Compressed sensing with prior information

Özgür Yılmaz

Department of Mathematics
University of British Columbia

SINBAD Consortium Meeting 2011

December 5, 2011

Compressed sensing: A revolution in sampling theory

- During the last 7 years, we have been witnessing a **revolution in sampling theory**.
- **Main conclusion:** sparse signals can be recovered from very few, "what appears to be incomplete" measurements in a tractable way.
- Initiated by the works of Donoho, and of Candès and Tao ($\sim$ 2004).
- Opened up a new field called **compressed sensing** or **compressive sampling**: Very active area. To follow:

  Compressive sensing resources at http://dsp.rice.edu/cs

  Nuit-Blanche Blog at http://nuit-blanche.blogspot.com
- Relies heavily on the theory of **sparse approximations** – around for more than two decades (transforms such as wavelets, curvelets, Gabor,...).
- Interesting and difficult mathematics and important applications such as seismic signal processing, imaging, and inversion.

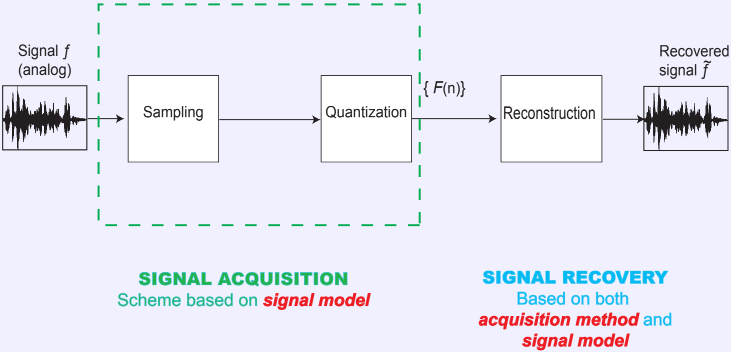
Sampling and Reconstruction: Big Picture

Inherently analog signals: Audio, images, seismic, etc.
Objective: Use digital technology to store and process analog signals – find efficient digital representation of analog signals.

Sampling and Reconstruction: Big Picture

Inherently analog signals: Audio, images, seismic, etc.
Objective: Use digital technology to store and process analog signals – find efficient digital representation of analog signals.

(Generalized) Sampling of Analog Signals



Outline of the talk

- Sampling theory and compressed sensing – an overview
- The use of prior information in compressed sensing
  - in the reconstruction stage: recovery via weighted ℓ_1 minimization
  - in the sampling stage: adaptive compressed sensing

Sampling and Reconstruction

Objectives:

- **Sampling scheme.** Specify how to obtain finitely many measurements of the signal f from which one can "recover" f. That is, the acquired samples should contain sufficient information to recover/approximate f.
- **Quantization.** Specify how to digitize the sample values (crucial for A/D conversion) in a way that is robustly implementable in analog hardware.
- **Reconstruction scheme.** Specify how to recover f from the samples.

Sampling and Reconstruction

Objectives:

- **Sampling scheme.** Specify how to obtain finitely many measurements of the signal f from which one can "recover" f. That is, the acquired samples should contain sufficient information to recover/approximate f.
- **Quantization.** Specify how to digitize the sample values (crucial for A/D conversion) in a way that is robustly implementable in analog hardware.
- **Reconstruction scheme.** Specify how to recover f from the samples.

Main Problems and Challenges

- Must model the signals of interest, e.g., bandlimited, sparse etc... Note that without modeling, there is no hope of a "sampling theory".
- Specify when we have exact recovery.
- When we don't have exact recovery, tie the resolution of the approximation to the sampling density (i.e., grid size, total number of samples etc.).
- Quantization has its own challenges, e.g., see work by Saab, OY et al.
- In any case, the schemes must be practicable.

Classical Sampling Theorem

We all now the “classical sampling theorem” of Shannon, Nyquist, Whittaker, Kotelnikov, Ogura, Borel, even Cauchy...

Signal model. The set of all bandlimited functions with bandlimit Ω – denote this set by B_Ω.

Sampling scheme. Collect values of $f \in B_\Omega$ on a sufficiently dense uniform grid, i.e., $\{f(n\tau) : \ n = \dots, -2, -1, 0, 1, 2, \dots\}$. Specifically, $\tau < \frac{1}{2\Omega}$.

Reconstruction scheme. Exact reconstruction via

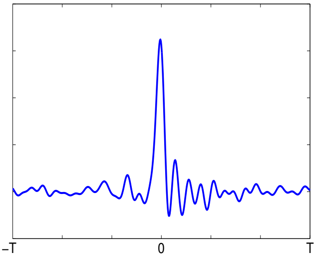
$$f(t) = \sum_n f(n\tau)\phi(t - n\tau), \quad \forall t.$$

Here ϕ is an appropriate low-pass filter.

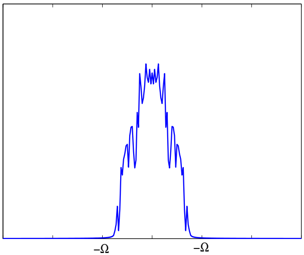
Practicability. If we slightly oversample, we can use a filter ϕ with fast decay, so obtain local reconstruction. This way, we also get robustness w.r.t. noise.

Classical Sampling Theorem: The picture

A bandlimited f

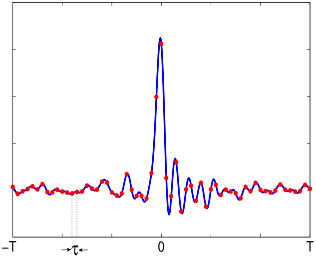


Fourier transform of f

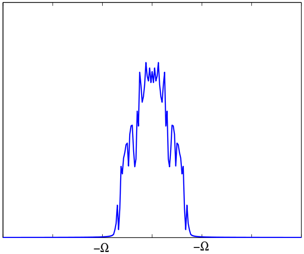


Classical Sampling Theorem: The picture

A bandlimited f



Fourier transform of f



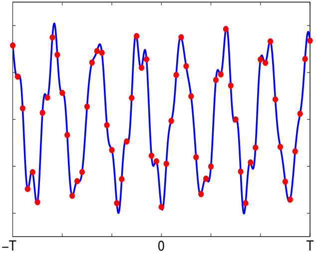
Need $N \approx 2\Omega \times 2T$ samples to reconstruct f on $[-T, T]$.
Equivalently: Every bandlimited function $f \in B_\Omega$, restricted to $[-T, T]$, can be represented by a vector $\mathbf{f} \in \mathbb{R}^N$ which we obtain by collecting N **measurements.**

Compressive Sampling Theory

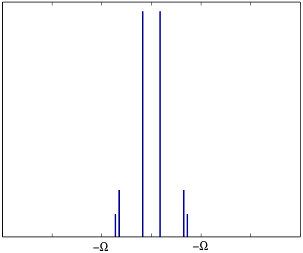
Above: Reduced a bandlimited function f to a vector $\mathbf{f}$ in $\mathbb{R}^N$.

Question: Can we reduce the dimensionality of the problem by **restricting the signal class further**?

Another bandlimited f



Fourier transform of f



Do we still need $N \approx 4\Omega T$ samples to reconstruct $\mathbf{f} \in \mathbb{R}^N$?

Compressive Sampling Theory

Rephrase the question. Suppose we have:

Signal model. $f \in B_\Omega$ and f has a **sparse Fourier transform**.

Still need to sample at the Nyquist rate for a good (perfect) reconstruction?

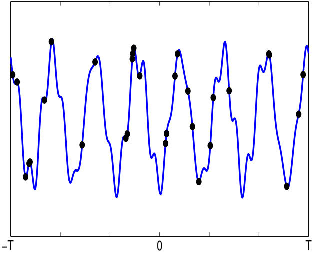
Compressive Sampling Theory

Rephrase the question. Suppose we have:

Signal model. $f \in B_\Omega$ and f has a **sparse Fourier transform**.

Still need to sample at the Nyquist rate for a good (perfect) reconstruction?

New sampling scheme. Consider the following set of $m < N$ samples at (random) irregular points.



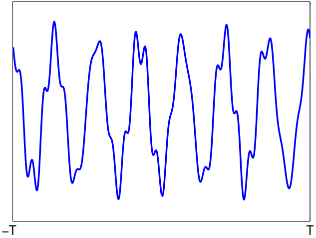
Average sampling density is **only 50% of Nyquist rate**, i.e., $m \approx N/2$.

Claim: We can recover f from these samples!

Recovery scheme. Find signal with matching samples that has the "sparsest" Fourier transform.

Compressive Sampling Theory

Here is the reconstruction obtained from the above samples (approx. 50% of Nyquist rate)



- We get essentially perfect reconstruction!
- How did we reconstruct? Next...

Compressive Sampling Theory – general framework

- Signal $f \in \mathbb{R}^N$. We want to collect information on f (*in the example: f is the full signal.*).
- **Model the signal class:** f admits a sparse representation w.r.t. a known basis B: $f = B^* x$ where x is sparse. (*in the example: B is the Fourier basis.*)
- **Specify a measurement scheme:** Construct an $m \times N$ measurement matrix M with $m \ll N$

$$\mathbf{f}_{\text{meas}} = Mf = MB^* x$$

(*in the example: $\mathbf{f}_{meas}$ is the vector of non-uniform samples and M is the random restriction matrix in the example.*)
- **Reconstruction method:** Solve the underdetermined sparse recovery problem:

$$x_{\text{approx}} = \text{“sparsest” } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{"sparsest" } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{“sparsest” } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Compressive Sampling Theory - sparsity transforms

First address question 1: How do we find sparsity transforms?

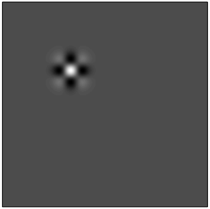
- Note that this is dependent heavily on the class of signals of interest.
- In the above example, the sparsity transform was Fourier transform.
- Applied and computational harmonic analysis community has been developing such transforms during the last three decades that are tailored to important signal classes such as: audio, natural images, seismic data and images.
- Rich area with interesting mathematics, directly applicable constructive results such as wavelet transform, curvelet transform etc.
- Next, we give examples of some important sparsity transforms.

Sparsity transform - natural images

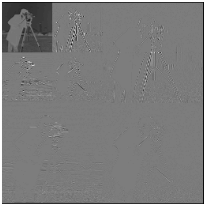
Wavelet transform sparsifies natural images.

image

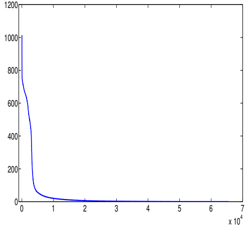
a wavelet atom



wavelet transform



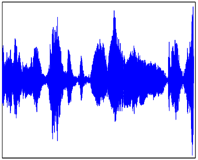
sorted coefficients



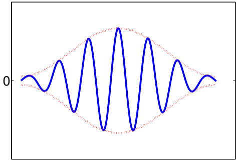
Sparsity transform - audio

Short-time Fourier (Gabor) transform sparsifies audio signals.

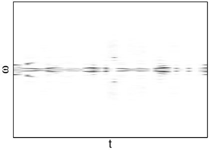
audio signal



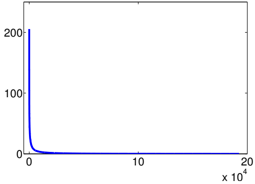
a Gabor atom



STFT transform



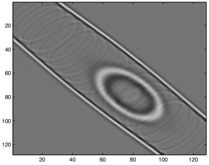
sorted coefficients



Sparsity transform - seismic

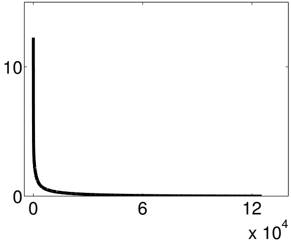
Curvelet transform sparsifies seismic data and images.

sampled Green's function



a curvelet atom

sorted coefficients



Compressive Sampling Theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{"sparsest" } z \text{ such that } \mathbf{f}_{\text{meas}} = MB^* z.$$

Main questions:

1. How do we find the sparsifying basis B?
2. How do we construct the measurement matrix M?
3. How many measurements do we need to have $x_{\text{approx}} = x$?
4. How do we solve the sparse recovery problem?

Reconstruction: sparse recovery problem

Want to reconstruct f from the measurements

$$b = Mf = \underbrace{MB^*}_{A} x.$$

Some design goals:

1. **Exact recovery for all sufficiently sparse signals.** Want to recover **every** k**-sparse** x from the measurements $b = Ax$ with n as small as possible (say we fix k, N).
2. **Close to the best** k**-term approximation for compressible signals.** Want good estimates if x is not sparse but can be well-approximated by a sparse signal.
3. **Robustness to noise in either case above.** Want good estimates when the measurements are contaminated by noise, i.e., $\hat{b} = Ax + e$ where e is additive noise.
4. **Computationally tractable recovery method.**

CS – many surprises since 2004!

We can achieve all the goals above (main results by Donoho, and Candes, Romberg, Tao) – just use a recovery algorithm based on ℓ_1 minimization:

$$\Delta_1(b) := \arg\min \|z\|_1 \;\; \text{subject to } Az = b \qquad \text{no noise case}$$

$$\Delta_1^\epsilon(\hat{b}) := \arg\min \|z\|_1 \;\; \text{subject to } \|Az - b\|_2 \le \epsilon \qquad \text{noisy case}$$

CS – many surprises since 2004!

We can achieve all the goals above (main results by Donoho, and Candes, Romberg, Tao) – just use a recovery algorithm based on ℓ_1 minimization:

$$\Delta_1(b) := \arg\min \|z\|_1 \text{ subject to } Az = b \qquad \text{no noise case}$$

$$\Delta_1^\epsilon(\hat{b}) := \arg\min \|z\|_1 \text{ subject to } \|Az - b\|_2 \leq \epsilon \qquad \text{noisy case}$$

In particular:

- If $A \in \mathbb{R}^{n \times N}$ is a sufficiently "incoherent matrix" and k is sufficiently small

$$\Delta_1(b) = x, \quad \text{i.e., exact recovery, } \textbf{for every } k\textbf{-sparse } x\textbf{.}$$

CS – many surprises since 2004!

We can achieve all the goals above (main results by Donoho, and Candes, Romberg, Tao) – just use a recovery algorithm based on ℓ_1 minimization:

$$\Delta_1(b) := \arg\min \|z\|_1 \ \text{ subject to } Az = b \qquad \text{no noise case}$$

$$\Delta_1^\epsilon(\hat{b}) := \arg\min \|z\|_1 \ \text{ subject to } \|Az - b\|_2 \leq \epsilon \qquad \text{noisy case}$$

In particular:

- If $A \in \mathbb{R}^{n \times N}$ is a sufficiently "incoherent matrix" and k is sufficiently small

$$\Delta_1(b) = x, \quad \text{i.e., exact recovery, } \textbf{for every } k\textbf{-sparse } x\textbf{.}$$

- For such A, Δ_1 provides a good approximation for arbitrary $x \in \mathbb{R}^N$:

$$\|x - \Delta_1(b)\| \lesssim \sigma_k(x)_{\ell_1}/\sqrt{k}, \quad \text{i.e., good recovery for compressible } x.$$

CS – many surprises since 2004!

We can achieve all the goals above (main results by Donoho, and Candes, Romberg, Tao) – just use a recovery algorithm based on ℓ_1 minimization:

$$\Delta_1(b) := \arg\min \|z\|_1 \;\; \text{subject to } Az = b \qquad \text{no noise case}$$

$$\Delta_1^\epsilon(\hat{b}) := \arg\min \|z\|_1 \;\; \text{subject to } \|Az - b\|_2 \leq \epsilon \qquad \text{noisy case}$$

In particular:

- If $A \in \mathbb{R}^{n \times N}$ is a sufficiently "incoherent matrix" and k is sufficiently small

  $$\Delta_1(b) = x, \;\; \text{i.e., exact recovery, } \textbf{for every } k\textbf{-sparse } x\textbf{.}$$

- For such A, Δ_1 provides a good approximation for arbitrary $x \in \mathbb{R}^N$:

  $$\|x - \Delta_1(b)\| \lesssim \sigma_k(x)_{\ell_1}/\sqrt{k}, \;\; \text{i.e., good recovery for compressible } x.$$

- For such A, the recovery results above stay within noise level if the measurements are contaminated by noise.

CS – many surprises since 2004!

We can achieve all the goals above (main results by Donoho, and Candes, Romberg, Tao) – just use a recovery algorithm based on ℓ_1 minimization:

$$\Delta_1(b) := \arg\min \|z\|_1 \text{ subject to } Az = b \qquad \text{no noise case}$$

$$\Delta_1^\epsilon(\hat{b}) := \arg\min \|z\|_1 \text{ subject to } \|Az - b\|_2 \le \epsilon \qquad \text{noisy case}$$

In particular:

- If $A \in \mathbb{R}^{n\times N}$ is a sufficiently "incoherent matrix" and k is sufficiently small

  $$\Delta_1(b) = x, \quad \text{i.e., exact recovery, } \textbf{for every } k\textbf{-sparse } x\textbf{.}$$

- For such A, Δ_1 provides a good approximation for arbitrary $x \in \mathbb{R}^N$:

  $$\|x - \Delta_1(b)\| \lesssim \sigma_k(x)_{\ell_1}/\sqrt{k}, \quad \text{i.e., good recovery for compressible } x.$$

- For such A, the recovery results above stay within noise level if the measurements are contaminated by noise.

There are other algorithms for CS recovery—e.g., Δ_p with $0 < p < 1$, OMP, CoSamp, . . .

How to choose the measurement matrix

- There are precise conditions on A (in terms of its RIP constants) that guarantee that the above results hold.
- For example, if A is a random matrix with iid Gaussian entries, then

$$m \gtrsim k \log(N/k)$$

  will suffice. **Num. of measurements scales only logarithmically with the ambient dimension: grid size in our previous example.**
- This is theoretically optimal (deep results in geometric functional analysis).
- Other classes (Bernoulli, partial Fourier, ...) of random matrices will do, too!

Choosing the measurement matrix — more remarks

- Gaussian and sub-Gaussian matrices are unitarily invariant, so the dimension relation is independent of the sparsity basis. These are **universal measurement matrices**:

$$M \text{ is Gaussian and } B \text{ is unitary} \implies A = MB^* \text{ is Gaussian.}$$

- Ideal for dimension reduction in simulations. Also, acquisition with simultaneous sources.

Choosing the measurement matrix — more remarks

- Gaussian and sub-Gaussian matrices are unitarily invariant, so the dimension relation is independent of the sparsity basis. These are **universal measurement matrices**:

$$M \text{ is Gaussian and } B \text{ is unitary} \implies A = MB^* \text{ is Gaussian.}$$

- Ideal for dimension reduction in simulations. Also, acquisition with simultaneous sources.
- Difficult to implement depending on the physics—e.g., in the sampling example. In such cases:
  - sample in a domain that is incoherent with the sparsity domain: e.g.,

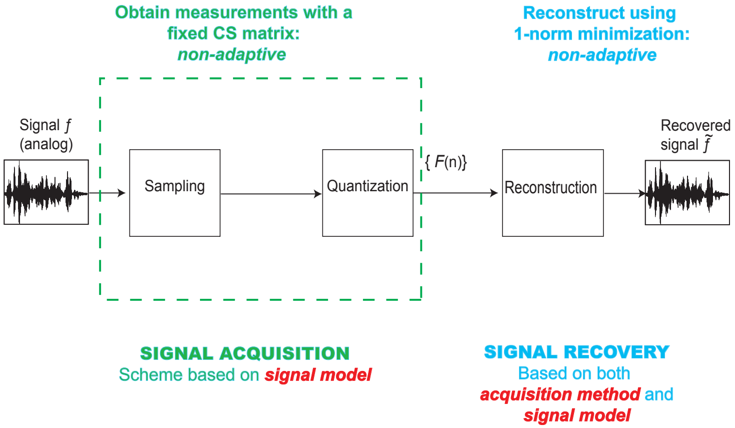
$$\text{sparse in Fourier} \implies \text{sample in time}$$

  - Randomly sub-sample (possibly on a jittered grid), i.e., "apply" a restriction matrix R.

The corresponding $A = RM$ will be a "good" compressive sampling matrix.

CS – incorporating prior info

Note that, like classical sampling, CS is a non-adaptive sampling paradigm:



CS – incorporating prior info

Remainder of the talk: We investigate methods of incorporating prior information on the support of the specific signal of interest to our sampling and reconstructions schemes.

1. **Recovery using weighted ℓ_1 minimization.**
   - Sensing is non-adaptive: Collect the measurements b (or $\hat{b}$ if there is noise) using an arbitrary CS matrix. On the other hand:
   - Recovery is adaptive: Use prior support info to come up with better recovery methods.

CS – incorporating prior info

Remainder of the talk: We investigate methods of incorporating prior information on the support of the specific signal of interest to our sampling and reconstructions schemes.

1. **Recovery using weighted ℓ_1 minimization.**
   - Sensing is non-adaptive: Collect the measurements b (or $\hat{b}$ if there is noise) using an arbitrary CS matrix. On the other hand:
   - Recovery is adaptive: Use prior support info to come up with better recovery methods.
2. **Adaptive Compressed Sensing.**
   - Sampling scheme is adaptive and incorporates a "compressive antialiasing" stage.
   - Recovery is also adaptive (using weighted ℓ_1).

Part I: CS – recovery via weighted ℓ_1

Setting: Suppose we have prior information on the support of x. In particular we have a support estimate that is generally partial and possibly inaccurate.

Want: incorporate such information in the recovery algorithm to get better results than those obtained via ℓ_1 minimization.

Why is this relevant?

Signals with Prior Information

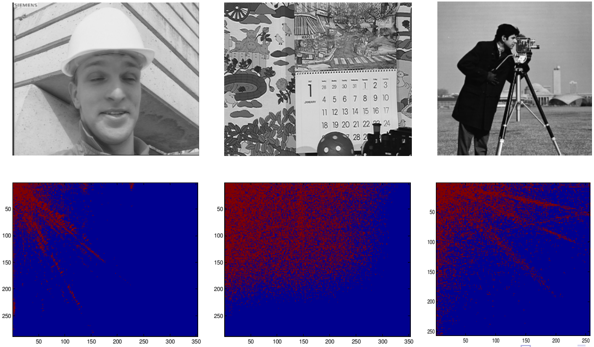
- In many applications, it is possible to draw an estimate of the support of the signal, for example:

Signals with Prior Information

- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.

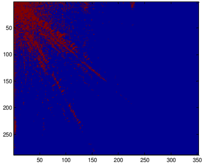
Signals with Prior Information

- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.



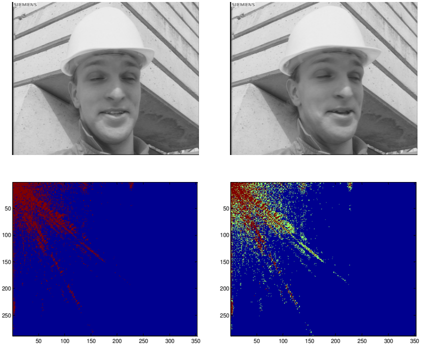
Signals with Prior Information

- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.



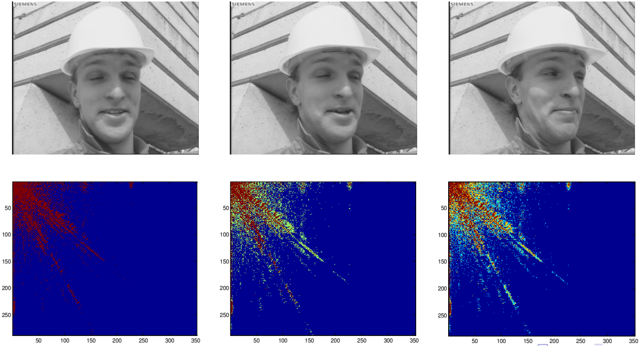
Signals with Prior Information

- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.



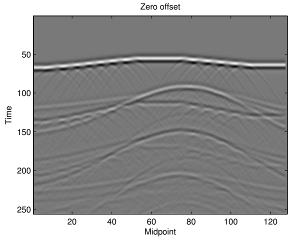
Signals with Prior Information

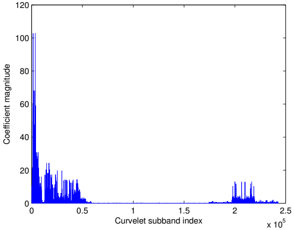
- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.



Signals with Prior Information

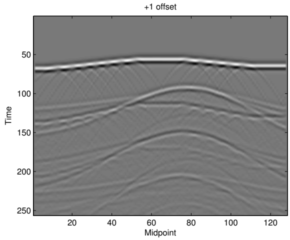
- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.
  - Other signals such as seismic data, ...

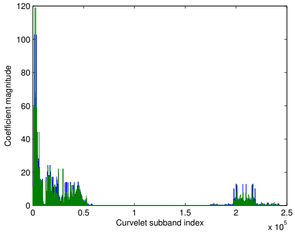




Signals with Prior Information

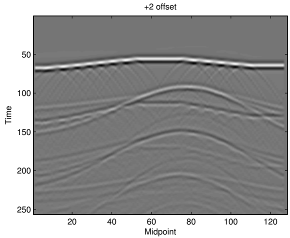
- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.
  - Other signals such as seismic data, ...

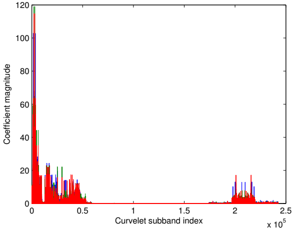




Signals with Prior Information

- In many applications, it is possible to draw an estimate of the support of the signal, for example:
  - Natural images have large DCT coefficients that are localized in the low frequency subbands.
  - Video sequences are temporally correlated, resulting in a shared subset of their support.
  - Other signals such as seismic data, ...





Problem formulation

Suppose that x is a k-sparse signal with unknown support T_0.

Given:

1. CS measurements of x (i.e., $b = Ax$, or $\hat{b} = Ax + e$ with $\|e\|_2 \leq \epsilon$).
2. A partially accurate support estimate $\widetilde{T}$. Let's quantify—two important parameters:

$$\rho := \frac{\#\widetilde{T}}{\#T_0} \qquad \text{relative size of the estimated support}$$

$$\alpha := \frac{\#T_0 \cap \widetilde{T}}{\#\widetilde{T}} \qquad \text{accuracy of the estimate}$$

In general, we have $0 \leq \rho \leq \frac{N}{k}$ and $0 \leq \alpha \leq \min\{1, \frac{1}{\rho}\}$.

Problem formulation

Suppose that x is a k-sparse signal with unknown support T_0.

Given:

1. CS measurements of x (i.e., $b = Ax$, or $\hat{b} = Ax + e$ with $\|e\|_2 \leq \epsilon$).
2. A partially accurate support estimate $\widetilde{T}$. Let's quantify—two important parameters:

$$\rho := \frac{\#\widetilde{T}}{\#T_0} \qquad \text{relative size of the estimated support}$$

$$\alpha := \frac{\#T_0 \cap \widetilde{T}}{\#\widetilde{T}} \qquad \text{accuracy of the estimate}$$

In general, we have $0 \leq \rho \leq \frac{N}{k}$ and $0 \leq \alpha \leq \min\{1, \frac{1}{\rho}\}$.

Goals:

- Incorporate $\widetilde{T}$ into the recovery algorithm (to get better recovery),
- Obtain theoretical recovery guarantees depending on the size and accuracy of $\widetilde{T}$ (i.e., ρ and α).

Proposed Algorithm – weighted ℓ_1 minimization

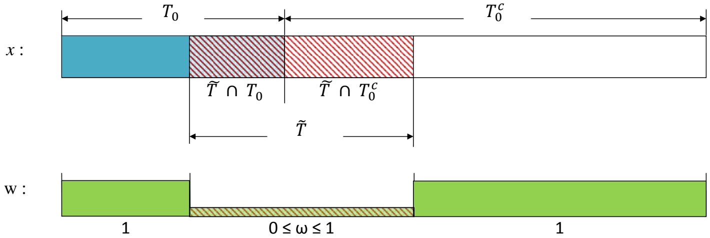
Given a set of (noisy) measurements $\hat{b}$, define

$$\Delta^{\epsilon}_{1,\mathrm{w}}(\hat{b}) := \arg\min_{x} \ \|x\|_{1,\mathrm{w}} \text{ subject to } \|Ax - \hat{b}\|_2 \leq \epsilon$$

where

$$\mathrm{w}_i = \begin{cases} 1, & i \in \widetilde{T}^c, \\ \omega, & i \in \widetilde{T}, \end{cases} \qquad \text{for some } 0 \leq \omega \leq 1.$$

Above $\|x\|_{1,\mathrm{w}} := \sum_i \mathrm{w}_i |x_i|$, and $\|e\|_2^2 \leq \epsilon$.



Improved sufficient conditions for weighted ℓ_1

We prove the following theorem in the case of weighted ℓ_1:

Theorem [FMSY]

Suppose for some $a > \max\{1, (1-\alpha)\rho\}$, $\delta_{ak} + a\gamma\delta_{(a+1)k} < a\gamma - 1$. Then

$$\|\Delta^{\epsilon}_{1,\mathrm{w}}(\hat{b}) - x\|_2 \le C_0'\epsilon + C_1' k^{-1/2}(\omega\|x_{T_o^c}\|_1 + (1-\omega)\|x_{\widetilde{T}^c \cap T_0^c}\|_1)$$

where $\gamma = \left(\omega + (1-\omega)\sqrt{1+\rho-2\alpha\rho}\right)^{-2}$.

Remarks.

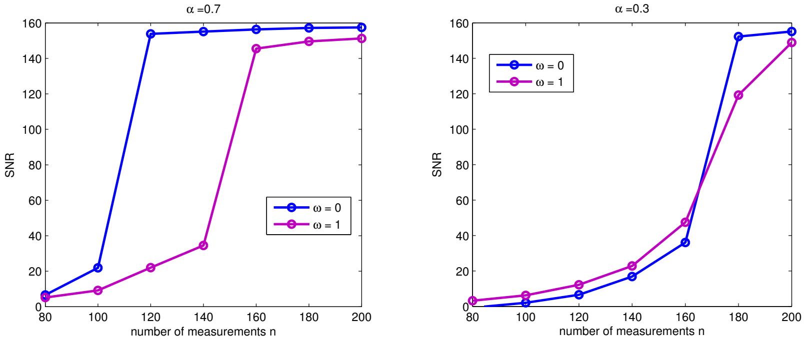
1. Above, $0 \le \omega \le 1$ is a fixed weight. If we set $\omega = 1$, our theorem reduces to the robust recovery theorem of CRT.
2. Recall $0 \le \alpha \le 1$ describes the accuracy of $\widetilde{T}$ and ρ describes its size.
3. **The sufficient conditions above are weaker than those for ℓ_1 minimization iff $\alpha > 0.5$. (Same holds for the constants.)**
4. Earlier work on the case $\omega = 0$: e.g., Borries, Vaswani and Lu; Jacques. Our results, to our knowledge, provide weakest sufficient cond. and smallest constants.

Numerical Experiments – sparse signals

- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.

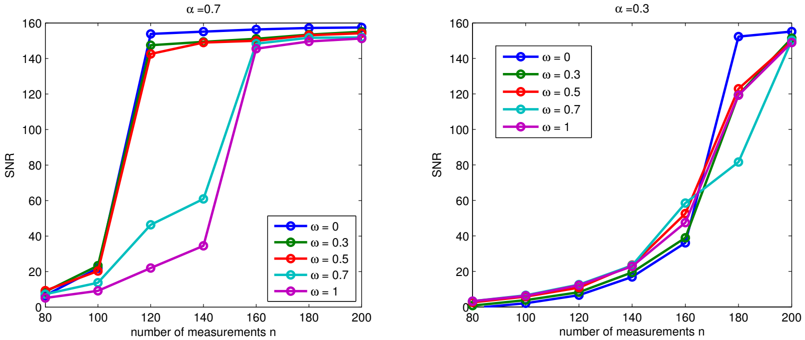
Numerical Experiments – sparse signals

- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.
- **The noise free case:**



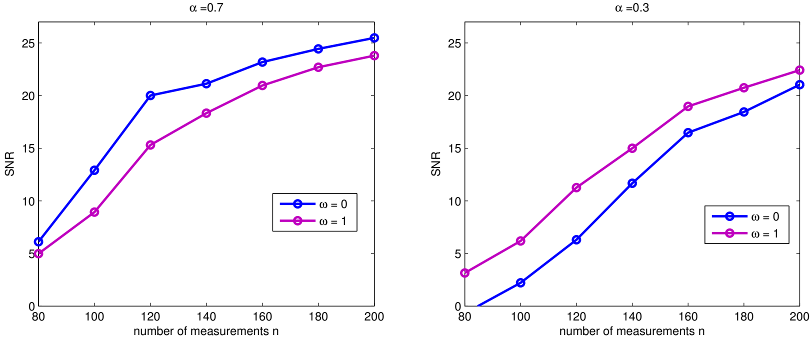
Numerical Experiments – sparse signals

- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.
- **The noise free case:**



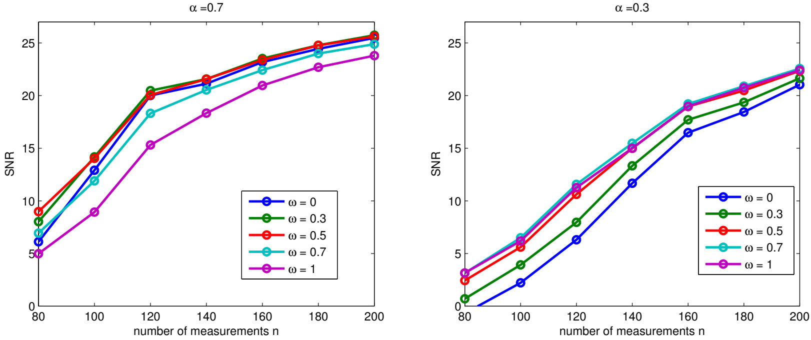
Numerical Experiments – sparse signals

- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.
- **The noisy measurement vector case:**



Numerical Experiments – sparse signals

- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.
- **The noisy measurement vector case:**

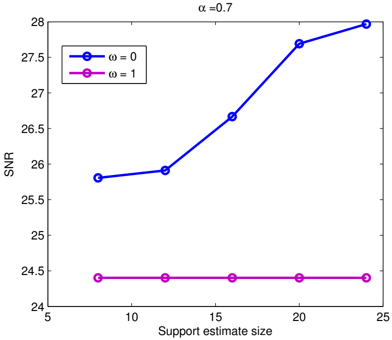


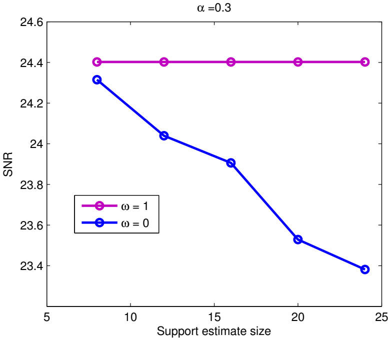
Numerical Experiments – compressible signals

- SNR averaged over 10 experiments for signals x whose coefficients decay like j^{-p} where $j \in \{1, \ldots N\}$ and $p = 1.5$. We take $n = 100$ and $N = 500$.

Numerical Experiments – compressible signals

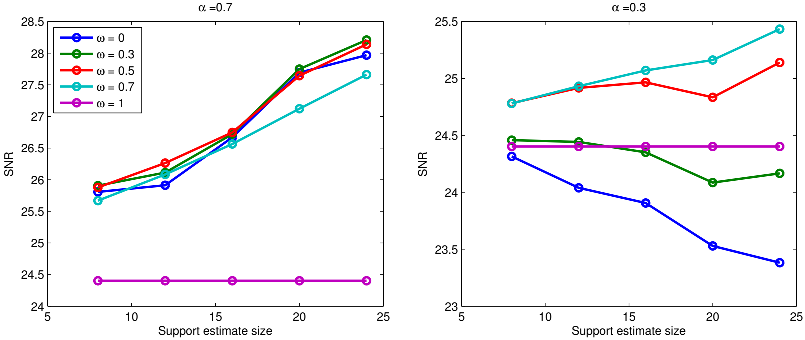
- SNR averaged over 10 experiments for signals x whose coefficients decay like j^{-p} where $j \in \{1, \dots N\}$ and $p = 1.5$. We take $n = 100$ and $N = 500$.
- **The noise free case:**





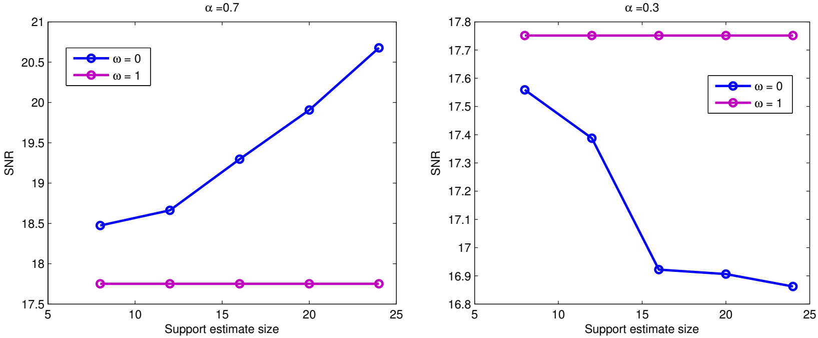
Numerical Experiments – compressible signals

- SNR averaged over 10 experiments for signals x whose coefficients decay like j^{-p} where $j \in \{1, \ldots N\}$ and $p = 1.5$. We take $n = 100$ and $N = 500$.
- **The noise free case:**



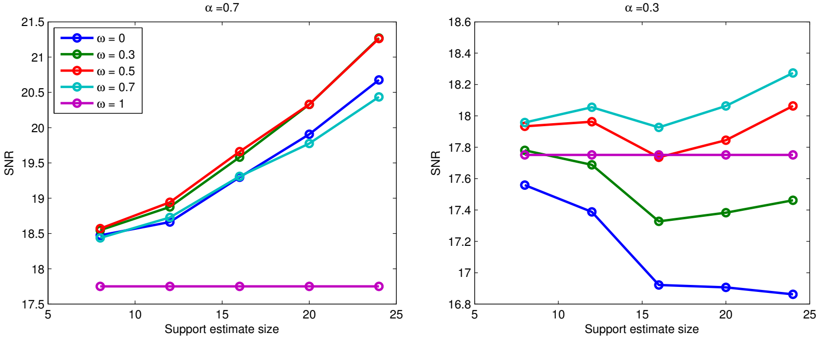
Numerical Experiments – compressible signals

- SNR averaged over 10 experiments for signals x whose coefficients decay like j^{-p} where $j \in \{1, \ldots N\}$ and $p = 1.5$. We take $n = 100$ and $N = 500$.
- **The noisy measurement vector case:**



Numerical Experiments – compressible signals

- SNR averaged over 10 experiments for signals x whose coefficients decay like j^{-p} where $j \in \{1, \ldots N\}$ and $p = 1.5$. We take $n = 100$ and $N = 500$.
- **The noisy measurement vector case:**



Perspective: how to choose ω

- **Numerical experiments suggest:** Intermediate values of the weight $\omega \approx 0.5$ are more robust to model mismatch and result in the highest SNR even when $\alpha < 0.5$.

Perspective: how to choose ω

- **Numerical experiments suggest:** Intermediate values of the weight $\omega \approx 0.5$ are more robust to model mismatch and result in the highest SNR even when $\alpha < 0.5$.
- If we optimize the error bound constants:

$$\alpha > 0.5 \implies \omega = 0, \quad \alpha < 0.5 \implies \omega = 1.$$

Perspective: how to choose ω

- **Numerical experiments suggest:** Intermediate values of the weight $\omega \approx 0.5$ are more robust to model mismatch and result in the highest SNR even when $\alpha < 0.5$.
- If we optimize the error bound constants:

$$\alpha > 0.5 \implies \omega = 0, \quad \alpha < 0.5 \implies \omega = 1.$$

- This does not take into account the other terms in the error bound:

$$\|x^* - x\|_2 \le C_0'(\omega)\epsilon + C_1'(\omega)k^{-1/2}\left(\omega\|x_{T_0^c}\|_1 + (1-\omega)\|x_{\widetilde{T}^c \cap T_0^c}\|_1\right).$$

Perspective: how to choose ω

- **Numerical experiments suggest:** Intermediate values of the weight $\omega \approx 0.5$ are more robust to model mismatch and result in the highest SNR even when $\alpha < 0.5$.
- If we optimize the error bound constants:

$$\alpha > 0.5 \implies \omega = 0, \quad \alpha < 0.5 \implies \omega = 1.$$

- This does not take into account the other terms in the error bound:

$$\|x^* - x\|_2 \leq C_0'(\omega)\epsilon + C_1'(\omega)k^{-1/2}\Big(\omega\|x_{T_0^c}\|_1 + (1-\omega)\|x_{\widetilde{T}^c \cap T_0^c}\|_1\Big).$$

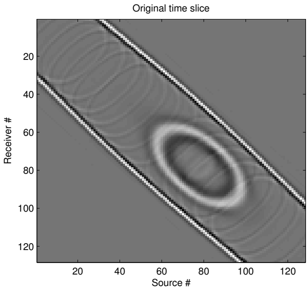
- As ω goes to zero,
  - the constant $C_1'(\omega)$ increases
  - the term $\omega\|x_{T_0^c}\|_1 + (1-\omega)\|x_{\widetilde{T}^c \cap T_0^c}\|_1$ decreases
- There may exist $0 < \omega < 1$ that minimizes their product (depending on the signal class as well as properties of the measurement matrix A).

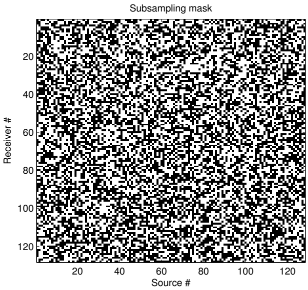
Application: Compressed sensing of seismic lines

- Full seismic line with 128 shots, 128 receivers, and 256 time samples.

Application: Compressed sensing of seismic lines

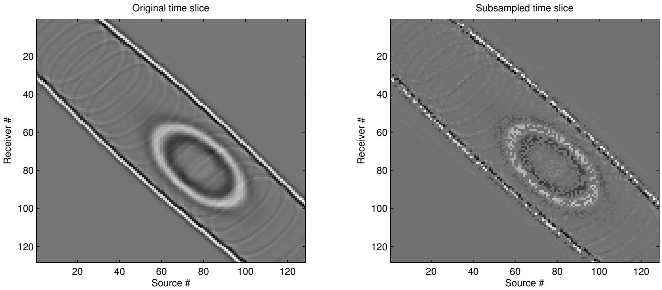
- Full seismic line with 128 shots, 128 receivers, and 256 time samples.
- Due to budgetary requirements or device malfunctioning, receivers and shots are randomly sampled (e.g., time slice 100), keeping only % 50 of the data.





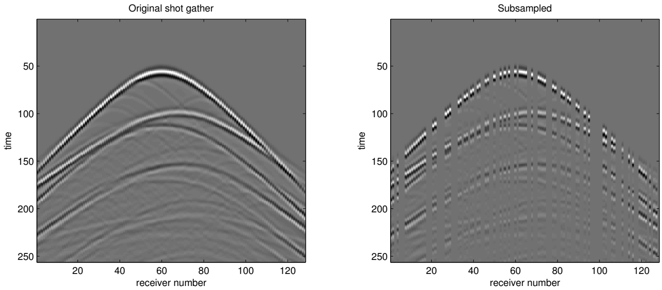
Application: Compressed sensing of seismic lines

- Full seismic line with 128 shots, 128 receivers, and 256 time samples.
- Due to budgetary requirements or device malfunctioning, receivers and shots are randomly sampled (e.g., time slice 100), keeping only % 50 of the data.



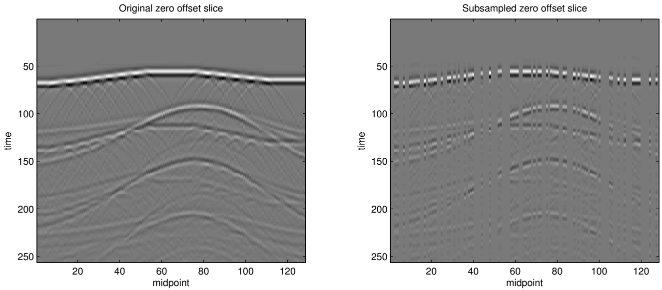
Application: Compressed sensing of seismic lines

- Full seismic line with 128 shots, 128 receivers, and 256 time samples.
- Due to budgetary requirements or device malfunctioning, receivers and shots are randomly sampled (e.g., time slice 100), keeping only % 50 of the data.
- Results in missing data along entire time axis (eg: common shot gather # 60)



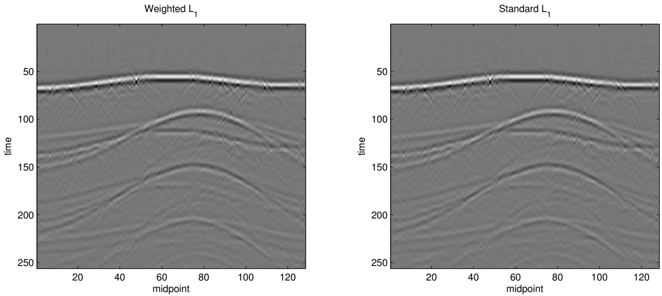
Recovery in offset domain

- Seismic line data is correlated in the midpoint offset domain.
- Map the subsampling mask to act on offset slices.



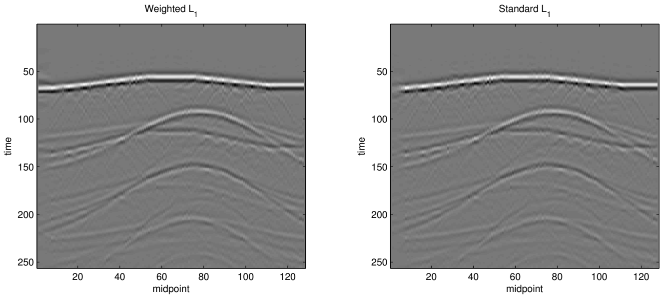
Recovery in offset domain

- Seismic line data is correlated in the midpoint offset domain.
- Map the subsampling mask to act on offset slices.
- Recover the zero offset using standard ℓ_1 minimization (same quality for wL1 and L1).



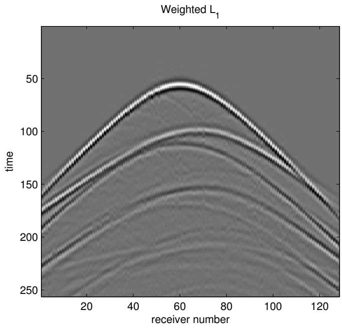
Recovery in offset domain

- Seismic line data is correlated in the midpoint offset domain.
- Map the subsampling mask to act on offset slices.
- Recover the zero offset using standard ℓ_1 minimization (same quality for wL1 and L1).
- Use the support of the zero offset slice to weight the recovery of other offset slices (eg: $+2$ offset).



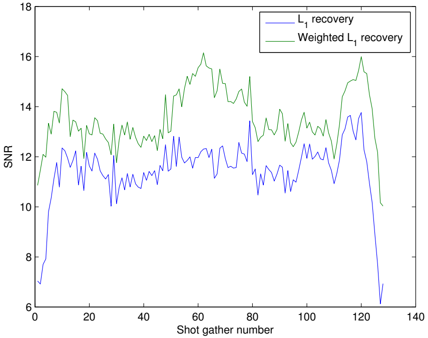
Performance of weighted ℓ_1 vs standard ℓ_1

- Map the data back to the source receiver domain (eg: shot gather 60).



Performance of weighted ℓ_1 vs standard ℓ_1

- Map the data back to the source receiver domain (eg: shot gather 60).
- Signal to noise ratio (SNR) of all 128 shot gathers.



Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.
- We also showed that if at least 50% of the support estimate is accurate, then the recovery is guaranteed with weaker RIP conditions and smaller error bounds.

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.
- We also showed that if at least 50% of the support estimate is accurate, then the recovery is guaranteed with weaker RIP conditions and smaller error bounds.
- Our algorithm improves compressive recovery of video sequences.

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.
- We also showed that if at least 50% of the support estimate is accurate, then the recovery is guaranteed with weaker RIP conditions and smaller error bounds.
- Our algorithm improves compressive recovery of video sequences.
- **Some questions:**
  - How/when can we find the support estimate $\tilde{T}$?

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.
- We also showed that if at least 50% of the support estimate is accurate, then the recovery is guaranteed with weaker RIP conditions and smaller error bounds.
- Our algorithm improves compressive recovery of video sequences.
- **Some questions:**
  - How/when can we find the support estimate $\tilde{T}$?
  - Can we draw a more accurate $\tilde{T}$ after solving the weighted ℓ_1 minimization problem?

Part I: Summary, conclusions, and further problems

- Weighted ℓ_1 minimization can recover less sparse signals than standard ℓ_1 when enough prior information is available.
- We showed that the recovery is stable and robust.
- We also showed that if at least 50% of the support estimate is accurate, then the recovery is guaranteed with weaker RIP conditions and smaller error bounds.
- Our algorithm improves compressive recovery of video sequences.
- **Some questions:**
  - How/when can we find the support estimate $\tilde{T}$?
  - Can we draw a more accurate $\tilde{T}$ after solving the weighted ℓ_1 minimization problem?
  - How would an iterative weighted ℓ_1 algorithm with **fixed** weights perform compared to IRL1 of Boyd and Candès?

Part II: Adaptive CS

Disclaimer. This is very recent work and the results we will present are preliminary.

Main Problem. "Compressive aliasing"... To explain, recall the classical sampling theory setup:

Ideal signal model: Bandlimited with known bandwidth, say maximum frequency ω_1.

Sampling scheme: Sample with frequency $\omega_s \geq 2\omega_1$

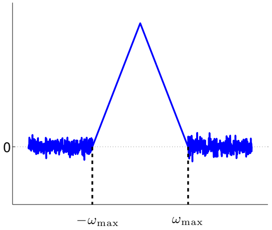
In practice: Signals may have higher bandwidth $\omega_{\max} > \frac{\omega_s}{2}$; moreover there may be off-band noise.

Problem: This would cause aliasing as well as noise in the sampled & reconstructed signal.

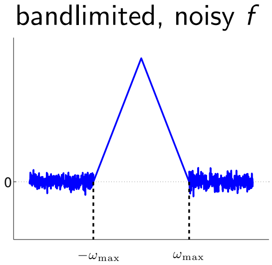
Remedy: Use a front-end low-pass filter and force the signal to obey the ideal signal model. Resulting approximation is the best approximation with bandwidth $\omega_s/2$.

Antialiasing in classical sampling

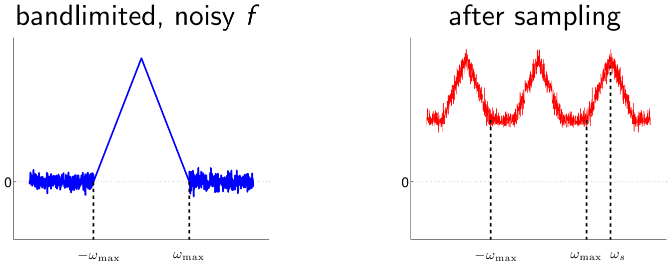
bandlimited, noisy f



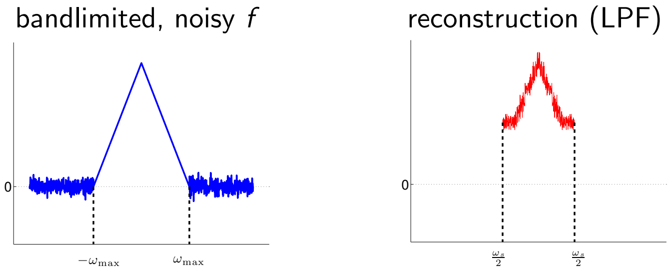
Antialiasing in classical sampling



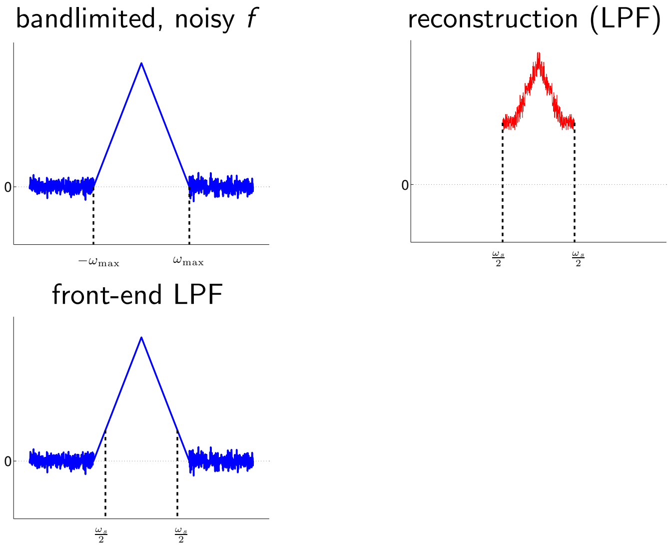
Antialiasing in classical sampling



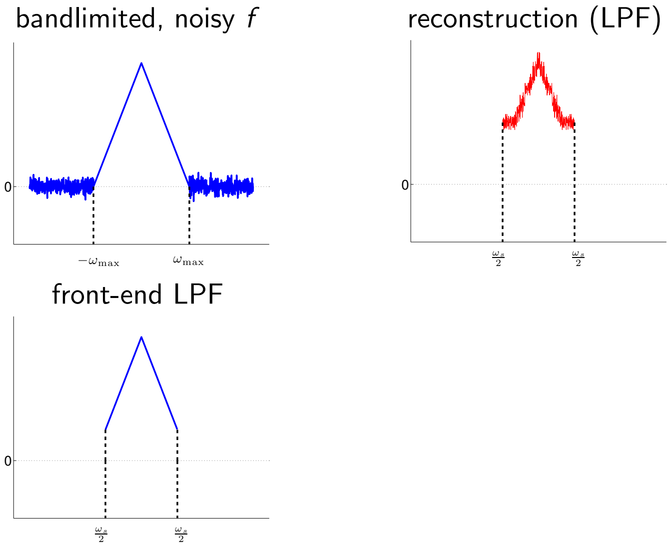
Antialiasing in classical sampling



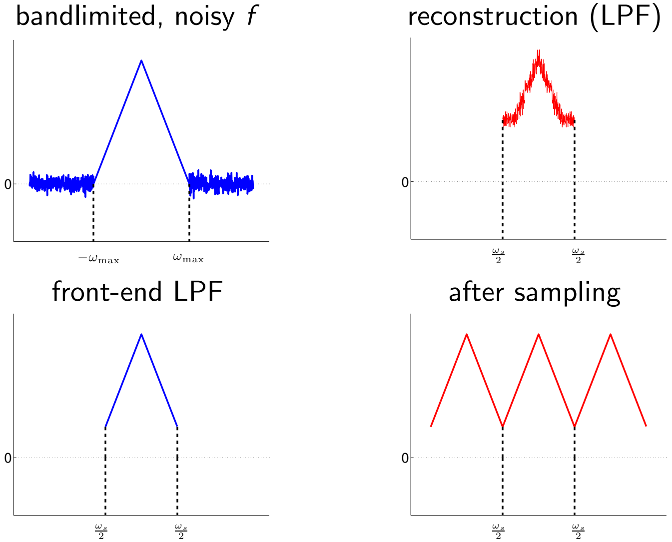
Antialiasing in classical sampling



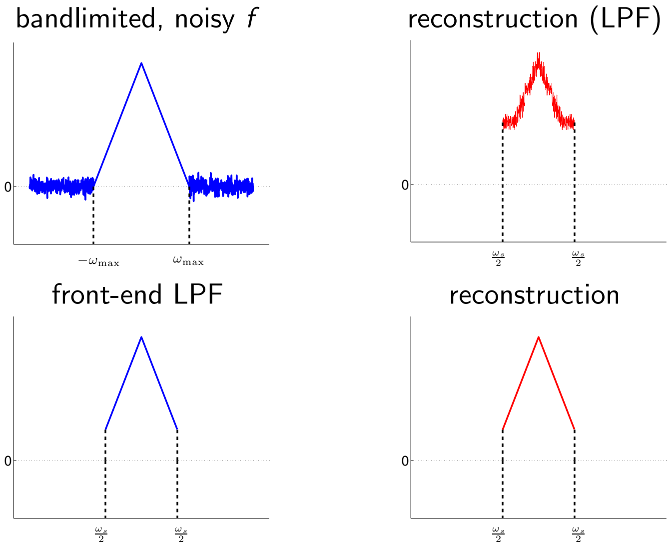
Antialiasing in classical sampling



Antialiasing in classical sampling



Antialiasing in classical sampling



Compressive aliasing

Parallel to our discussion above:

- **Ideal signal model in CS:** k-sparse signals in $\mathbb{R}^N$ ($k \sim$ bandwidth)
- **Sampling scheme:** Hit the signal by an appropriate $m \times N$ measurement matrix ($m \sim$ sampling density, is suited to recover all k-sparse signals)

Compressive aliasing

Parallel to our discussion above:

- **Ideal signal model in CS:** k-sparse signals in $\mathbb{R}^N$ ($k \sim$ bandwidth)
- **Sampling scheme:** Hit the signal by an appropriate $m \times N$ measurement matrix ($m \sim$ sampling density, is suited to recover all k-sparse signals)
- **In practice:** Signals are not k-sparse, but they are "compressible", i.e., coefficients decay (and are possibly noisy).
- **Recovery:** via ℓ_1, weighted ℓ_1, etc... These are suited to recover k-sparse signals exactly.

Compressive aliasing

Parallel to our discussion above:

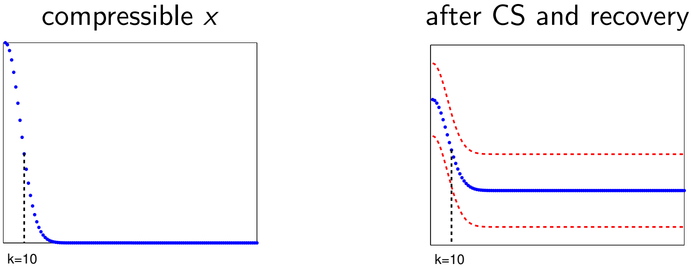
- **Ideal signal model in CS:** k-sparse signals in $\mathbb{R}^N$ ($k \sim$ bandwidth)
- **Sampling scheme:** Hit the signal by an appropriate $m \times N$ measurement matrix ($m \sim$ sampling density, is suited to recover all k-sparse signals)
- **In practice:** Signals are not k-sparse, but they are "compressible", i.e., coefficients decay (and are possibly noisy).
- **Recovery:** via ℓ_1, weighted ℓ_1, etc... These are suited to recover k-sparse signals exactly.
- **Good news:** CS robust recovery theorems guarantee that the approximation is "almost" as good as the best we could hope for.

Compressive aliasing

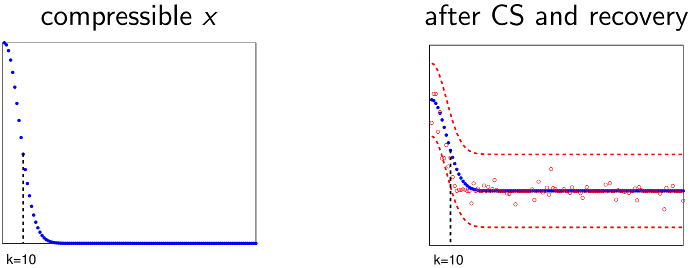
Parallel to our discussion above:

- **Ideal signal model in CS:** k-sparse signals in $\mathbb{R}^N$ ($k \sim$ bandwidth)
- **Sampling scheme:** Hit the signal by an appropriate $m \times N$ measurement matrix ($m \sim$ sampling density, is suited to recover all k-sparse signals)
- **In practice:** Signals are not k-sparse, but they are "compressible", i.e., coefficients decay (and are possibly noisy).
- **Recovery:** via ℓ_1, weighted ℓ_1, etc... These are suited to recover k-sparse signals exactly.
- **Good news:** CS robust recovery theorems guarantee that the approximation is "almost" as good as the best we could hope for.
- **Problem:** If the coefficient decay is not fast enough, large coefficients can still get significantly distorted : compressive aliasing.
- **Goal:** Find an antialiasing method when there is prior support information (as in Part I).

Compressive aliasing

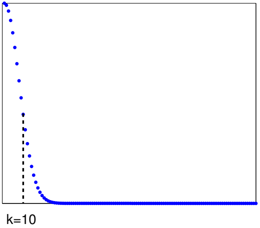


Compressive aliasing

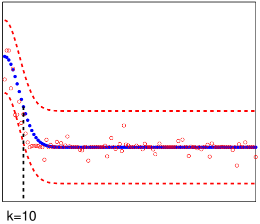


Compressive aliasing

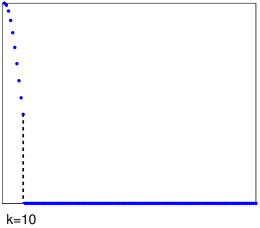
compressible x



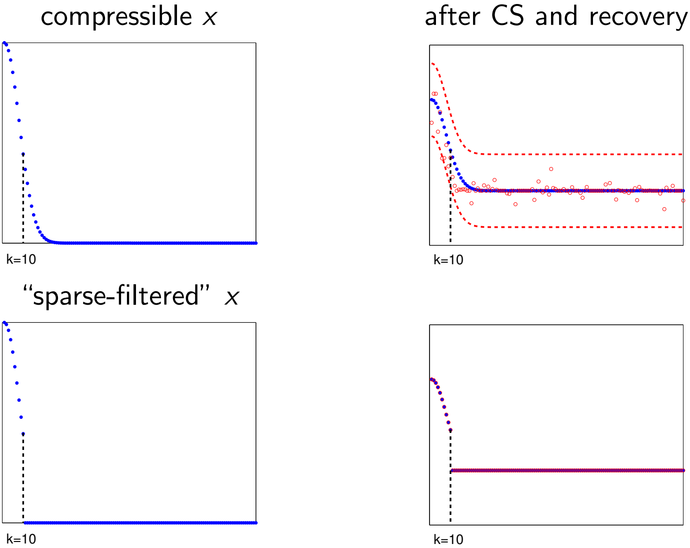
after CS and recovery



“sparse-filtered” x



Compressive aliasing



CS – sparse-filtering

Moral: If we knew the support of the sparse signal, we could sparsify it (analogous to low-pass filtering in classical sampling) and force it to obey the signal model ideal for CS.

Problem: If we knew the support, we don't need CS!

Compromise: What if we have a partial and possibly inaccurate support estimate?

Adaptive CS setup

Let f be a signal that is compressible with respect to a transform B, i.e.,

$$f = B^* x, \quad x \text{ decays fast.}$$

Suppose we knew $T_k = \text{supp}(x_k)$ – indices of largest k coefficients of x.

"sparse-filter" f: $f_{\text{sp}} = B^* W^2 \underbrace{Bf}_{x}.$

Here W is a diagonal matrix such that $W_{jj} = \begin{cases} 1 & \text{if } j \in T_k \\ 0 & \text{if } j \notin T_k \end{cases}$

Adaptive CS setup

Let f be a signal that is compressible with respect to a transform B, i.e.,

$$f = B^* x, \quad x \text{ decays fast.}$$

Suppose we knew $T_k = \mathrm{supp}(x_k)$ – indices of largest k coefficients of x.

"sparse-filter" f: $$f_{\mathrm{sp}} = B^* W^2 \underbrace{Bf}_{x}.$$

Here W is a diagonal matrix such that $W_{jj} = \begin{cases} 1 & \text{if } j \in T_k \\ 0 & \text{if } j \notin T_k \end{cases}$

The corresponding CS scheme—with CS matrix A—would be:

$$A \underbrace{B^* W^2 Bf}_{f_{\mathrm{sp}}} = y \quad \leftarrow \text{measurements of "sparse-filtered" } f$$

and we can recover x_k (thus $f_k = B^* x_k$) by solving

$$\min \|z\|_1 \text{ subject to } AB^* z = y.$$

Adaptive CS setup

Let f be a signal that is compressible with respect to a transform B, i.e.,

$$f = B^* x, \quad x \text{ decays fast.}$$

Suppose we knew $T_k = \text{supp}(x_k)$ – indices of largest k coefficients of x.

"sparse-filter" f: $$f_{\text{sp}} = B^* W^2 \underbrace{Bf}_{x}.$$

Here W is a diagonal matrix such that $W_{jj} = \begin{cases} 1 & \text{if } j \in T_k \\ 0 & \text{if } j \notin T_k \end{cases}$

The corresponding CS scheme—with CS matrix A—would be:

$$\underbrace{AB^* W^2 B}_{\text{meas. matrix}} f = y \;\leftarrow \text{measurements of "sparse-filtered" } f$$

and we can recover x_k (thus $f_k = B^* x_k$) by solving

$$\min \|z\|_1 \text{ subject to } AB^* z = y.$$

Adaptive CS setup

Let f be a signal that is compressible with respect to a transform B, i.e.,

$$f = B^* x, \quad x \text{ decays fast.}$$

Suppose we knew $T_k = \text{supp}(x_k)$ – indices of largest k coefficients of x.

"sparse-filter" f: $$f_{\text{sp}} = B^* W^2 \underbrace{Bf}_{x}.$$

Here W is a diagonal matrix such that $W_{jj} = \begin{cases} 1 & \text{if } j \in T_k \\ 0 & \text{if } j \notin T_k \end{cases}$

The corresponding CS scheme—with CS matrix A—would be:

$$AB^* \underbrace{W^2 Bf}_{x_k} = y \leftarrow \text{measurements of "sparse-filtered" } f$$

and we can recover x_k (thus $f_k = B^* x_k$) by solving

$$\min \|z\|_1 \text{ subject to } AB^* z = y.$$

Adaptive CS – more realistic approach

Same setup as in the previous slide: $f = B^* x$, $\quad x$ compressible
However, instead of T_k, we have a partial and inaccurate estimate $\widetilde{T}_k$.

Proposed Method.

- Let W be a diagonal weighting matrix such that

$$W_{jj} = \begin{cases} 1 & \text{if } j \in \widetilde{T}_k \\ \omega & \text{if } j \notin \widetilde{T}_k \end{cases}$$

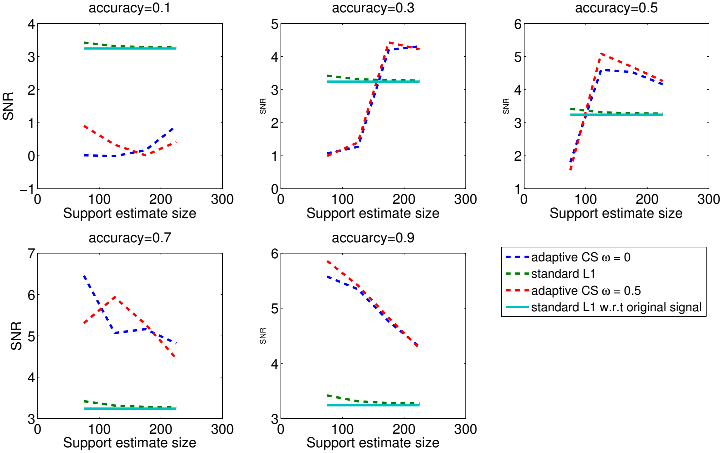
  where $0 < \omega < 1$ (some intermediate value for robustness).
- **Adaptive sampling:** Collect the (noisy) samples

$$AB^* W^2 Bf + e = y.$$

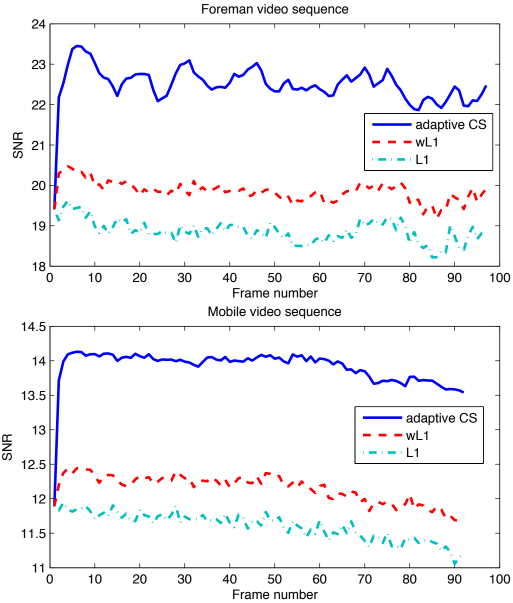
- **(Adaptive) Reconstruction** via
  - ℓ_1 minimization: $\min \|z\|_1$ subject to $\|AB^* z - y\| \leq \epsilon$, or
  - weighted ℓ_1: $\min \|z\|_1$ subject to $\|AB^* Wz - y\| \leq \epsilon$.

Example with synthetic signals

Signals with coefficients decaying like $j^{-0.7}$; $N = 500$, $m = 50$.

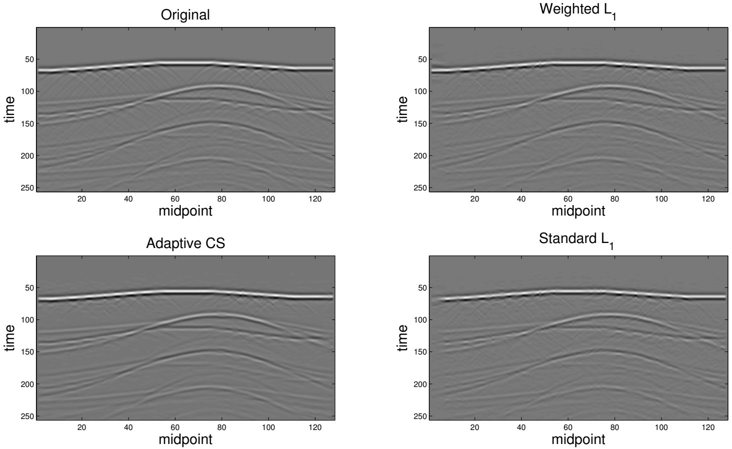


Example with video frame sequences



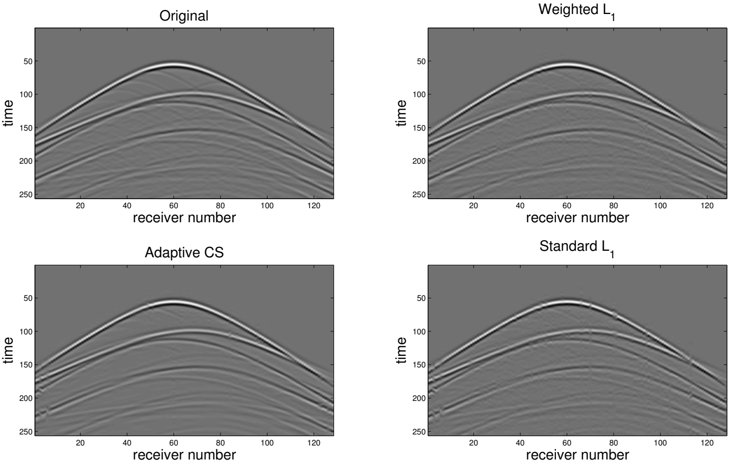
Example: adaptive CS of seismic lines

Same experiment as before – %50 of source/receiver pairs are missing. Below, we show offset=+2...



Example: adaptive CS of seismic lines

Same experiment as before – %50 of source/receiver pairs are missing.
Below, we show shot gather 60...



Concluding remarks

- CS provides a powerful sampling theory for the acquisition of signals that admit a sparse or compressible representation in some transform domain.
- CS is a non-adaptive sampling paradigm.
- If prior information is available, it can be effectively used in both the sampling stage and the reconstruction stage.
- We are currently working on fine-tuning the adaptive CS approach and using it both for seismic acquisition and in computations.

Acknowledgements

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.