Empirical recovery conditions for seismic sampling

Felix J. Herrmann, University of British Columbia, Canada

SUMMARY

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 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 acquisition. 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.

COMPRESSIVE-SENSING DESIGN AND ASSESSMENT

In summary, according to the theory of compressive sensing (CS - Candès et al., 2006; Donoho, 2006), the recovery of a signal with a sparse transform-domain representation is guaranteed when two conditions are met, namely 1) the transform-domain vector is sufficiently sparse and 2) the subsampling artifacts are incoherent, which is a direct consequence of measurements with a sampling matrix whose action mimics that of a Gaussian matrix. The first condition requires that the energy of the signal is well concentrated in the sparsifying domain. The second condition of incoherence requires understanding of interactions between the sparsifying transform, the basis of the domain in which the samples are taken, and the sampling operator. 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: **I: 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; **II: 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; **III: 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 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.

SPARSE RECOVERY

After organizing high-resolution samples into the vector $\mathbf{f} := \{f[q]\}_{q=0\cdots N-1} \in \mathbb{R}^N$, we arrive at the following expression for compressively-sampled measurements:

$$\mathbf{b} = \mathbf{RMf} \in \mathbb{R}^n \quad \text{with} \quad n \ll N. \tag{1}$$

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

$$\mathbf{A} := \overbrace{\mathbf{R}}^{\text{restriction}} \underbrace{\mathbf{M}}_{\text{measurement}} \overbrace{\mathbf{S}^H}^{\text{synthesis}} \tag{2}$$

with a sparsity-promoting program—i.e., $\widetilde{\mathbf{f}} = \mathbf{S}^H \widetilde{\mathbf{x}}$ with

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 := \sum_{p=0}^{P-1} |x[i]| \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}. \tag{3}$$

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 $P \geq N$. Our hope is—and there is unfortunately no formal proof—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 $\widetilde{\mathbf{f}} \in \mathbb{R}^N$.

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

$$\sigma(k) = \|\mathbf{f} - \mathbf{f}_k\| = \mathscr{O}(k^{1/2-s}), \tag{4}$$

with $\mathbf{f}_k$ the reconstruction from the largest k - coefficients. 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 these 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 (Candes et al., 2006) 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 $\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{S}^H\mathbf{x} = \mathbf{f}$. 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

$$\text{SNR}(\rho) = 20\log\frac{\|\mathbf{f} - \mathbf{f}_\rho\|}{\|\mathbf{f}\|}. \tag{5}$$

The smaller this ratio, the more coefficients we ignore and the sparser the transform-coefficient vector becomes. 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.

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 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.

Our results are summarized in Figure 1 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.

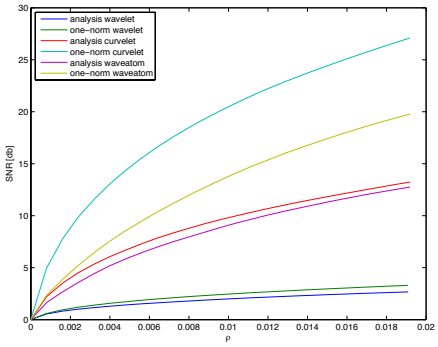

Figure 1: Signal-to-noise ratios (SNRs) for the nonlinear approximation errors of the common-receiver. The SNRs are plotted as a function of the sparsity ratio $\rho \in (0, 0.02]$. The plots include curves for the errors obtained from the analysis and one-norm minimized synthesis coefficients. Notice the significant improvement in SNRs for the synthesis coefficients obtained by solving a ℓ_1-norm problem.

SUBSAMPLING OF SHOTS

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., $\mathbf{I} := \mathbf{I}_{N_s} \otimes \mathbf{I}_{N_t}$, 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 $\mathbf{M} := \mathbf{G}_{N_s} \otimes \mathbf{I}_{N_t}$ 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 $\mathbf{R} := \mathbf{R}_{n_s} \otimes \mathbf{I}_{N_t}$ 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 mutual coherence, which 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 sources missing. Examples of both measurements for the real common-receiver gather are plotted in Figure 2(a) and 2(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 3 from data plotted in Figures 2(a) and 2(b).

Despite the fact that simultaneous acquisition with all sources firing simultaneously may not be easily implementable in practice, this approach has been applied successfully to reduce simulation and imaging costs (Herrmann et al., 2009; Herrmann, 2009; Lin and Herrmann, 2009a,b). In the "eyeball norm", the recovery from the simultaneously data is, as expected, clearly superior (cf. Figures 2(c) and 2(d)). This behavior 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.

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-source gather 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

$$\mathrm{SNR}(\delta) = 20 \log \frac{\|\mathbf{f} - \widetilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|}, \tag{6}$$

where $\widetilde{\mathbf{f}}_\delta = \mathbf{S}^H \widetilde{\mathbf{x}}_\delta$ and $\widetilde{\mathbf{x}}_\delta = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}_\delta \mathbf{x} = \mathbf{b}$. 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 $\mathbf{A}_\delta := \mathbf{R}_\delta \mathbf{M}_\delta \mathbf{S}^H$, where $\mathbf{R}_\delta := \mathbf{R}_{n_s} \otimes \mathbf{I}_{N_t}$ with $\delta = n_s/N_s$. For each simultaneous experiment, we also generate different realizations of the measurement matrix $\mathbf{M} := \mathbf{G}_{N_s} \otimes \mathbf{I}_{N_t}$.

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 3. From these plots it becomes immediately clear that simultaneous acquisition greatly improves 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 1, 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

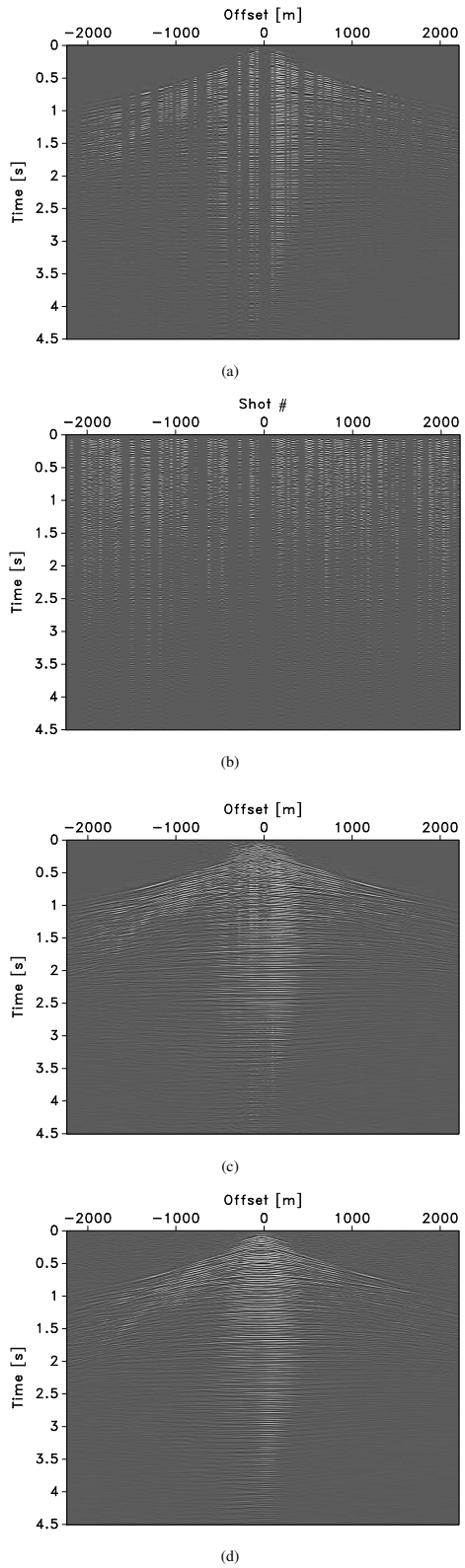

Figure 2: Recovery from a compressively-sampled common receiver gather with 50% ($\delta = 0.5$) of the sources missing. **(a)** Receiver gather with sequential shots selected uniformly at random. **(b)** The same but for random simultaneous shots. **(c)** Recovery from incomplete data in **(a)**. **(d)** The same but now for the data in **(b)**. Notice the remarkable improvement in the recovery from simultaneous data.

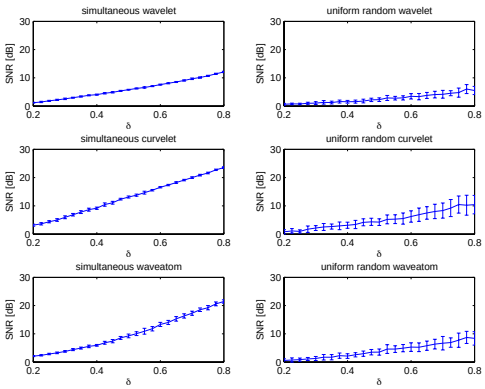

Figure 3: SNRs (cf. Equation 6) for nonlinear sparsity-promoting recovery from compressively sampled data with $20\% - 80\%$ of the sources missing ($\delta \in [0.2, 0.8]$). The results summarize 25 experiments for 25 different values of $\delta \in [0.2, 0.8]$. The plots include estimates for the standard 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.

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 5). 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 $\overline{\text{SNR}}(\delta)$. Then, we find the smallest ρ for which the nonlinear recovery SNR is greater or equal to $\overline{\text{SNR}}(\delta)$. 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 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 $\delta \in [0.2, 0.8]$

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\text{SNR}}(\delta) \leq \text{SNR}(\tilde{\rho})\}. \quad (7)$$

When the sampling ratio approaches one from below ($\delta \to 1$), the data becomes 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 only 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 4. 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 then 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 3), the results for the recovery from missing sequential sources are less clear. General trends predicted by CS are also observable for this type of acquisition, but the performance is significantly worse then for recovery with simultaneous sources. Finally, the observed oversampling ratios are reasonable for both curvelet and wave atoms.

CONCLUSIONS

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. Our observation of significant 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 steps has profound implications for exploration seismology that include: a new randomized sampling paradigm, where the cost of acquisition are no longer determined 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 work was financed through the SINBAD II project with support from BG, BP, Chevron, Petrobras, and WesternGeco and was partly supported by a NSERC Discovery Grant (22R81254) and CRD Grant DNOISE (334810-05).

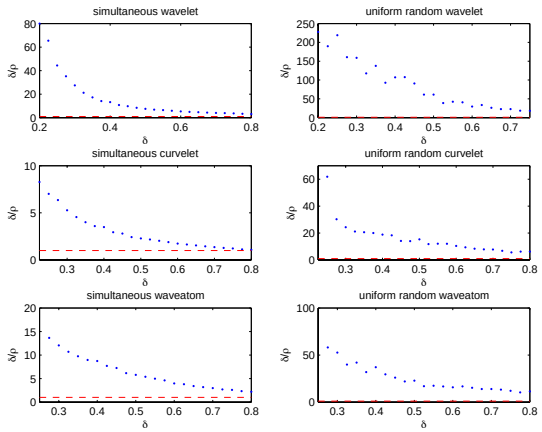

Figure 4: Oversampling ratio δ/ρ as a function of the sampling ratio δ (cf. Equation 7) for sequential- and simultaneous-source experiments. As expected, the overhead is smallest for simultaneous acquisition and curvelet-based recovery.