problem into a series of much smaller subproblems that work on different subsets of random source-encoded supershots [41]. (The best sampling strategy itself is still an open research problem; see, for example, [30], [58].) This approach corresponds to drawing a collection of supershots, followed by imaging, and using this image as a warm start for a new inversion with a new independently drawn collection of supershots. This process is repeated until progress towards the solution stalls. In Algorithm 1, we outline this approach for a generic solver $\mathbb{P}(RM; \Delta x_0)$ that uses warm starts Δx_0.

Algorithm 1: Stochastic approximation with warm starts

```
Δx_0 ← 0; k ← 0 ;                                  // initialize
while ‖Δx_0 − Δx̃‖_2 ≥ ϵ do
    k ← k + 1;                                     // increase counter
    Δx̃ ← Δx_0;                                     // update warm start
    RM ← Draw(RM);                                 // draw supershots
    Δx_0 ← Solve(ℙ(RM); Δx̃);                       // solve subproblem
end
```

To establish a baseline for comparison, we first compute an image by solving (5) for all 192 shots and 10 frequencies and 10 iterations of LSQR [51]. The result of this exercise, for data generated from the synthetic Marmoussi model [6] is included in Figure 6(a). We also solve a series of dimensionality-reduced subproblems with 8 supershots and 3 frequencies for 20 stochastic approximations, with LSQR ($\mathbb{P}_{\ell_2}(RM)$), with SPGL1 ($\mathbb{P}_{\ell_1}(RM)$), each with and without independent renewals of RM. These experiments are summarized in Figure 6(b-e) from which we make the following observations. First, redrawing the supershots after solving each subproblem clearly improves the performance of both solvers. We can understand this observation because these renewals remove possible correlations between RM and the current estimate for the (curvelet-domain) perturbation. Second, the image obtained by sparsity promotion is clearly superior in quality compared to the least-squares result, which we can explain with insights from CS. Third, this sparsifying result albeit noisy also compares favorably to the baseline image; it has higher resolution and better resolved amplitudes at depth.

We obtained a remarkably good result with a significantly reduced computational cost. We attribute this performance to curvelet-domain compressibility, which serves as a strong prior that mitigates source crosstalk and regularizes the inversion.

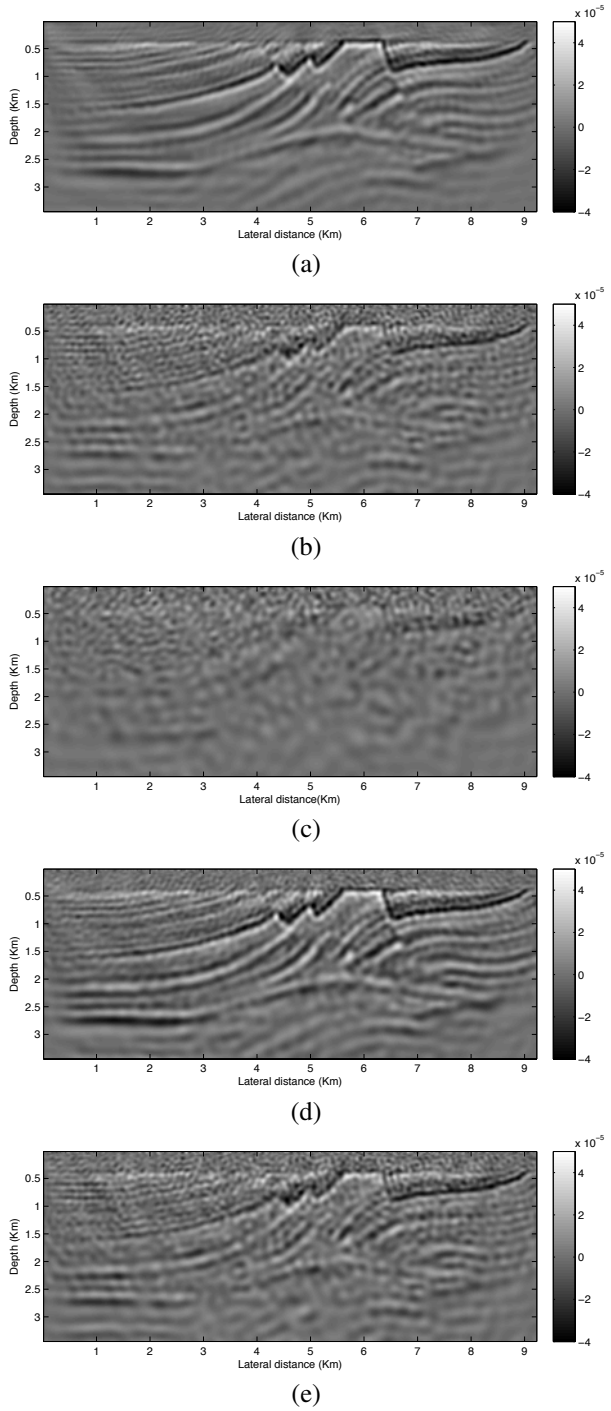


Fig. 6: Comparison between images obtained by linearized with and without dimensionality reduction. (a) Baseline image calculated for all data with 10 iterations of LSQR. (b) Image obtained by Algorithm 1 with $\mathbb{P}_{\ell_2}(RM)$ with 10 restarts independent redraws for RM. (c) The same but with the same RM. (d) Image obtained by Algorithm 1 with $\mathbb{P}_{\ell_1}(RM)$ for approximately the same number of PDE solves. (e) The same but with the same RM.

C. Compressive inversion

As we reported in earlier work (e.g., see [35], [42]), the cost of computing gradient and Newton updates in full-waveform inversion (FWI) is one of the major impediments that prevents successful adaptation of this industry-size problems. FWI involves the solution of an multi-experiment unconstrained optimization problem (cf. (5) for the linearized case):

$$\underset{m}{\text{minimize}} \ \frac{1}{2K}\sum_{i=1}^{K}\|b_i - \mathcal{F}_i[m,q_i]\|_2^2, \tag{6}$$

with b_i monochromatic shot records with the Earth response to monochromatic sources q_i, and $\mathcal{F}_i[m,q_i]$ represents monochromatic nonlinear forward operators. This operator is parameterized by the velocity m.

To overcome the computational burden of solving (6), we follow a similar procedure as outlined in Section IV-B but with the difference that we do a new linearization after each of the SPGL1 subproblems.

We test our algorithm on the synthetic model plotted in Fig. 7(a), which we use to generate data with a source signature given by a $12\,\mathrm{Hz}$ Ricker wavelet. To mimic practice, we use a smooth starting model without lateral information (Fig. 7(b)) and we start the inversion at $2.9\,\mathrm{Hz}$. This means that the seismic data carries relatively little low-frequency information. All simulations are carried out with 350 shot positions sampled at a $20\mathrm{m}$ interval and 701 receiver positions sampled at a $10\mathrm{m}$ interval, yielding an maximum offset of $7\mathrm{km}$. To improve convergence, the inversions are carried out sequentially in 10 overlapping frequency bands on the interval $2.9 - 22.5\mathrm{Hz}$ [7], each using 7 different simultaneous shots and 10 selected frequencies. For each subproblem, we use roughly 20 iterations of SPGL1 at a cost roughly equivalent to one tenth of the cost of a gradient calculation with all of sources. The result for each frequency band after 10 SPGL1 subproblems is depicted in Fig. 7(c). We can see from this result that our inversion captures more or less all discontinuities with a resolution commensurate the frequency range over which we carry out the inversion. This is remarkable and again the result of combining randomized-dimensionality results and sparse recovery from CS with recent insights from stochastic optimization. As before, drawing independent supershots after solving each SPGL1 subproblem benefited our results [41], [57].

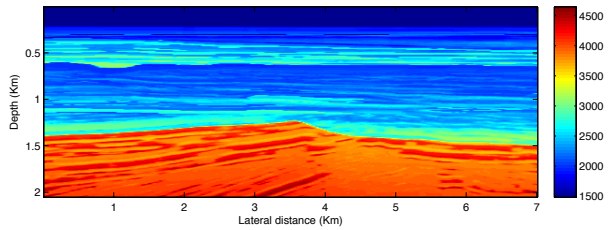
As before, we reduce the computation costs of minimizing or of solving problem (6) by randomizing source superposition. Choosing a different collection of supershots for each subproblem gives superior results.

V. Solving the sparse-optimization problem

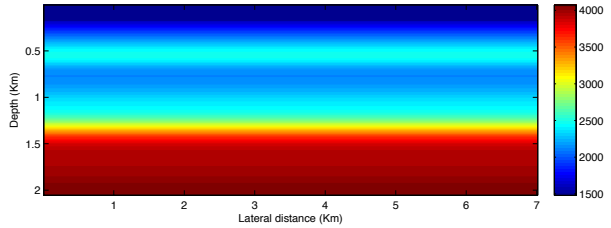
The main computational challenge in the dimensionality-reduction techniques that we describe are rooted in solving convex optimization problems. Almost all of these problems have the form

$$\mathrm{BP}_\sigma: \quad \underset{x}{\text{minimize}} \ \|x\|_1 \text{ subject to } \|Ax-b\|_2 \le \sigma,$$

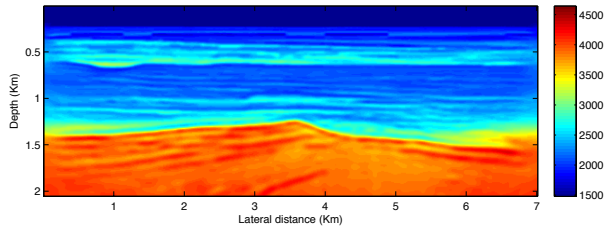
where σ is an estimate of the required data misfit, often related to the noise level and model mismatch. (The value $\sigma = 0$



(a)



(b)



(c)

Fig. 7: Full-waveform inversion result. (a) Initial model. (b) True model. (c) Inverted result starting from 2.9Hz with 7 simultaneous shots and 10 frequencies.

yields the *basis pursuit* problem [15].) The nonsmoothness of the objective is the essential difficulty. If these problems were relatively small, then many of the current workhorse algorithms for convex optimization (e.g., simplex and interior-methods) could be used off-the-shelf. However, these methods typically rely on explicit matrix representations of A. Thus there is now significant effort devoted to developing matrix-free algorithms tailored to these problems, which are typically characterized by large dense matrices. The terrific problem sizes in the seismic context is yet a further challenge: a "small" seismic problem can easily have 2^{21} variables.

One of our aims here is to give a broad view of the main approaches, and to describe the approach used by the SPGL1 software package [60], [61], which we use routinely for tackling seismic sparse recovery problems.

A. Main approaches

Most approaches for solving BP_σ are based on its "Lagrangian" reformulation

$$\text{QP}_\lambda: \quad \underset{x}{\text{minimize}} \ \tfrac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1.$$

The positive parameter λ is related to the Lagrange multiplier of the constraint in BP_σ, and it balances the tradeoff between the two norm of the data misfit and the one norm of the solution, which promotes sparsity. (When x is real-valued, $\|x\|_1 = \sum_j |x_j|$ is a piecewise linear function, and QP_λ is equivalent to a quadratic program, which is an established technology; when x is complex-valued, $\|x\|_1 = \sum_j \sqrt{\Re(x_j)^2 + \Im(x_j)^2}$, which is not piecewise linear, and the equivalence to quadratic programming is lost.) For an appropriate choice of λ, this formulation has the same solution to BP_σ, and thus in some sense these two problems are equivalent. However, except for very special cases, the value of λ that induces the equivalence cannot be determined without first solving BP_σ. The typical approach is thus based on solving a sequence of problems QP_λ defined by a decreasing sequence of parameters λ [1], [37]. This gradually decreases the data misfit, which usually allows more nonzeroes into the solution. The overall process terminates when the data mismatch reaches a prescribed accuracy. As we illustrate later, this can be an inefficient approach that requires the solution of too many subproblems.

Many algorithms are available for solving QP_λ or closely related variations, including iteratively-reweighted least-squares [14], [20], iterative soft-thresholding (IST), which has been rederived from various perspectives [18], [28], and projected gradient (PG) [19], [27]. A variation of the PG algorithm features prominantly in the SPGL1 software package, where it is used to approximately solve a sequence of subproblems [61, §4.1].

B. Proximal splitting

It is increasingly clear that many of the seemingly different algorithms that are being proposed for problems such as QP_λ are founded in classical ideas from convex analysis and optimization, and can be often described within the concise framework of proximal forward-backward splitting methods. These approaches are now prevalent in the new field of sparse optimization, which is concerned with exploiting sparse structure.

Consider the problem

$$\underset{x}{\text{minimize}} \ \phi(x) + \rho(x),$$

where ϕ is convex and smooth, and ρ is convex and—of central importance in sparse optimization—nondifferentiable. This formulation captures a range of problems that arise in signal-processing applications. For example, $\phi(x) = \frac{1}{2}\|Ax-b\|_2^2$ and $\rho(x) = \lambda\|x\|_1$ yields QP_λ; alternatively setting

$$\rho(x) = \begin{cases} 0 & \text{if } \|x\|_1 \le \tau, \\ +\infty & \text{otherwise,} \end{cases}$$

yields the Lasso [56] variation

$$\text{LS}_\tau: \quad \underset{x}{\text{minimize}} \ \tfrac{1}{2}\|Ax-b\|^2 \text{ subject to } \|x\| \le \tau.$$

The iteration scheme

$$x_{k+1} = \text{prox}_{\alpha_k \rho}\{x_k - \alpha_k \nabla\phi(x_k)\}, \tag{7}$$

where prox is the proximity operator corresponding to ρ and α_k is a step length, constitutes the basic form of the proximal-splitting method. When $\rho = \lambda\|x\|_1$, the corresponding proximity operator is the soft-thresholding operator [43, Ch. 11]. We see that there is a steepest descent step on the smooth component of the objective (i.e., ϕ), followed by a proximal

step based on the nonsmooth component of the objective (i.e., ρ). The resulting sequence $\{x_k\}$ converges to a solution of (7) under very mild conditions. A comprehensive overview is given by Combettes and Wajs [16]. Various algorithms, including PG (also known as projected Landweber), and IST, can be immediately cast in this framework.

The strategy that defines the sequence of step lengths α_k has a enormous effect on the effectiveness of these splitting approaches. Classical approaches, such as steepest descent, typically insist on selecting a steplength such that the objective decreases monotonically. However, there is now significant numerical and theoretical evidence to suggest that nonmonotonic approaches that consider one or more previous iterations in determining the step in the next iteration, often preform remarkably better than their monotonic counterparts; a short selection of references includes [4], [27], [50].

C. Pareto curve

A fundamental problem remains: even if we do have an effective algorithm for QP_λ (or LS_τ), how do we best choose a parameter λ (or τ) that yields a required data misfit? The Pareto curve, which traces the optimal trade-off between the two-norm of the residual $r = b - Ax$ and the one-norm of the solution x, is a helpful tool for visualizing the effect of regularization. Fig. 8 gives a schematic illustration of a the curve and some of its features. Points below the curve are not attainable. Any point on the curve, which is uniquely defined by a given A and b, gives the corresponding values σ (vertical axis) and τ (horizontal axis) that cause BP_σ and LS_τ to have the same solution. The negative of the slope at that point gives the corresponding value of λ that causes QP_λ to have the same solution; e.g., see point ①. Point ② coincides with the solution of BP_σ with $\sigma = \|b\|_2$ and of LS_τ with $\tau = 0$; point ③ coincides with the solution of BP_σ with $\sigma = 0$ and of LS_τ with $\tau = 0$. Left- and right-hand limits can be used to define the value of λ at points ② and ③. The relevance of this curve in the seismic context is discussed by [34].

The Pareto curve can be interpreted as the graph of the value function

$$\phi(\tau) = \inf_{\|x\|_1 \le \tau} \{\|Ax - b\|_2\}.$$

Let x_τ be the optimal solution of LS_τ, and let $r_\tau = b - Ax_\tau$ be the corresponding residual. Let $\bar{\tau}$ be the smallest value of τ at which the graph first touches the horizontal axis. (This is guaranteed if A has full rank.) The function ϕ and the Pareto curve is characterized by the following theorem, due to van den Berg and Friedlander [61].

Theorem 3. *Suppose that A is full rank. Then*

1) *The function ϕ is convex and nonincreasing.*
2) *For all $\tau \in (0, \bar{\tau})$, ϕ is continuously differentiable, $\phi'(\tau) = -\lambda_\tau$, where $\lambda_\tau = \|A^H y_\tau\|_\infty$ and $y_\tau = r_\tau / \|r_\tau\|_2$.*
3) *For $\tau \in [0, \bar{\tau}]$, $\|x_\tau\|_1 = \tau$, and ϕ is strictly decreasing.*

The solid curve in Fig. 9(a) graphs the Pareto curve for a seismic interpolation problem similar to that shown in Fig. 3.

Although QP_λ has proven to be the most used approach, it is generally not clear how to choose the parameter λ such its solution gives a desired misfit. This difficulty is illustrated by Fig. 9(b). The solid (black) curve is the true Pareto curve; the three solid (red) dots are solution/residual pairs of QP_λ that correspond to equally spaced values of λ between $\|A^H b\|_\infty$ and 0. This is typical behavior: even though the values of λ are equally spaced, the resulting samples are not at all equally spaced, and a quadratic interpolation/extrapolation (dotted red line) based on these samples severly underestimates the curve. Similar behavior in other problems that directly weight competing objects has been objserved by [17], [40]. On the other hand, the circles (blue) are solution/residual pairs of BP_σ for equally spaced samples of σ between $\|b\|_2$ and 0 yields good coverage, and an estimate of the curve based on these samples (blue solid line) closely approximates the true curve.

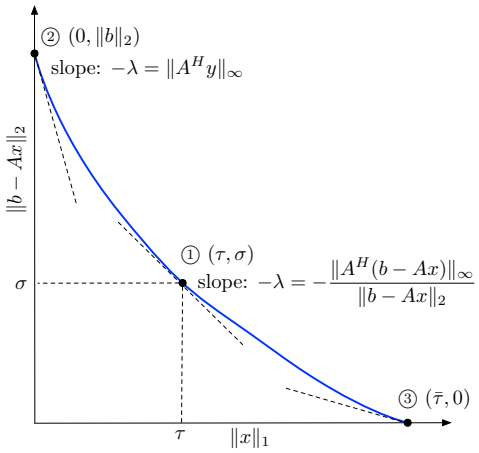
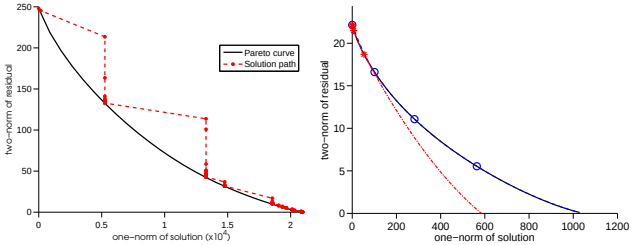


Fig. 8: (Adapted from [34].) Schematic illustration of a Pareto curve. Point ① exposes the connection between the three parameters of QP_λ, BP_σ, and LS_τ. Point ③ corresponds to a solution of BP_σ with $\sigma = 0$.



(a) (b)

Fig. 9: (Adapted from [61].) (a) A typical Pareto curve, and the path taken by the SPGL1 algorithm; (b) approximating the Pareto curve from a few samples.

D. Pareto root-finding

The SPGL1 approach for BP_σ is based on approximately solving a sequence of subproblems LS_τ, using a spectral PG method [4]; at each iteration k it refines the estimate τ_k such that $\tau_k \to \tau_\sigma$, which causes LS_τ and BP_σ to share a solution. The sequence of estimates τ_k is derived by simply applying

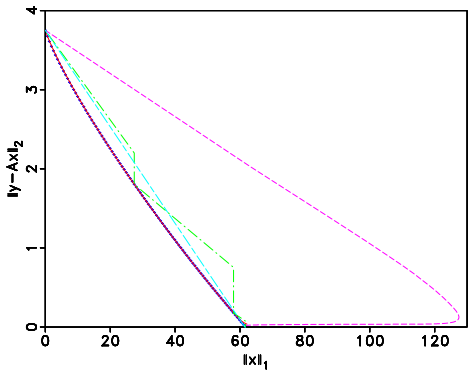


Fig. 10: (Adapted from [34].) Pareto curve and solution paths of four solvers for a BP_σ, with $\sigma = 0$. The symbols + represent a sampling of the Pareto curve. The solid (—) line, obscured by the Pareto curve, is the solution path of IST with cooling, the chain (– · –) line the path of SPGL1, the dashed (– –) line the path of IST, and the dotted $(\cdots)$ line the path of IRLS.

Newton's method to find a root of the nonlinear equation

$$\phi(\tau) = \sigma.$$

Theorem 3 is central to this approach because it describes how the gradient of ϕ, needed for Newton's method, is related to the solutions of the LS_τ subproblems. Practical refinements are needed that allow for LS_τ to be solved only approximately [61, §3], [62, §3]. Fig. 10 shows how SPGL1 and IST (with cooling) closely follow the Pareto curve; however, SPGL1 requires significantly fewer matrix multiplies.

VI. Discussion and conclusions

We discussed possible adaptations of CS to solve outstanding problems in exploration seismology including measures to make acquisition and computations more efficient. The presented results illustrate that we are at the cusp of exciting new developments where acquisition and processing workflows are not hampered by the fear of creating coherent artifacts related to periodic subsampling. Instead, we arrive at a workflow with control over these artifacts. This is accomplished by the following three new design principles, and the slogan "randomize, sparsify, and convexify":

- randomize—break coherent aliases by introducing randomness, e.g., by designing randomly perturbed acquisition grids, or by designing randomized simultaneous sources;
- sparsify—use sparsifying transforms in conjunction with sparsity-promoting programs that separate signal from subsampling artifacts, and that restore amplitudes;
- convexify—relax difficult combinatorial problems into tractable convex optimization problems.

The potential benefits of CS are real and significant. But to realize them, several obstacles need to be surmounted, including the need to overcome the inertia of entrenched engineering practices, and adapting the theoretical framework to practical acquisition schemes and workflows for imaging and inversion.

Field-data acquisition presents an especially thorny problem. It needs to address physical limitations on the characteristics of sources—e.g., limited control over source signatures of airguns—and on the repeatability and placement accuracy of the sources. In addition, analog-to-digital conversion by the hydrophones and geophones limits their dynamic range, which can have detrimental effects on the recovery quality, unless special measures are taken [31].

On the field and computational sides, we are also faced with the fact that we are not able to use measurement matrices that we know are CS-compliant, but those that are imposed on us by the nature of the application. Often, these matrices do not—or cannot be verified to—satisfy the theortical requirements of CS. But as the seismic examples have illustrated, there is strong evidence that the principles of CS continue to apply outside of its formal theoreical framework. Until our theoretical tools can be extended to cover the situations that we envisage, we continue to rely on the insight provided by the current framework [38], [44].

The seismic application of CS and its extensions rely on solving extremely large system of equations that arise from the physical setting of exploration seismology. This puts pressures on developing large-scale solvers that can handle massive data volumes with billions of unknowns that are omnipresent in industrial exploration seismology. Unless we develop sparsity-promoting solvers that limit the number of passes through the data, widespread adaptation of CS to exploration seismology stay beyond the computational capabilities of industry. However, as the extensions of CS presented in this paper indicate, the ideas of CS can be incorporated into imaging and inversion. This allows us to work in model space, which is significantly smaller in dimension than data space. In addition, integration of CS-inspired dimensionality-reduction ideas may allow us to invert data as its acquired, which may be more efficient and cost effective.

VII. Acknowledgments

We are grateful to Nick Moldoveanu (WesternGeco) for making the coil dataset available and Charles Jones (BG) for providing us with the BG Compass velocity model. We also would like to thank Haneet Wason, Hassan Mansour, Tim Lin, and Xiang Li for preparing the figures. This publication was prepared using CurveLab, a toolbox implementing the Fast Discrete Curvelet Transform, WaveAtom a toolbox implementing the Wave Atom transform, Madagascar, a package for reproducible computational experiments, SPGL1, and SPOT, a suite of linear operators for sparse signal reconstruction. The authors were financially supported by CRD Grant DNOISE 334810-05. The industrial sponsors of the Seismic Laboratory for Imaging and Modelling BG Group, BP, Chevron, Petrobras, and WesternGeco are also gratefully acknowledged.

Authors

Felix J. Herrmann received in 1997 a Ph.D. degree in Engineering Physics from the Delft University of Technology, the Netherlands. He was a visiting scholar at Stanford in 1998, a post-doctoral fellow at MIT's Earth Resources Laboratory

from 1999 to 2002, and was a senior Fellow at the UCLA's Institute for Pure and Applied Mathematics. He is currently associate professor at the Department of Earth & Ocean Sciences at the University of British Columbia where he founded the Seismic Laboratory for Imaging and Modeling (SLIM). He is interested in theoretical and applied aspects of (exploration) reflection seismology. He runs the industry-supported research program SINBAD and is the PI of the DNOISE project. He serves on the editorial boards of the Journal of Applied Geophysics and Geophysical Prospecting. He is also a member of the EAGE, CSEG, SEG, SIAM, and AGU.

Michael P. Friedlander is Associate Professor of Computer Science at the University of British Columbia. He received his PhD in Operations Research from Stanford University in 2002, and his BA in Physics from Cornell University in 1993. From 2002 to 2004 he was the Wilkinson Fellow in Scientific Computing at Argonne National Laboratory. He was a senior fellow at UCLA's Institute for Pure and Applied Mathematics in 2010. He serves on the editorial boards of SIAM J. on Optimization, SIAM J. on Matrix Analysis and Applications, Mathematical Programming Computation, Optimization Methods and Software, and the Electronic Transactions on Numerical Analysis. His research is primarily in developing numerical methods for constrained optimization, their software implementation, and applications in signal processing and image reconstruction.

Özgür Yılmaz received the B.Sc. degrees in mathematics and in electrical engineering from Bogazii University, Istanbul, Turkey, in 1997, and the Ph.D. degree in Applied and Computational Mathematics from Princeton University in 2001. From 2002 to 2004, he was an Avron Douglis Lecturer at the University of Maryland, College Park. He joined the Mathematics Department at the University of British Columbia (UBC), Vancouver, BC, Canada, in 2004, where he is currently an Associate Professor. He is a member of the Institute of Applied Mathematics (IAM) and and an associate member of the Institute for Computing, Information, and Cognitive Systems (ICICS), UBC. His research interests include applied harmonic analysis, information theory, and signal processing.

REFERENCES

[1] Fixed-point continuation applied to compressed sensing: implementation and numerical experiments. *J. Comp. Math.*, 28(2):170194, 2010.

[2] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28:253–263, 2008. 10.1007/s00365-007-9003-x.

[3] DP Bertsekas and JN Tsitsiklis. *Neuro-Dynamic Programming (Belmont, MA: Athena Scientific)*. 1996.

[4] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. 10(4):1196–1211, 2000.

[5] T. Blumensath and M.E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[6] A. Bourgeois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg. *The Marmousi experience*, volume 5-9, chapter Marmousi data and model. EAGE, 1991.

[7] C. Bunks, F. Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

[8] E. J. Candès and L. Demanet. Curvelets and fourier integral operators. *C. R. Acad. Sci. Paris*, 2003.

[9] E. J. Candes, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5(3):861–899, 2006.

[10] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[11] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406 –5425, dec. 2006.

[12] Emmanuel J. Cands, Yonina C. Eldar, Deanna Needell, and Paige Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59 – 73, 2011.

[13] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters, IEEE*, 14(10):707–710, 2007.

[14] R. Chartrand and Wotao Yin. Iteratively reweighted algorithms for compressive sensing. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3869 –3872, 31 2008-april 4 2008.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. 20(1):33–61, 1998.

[16] Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. 4(4):1168–1200, 2005.

[17] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct. Optim.*, 14:63–69, 1997.

[18] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57:1413–1457, 2004.

[19] I. Daubechies, M. Fornasier, and I. Loris. Accelereated projected gradient method for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl.*, 2007. To appear.

[20] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Gntrk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010.

[21] L. Demanet and L. Ying. Wave atoms and sparsity of oscillatory patterns. *Applied and Computational Harmonic Analysis*, 23(3):368–387, 2007.

[22] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[23] D.L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[24] Matthias Chung Eldad Haber and Felix J. Herrmann. An effective method for parameter estimation with pde constraints with multiple right hand sides. Technical Report TR-2010-4, UBC-Earth and Ocean Sciences Department, 2010.

[25] Mike Fehler and Ken Larner. Seg advanced modeling (seam): Phase i first year update. *The Leading Edge*, 27(8):1006–1007, 2008.

[26] Xiang Li Felix J. Herrmann. Efficient least-squares migration with sparsity promotion. EAGE, EAGE Technical Program Expanded Abstracts, 2011.

[27] M. Figueiredo, R. Nowak, and S. J. Wright. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *Sel. Top. in Signal Process., IEEE J.*, 1(4):586–597, 2007.

[28] Mário A. T. Figueiredo and Robert D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Sign. Proc.*, 12(8):906–916, August 2003.

[29] S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, 2009.

[30] M. P. Friedlander and M. Schmidt. Hybrid deterministic-stochastic methods for data fitting. Tech. rep., Department of Computer Science, University of British Columbia, Vancouver, April 2011.

[31] C. Sinan Güntürk, Alexander M. Powell, Rayan Saab, and Ö. Yilmaz. Sobolev duals for random frames and sigma-delta quantization of compressed sensing measurements. *CoRR*, abs/1002.0182, 2010.

[32] G. Hennenfent and F. J. Herrmann. Seismic denoising with non-uniformly sampled curvelets. *IEEE Comp. in Sci. and Eng.*, 8(3):16–25, 2006.

[33] G. Hennenfent and F. J. Herrmann. Simply denoise: wavefield reconstruction via jittered undersampling. *Geophysics*, 73(3), May-June 2008.

[34] Gilles Hennenfent, Ewout van den Berg, Michael P. Friedlander, and Felix J. Herrmann. New insights into one-norm solvers from the pareto curve. *Geophysics*, 73(4):A23–26, 2008.

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

[36] F. J. Herrmann, Y. A. Erlangga, and T.T.Y. Lin. Compressive simultaneous full-waveform simulation. *Geophysics*, 74:A35, 2009.

[37] F. J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 2008. doi:10.1111/j.1365-246X.2007.03698.x.

[38] Felix J. Herrmann. Randomized sampling and sparsity: Getting more information from fewer samples. *Geophysics*, 75(6):WB173–WB187, 2010.

[39] Felix J. Herrmann, Peyman P. Moghaddam, and Chris C. Stolk. Sparsity- and continuity-promoting seismic imaging with curvelet frames. *Journal of Applied and Computational Harmonic Analysis*, 24(2):150–173, 2008. doi:10.1016/j.acha.2007.06.007.

[40] S. Leyffer. A note on multiobjective optimization and complementarity constraints. Preprint ANL/MCS-P1290-0905, Mathematics and Computer Science Division, Argonne National Laboratory, Illinois, September 2005.

[41] Xiang Li, Aleksandr Aravkin an Tristan van Leeuwen, and Felix J. Herrmann. Modified gauss-newton with sparse updates. SBGF, 2011.

[42] Xiang LI and Felix J. Herrmann. Full-waveform inversion from compressively recovered model updates. volume 29, pages 1029–1033. SEG, 2010.

[43] S. G. Mallat. *A Wavelet Tour of Signal Processing: the Sparse Way*. Academic Press, 2009.

[44] Hassan Mansour, Haneet Wason, Tim Lin, and Felix J. Herrmann. A compressive sensing perspective on simultaneous marine acquisition. SBGF, SEG Technical Program Expanded Abstracts, 2011.

[45] Nick Moldoveanu. Random sampling: A new strategy for marine acquisition. *SEG Technical Program Expanded Abstracts*, 29(1):51–55, 2010.

[46] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, April 1995.

[47] D. Needell and J.A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

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

[49] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[50] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. *Doklady AN SSSR*, 269, 1983. translated as Soviet Math. Dokl.

[51] C. C. Paige and M. A. Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM TOMS*, 8(1):43–71, 1982.

[52] L. A. Romero, D. C. Ghiglia, C. C. Ober, and S. A. Morton. Phase encoding of shot records in prestack migration. *Geophysics*, 65(2):426–436, 2000.

[53] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[54] R. Saab and O. Yilmaz. Sparse recovery by non-convex optimization – instance optimality. *Applied and Computational Harmonic Analysis*, 29(1):30–48, 2010.

[55] M. D. Sacchi, T. J. Ulrych, and C. J. Walker. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE Transactions on Signal Processing*, 46(1):31 – 38, January 1998.

[56] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, 58(1):267–288, 1997.

[57] Felix Herrmann Tim T.Y. Lin. Estimating primaries by sparse inversion in a curvelet-like representation domain. EAGE, EAGE Technical Program Expanded Abstracts, 2011.

[58] Michael Friedlander Tristan van Leeuwen, Mark Schmidt and Felix Herrmann. A hybrid stochastic-deterministic optimization method for waveform inversion. EAGE, EAGE Technical Program Expanded Abstracts, 2011.

[59] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[60] E. van den Berg and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction. Available at http://www.cs.ubc.ca/labs/scl/index.php/Main/Spgl1, June 2007.

[61] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. Technical Report TR-2008-01, UBC Computer Science Department, January 2008.

[62] E. van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. Tech. Rep. TR-2010-02, Department of Computer Science, University of British Columbia, Vancouver, January 2010. To appear in SIAM J. Optimization.

[63] S. Xu, Y. Zhang, D. Pham, and G. Lambare. Antileakage Fourier transform for seismic data regularization. *Geophysics*, 70(4):V87 – V95, 2005.