Weighted methods in sparse recovery

Özgür Yılmaz
UBC Mathematics

SINBAD Consortium Meeting 2012

December 3, 2012

Outline – the question

- This talk is about the sparse recovery problem: Find (a sparse) the sparsest (approximate) solution of an underdetermined system of linear equations.
- Two similar, yet distinct, settings: sparse approximation problem and compressed sensing. Our focus will be on compressed sensing as that is more relevant for our problems in this project.
- In compressed sensing (an overview will follow), the goal is to recover signals from seemingly incomplete measurements (or sub-Nyquist rate samples).
- **Model:** Signals are sparse in some transform domain. *This is all!*
- **Our focus in this talk:** *If we have additional prior information*, can we improve the recovery performance **without changing the sampling procedure**?

Outline – relevance and some answers

We will present various algorithms that improve recovery if we have some prior information about the *locations of non-zero coefficients*, i.e., the **support**, of the original signal. This is often the case in practice:

- Video signals (correlations among consecutive frames)
- Seismic data (correlations, e.g., among adjacent offset gathers)
- Reweighting?

In this talk:

- Weighted ℓ_1 (Friedlander-**Mansour**-Saab-OY)
- Weighted ℓ_p, $0 < p < 1$ (**Ghadermarzy**-Mansour-OY)
- Weighted ℓ_1–analysis (**Hargreaves**-OY)
- Weighted Kaczmarz and a row-action based reweighted sparse recovery algorithm (**Mansour**-OY)

Compressed sensing (CS): an overview

- During the last 8 years: a **revolution in sampling theory**.
- **Main conclusion:** sparse signals can be recovered from few, "seemingly 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** : Very active area. To follow:

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

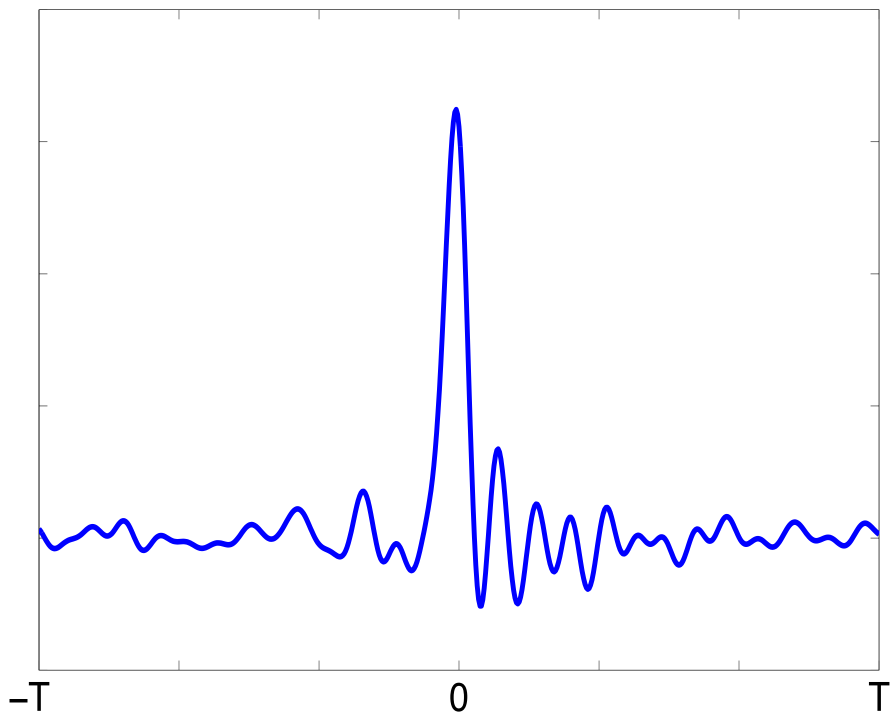
  Nuit-Blanche Blog at http://nuit-blanche.blogspot.com
- Relies heavily on **sparse approximations** that has been around for more than two decades (transforms like wavelets, curvelets, Gabor).
- Interesting and difficult mathematics and important applications such as seismic signal processing, imaging,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

Classical sampling vs. compressed sensing

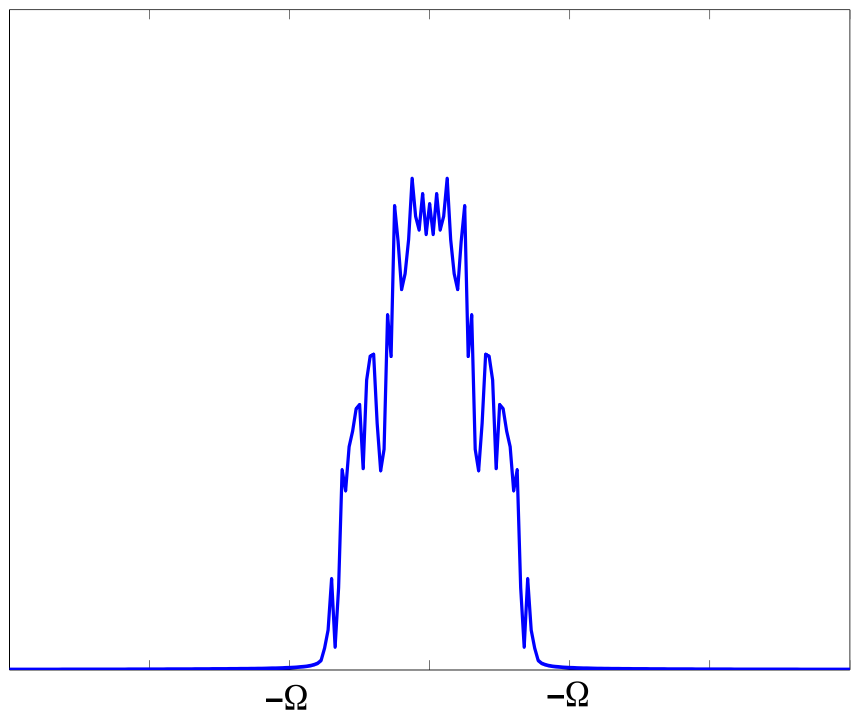
We now illustrate how compressed sensing differs from classical sampling (a la Shannon-Nyquist). **First, classical sampling:**

Let $f \in B_\Omega$: bandlimited with bandlimit Ω. Then $\tau_{\text{Nyquist}} = 1/2\Omega$.

A bandlimited f



Fourier transform of f

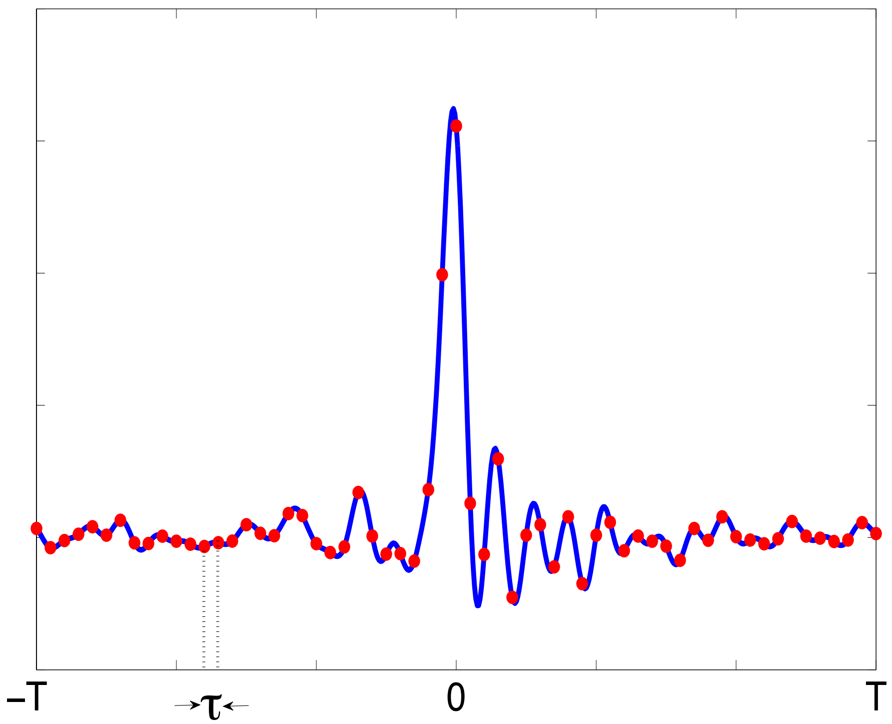


Classical sampling vs. compressed sensing

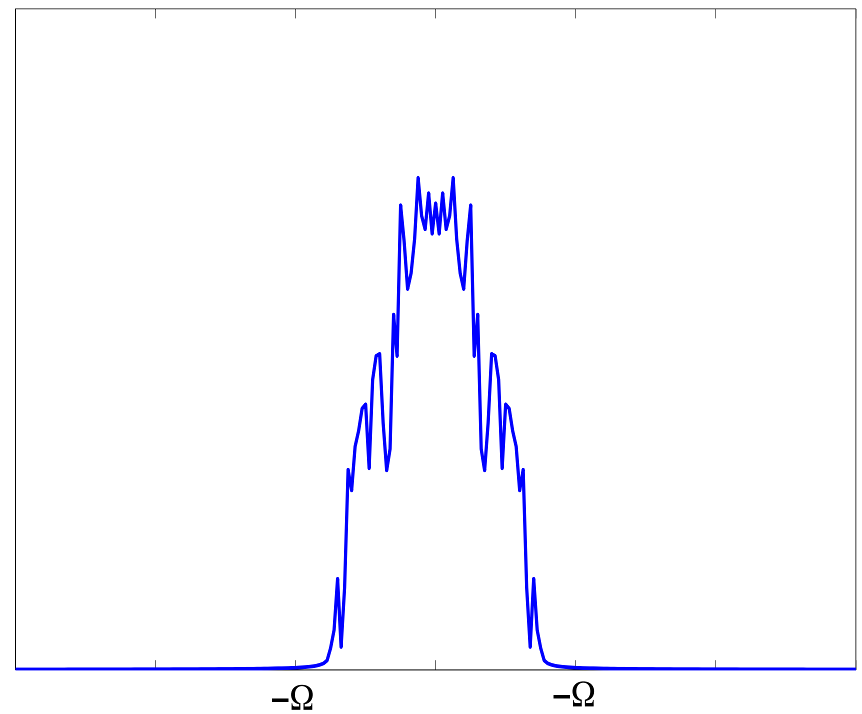
We now illustrate how compressed sensing differs from classical sampling (a la Shannon-Nyquist). **First, classical sampling:**

Let $f \in B_\Omega$: bandlimited with bandlimit Ω. Then $\tau_{\text{Nyquist}} = 1/2\Omega$.

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 $f \in B_\Omega$, on $[-T, T]$ can be represented by a vector $\mathbf{f} \in \mathbb{R}^N$ obtained by collecting N **measurements.**

Classical sampling vs. compressed sensing

What makes the classical sampling approach work?

1. $f \in B_\Omega \longleftarrow$ "model the signal class".
2. We measure f by obtaining its samples on a regular grid $\longleftarrow$ "specify the measurement scheme".
3. Use Shannon-Nyquist sampling formula to reconstruct $\longleftarrow$ "find a reconstruction method".

Note that:

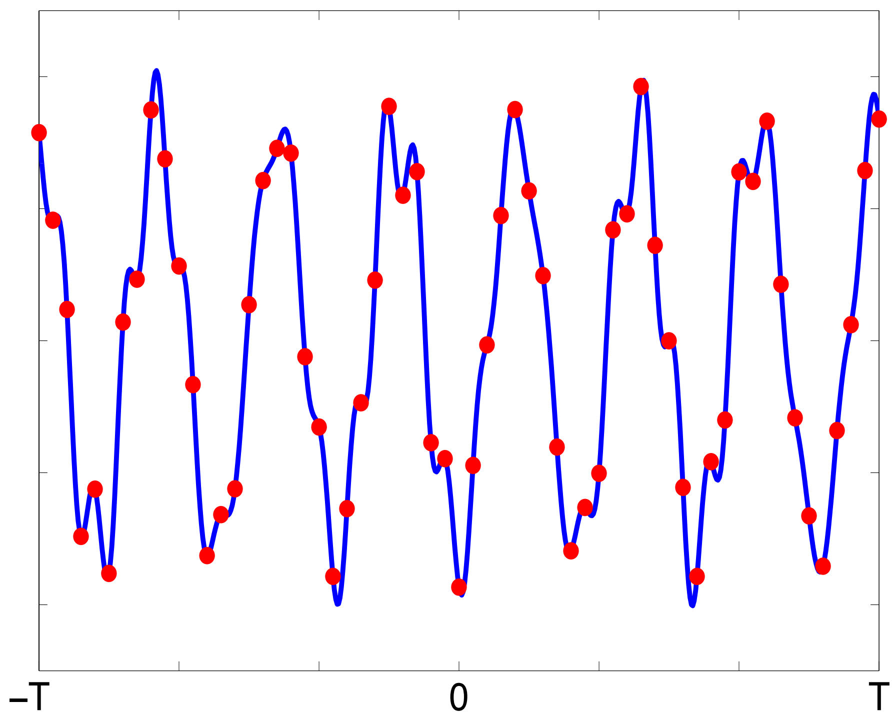
- Ambient dimension of the corresponding representation is $N \sim \Omega T$.
- We get perfect reconstruction if we collect these N samples.
- Different N-dimensional vectors correspond to samples of different bandlimited functions – so **no hope for dimension reduction**—i.e., we need N independent measurements— **under this signal model**.

Classical sampling vs. compressed sensing

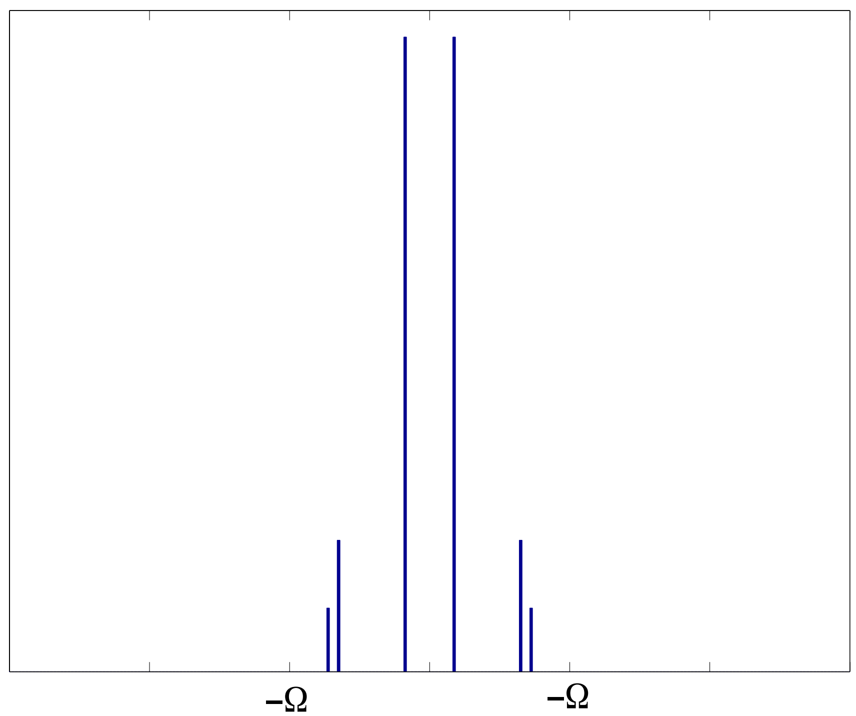
Above: Reduced a bandlimited function f to a vector $\mathbf{f}$ in $\mathbb{R}^N$.

Question: Can we reduce the dimensionality of the problem by **restricting the signal class further**? Say f is sparse in Fourier.

Another bandlimited f



Fourier transform of f



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

Compressed sensing – general framework

- Signal $f \in \mathbb{R}^N$, want to collect information on f.
- **Model the signal class:** f is sparse w.r.t. a known basis B:

$$f = B^* x \quad \text{where } x \text{ is sparse.}$$

(Above B is the $N \times N$ DFT matrix.)

- **Specify a measurement scheme:** Construct an $m \times N$ measurement matrix M with $m \ll N$

$$f_{\text{meas}} = Mf = MB^* x$$

Note: f_{meas} is m dimensional with $m \ll N$!

- **Reconstruction method:** Solve the underdetermined sparse recovery problem:

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

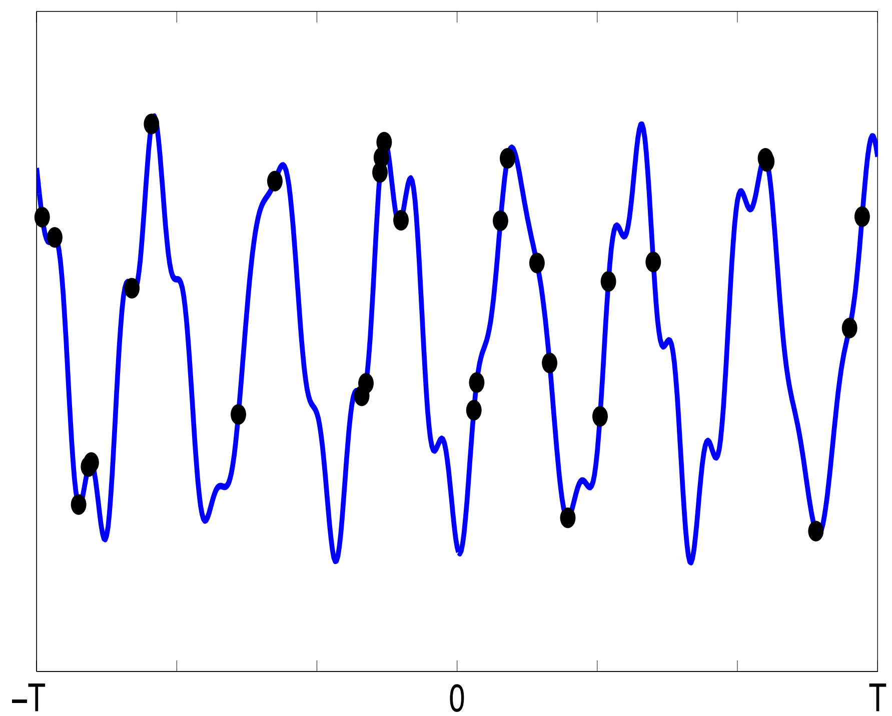
Compressed sensing: back to our example

CS signal model: $f \in B_\Omega$ and f has a **sparse Fourier transform**:

$$\hat{f} = F^* f \quad \text{has few non-zero entries } (B = F, \text{ the DFT matrix}).$$

CS measurement scheme: Collect $m \approx N/2$ samples at **irregular points**, i.e., average sampling density is **only 50% of Nyquist rate**.

$$f_{\text{meas}} = Rf = RF^* \hat{f} \quad (M = R,\ m \text{ random rows of identity}).$$

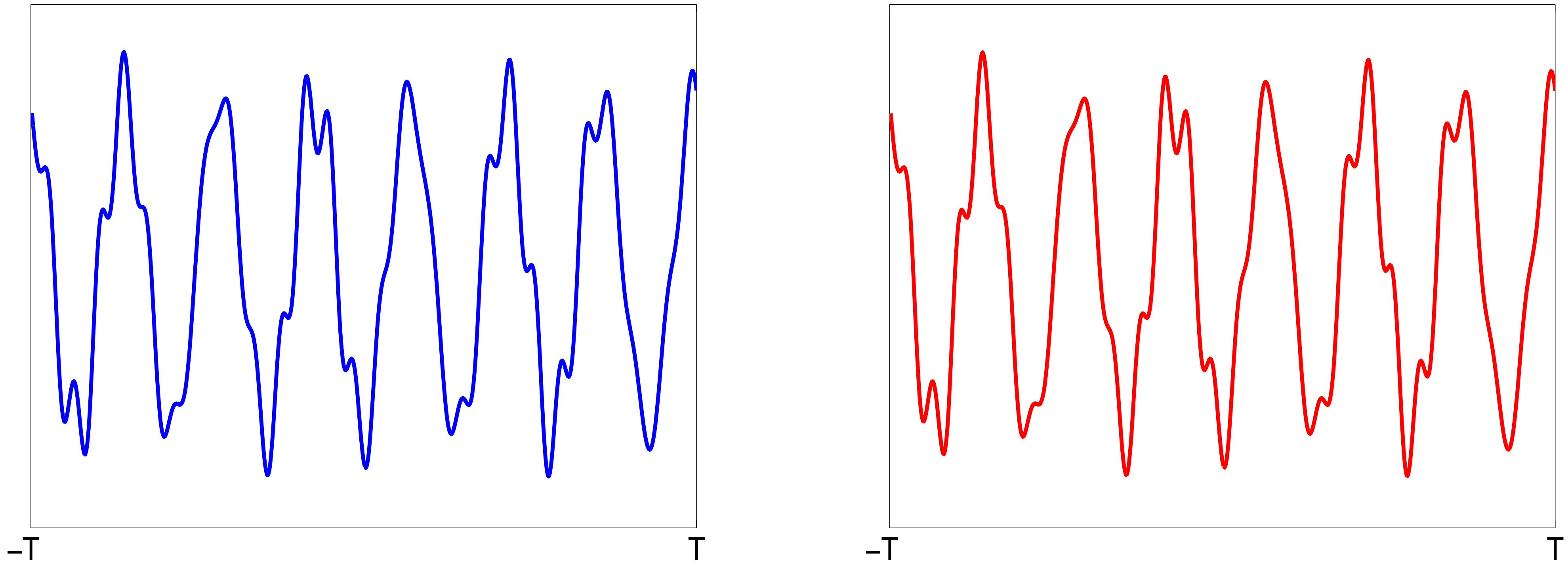


CS Reconstruction: We can recover $\hat{f}$ (thus f) from these samples via:

$$\hat{f}_{\text{approx}} = \arg\min \|z\|_0 \text{ subject to } RF^* z = f_{\text{meas}}.$$

Compressed sensing theory – imposing sparsity

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



- We get essentially perfect reconstruction!
- How did we solve the combinatorial optimization problem:

$$\min \|z\|_0 \text{ subject to } RF^* z = f_{\text{meas}}?$$

We will come back to this later.

Compressed sensing theory – imposing sparsity

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?

Compressed sensing theory – imposing sparsity

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?

Compressed sensing theory - sparsity transforms

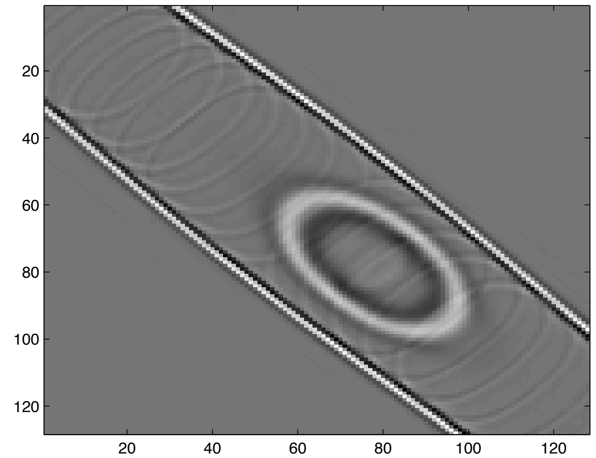
How do we find sparsity transforms?

- 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...
- In the seismic setting, curvelet transform is the transform of our choice.

Sparsity transform - seismic

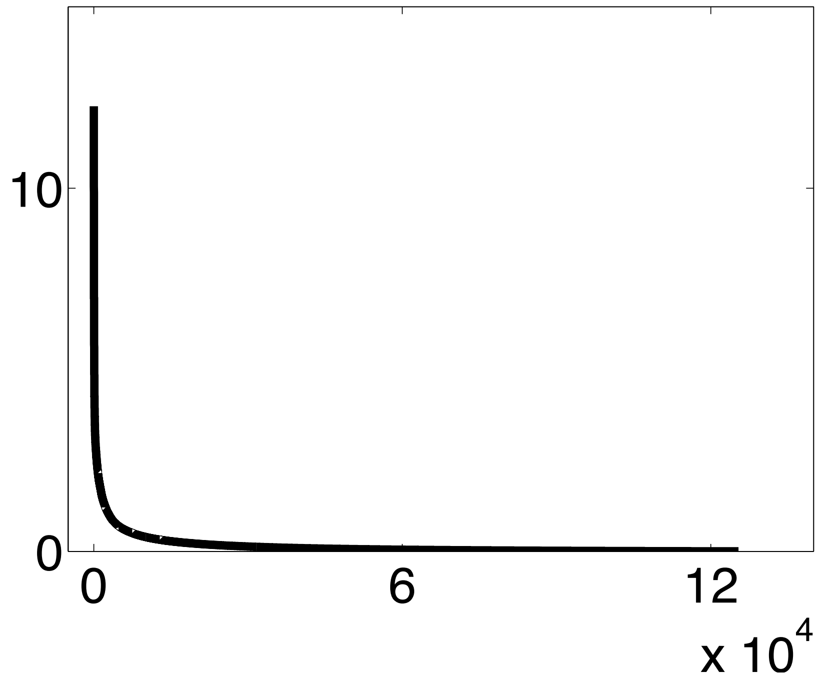
Curvelet transform sparsifies seismic data and images.

sampled Green's function



a curvelet atom

sorted coefficients



Compressed sensing theory: main questions

Sparse recovery problem:

$$x_{\text{approx}} = \text{``sparsest'' } z \text{ such that } f_{\text{meas}} = MB^*z.$$

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

$$b = Mf = \underbrace{MB^*}_{A} x \tag{1}$$

$$\text{OR} \quad \hat{b} = Ax + e \quad \text{(here } e \text{ is additive noise).} \tag{2}$$

Let $\Delta : \mathbb{R}^m \mapsto \mathbb{R}^N$ be a reconstruction map (or "decoder").

Some design goals:

C1. $\Delta(Ax) = x$ whenever x is k-sparse (**exact reconstruction for sufficiently small** k).

C2. $\|x - \Delta(Ax + e)\| \lesssim \|e\| + \|x - x_k\|$. **Reconstruction works for noisy measurements and approx. sparse signals.**

C3. $\Delta(\cdot)$ can be **computed efficiently** (in some sense).

C4. Number of measurements m is as small as possible (depending on k,N, and the choise of the measurement matrix M).

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:

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

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

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:

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

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

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:

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

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

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:

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

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

In particular:

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

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

- 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.

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:

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

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

In particular:

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

  $\Delta_1(b) = x$, i.e., exact recovery, **for every k-sparse x.**

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

  $\|x - \Delta_1(b)\| \lesssim \sigma_k(x)_{\ell_1}/\sqrt{k}$, 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.

  # measurements $\sim$ log of the ambient dimension (grid size)

- 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 $A = RF^*$ is a "good" compressive sampling matrix. **(See Enrico Au-Yeung's talk.)**

CS – incorporating prior info

CS is a non-adaptive sampling paradigm: Measurement matrix is fixed once and for all, regardless of the signal to be acquired.

CS – incorporating prior info

CS is a non-adaptive sampling paradigm: Measurement matrix is fixed once and for all, regardless of the signal to be acquired.

Remainder of the talk: Methods of incorporating prior information on the support of the specific signal of interest to sparse recovery. In all:

- Sensing is non-adaptive: Collect the measurements b (or $\hat{b}$ if there is noise) using an arbitrary CS matrix.
- Recovery is adaptive:
  - Suppose we have prior information on the support of x. In particular we have a support estimate that is generally partial and possibly inaccurate.
  - Use such prior support info to improve sparse recovery.
- **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:

- 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

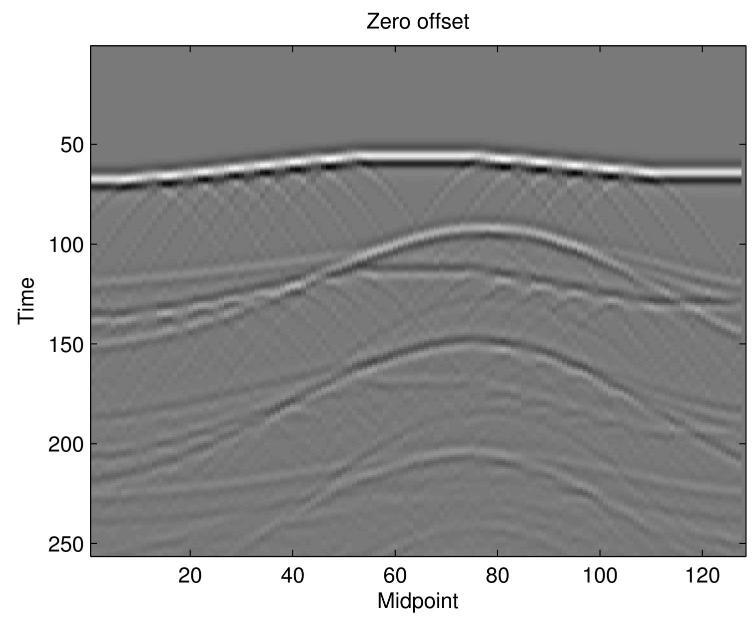
In many applications, it is possible to draw an estimate of the support of the signal, for example:

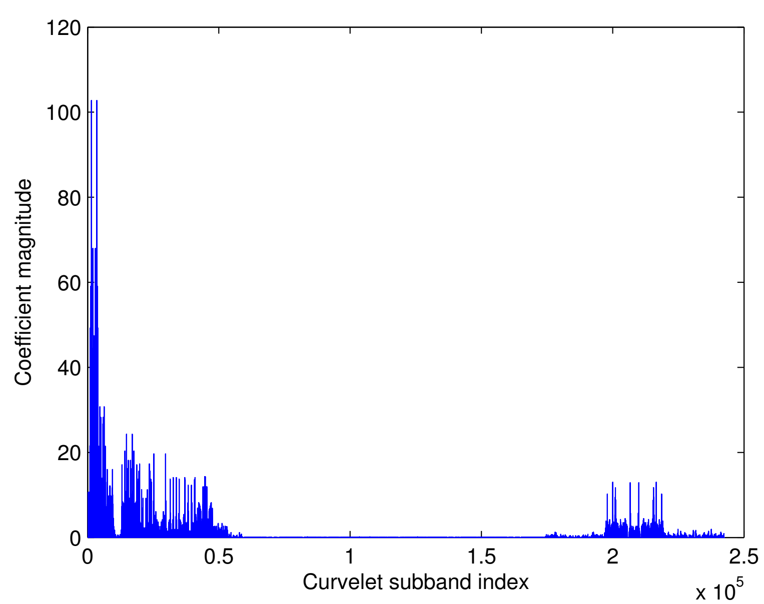
- Natural images have large DCT coefficients that are localized in the low frequency subbands.
- Video sequences are temporally correlated, resulting in a shared subset of their support.
- **Seismic data: adjacent frequency slices or offset gathers have correlated curvelet support.**

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.
- **Seismic data: adjacent frequency slices or offset gathers have correlated curvelet support.**

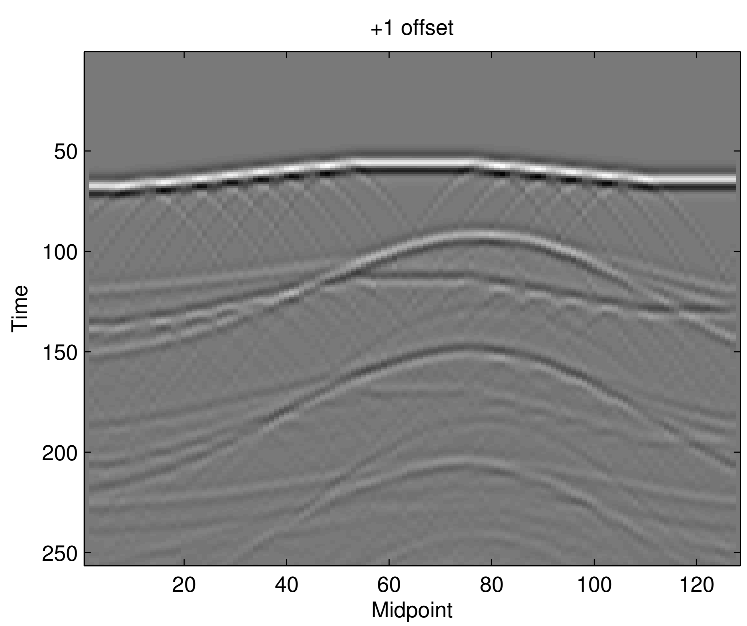


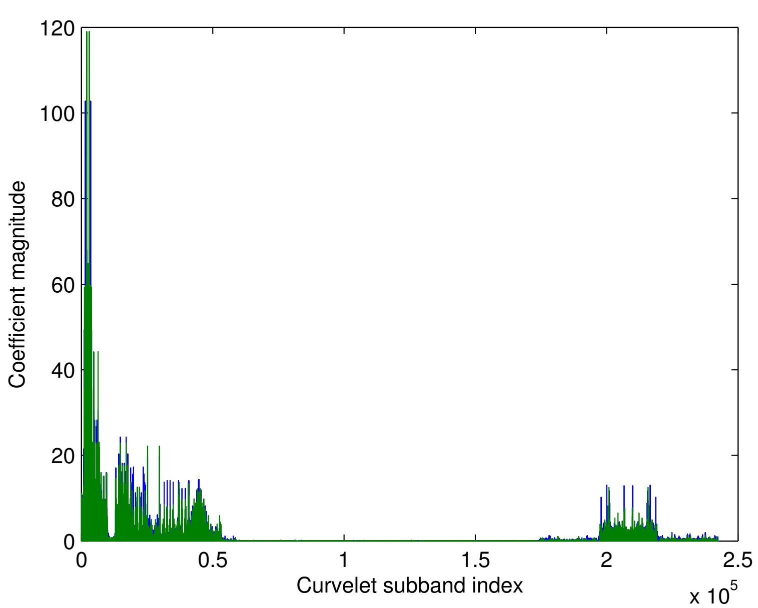


Signals with Prior Information

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.
- **Seismic data: adjacent frequency slices or offset gathers have correlated curvelet support.**

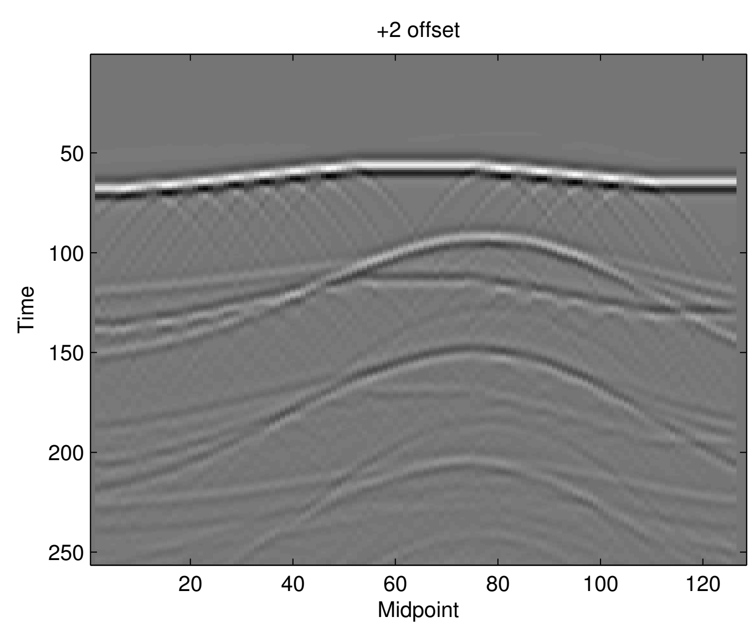


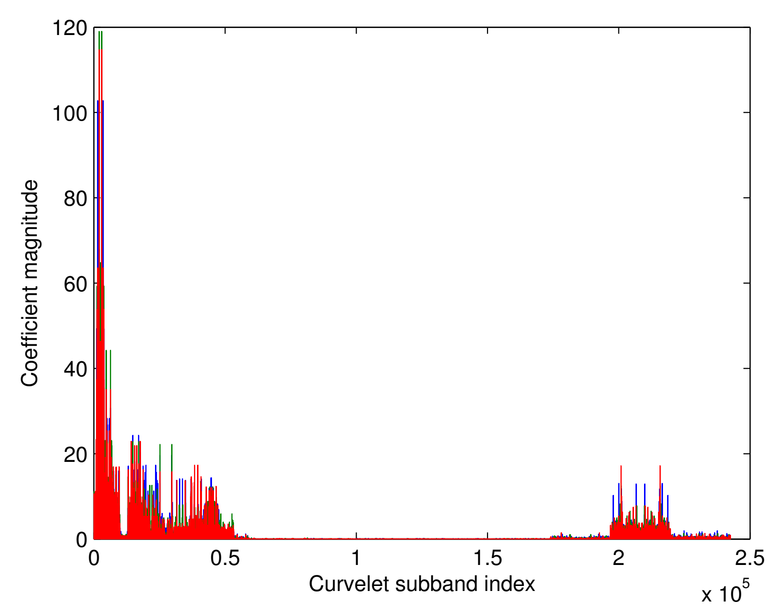


Signals with Prior Information

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.
- **Seismic data: adjacent frequency slices or offset gathers have correlated curvelet support.**





CS – incorporating prior info

Various methods we will discuss:

1. *Recovery using weighted ℓ_1 minimization.* (Mansour)
    - Choose appropriate weights “on-support” and “off-support”.
2. *Recovery using weighted ℓ_p minimization,* $0 < p < 1$ (Ghadermarzy)
    - Similar to above, but now based on non-convex optimization.
3. *Recovery using weighted ℓ_1 minimization of analysis coefficients* (Hargreaves)
    - Analysis formulation when sparsity transform is redundant (e.g., curvelets) with a novel weighting scheme.
4. *Weighted randomized Kaczmarz for sparse solutions of overdetermined linear systems* (Mansour)
    - A row-action method for solving overdetermined systems with sparse solutions.
    - Surprisingly effective for CS (underdetermined systems) as well.

Problem formulation – synthesis

The following applies to weighted ℓ_1 and weighted ℓ_p.

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 – synthesis

The following applies to weighted ℓ_1 and weighted ℓ_p.

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 I – weighted ℓ_1 minimization

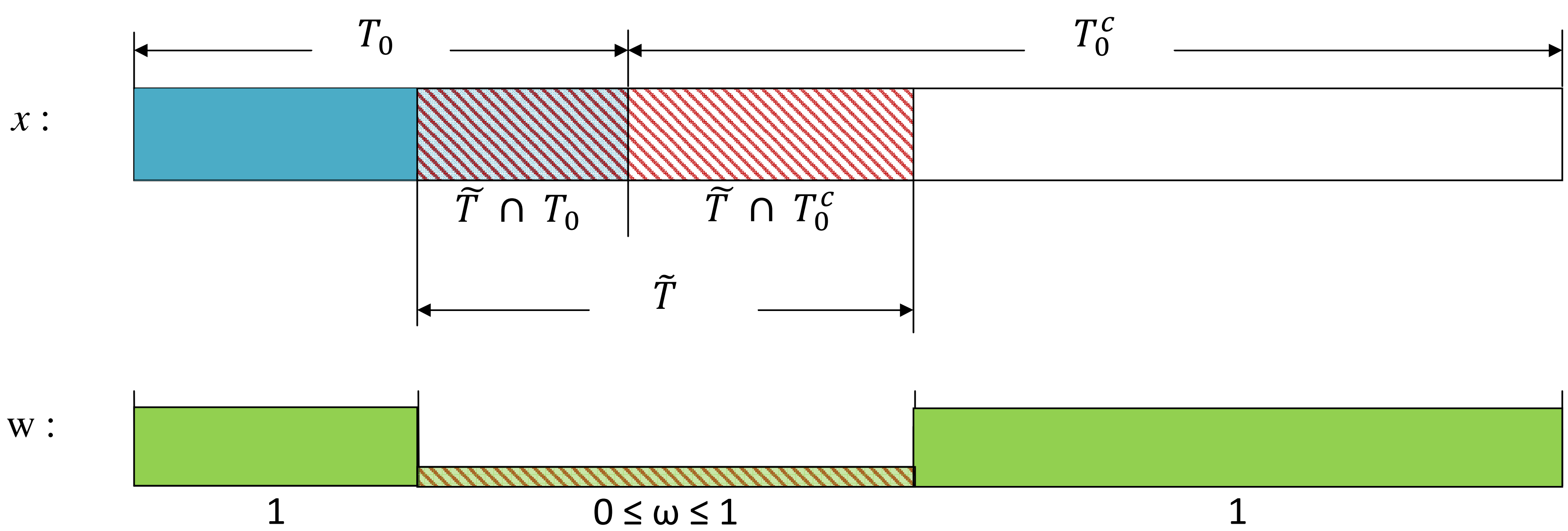
Given a set of (noisy) measurements $\hat{b}$, define

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

where

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

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

$$\|\Delta^{\epsilon}_{1,\mathrm{w}}(\hat{b}) - x\|_2 \leq C'_0\epsilon + C'_1 k^{-1/2}(\omega\|x_{T_o^c}\|_1 + (1-\omega)\|x_{\widetilde{T}^c \cap T_0^c}\|_1)$$

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

Remarks.

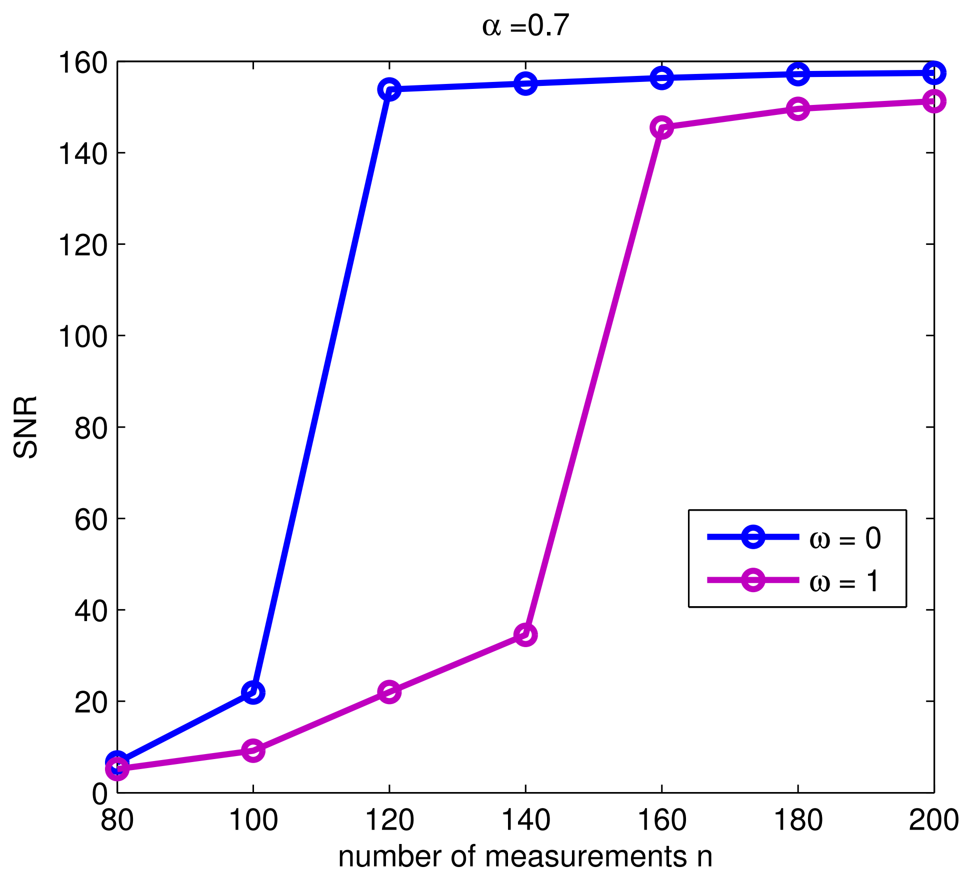
1. Above, $0 \leq \omega \leq 1$ is a fixed weight. If we set $\omega = 1$, our theorem reduces to the robust recovery theorem of CRT.
2. Recall $0 \leq \alpha \leq 1$ describes the accuracy of $\widetilde{T}$ and ρ describes its size.
3. **The sufficient conditions above are weaker than those for ℓ_1 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

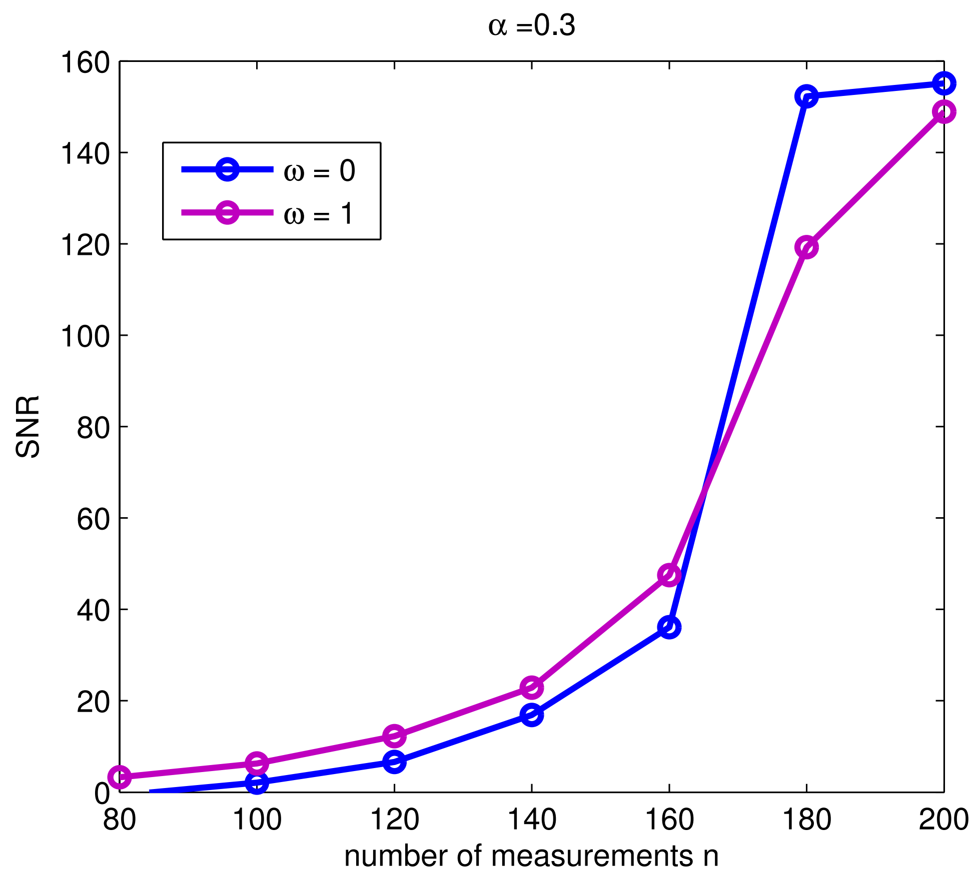
Numerical experiments – sparse signals

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

Numerical experiments – sparse signals

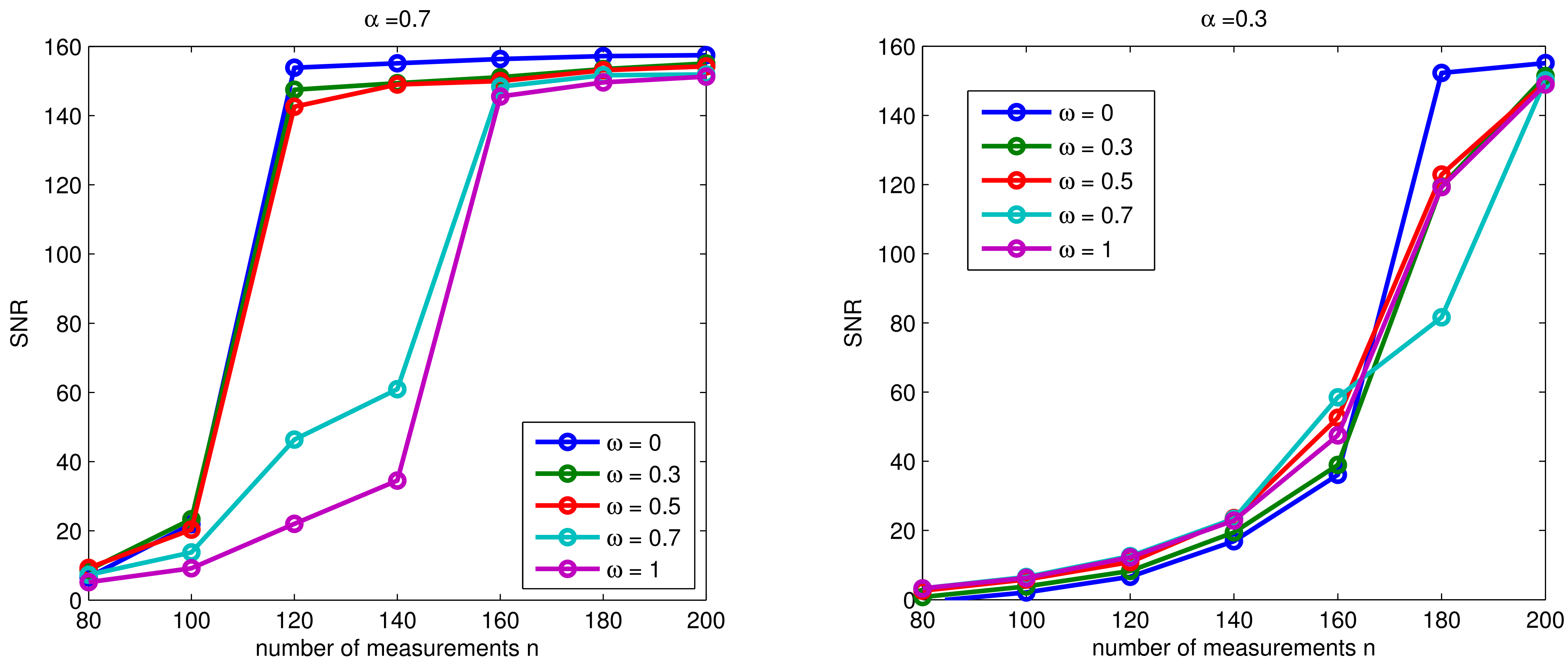
- SNR averaged over 20 experiments for k-sparse signals x with $k = 40$, and $N = 500$.
- **The noise free case:**





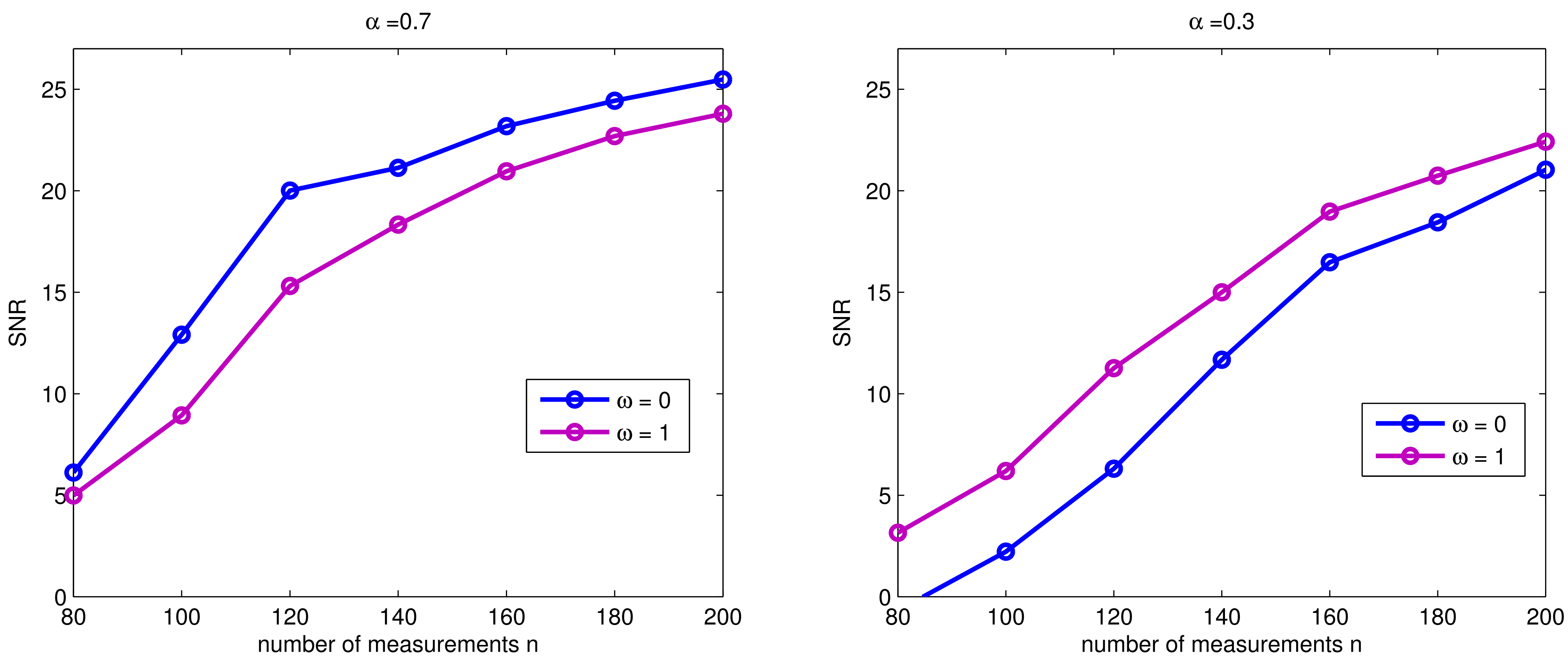
Numerical experiments – sparse signals

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



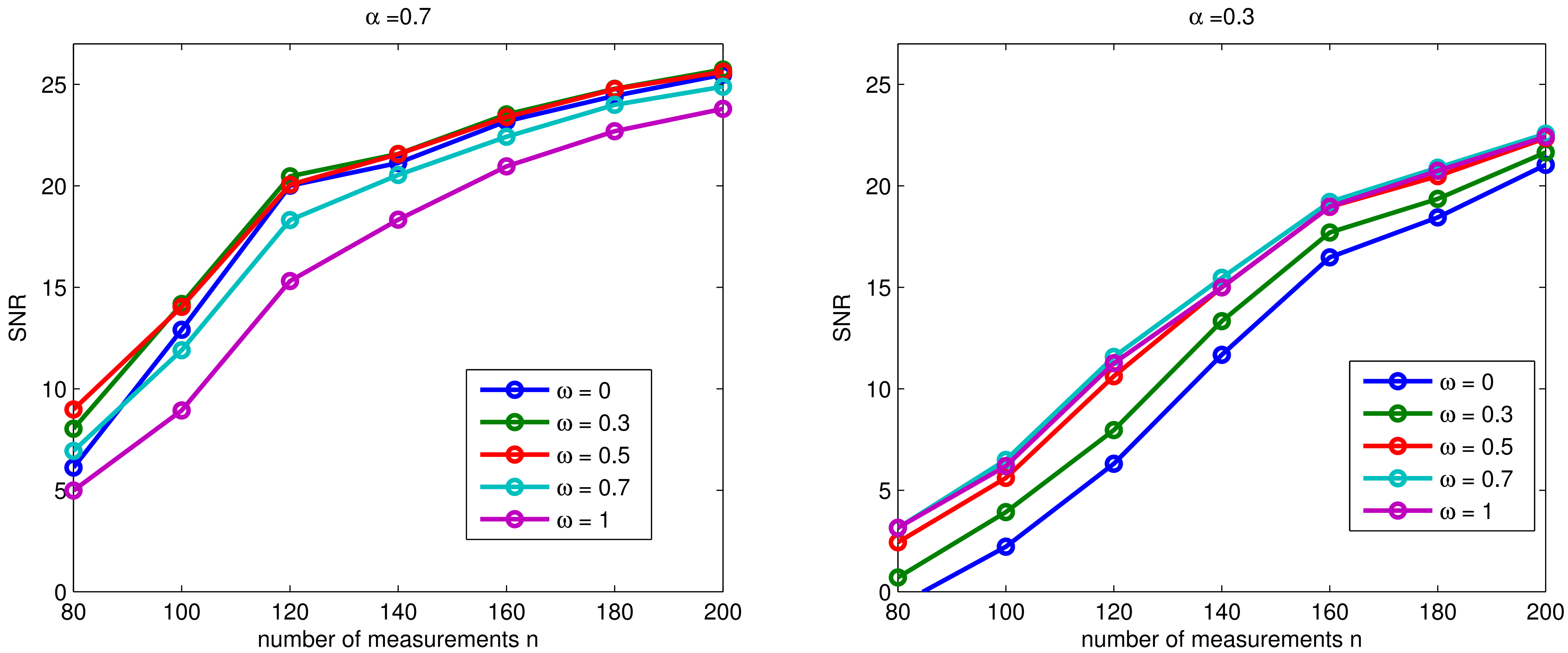
Numerical experiments – sparse signals

- 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:**

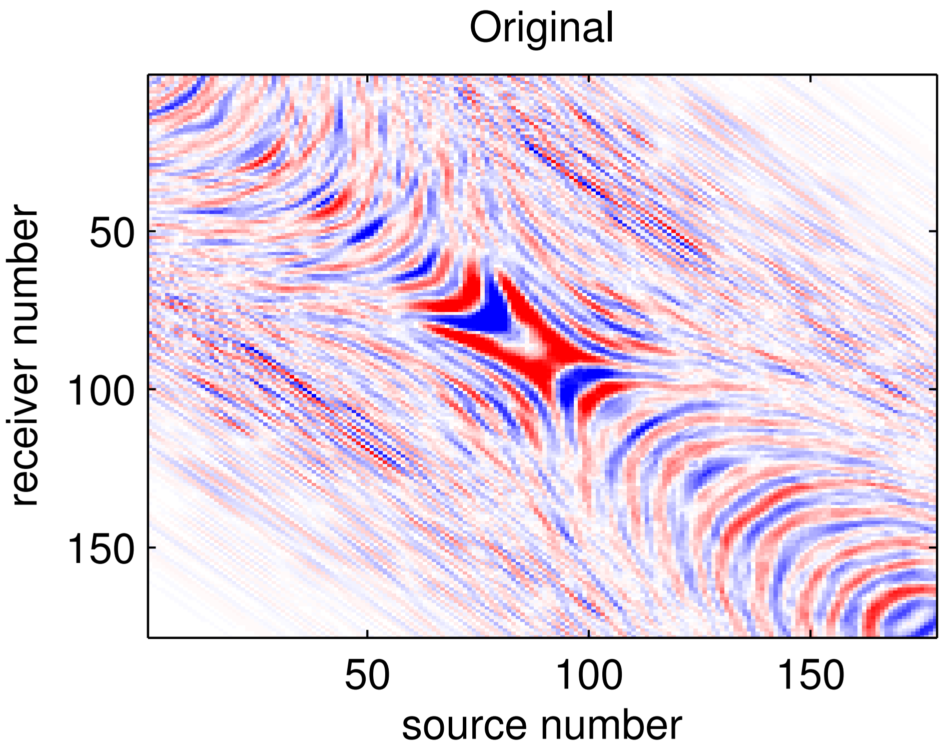


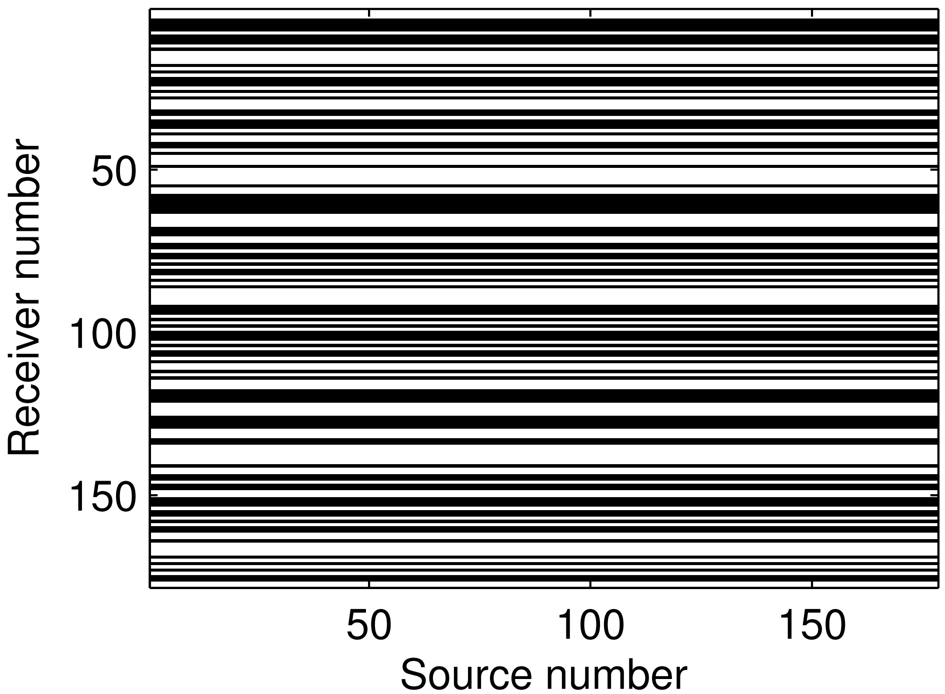
Compressed sensing of seismic lines

- Full seismic line (Gulf of Suez) with 178 shots, 178 receivers, and 500 time samples.

Compressed sensing of seismic lines

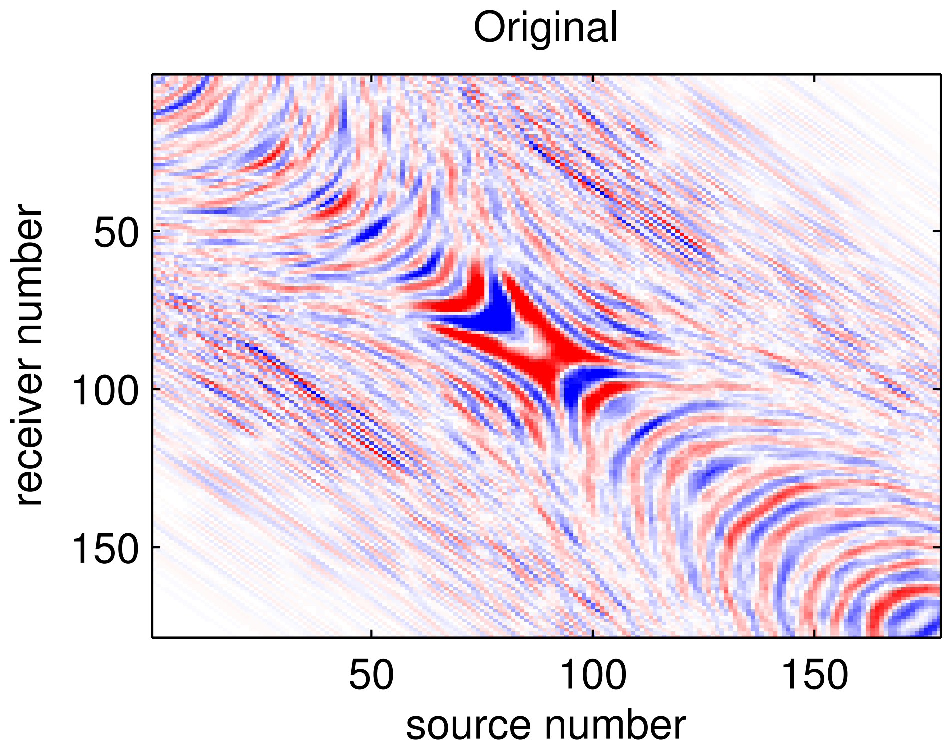
- Full seismic line (Gulf of Suez) with 178 shots, 178 receivers, and 500 time samples.
- Due to budgetary requirements or device malfunctioning, some receivers are inactive (e.g.: time slice 350).

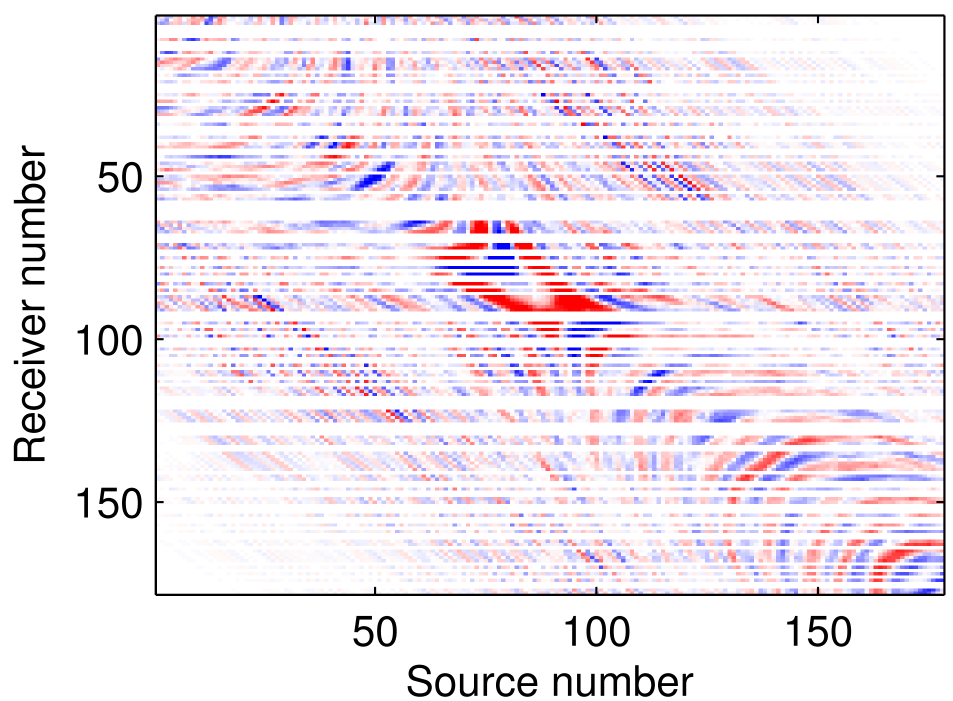




Compressed sensing of seismic lines

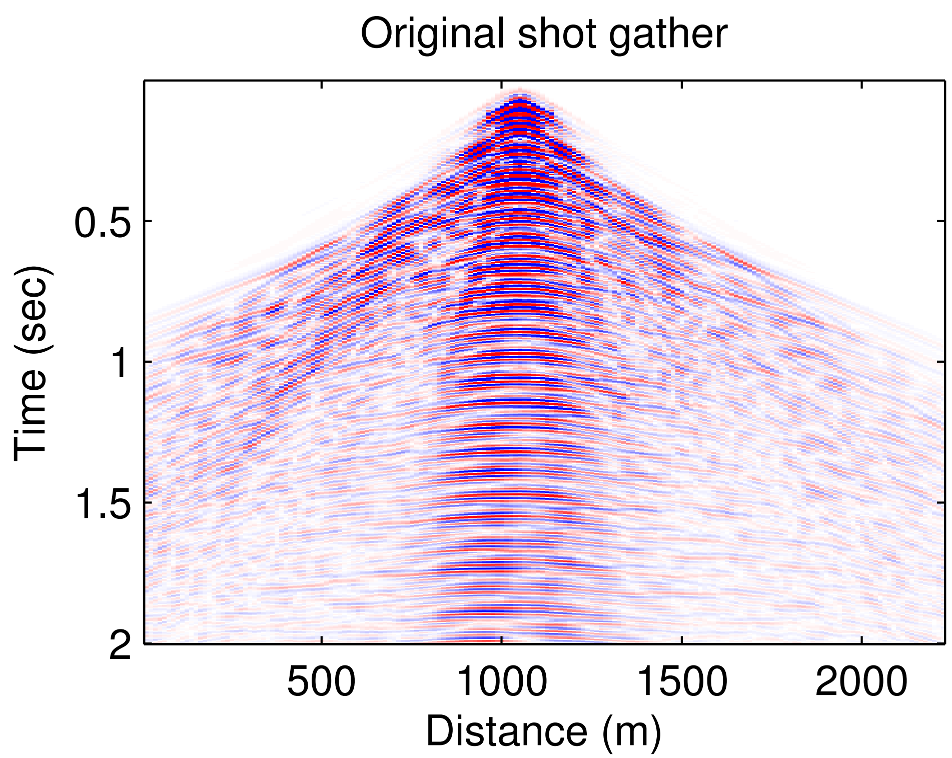
- Full seismic line (Gulf of Suez) with 178 shots, 178 receivers, and 500 time samples.
- Due to budgetary requirements or device malfunctioning, some receivers are inactive (e.g.: time slice 350).

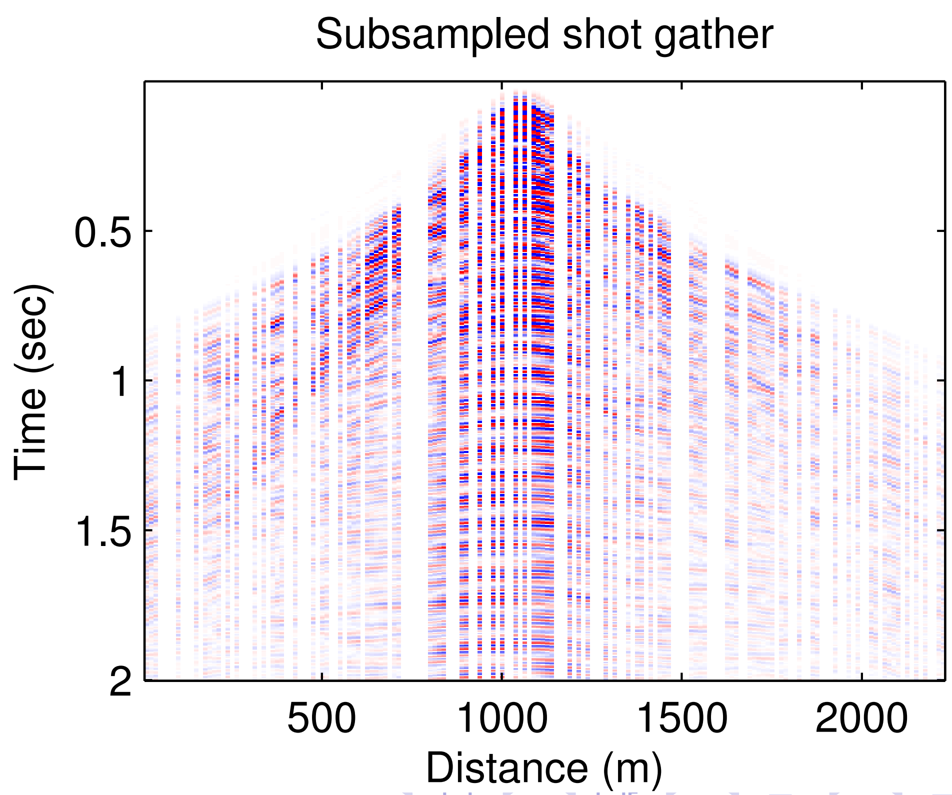




Compressed sensing of seismic lines

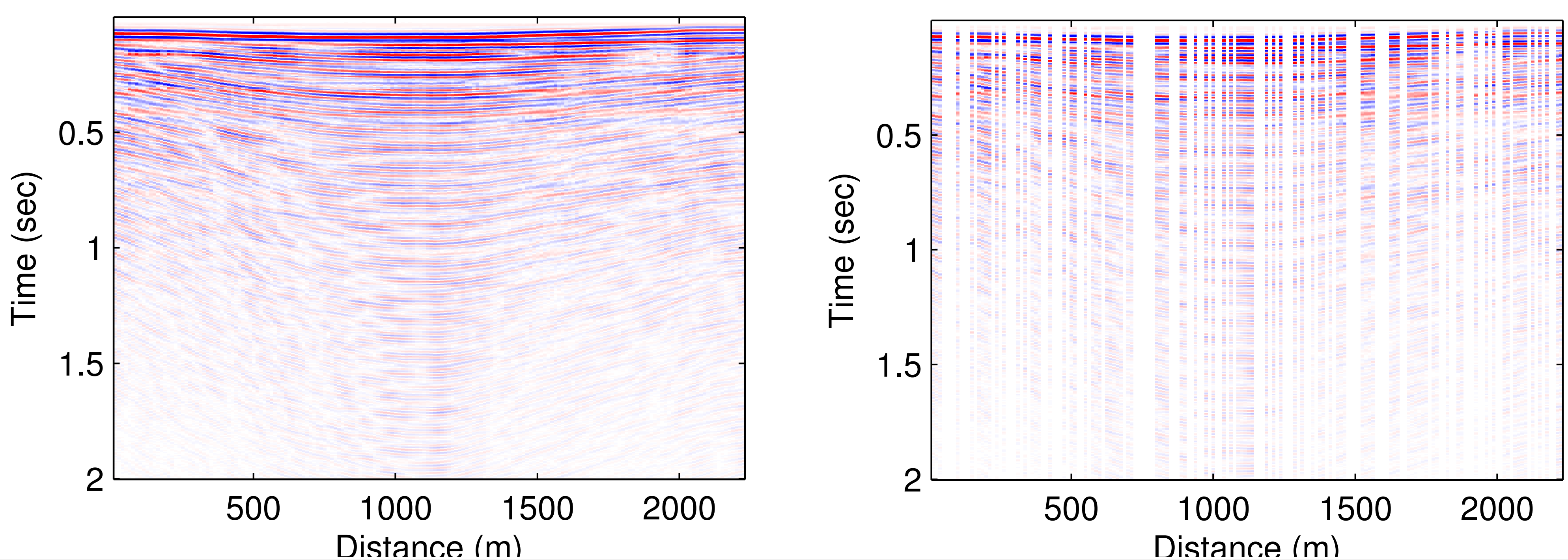
- Full seismic line (Gulf of Suez) with 178 shots, 178 receivers, and 500 time samples.
- Due to budgetary requirements or device malfunctioning, some receivers are inactive (e.g.: time slice 350).
- Results in missing data along entire time axis (eg: common shot gather # 84)





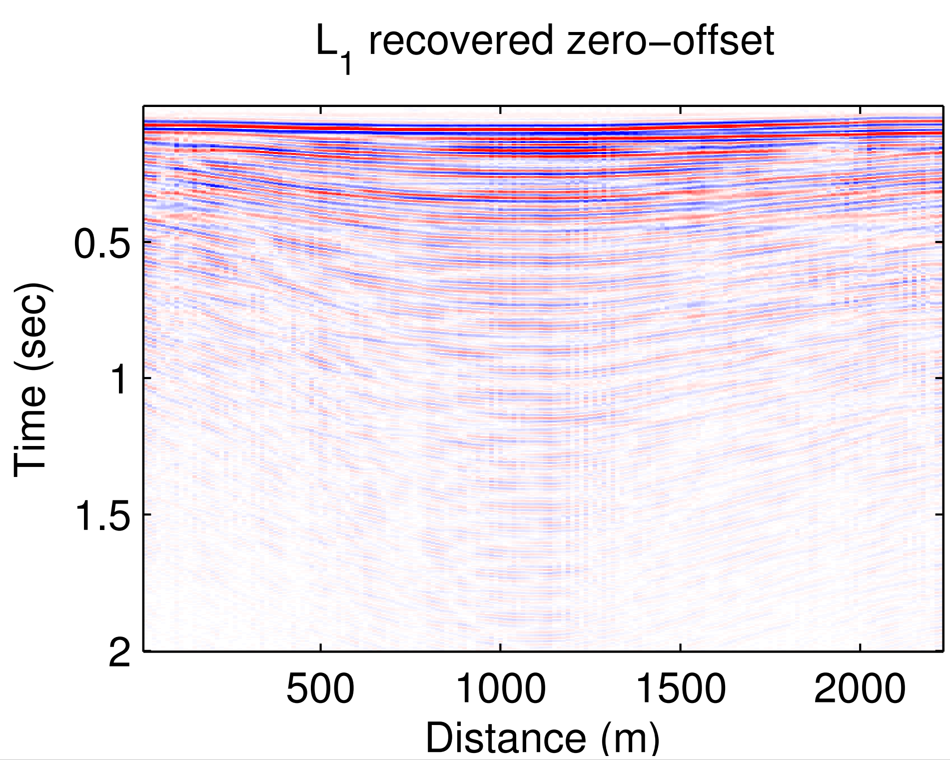
Recovery in offset domain

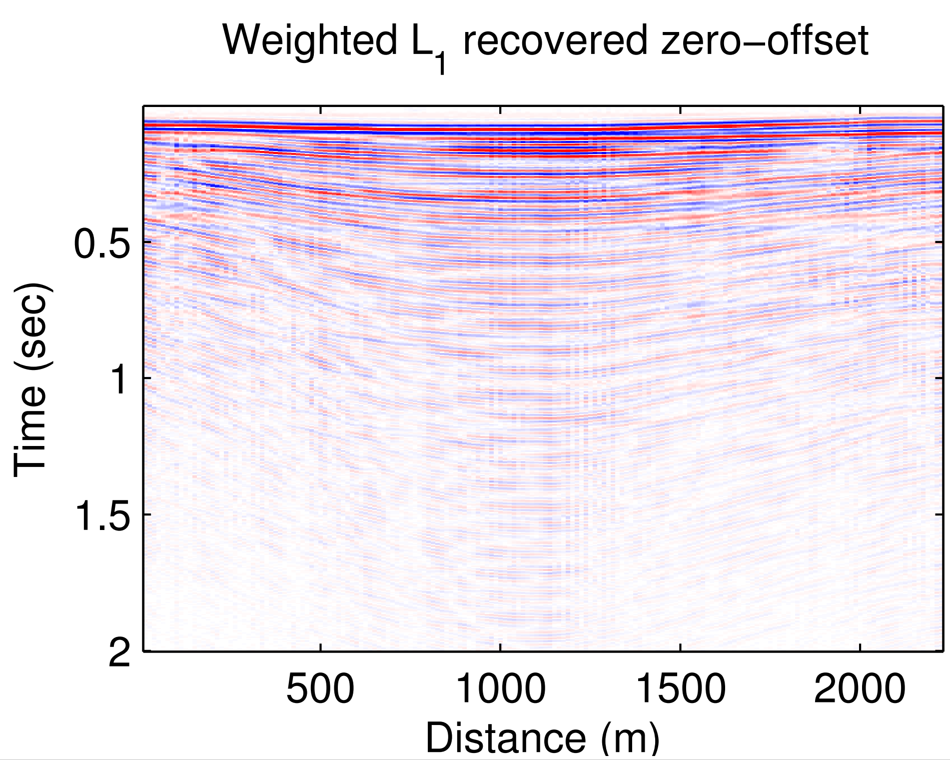
- Seismic line data is correlated in the midpoint-offset domain.
- Map the subsampling mask to act on offset slices (e.g., see zero offset slice below).



Recovery in offset domain

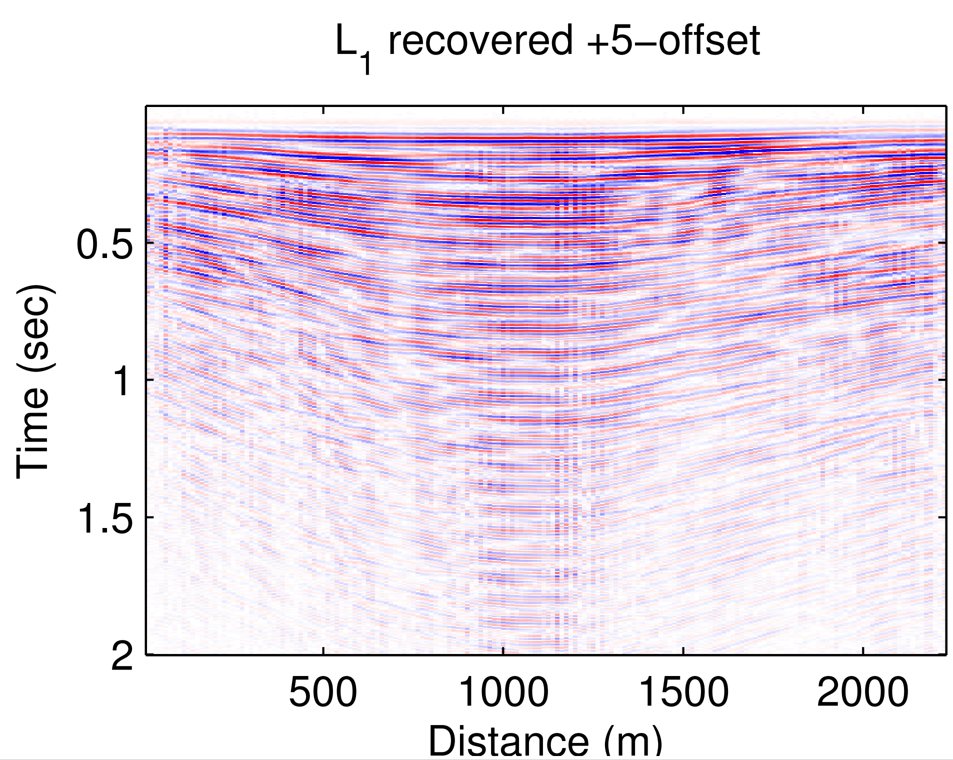
- Seismic line data is correlated in the midpoint-offset domain.
- Map the subsampling mask to act on offset slices (e.g., see zero offset slice below).
- Recover the zero offset using standard ℓ_1 minimization (same quality for wL1 and L1).

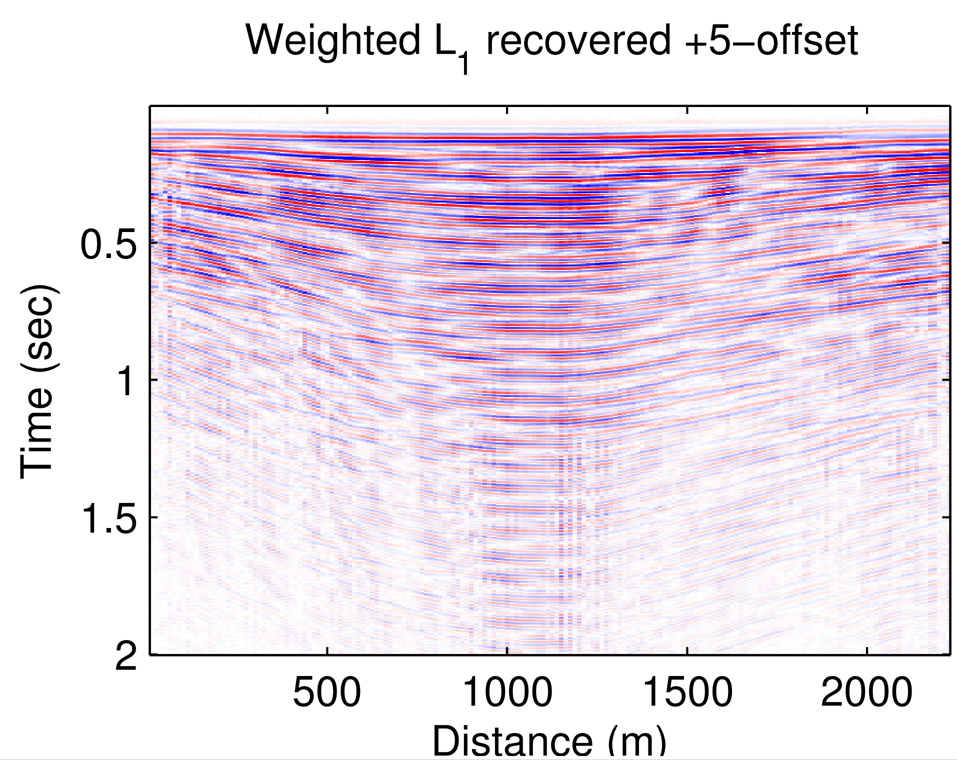




Recovery in offset domain

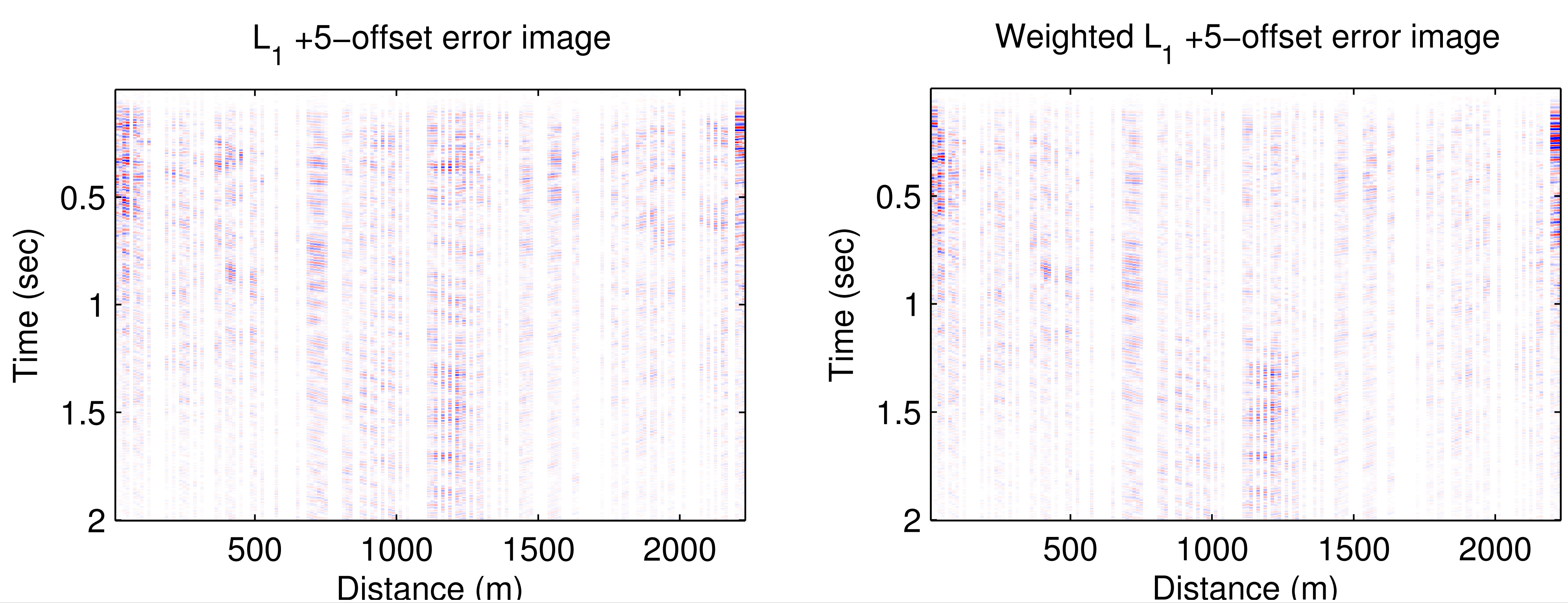
- Seismic line data is correlated in the midpoint-offset domain.
- Map the subsampling mask to act on offset slices (e.g., see zero offset slice below).
- Recover the zero offset using standard ℓ_1 minimization (same quality for wL1 and L1).
- Use the support of the zero offset slice to weight the recovery of other offset slices (eg: $+5$ offset).





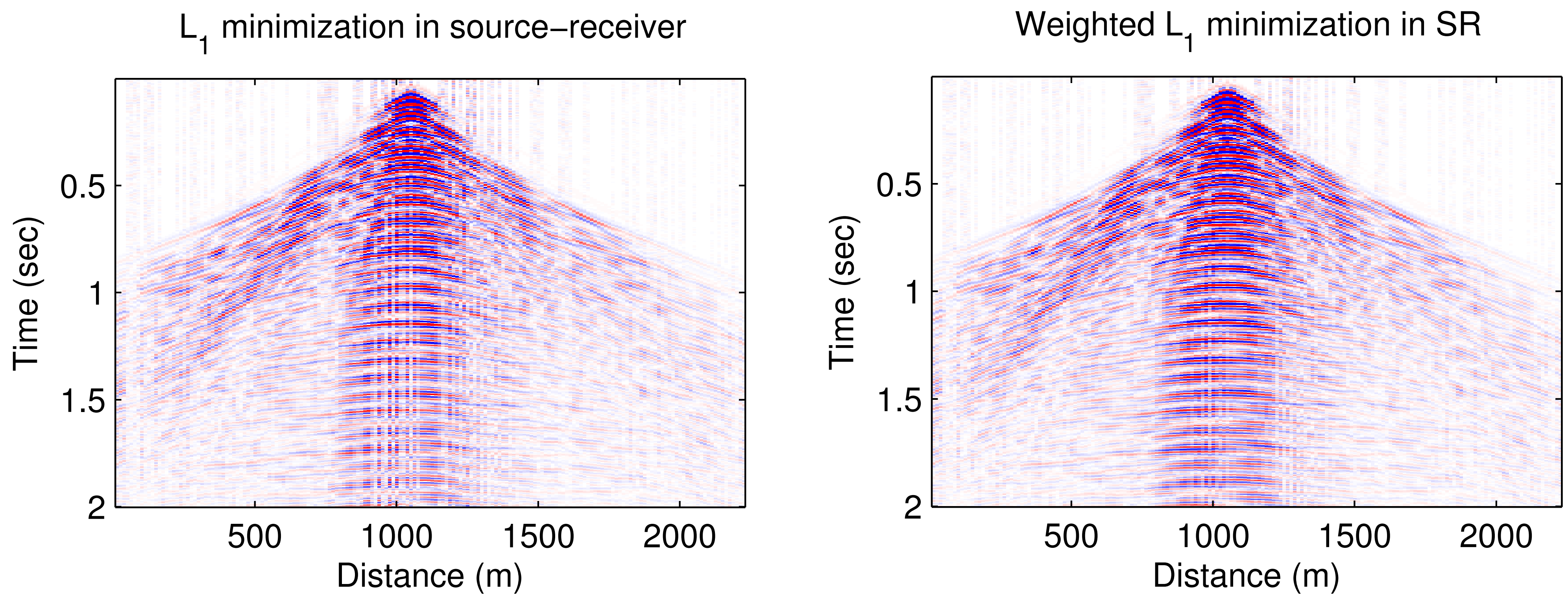
Recovery in offset domain

- Seismic line data is correlated in the midpoint-offset domain.
- Map the subsampling mask to act on offset slices (e.g., see zero offset slice below).
- 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: $+5$ offset).



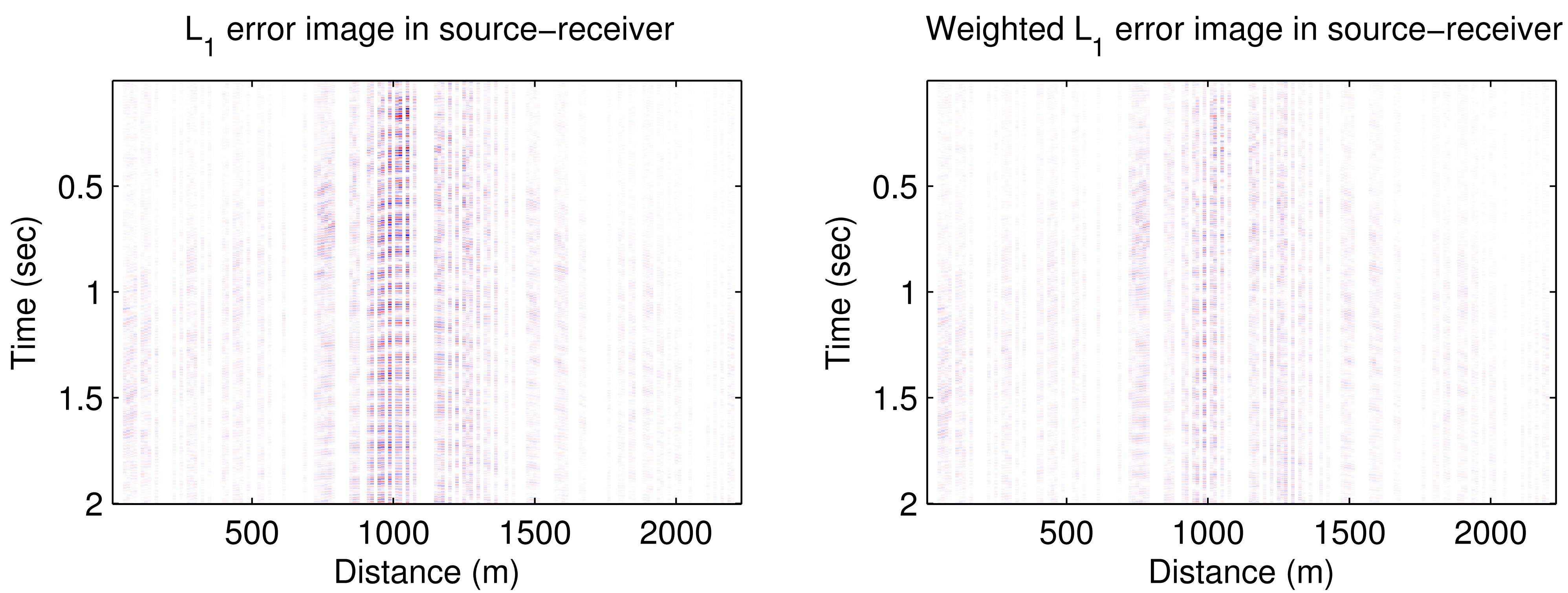
Performance of weighted ℓ_1 vs standard ℓ_1

- Map the data back to the source receiver domain (eg: shot gather # 84).



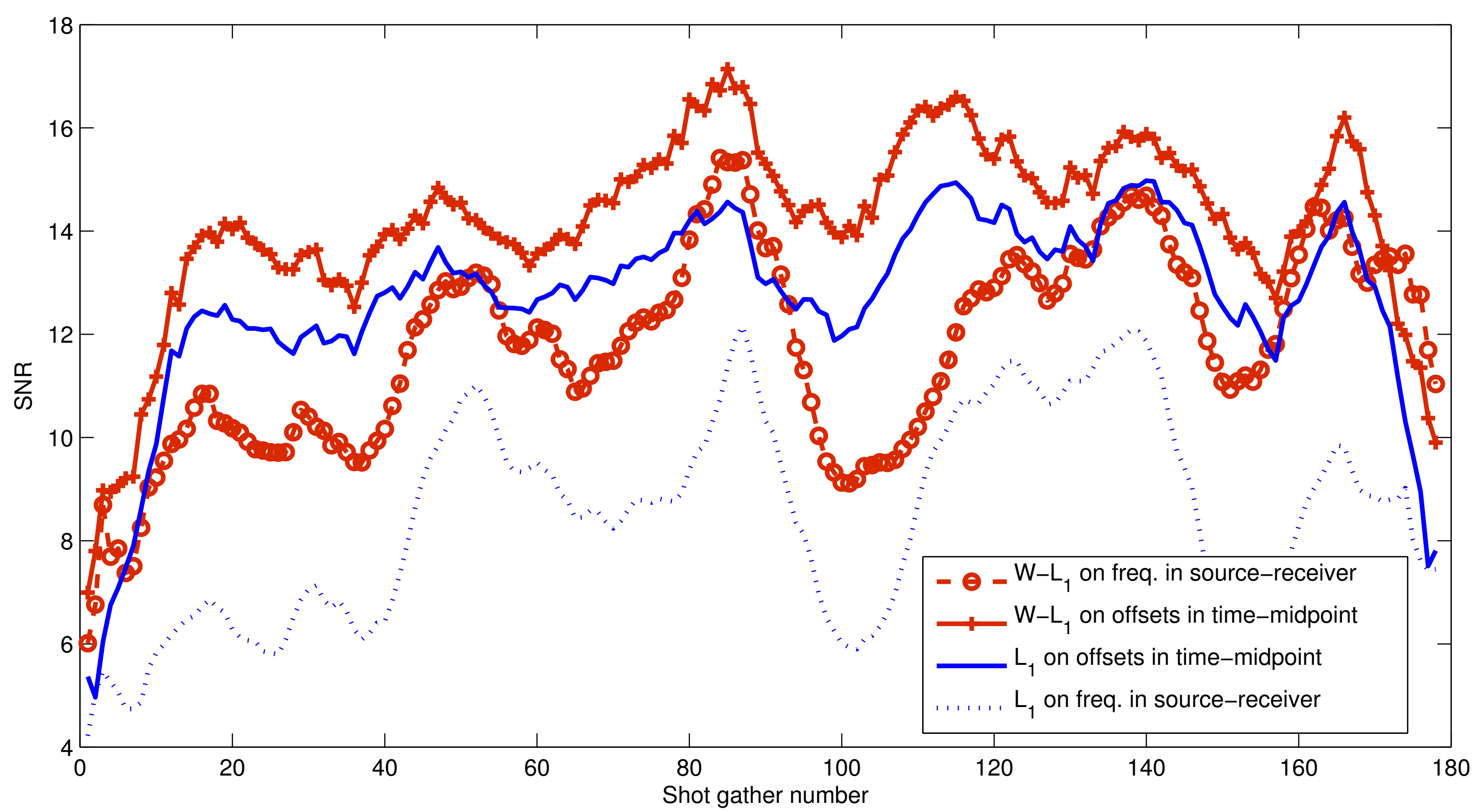
Performance of weighted ℓ_1 vs standard ℓ_1

- Map the data back to the source receiver domain (eg: shot gather # 84).



Performance of weighted ℓ_1 vs standard ℓ_1

- Map the data back to the source receiver domain (eg: shot gather # 84).
- Signal to noise ratio (SNR) of all 128 shot gathers.



Proposed Algorithm II – weighted ℓ_p minimization

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

$$\Delta^{\epsilon}_{p,\mathrm{w}}(\hat{b}) := \arg\min_x \ \|x\|^p_{p,\mathrm{w}} \text{ subject to } \|Ax - \hat{b}\|_2 \le \epsilon$$

where

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

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

Remarks:

1. This is a non-convex optimization problem because $0 < p < 1$.
2. We know that ℓ_p minimization can outperform ℓ_1 minimization significantly, e.g., Saab-Yilmaz 2010. This motivates us to consider a weighted version.
3. We can prove better sufficient conditions for recovery compared to weighted ℓ_1. See Ghadermarzy's talk for details.

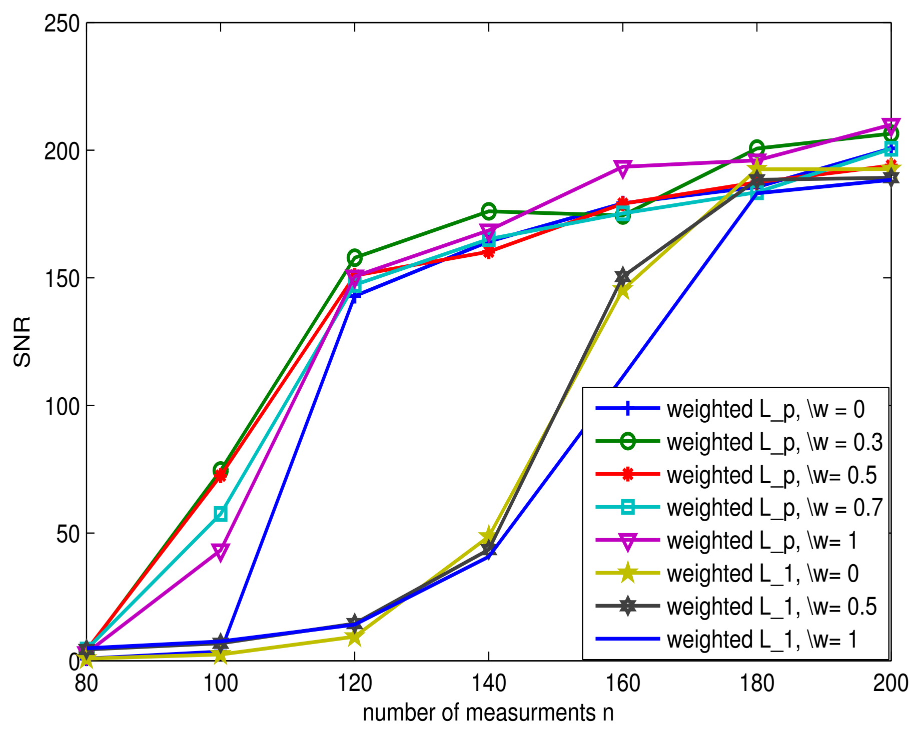
Weighted ℓ_p – numerical experiments

- SNR averaged over 10 experiments for k-sparse signals x with $k = 40$, $N = 500$, and $p = 0.5$.

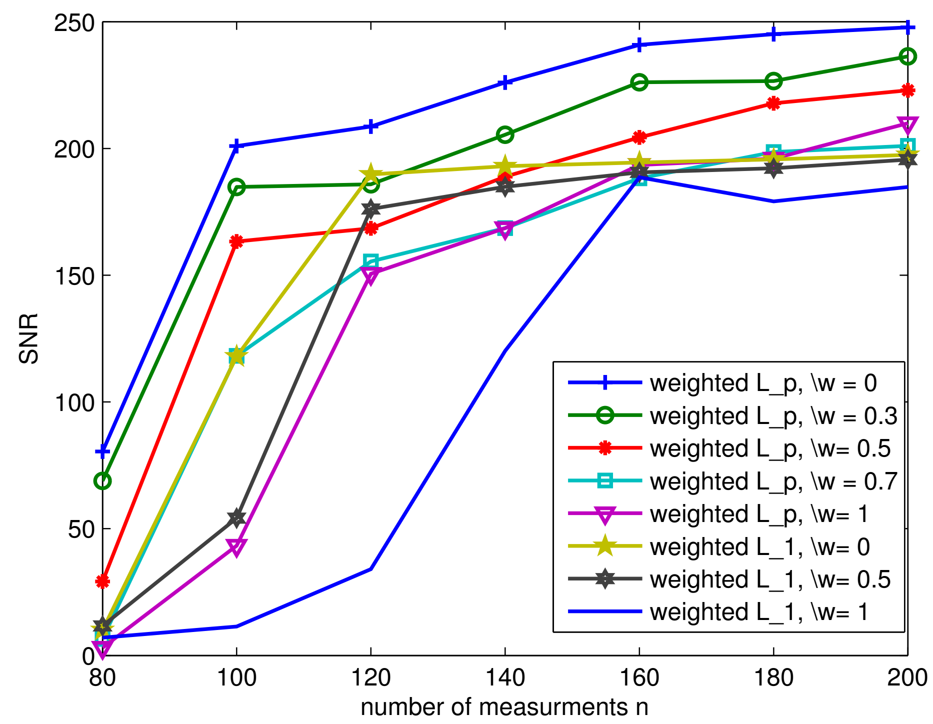
Weighted ℓ_p – numerical experiments

- SNR averaged over 10 experiments for k-sparse signals x with $k = 40$, $N = 500$, and $p = 0.5$.
- **The noise free case:**

$\alpha = 0.3$



$\alpha = 0.7$



Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|z\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = MBz \quad \Longrightarrow \quad \tilde{f} = B\tilde{x} \tag{3}$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = Mg. \tag{4}$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.
- **Main difference:** B ($N \times N$) is invertible, D ($N \times L$, $L > N$) is not! So: infinitely many ways of choosing transform coefficients.
- **Main implication:** Can replace B^* in (4) with any right inverse of D. Each choice will result in a different optimization problem (4′)

$$\tilde{x} = \arg\min_z \|z\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = MDz \quad \Longrightarrow \quad \tilde{f} = D\tilde{x} \tag{SY}$$

$$\tilde{\tilde{f}} = \arg\min_g \|D_{\text{RI}}g\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = Mg. \tag{AN}$$

Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|\overbrace{z}^{B^*g}\|_1 \text{ s.t. } f_{\text{meas}} = M\overbrace{Bz}^{g} \implies \tilde{f} = B\tilde{x} \quad (3)$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \text{ s.t. } f_{\text{meas}} = Mg. \quad (4)$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.
- **Main difference:** B $(N \times N)$ is invertible, D $(N \times L,\ L > N)$ is not! So: infinitely many ways of choosing transform coefficients.
- **Main implication:** Can replace B^* in (4) with any right inverse of D. Each choice will result in a different optimization problem (4′)

$$\tilde{x} = \arg\min_z \|z\|_1 \text{ subject to } f_{\text{meas}} = MDz \implies \tilde{f} = D\tilde{x} \quad \text{(SY)}$$

$$\tilde{\tilde{f}} = \arg\min_g \|D_{\text{RI}}g\|_1 \text{ subject to } f_{\text{meas}} = Mg. \quad \text{(AN)}$$

Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|\overbrace{z}^{B^*g}\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = M\overbrace{Bz}^{g} \implies \tilde{f} = B\tilde{x} \tag{3}$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \quad \text{s.t.} \quad f_{\text{meas}} = Mg. \tag{4}$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.
- **Main difference:** B ($N \times N$) is invertible, D ($N \times L$, $L > N$) is not! So: infinitely many ways of choosing transform coefficients.
- **Main implication:** Can replace B^* in (4) with any right inverse of D. Each choice will result in a different optimization problem (4′)

$$\tilde{x} = \arg\min_z \|z\|_1 \quad \text{subject to} \quad f_{\text{meas}} = MDz \implies \tilde{f} = D\tilde{x} \tag{SY}$$

$$\tilde{\tilde{f}} = \arg\min_g \|D_{\mathrm{RI}}g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg. \tag{AN}$$

Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|z\|_1 \text{ subject to } f_{\text{meas}} = MBz \implies \tilde{f} = B\tilde{x} \tag{3}$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \text{ subject to } f_{\text{meas}} = Mg. \tag{4}$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.

Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|z\|_1 \text{ subject to } f_{\text{meas}} = MBz \implies \tilde{f} = B\tilde{x} \tag{3}$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \text{ subject to } f_{\text{meas}} = Mg. \tag{4}$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.
- **Main difference:** B $(N \times N)$ is invertible, D $(N \times L,\ L > N)$ is not! So: infinitely many ways of choosing transform coefficients.

Proposed Algorithm III – weighted ℓ_1 analysis

- **Signal model:** f is sparse w.r.t. a basis B: $f = Bx$, x is sparse. Then, $x = B^{-1}f = B^*f$ assuming B is an ONB.
- Sparse recovery: Let M be the measurement matrix. Two equivalent formulations.

$$\tilde{x} = \arg\min_z \|z\|_1 \text{ subject to } f_{\text{meas}} = MBz \implies \tilde{f} = B\tilde{x} \quad (3)$$

$$\tilde{f} = \arg\min_g \|B^*g\|_1 \text{ subject to } f_{\text{meas}} = Mg. \quad (4)$$

- **Signal model:** f is sparse w.r.t. a frame D: $f = Dx$, x is sparse.
- **Main difference:** B ($N \times N$) is invertible, D ($N \times L$, $L > N$) is not! So: infinitely many ways of choosing transform coefficients.
- **Main implication:** Can replace B^* in (4) with any right inverse of D. Each choice will result in a different optimization problem (4′)

$$\tilde{x} = \arg\min_z \|z\|_1 \text{ subject to } f_{\text{meas}} = MDz \implies \tilde{f} = D\tilde{x} \quad \text{(SY)}$$

$$\tilde{\tilde{f}} = \arg\min_g \|D_{\text{RI}}g\|_1 \text{ subject to } f_{\text{meas}} = Mg. \quad \text{(AN)}$$

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_g \|D_{\mathrm{RI}} g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Different, in general, from the synthesis problem when D is redundant!

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_{g} \|D_{\mathrm{RI}} g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Different, in general, from the synthesis problem when D is redundant!
- A special choice: $D_{\mathrm{RI}} = D^{\dagger} = D^*(DD^*)^{-1}$

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_g \|D_{\mathrm{RI}} g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Different, in general, from the synthesis problem when D is redundant!
- A special choice: $D_{\mathrm{RI}} = D^{\dagger} = D^{*}(DD^{*})^{-1}$
- Several preliminary theoretical results: Elad et al., Candès et al., Li et al. Still various fundamental questions open!

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_{g} \|D_{\mathrm{RI}} g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Different, in general, from the synthesis problem when D is redundant!
- A special choice: $D_{\mathrm{RI}} = D^{\dagger} = D^*(DD^*)^{-1}$
- Several preliminary theoretical results: Elad et al., Candès et al., Li et al. Still various fundamental questions open!
- In the case of curvelets: $D = C^H$, and $D^{\dagger} = C$ (as curvelet frames are "Parseval frames", i.e., $C^H C = I$.
- So: the **analysis formulation for seismic:**

$$\tilde{f} = \arg\min_{g} \|Cg\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_{g} \|D_{\mathrm{RI}}\, g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Different, in general, from the synthesis problem when D is redundant!
- A special choice: $D_{\mathrm{RI}} = D^{\dagger} = D^{*}(DD^{*})^{-1}$
- Several preliminary theoretical results: Elad et al., Candès et al., Li et al. Still various fundamental questions open!
- In the case of curvelets: $D = C^H$, and $D^{\dagger} = C$ (as curvelet frames are "Parseval frames", i.e., $C^H C = I$.
- So: the **analysis formulation for seismic:**

$$\tilde{f} = \arg\min_{g} \|Cg\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- Can we use a "weighted" approach again?

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_{g} \|D^{\dagger} g\|_1 \text{ subject to } f_{\text{meas}} = Mg.$$

- A **weighted approach** again if we have a "support estimate"?
- Suppose $\exists$ sparse x such that $f = Dx$ with estimated support $\widetilde{T}$.
  1. We can mimick what we did before:

$$\tilde{f} = \arg\min_{g} \|WD^{\dagger} g\|_1 \text{ subject to } f_{\text{meas}} = Mg.$$

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_g \|D^\dagger g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- A **weighted approach** again if we have a "support estimate"?
- Suppose $\exists$ sparse x such that $f = Dx$ with estimated support $\widetilde{T}$.
  1. We can mimick what we did before:

  $$\tilde{f} = \arg\min_g \|WD^\dagger g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

  2. **Alternative approach:** We write $f = D\widetilde{W}z$ and claim z should also be sparse. Here $\widetilde{W}$ is diagonal with $\omega < 1$ and

  $$\widetilde{W}_{ii} = \begin{cases} 1 & \text{if } i \in \widetilde{T} \\ \omega & \text{if } i \notin \widetilde{T} \end{cases},$$

Proposed Algorithm III – weighted ℓ_1 analysis

Analysis formulation of sparse recovery problem:

$$\tilde{f} = \arg\min_g \|D^\dagger g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

- A **weighted approach** again if we have a "support estimate"?
- Suppose $\exists$ sparse x such that $f = Dx$ with estimated support $\widetilde{T}$.
  1. We can mimick what we did before:

$$\tilde{f} = \arg\min_g \|WD^\dagger g\|_1 \quad \text{subject to} \quad f_{\text{meas}} = Mg.$$

  2. **Alternative approach:** We write $f = D\widetilde{W}z$ and claim z should also be sparse. Here $\widetilde{W}$ is diagonal with $\omega < 1$ and

$$\widetilde{W}_{ii} = \begin{cases} 1 & \text{if } i \in \widetilde{T} \\ \omega & \text{if } i \notin \widetilde{T} \end{cases}, \quad \text{recall} \quad W_{ii} = \begin{cases} \omega & \text{if } i \in \widetilde{T} \\ 1 & \text{if } i \notin \widetilde{T} \end{cases},$$

Proposed Algorithm III – weighted ℓ_1 analysis

With this alternative approach: Given the support estimare $\widetilde{T}$, solve

$$\tilde{f} = \arg\min_g \|(D\widetilde{W})^\dagger g\|_1 \text{ subject to } \|f_{\text{meas}} - Mg\|_2 \leq \epsilon.$$

with

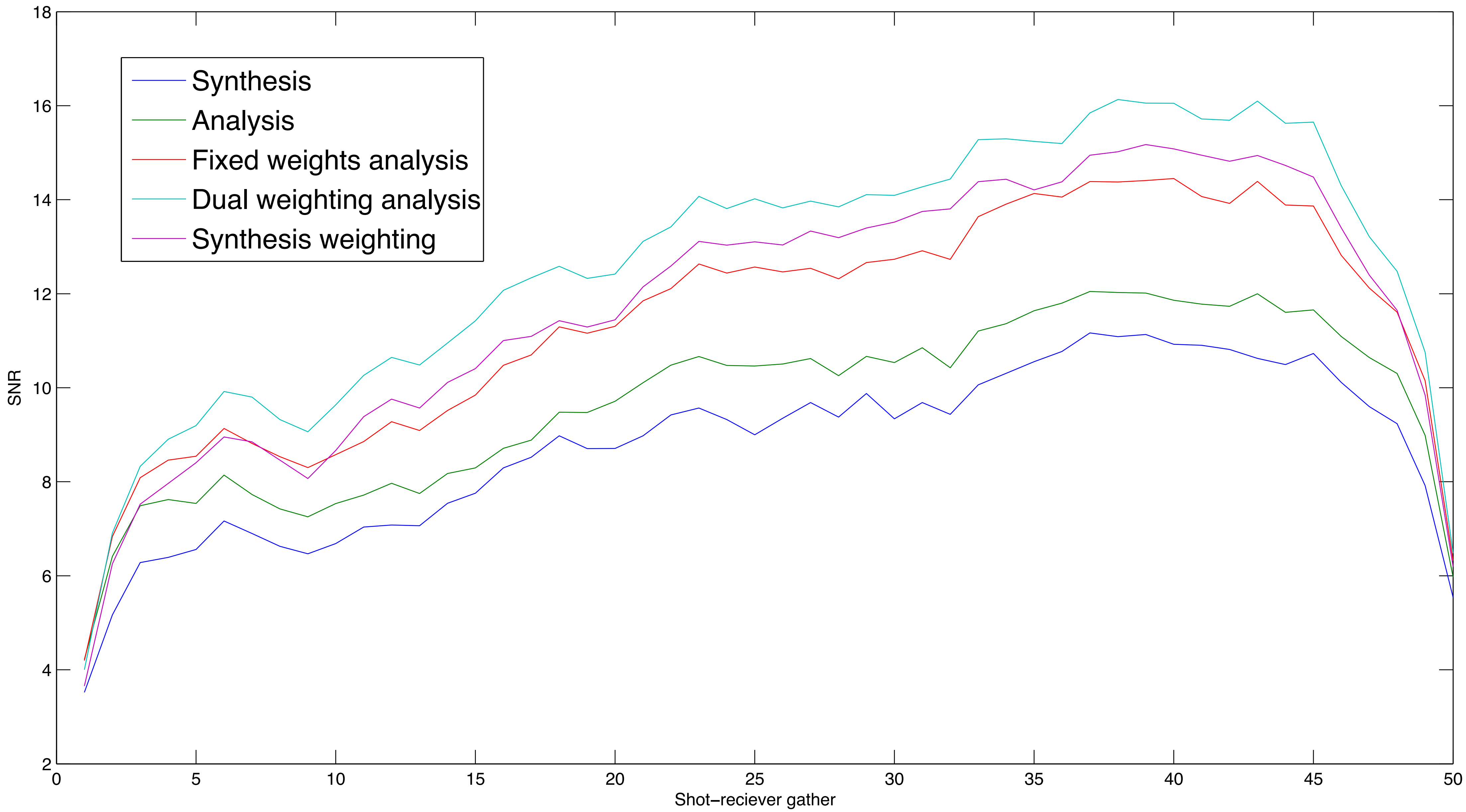
$$\widetilde{W}_{ii} = \begin{cases} 1 & \text{if } i \in \widetilde{T} \\ \omega & \text{if } i \notin \widetilde{T} \end{cases},$$

Various open theoretical and practical questions:

- Performance guarantees...
- How to estimate $\widetilde{T}$, value of ω?
- Other (potentially optimal) right inverse of $D\widetilde{W}$?
- Iterative reweighted versions?
- See the talk by Hargreaves for some answers and application to the above seismic interpolation problem.

Proposed Algorithm III – weighted ℓ_1 analysis

A snapshot from experimental results:



Proposed Algorithm IV: sparse randomized Kaczmarz

Kaczmarz Method (1937): Popular algorithm for solving **overdetermined** linear systems:

$$Ax = b + \text{noise}, \quad A : m \times n, \text{ with } \textit{commonly } m > n$$

Row-action method... Fast, simple, requires low memory...

- **The classical Kaczmarz algorithm:** sweep through the rows of A in an ordered manner: below, $i \equiv j \pmod{m}$.

Proposed Algorithm IV: sparse randomized Kaczmarz

Kaczmarz Method (1937): Popular algorithm for solving **overdetermined** linear systems:

$$Ax = b + \text{noise}, \quad A : m \times n, \text{ with } \textit{commonly } m > n$$

Row-action method... Fast, simple, requires low memory...

- **The classical Kaczmarz algorithm:** sweep through the rows of A in an ordered manner: below, $i \equiv j \pmod{m}$.

$$x_j = P_{a_i}(x) + P_{a_i^\perp}(x_{j-1}) \tag{5}$$

$$= \frac{\langle a_i, x\rangle}{\langle a_i, a_i\rangle} a_i^T + \left(x_{j-1} - \frac{\langle a_i, x_{j-1}\rangle}{\langle a_i, a_i\rangle} a_i^T \right), \tag{6}$$

Proposed Algorithm IV: sparse randomized Kaczmarz

Kaczmarz Method (1937): Popular algorithm for solving **overdetermined** linear systems:

$$Ax = b + \text{noise}, \quad A : m \times n, \text{ with } \textit{commonly } m > n$$

Row-action method... Fast, simple, requires low memory...

- **The classical Kaczmarz algorithm:** sweep through the rows of A in an ordered manner: below, $i \equiv j \pmod{m}$.

$$x_j = P_{a_i}(x) + P_{a_i^\perp}(x_{j-1}) \tag{5}$$

$$= \frac{b(i)}{\langle a_i, a_i \rangle} a_i^T + \left(x_{j-1} - \frac{\langle a_i, x_{j-1} \rangle}{\langle a_i, a_i \rangle} a_i^T \right), \tag{6}$$

Proposed Algorithm IV: sparse randomized Kaczmarz

Kaczmarz Method (1937): Popular algorithm for solving **overdetermined** linear systems:

$$Ax = b + \text{noise}, \quad A : m \times n, \text{ with } \textit{commonly } m > n$$

Row-action method... Fast, simple, requires low memory...

- **The randomized Kazcmarz (RK) algorithm (Strohmer-Vershynin, 2010)**: at each iteration, choose a_i randomly with probability $\frac{\|a_i\|_2^2}{\|A\|_F^2}$.

$$x_j = P_{a_i}(x) + P_{a_i^\perp}(x_{j-1}) \quad (5)$$

$$= \frac{b(i)}{\langle a_i, a_i \rangle} a_i^T + \left(x_{j-1} - \frac{\langle a_i, x_{j-1} \rangle}{\langle a_i, a_i \rangle} a_i^T \right), \quad (6)$$

Sparse randomized Kaczmarz (SRK)

The SRK algorithm (Mansour-Yilmaz):

- Assume the solution we seek is (approximately) sparse.
- At each iteration j, choose a_i randomly as above.
- Suppose we have a support estimate S. Set weights

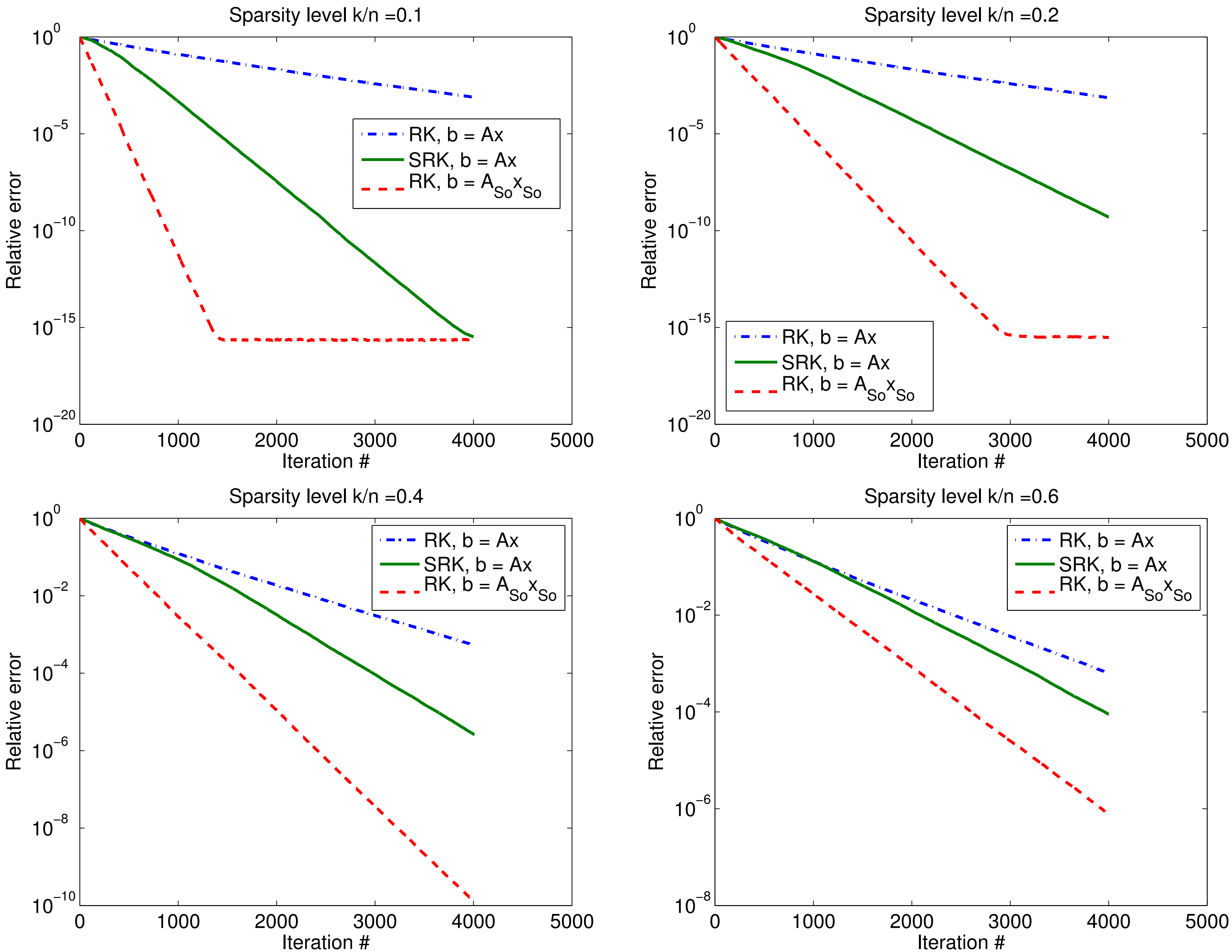
$$w_j(p) = \begin{cases} 1 & \text{if } p \in S, \\ \frac{1}{\sqrt{j}} & \text{if } p \notin S \end{cases}$$

- Update

$$\begin{aligned} x_j &= \frac{\langle a_i, x \rangle}{\langle \mathrm{w}_j \odot a_i, \mathrm{w}_j \odot a_i \rangle} (\mathrm{w}_j \odot a_i)^T + P_{(\mathrm{w}_j \odot a_i)^\perp}(x_{j-1}) \\ &= x_{j-1} + \frac{b(i) - \langle \mathrm{w}_j \odot a_i, x_{j-1} \rangle}{\|\mathrm{w}_j \odot a_i\|_2^2} (\mathrm{w}_j \odot a_i)^T. \end{aligned} \tag{7}$$

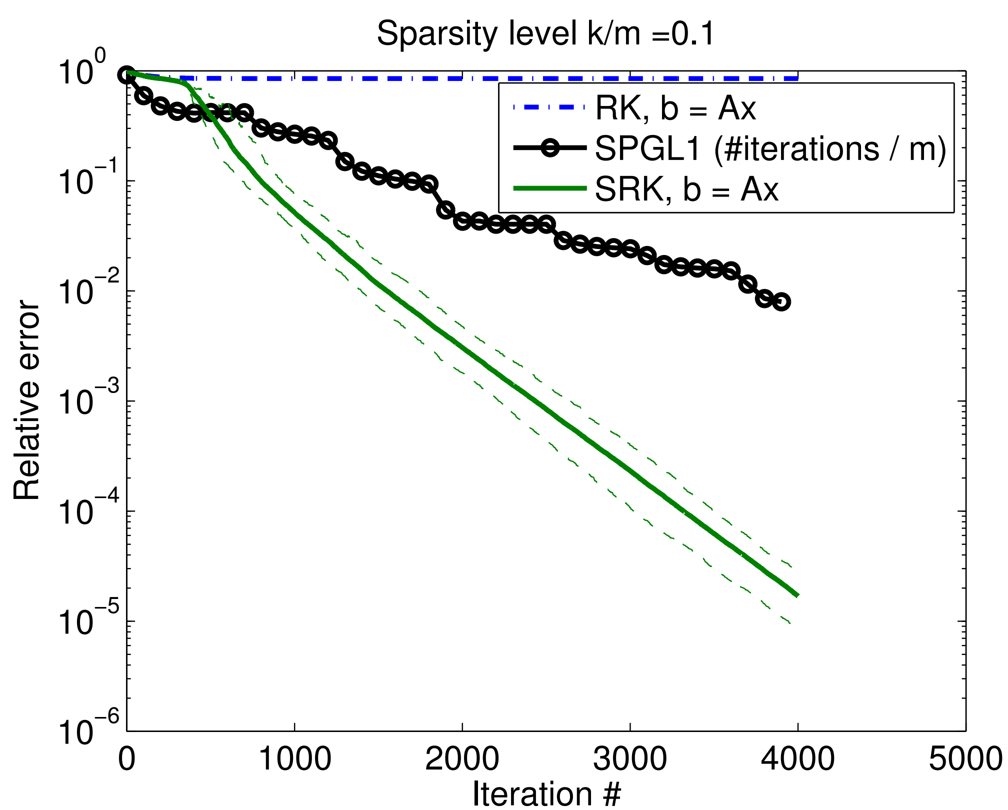
The SRK algorithm – overdetermined case

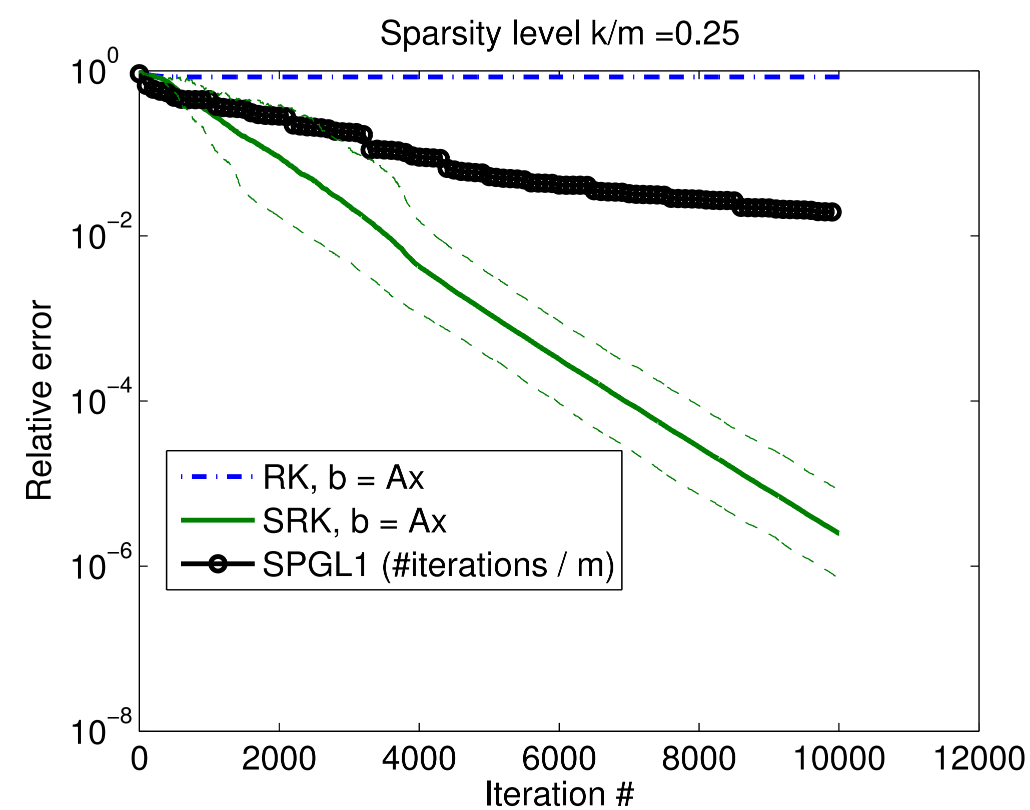
Average performance over 20 runs of SRK with A: 1000×200 Gaussian matrix.



The SRK algorithm – underdetermined case

Average performance over 20 runs of SRK with A: 100×400 Gaussian matrix. In other words, **the sparse recovery problem**!





The SRK algorithm

Some remarks:

- Empirical results are very encouraging for both overdetermined and underdetermined cases. A mathematical analysis is underway.
- Robust to noise. Also, works fine with compressible signals.
- Potential applications in full waveform inversion – see Mansour's talk.

Concluding remarks

- Compressive sampling theory: number of samples scales only logarithmically with the grid size!
- Theory helps us design effective (optimal) acquisition geometries.
- Transforming consequences for seismic (as well as other) signal acquisition and processing.
- Important problem: Incorporate prior information into the recovery algorithms.
- Proposed four ways to do this: each have pros and cons, but they all improve the recovery obtained by ℓ_1 minimization.

References

A fast randomized Kaczmarz algorithm for sparse solutions of consistent linear systems. H. Mansour and Ö. Yılmaz. Submitted.

Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization. H. Mansour, F.J. Herrmann, and Ö. Yılmaz. Submitted.

Non-convex compressed sensing using partial support information. N. Ghadermarzy, H. Mansour, and Ö. Yılmaz. In preparation.

Fighting the curse of dimensionality: compressive sensing in exploration seismology. F.J. Herrmann, M.P. Friedlander, and Ö. Yılmaz. IEEE Signal Processing Magazine, 29(3):88100, 2012.

Recovering Compressively Sampled Signals Using Partial Support Information. M.P. Friedlander, H. Mansour, R. Saab, and Ö. Yılmaz. IEEE Transactions on Information Theory, 58(2):1122-1134, 2012.

Support driven reweighted 1-norm minimization. H. Mansour and Ö. Yılmaz. Proc. ICASSP, 2012.

Weighted 1-norm minimization with multiple weighting sets. H. Mansour and Ö. Yılmaz. Proc. SPIE Wavelets and Sparsity XIV, 2011.

Acknowledgements

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R82411), the NSERC Discovery Accelerator Supplement Award (AID 411944-2011), and the NSERC 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.