What we know and what we do not know about practical compressive sampling

Deanna Needell

Claremont McKenna College

Jan. 13, 2014

FGMIA 2014, Paris, France

Outline

- Introduction
  - Mathematical Formulation & Methods
- Practical CS
  - Other notions of sparsity
  - Heavy quantization
  - Adaptive sampling

The mathematical problem

1. Signal of interest $f \in \mathbb{C}^d (= \mathbb{C}^{N \times N})$

2. Measurement operator $\mathscr{A} : \mathbb{C}^d \to \mathbb{C}^m$ $(m \ll d)$

3. Measurements $y = \mathscr{A} f + \xi$

$$\begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & \mathscr{A} & \\ & & \end{bmatrix} \begin{bmatrix} \\ f \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \xi \\ \\ \end{bmatrix}$$

4. *Problem:* Reconstruct signal f from measurements y

Sparsity

Measurements $y = \mathscr{A} f + \xi$.

$$\begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & \mathscr{A} & \\ & & \end{bmatrix} \begin{bmatrix} \\ f \\ \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \xi \\ \\ \end{bmatrix}$$

Assume f is *sparse*:

✧ In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \le s \ll d$

✧ In orthonormal basis: $f = Bx$ where $\|x\|_0 \le s \ll d$

In practice, we encounter *compressible* signals.

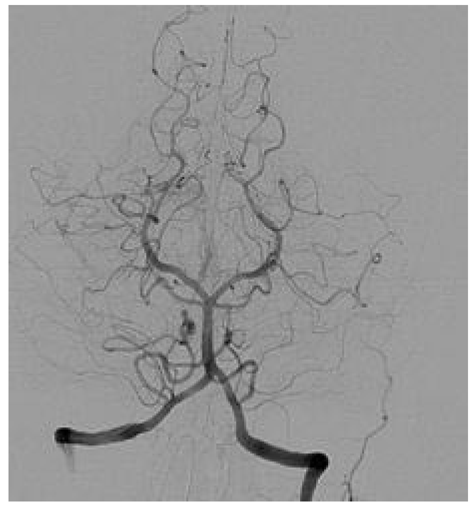
✦ f_s is the best s-sparse approximation to f

Many applications...

- Radar, Error Correction
- Computational Biology, Geophysical Data Analysis
- Data Mining, classification
- Neuroscience
- Imaging
- Sparse channel estimation, sparse initial state estimation
- Topology identification of interconnected systems
- ...

Sparsity...

Sparsity in coordinate basis: f=x



Reconstructing the signal f from measurements y

- ℓ_1-minimization [Candès-Romberg-Tao]

Let A satisfy the *Restricted Isometry Property* and set:

$$\hat{f} = \underset{g}{\operatorname{argmin}} \|g\|_1 \quad \text{such that} \quad \|\mathscr{A} f - y\|_2 \leq \varepsilon,$$

where $\|\xi\|_2 \leq \varepsilon$. Then we can stably recover the signal f:

$$\|f - \hat{f}\|_2 \lesssim \varepsilon + \frac{\|x - x_s\|_1}{\sqrt{s}}.$$

This error bound is optimal.

Restricted Isometry Property

- $\mathscr{A}$ satisfies the Restricted Isometry Property (RIP) when there is $\delta < c$ such that

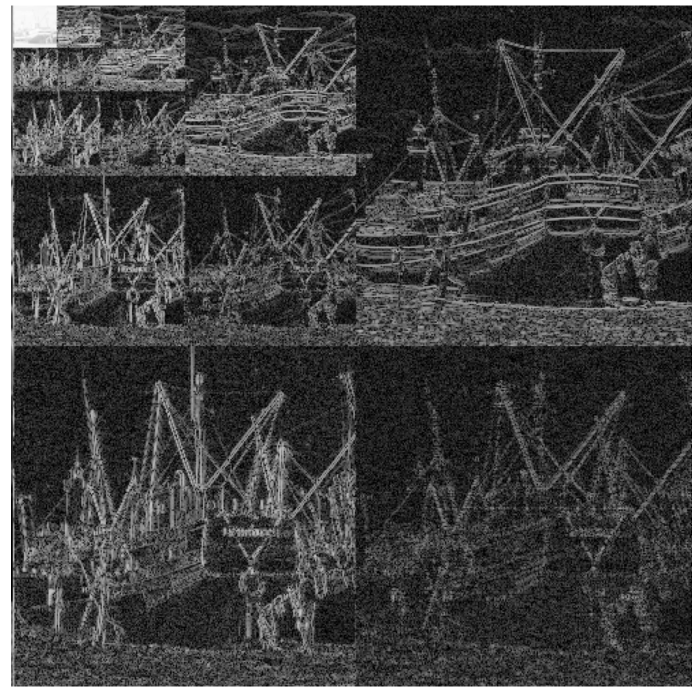
$$(1-\delta)\|f\|_2 \le \|\mathscr{A} f\|_2 \le (1+\delta)\|f\|_2 \quad \text{whenever } \|f\|_0 \le s.$$

- $m \times d$ Gaussian or Bernoulli measurement matrices satisfy the RIP with high probability when
$$m \gtrsim s \log d.$$

- Random Fourier and others with fast multiply have similar property: $m \gtrsim s \log^4 d$.

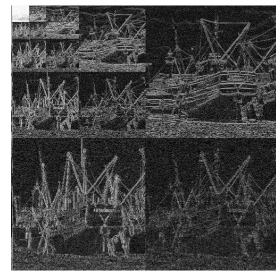
Sparsity...

In orthonormal basis: $f = Bx$



Natural Images

Images are compressible in *Wavelet bases*.



$$f = \sum_{j,k=1}^{N} x_{j,k} H_{j,k}, \qquad x_{j,k} = \langle f, H_{j,k} \rangle, \qquad \|f\|_2 = \|x\|_2,$$

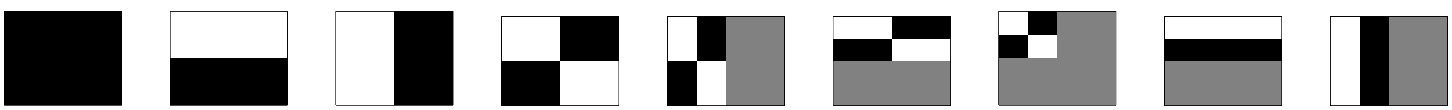


Figure 1: Haar basis functions

Wavelet transform is *orthonormal* and multi-scale. Sparsity level of image is higher on detail coefficients.

Sparsity in orthonormal basis B

✦ L1-minimization Method

For orthonormal basis B, $f = Bx$ with x sparse, one may solve the ℓ_1-minimization program:

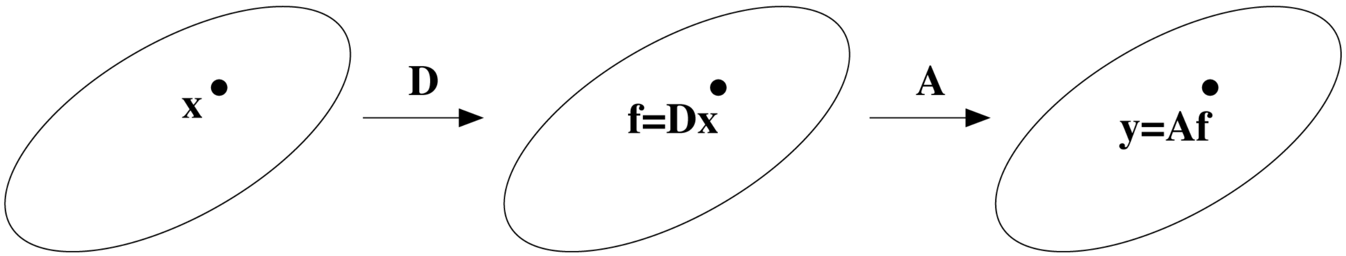
$$\hat{f} = \underset{\tilde{f}\in\mathbb{C}^n}{\operatorname{argmin}} \|B^{-1}\tilde{f}\|_1 \quad \text{subject to} \quad \|\mathscr{A}\tilde{f} - y\|_2 \le \varepsilon.$$

Same results hold.

Sparsity...

In arbitrary dictionary: $f = Dx$

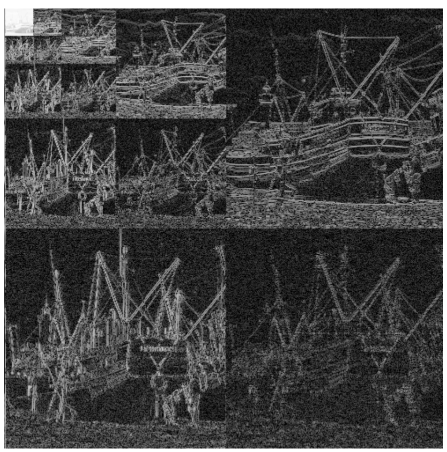
The CS Process

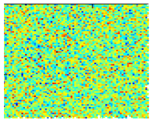


Coefficient Domain

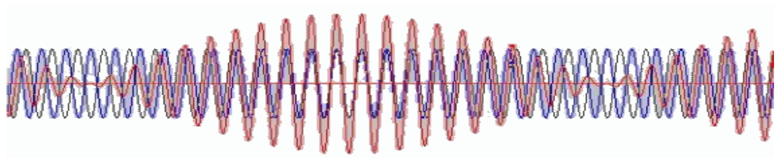
Signal Domain

Compressed Sensing Domain





Example: Oversampled DFT



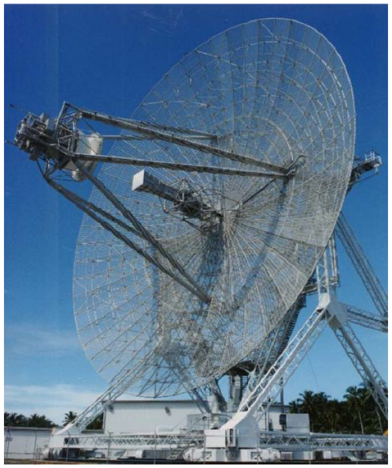
- $n \times n$ DFT: $d_k(t) = \frac{1}{\sqrt{n}} e^{-2\pi i k t/n}$

- Sparse in the DFT → superpositions of sinusoids with frequencies in the lattice.

- Instead, use the *oversampled DFT*:

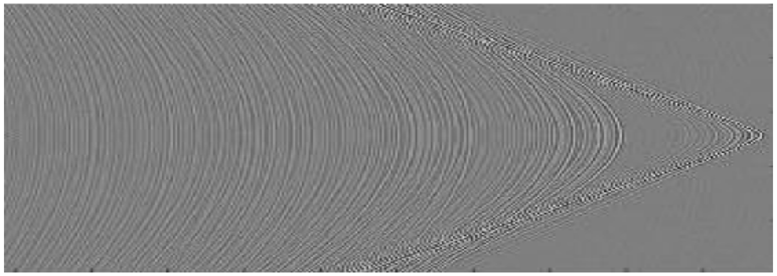
- Then D is an overcomplete frame with highly coherent columns → *conventional CS does not apply*.

Example: Gabor frames



- Gabor frame: $G_k(t) = g(t - k_2 a) e^{2\pi i k_1 b t}$
- Radar, sonar, and imaging system applications use Gabor frames and wish to recover signals in this basis.
- Then D is an overcomplete frame with possibly highly coherent columns $\rightarrow$ *conventional CS does not apply*.

Example: Curvelet frames

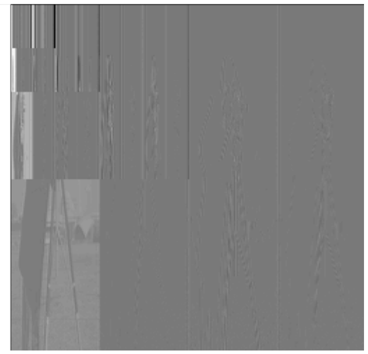


- A Curvelet frame has some properties of an ONB but is overcomplete.

- Curvelets approximate well the curved singularities in images and are thus used widely in image processing.

- Again, this means D is an overcomplete dictionary → *conventional CS does not apply*.

Example: UWT



- The undecimated wavelet transform has a translation invariance property that is missing in the DWT.
- The UWT is overcomplete and this redundancy has been found to be helpful in image processing.
- Again, this means D is a redundant dictionary $\rightarrow$ *conventional CS does not apply*.

ℓ_1-Synthesis Method

✦ For arbitrary tight frame D, one may solve the ℓ_1-synthesis program:

$$\hat{f} = D\left(\operatorname*{argmin}_{\tilde{x}\in\mathbb{C}^n} \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\mathscr{A} D\tilde{x} - y\|_2 \leq \varepsilon\right).$$

Some work on this method [Candès et.al., Rauhut et.al., Elad et.al.,...]

✦ *To do:* Understand the ℓ_1-synthesis problem, necessary assumptions, recovery guarantees.

ℓ_1-Analysis Method

- For arbitrary tight frame D, one may solve the ℓ_1-analysis program:

$$\hat{f} = \operatorname*{argmin}_{\tilde{f} \in \mathbb{C}^n} \| D^* \tilde{f} \|_1 \quad \text{subject to} \quad \| \mathscr{A} \tilde{f} - y \|_2 \leq \varepsilon.$$

Condition on A?

✦ D-RIP

We say that the measurement matrix $\mathscr{A}$ obeys the *restricted isometry property adapted to* D (D-RIP) if there is $\delta < c$ such that

$$(1-\delta)\|Dx\|_2^2 \leq \|\mathscr{A}Dx\|_2^2 \leq (1+\delta)\|Dx\|_2^2$$

holds for all s-sparse x.

✦ Similarly to the RIP, many classes of $m \times d$ random matrices satisfy the D-RIP with $m \approx s\log(d/s)$.

CS with tight frame dictionaries

✦ Theorem [Candès-Eldar-N-Randall]

Let D be an arbitrary tight frame and let $\mathscr{A}$ be a measurement matrix satisfying D-RIP. Then the solution $\hat{f}$ to ℓ_1-analysis satisfies

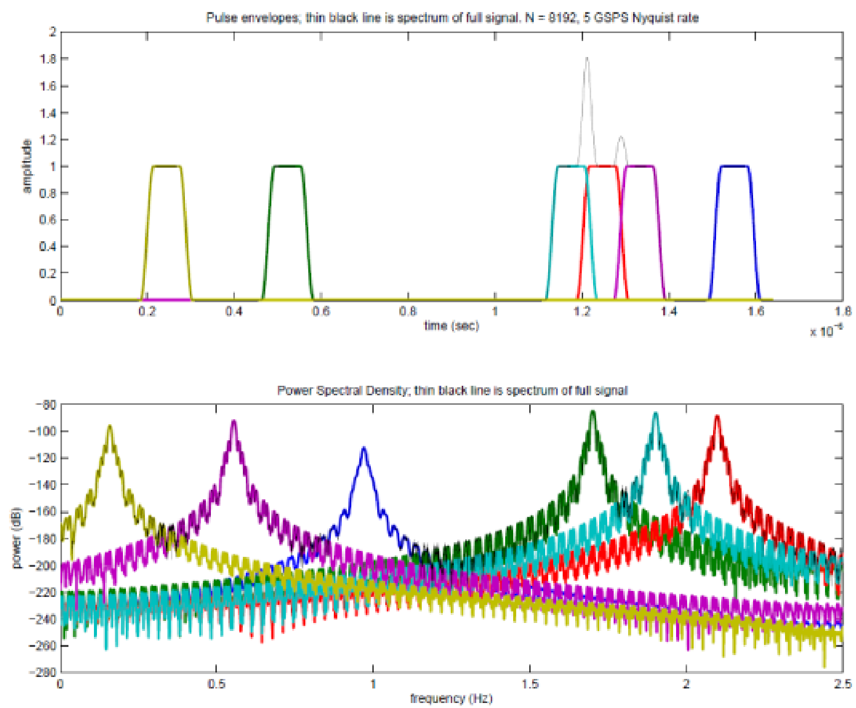
$$\|\hat{f} - f\|_2 \lesssim \varepsilon + \frac{\|D^* f - (D^* f)_s\|_1}{\sqrt{s}}.$$

✦ In other words, This result says that ℓ_1-analysis is very accurate when $D^* f$ has rapidly decaying coefficients and D is a tight frame.

ℓ_1-analysis: Experimental Setup

$n = 8192, m = 400, d = 491,520$

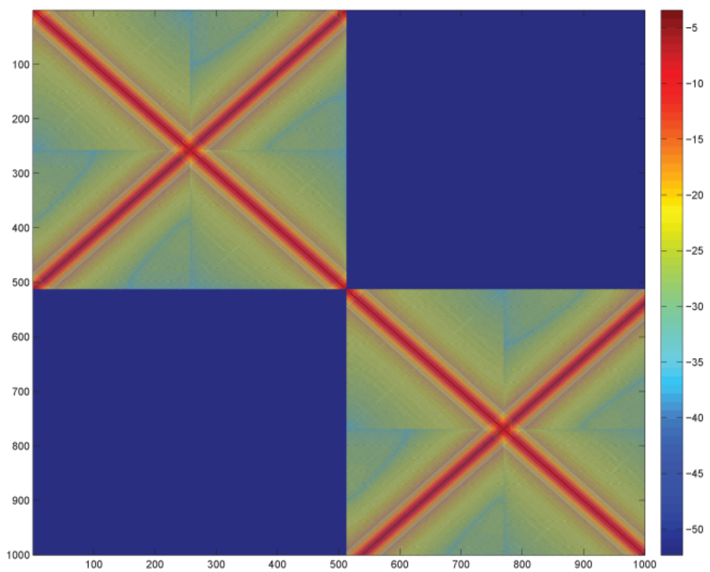
A: $m \times n$ Gaussian, D: $n \times d$ Gabor

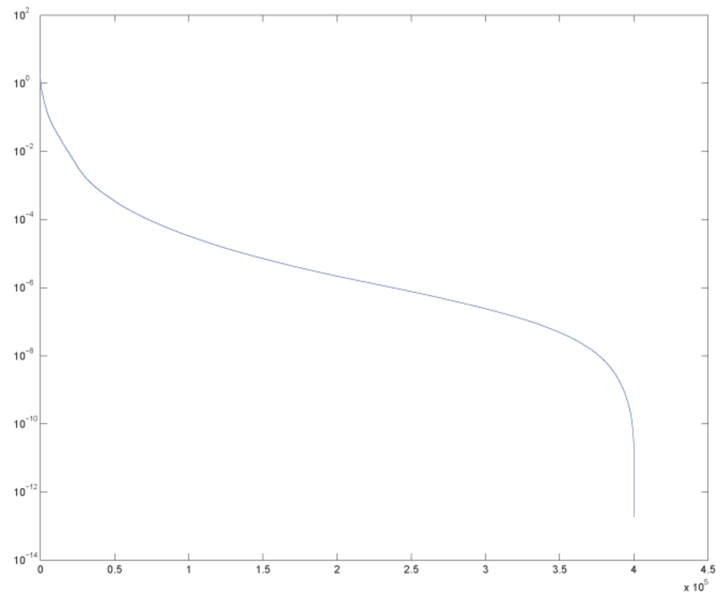


ℓ_1-analysis: Experimental Setup

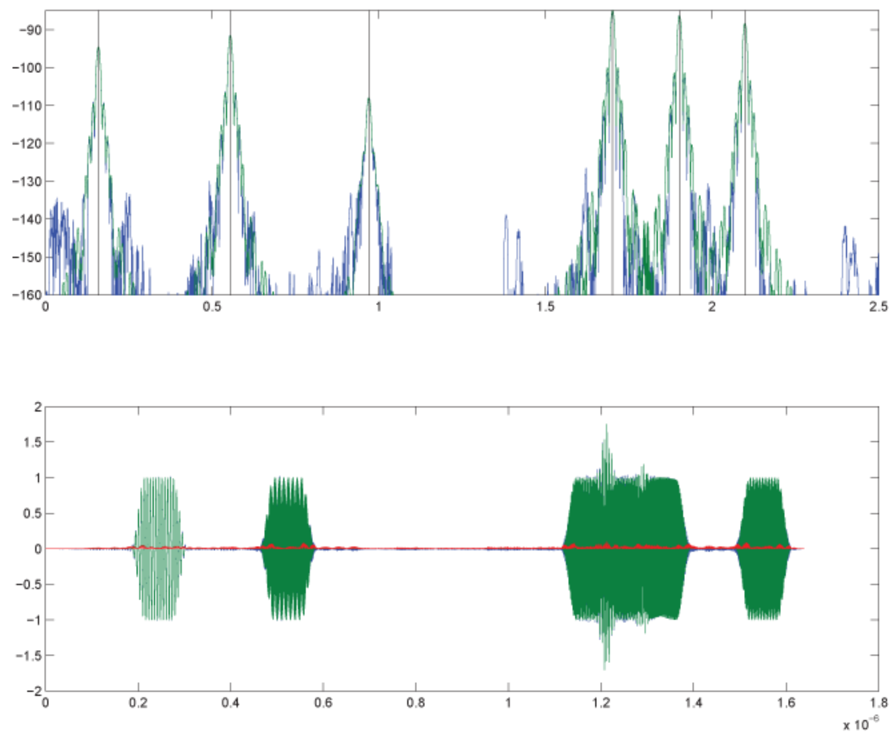
$n = 8192, m = 400, d = 491,520$

A: $m \times n$ Gaussian, D: $n \times d$ Gabor





ℓ_1-analysis: Experimental Results



Other algorithms

✦ ℓ_1-analysis is very accurate when $D^* f$ has rapidly decaying coefficients and D is a tight frame. This is precisely because this method operates in "analysis" space.

✦ *To do:* analysis methods for non-tight frames, without decaying analysis coefficients (concatenations), other models

✦ What about operating in signal or coefficient space?

Is it really a pipe?

(Thanks to M. Davenport for this clever analogy.)

CoSaMP

CoSaMP (N-Tropp)

input: Sampling operator A, measurements y, sparsity level s
initialize: Set $x^0 = 0$, $i = 0$.
repeat
signal proxy: Set $p = A^*(y - Ax^i)$, $\Omega = \mathrm{supp}(p_{2s})$, $T = \Omega \cup \mathrm{supp}(x^i)$.
signal estimation: Using least-squares, set $b|_T = A_T^\dagger y$ and $b|_{T^c} = 0$.
prune and update: Increment i and to obtain the next approximation, set $x^i = b_s$.
output: s-sparse reconstructed vector $\widehat{x} = x^i$

Signal Space CoSaMP

SIGNAL SPACE COSAMP (Davenport-N-Wakin)

input: $\boldsymbol{A}$, $\boldsymbol{D}$, $\boldsymbol{y}$, s, stopping criterion
initialize: $\boldsymbol{r} = \boldsymbol{y}$, $\boldsymbol{x}^0 = 0$, $\ell = 0$, $\Gamma = \emptyset$
repeat

proxy: $\boldsymbol{h} = \boldsymbol{A}^* \boldsymbol{r}$
identify: $\Omega = \mathscr{S}_{\boldsymbol{D}}(\boldsymbol{h}, 2s)$
merge: $T = \Omega \cup \Gamma$
update: $\widetilde{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{z}\|_2 \quad \text{s.t.} \quad \boldsymbol{z} \in \mathscr{R}(\boldsymbol{D}_T)$
$\Gamma = \mathscr{S}_{\boldsymbol{D}}(\widetilde{\boldsymbol{x}}, s)$
$\boldsymbol{x}^{\ell+1} = \mathscr{P}_{\Gamma} \widetilde{\boldsymbol{x}}$
$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{\ell+1}$
$\ell = \ell + 1$
output: $\widehat{\boldsymbol{x}} = \boldsymbol{x}^{\ell}$

Signal Space CoSaMP

- Here we must contend with

$$\Lambda_{\text{opt}}(\boldsymbol{z}, s) := \underset{\Lambda:|\Lambda|=s}{\operatorname{argmin}} \|\boldsymbol{z} - \mathscr{P}_\Lambda \boldsymbol{z}\|_2, \quad \mathscr{P}_\Lambda : \mathbb{C}^n \to \mathscr{R}(\boldsymbol{D}_\Lambda).$$

- Estimate by $\mathscr{S}_{\boldsymbol{D}}(\boldsymbol{z}, s)$ with $|\mathscr{S}_{\boldsymbol{D}}(\boldsymbol{z}, s)| = s$, that satisfies

$$\left\|\mathscr{P}_{\Lambda_{\text{opt}}(\boldsymbol{z},s)}\boldsymbol{z} - \mathscr{P}_{\mathscr{S}_{\boldsymbol{D}}(\boldsymbol{z},s)}\boldsymbol{z}\right\|_2 \leq \min\left(\epsilon_1 \left\|\mathscr{P}_{\Lambda_{\text{opt}}(\boldsymbol{z},s)}\boldsymbol{z}\right\|_2, \epsilon_2 \left\|\boldsymbol{z} - \mathscr{P}_{\Lambda_{\text{opt}}(\boldsymbol{z},s)}\boldsymbol{z}\right\|_2\right)$$

for some constants $\epsilon_1, \epsilon_2 \geq 0$.

Approximate Projection

- Practical choices for $\mathscr{S}_{\boldsymbol{D}}(\boldsymbol{z}, s)$:
  - Any sparse recovery algorithm!
  - OMP
  - CoSaMP
  - ℓ_1-minimization followed by hard thresholding

Signal Space CoSaMP

✦ Theorem [Davenport-N-Wakin] Let D be an arbitrary tight frame, A be a measurement matrix satisfying D-RIP, and f a sparse signal with respect to D. Then the solution $\hat{f}$ from *Signal Space CoSaMP* satisfies

$$\|\hat{f} - f\|_2 \lesssim \varepsilon.$$

(And similar results for approximate sparsity, depending on the approximation method used for $\Lambda_{\text{opt}}(z, s)$.)

✦ *To do:* Design approximation methods that satisfy necessary recovery bounds (sparse approximation).

Signal Space CoSaMP: Experimental Results

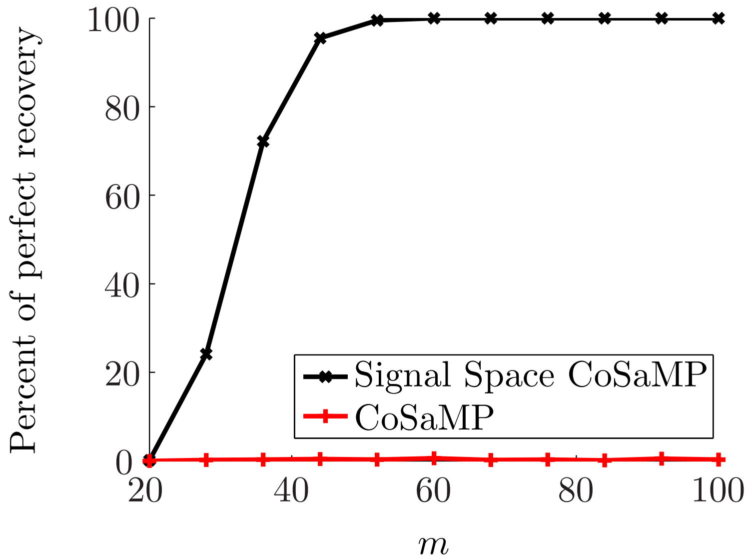


Figure 2: *Performance in recovering signals having a $s = 8$ sparse representation in a dictionary $\boldsymbol{D}$ with orthogonal, but not normalized, columns.*

Signal Space CoSaMP: Experimental Results

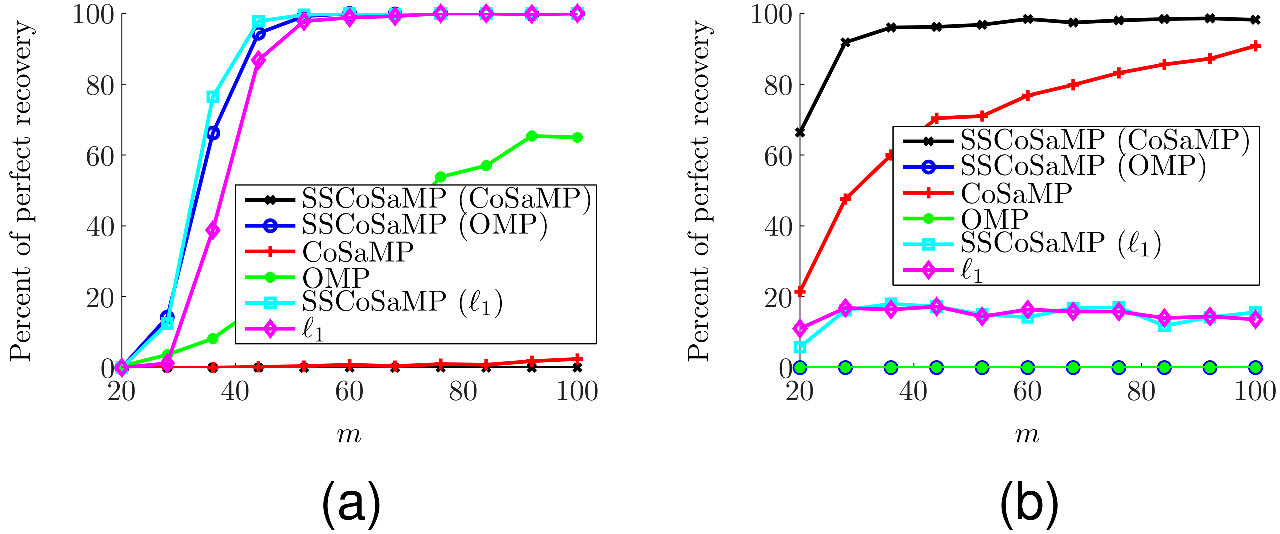


Figure 3: *Results with $s = 8$ sparse representation in a 4× overcomplete DFT dictionary: (a) well-separated coefficients, (b) clustered coefficients.*

Signal Space CoSaMP: Recent improvements

✦ Recently improved results [Giryes-N and Hegde-Indyk-Schmidt] which relax the assumptions on the approximate projections.

✦ These results show that at least for RIP/incoherent dictionaries, standard algorithms like CoSaMP/OMP/IHT suffice for the approximate projections.

To do:

✦ The interesting/challenging case is when the dictionary does *not* satisfy such a condition. Are there methods which provide these approximate projections? Or are they not even necessary?

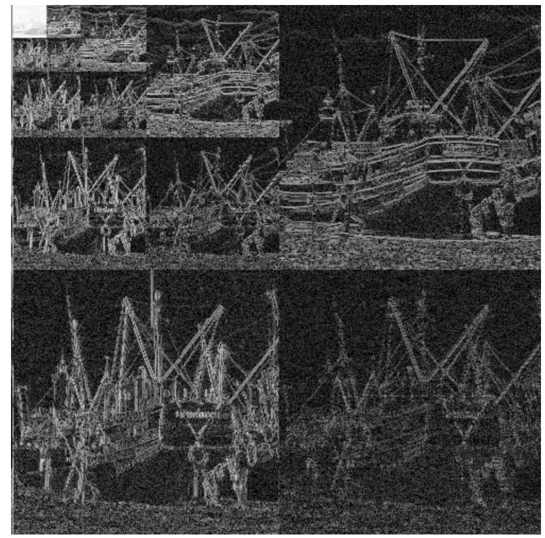
Natural images

Sparse...

256×256 “Boats" image

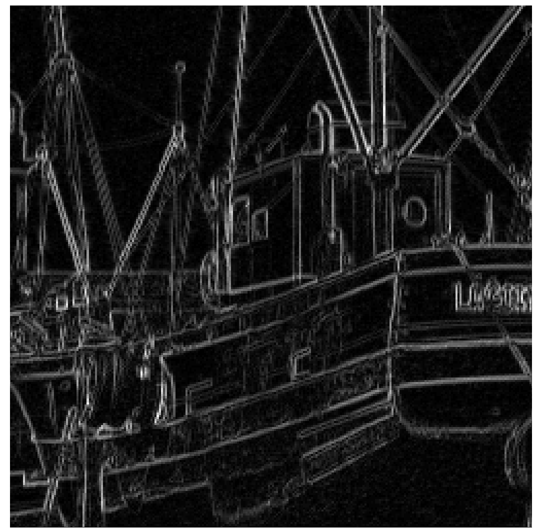
Natural images

Sparse wavelet representation...



Natural images

Images are compressible in *discrete gradient*.



Natural images

Images are compressible in *discrete gradient*.

The discrete directional derivatives of an image $f \in \mathbb{C}^{N \times N}$ are

$$f_x : \mathbb{C}^{N \times N} \to \mathbb{C}^{(N-1) \times N}, \qquad (f_x)_{j,k} = f_{j,k} - f_{j-1,k},$$

$$f_y : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times (N-1)}, \qquad (f_y)_{j,k} = f_{j,k} - f_{j,k-1},$$

the discrete gradient operator is

$$\nabla[f] = (f_x, f_y)$$

Sparsity in gradient

- CS Theory

The gradient operator ∇ is not an orthonormal basis or a tight frame. In fact, it is extremely ill-conditioned!

Comparison of two compressed sensing reconstruction algorithms

- Haar-minimization (L_1-Haar)

$$\hat{f}_{Haar} = \text{argmin} \|H(Z)\|_1 \quad \text{subject to} \quad \|\mathscr{A} Z - y\|_2 \leq \varepsilon$$

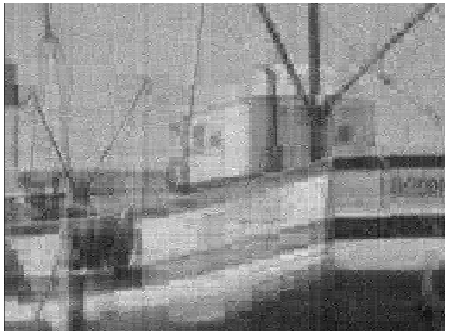
- Total Variation minimization (TV)

$\hat{f}_{TV} = \text{argmin} \|\nabla[Z]\|_1 \quad \text{subject to} \quad \|\mathscr{A} Z - y\|_2 \leq \varepsilon$, where $\|Z\|_{TV} = \|\nabla[Z]\|_1$ is the *total-variation norm*.

Imaging via compressed sensing

(a) Original

(b) TV



(c) L_1-Haar

Figure 4: Reconstruction using $m = .2N^2$

Imaging via compressed sensing

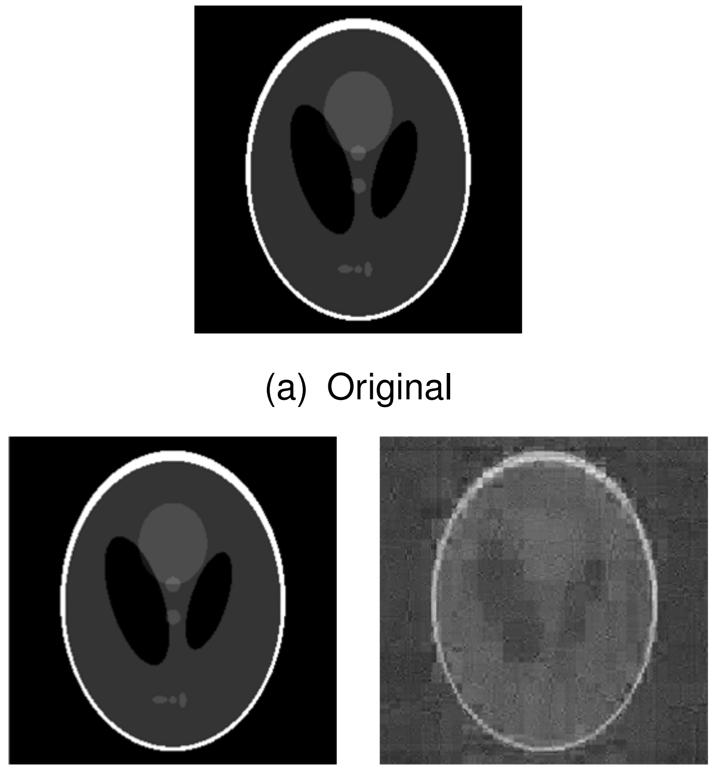
(a) Original

(b) TV

(c) L_1-Haar

Figure 5: Reconstruction using $m = .2N^2$ measurements

Imaging via compressed sensing



(a) Original

(b) TV

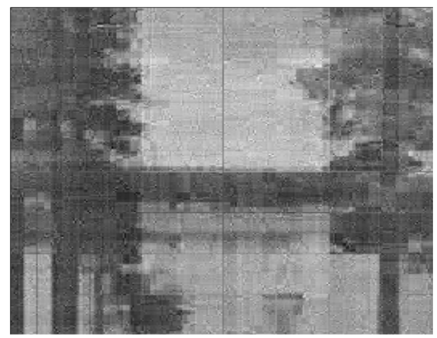
(c) L_1-Haar

Figure 6: Reconstruction using $m = .2N^2$ measurements.

Imaging via compressed sensing

(a) (Quantization)

(b) TV



(c) L_1-Haar

Figure 7: Reconstruction using $m = .2N^2$ measurements

Imaging via compressed sensing

InView (Austin TX)

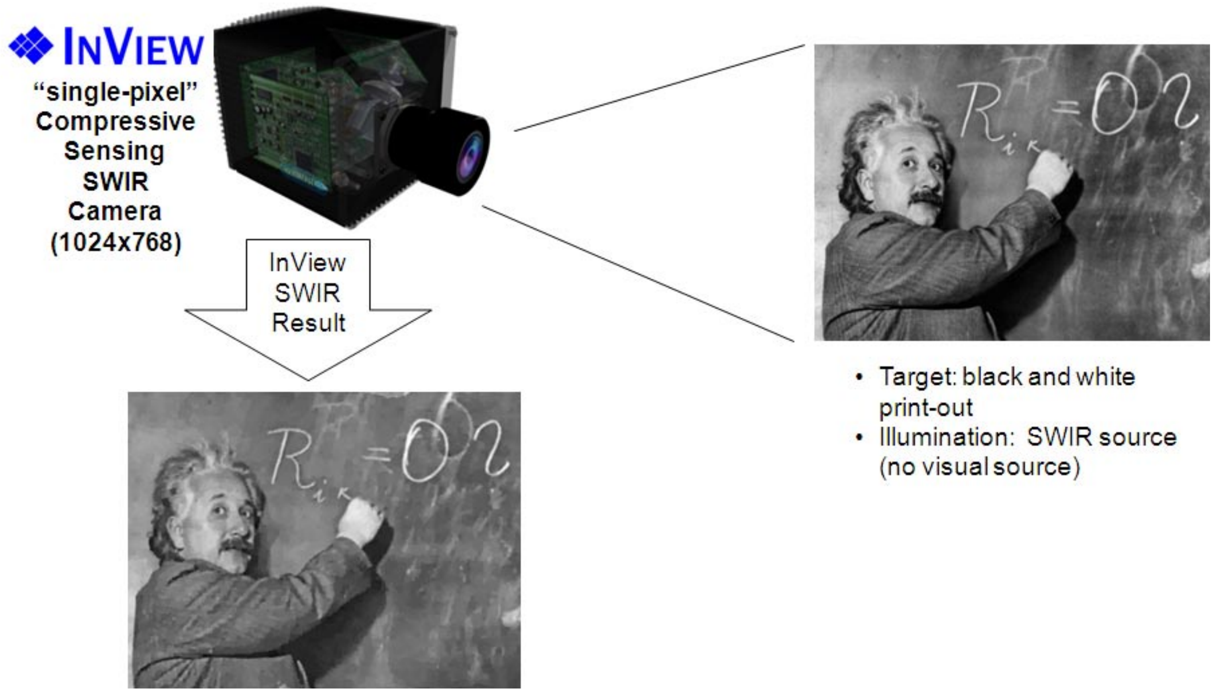


Figure 8: SWIR Reconstruction using $m = .5N^2$ measurements

Imaging via compressed sensing

InView (Austin TX)

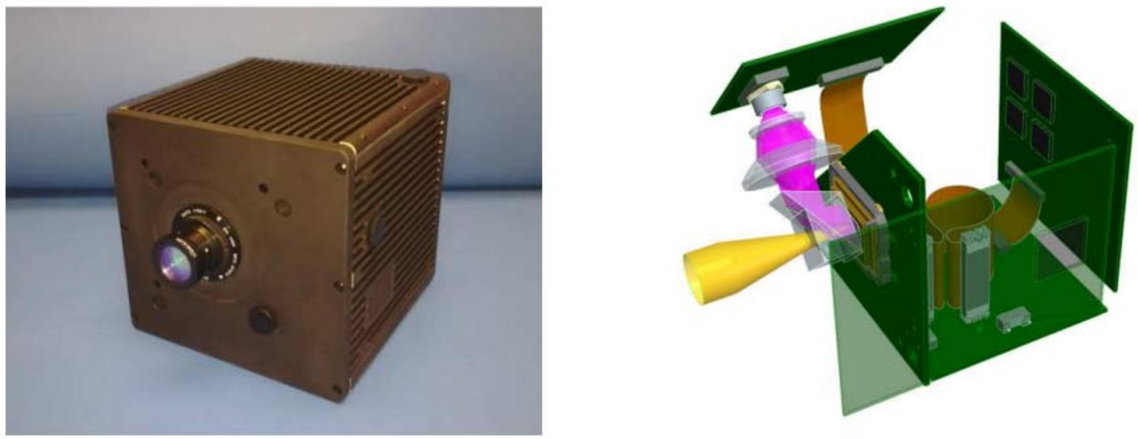


Figure 9: InView SWIR camera

Empirical → Theoretical?

✦ TV Works

Empirically, it has been well known that

$$\hat{f}_{TV} = \operatorname{argmin} \|Z\|_{TV} \quad \text{subject to} \quad \|\mathscr{A}Z - y\|_2 \leq \varepsilon, \qquad (TV)$$

provides quality, stable image recovery.

✦ No provable stability guarantees.

Stable signal recovery using total-variation minimization

Theorem 1. [N-Ward] *From* $m \gtrsim s\log(N)$ *linear RIP measurements, for any* $f \in \mathbb{C}^{N\times N}$*,*

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathscr{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N)$ *factor*

Higher dimensional objects

Movies, higher dimensional objects?

Theorem 2. [N-Ward] *From $m \gtrsim s\log(N^d)$ linear RIP measurements, for any $f \in \mathbb{C}^{N^d}$,*

$$\hat{f} = \operatorname{argmin} \|Z\|_{TV} \quad \textit{such that} \quad \|\mathscr{A}(Z) - y\|_2 \leq \varepsilon,$$

satisfies

$$\|f - \hat{f}\|_{TV} \lesssim \|\nabla[f] - \nabla[f]_s\|_1 + \sqrt{s}\varepsilon \qquad \text{(gradient error)}$$

and

$$\|f - \hat{f}\|_2 \lesssim \log(N^d/s) \cdot \left[\frac{\|\nabla[f] - \nabla[f]_s\|_1}{\sqrt{s}} + \varepsilon\right] \qquad \text{(signal error)}$$

This error guarantee is optimal up to the $\log(N^d/s)$ factor

Stable signal recovery using total-variation minimization

Method of proof:

✧ First prove stable *gradient* recovery

✧ Translate stable *gradient* recovery to stable *signal* recovery using the strengthened Sobolev inequality.

To do:

✦ Remove logarithmic factors, design more efficient measurement schemes.

✦ Incorporate wavelets, Laplacian, etc. for optimal performance.

✦ Prove for 1-d signals!

1-bit compressive sensing

- Measurements: $y = \text{sign}(Af)$ (extreme quantization)
- Noise: Random or adversarial bit flips
- Assumption: signal f lies in some (convex) set K
- $\hat{f} = \max_x \langle y, Ax \rangle \quad \text{s.t.} \quad x \in K$
- (Plan-Vershynin): $\|\hat{f} - f\|_2 \lesssim w(K)/\sqrt{m}$
- Greedy methods for accurate recovery from optimal number of (e.g. Gaussian) measurements [Baraniuk et al.]

1-bit compressive sensing

- In general, results are of the form:

$$\|\hat{f} - f\|_2 \lesssim \lambda^{-c},$$

where $\lambda = \frac{m}{s\log(n/s)}$ is the oversampling factor.

- New results [Baraniuk-Foucart-N-Plan-Wootters]: Provide a reconstruction method to obtain

$$\|\hat{f} - f\|_2 \lesssim exp(-\lambda),$$

(in preparation).

1-bit compressive sensing

- To do:

- Optimal greedy methods for recovery (what is optimal?)

- Methods for recovery when sparsity is w.r.t. aribtrary dictionary D

- Mixed models of quantization – unified framework for all precision

Adaptive measurement schemes

- Design measurement operator on the fly
- Fundamental limitations on improved recovery [Candès-Davenport]
- However, improvements still possible (such as reduced number of measurements needed) [Aldroubi et al., Iwen-Tewfik, Indyk et al.]
- Adaptive measurement schemes for fixed sampling structures, total variation, sparsity in dictionaries, average case results, ...

Adaptive measurement schemes

- Sampling from constrained measurements
  - Certain constrained settings don't afford improvements via adaptivity (Davenport-N)
  - Identify geometric properties of constraints that offer adaptive improvements
  - Design adaptive measurement schemes and recovery algorithms for those that do

Thank you!

E-mail:

✧ dneedell@cmc.edu

Web:

✧ www.cmc.edu/pages/faculty/DNeedell

References:

- E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 59(8):1207Ű1223, 2006.
- E. J. Candès, Y. C. Eldar, D. Needell and P. Randall. Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic Analysis, 31(1):59-73, 2010.
- M. A. Davenport, D. Needell and M. B. Wakin. Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries, submitted.
- P. Indyk, E. Price and D. Woodruff. On the Power of Adaptivity in Sparse Recovery, FOCS 2011.
- D. Needell and R. Ward. Stable image reconstruction using total variation minimization. J. Fourier Analysis and Applications, to appear.
- D. Needell and R. Ward. Total variation minimization for stable multidimensional signal recovery, submitted.
- Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming, Comm. Pure Appl. Math., to appear.