Less is More: Compressed Sensing

SIAM Annual Meeting

Deanna Needell
Mathematics, UCLA

So much data…


The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

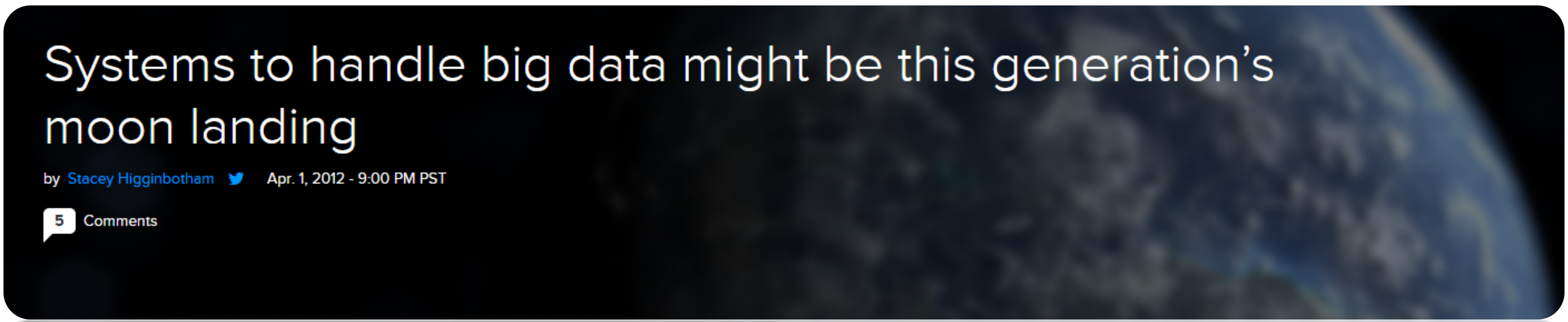
So much data…


COMPUTER
WHY
IS A
BIG DAT
BIG DEAL
nature
The Economist
The data delu
POPULAR SCIENCE
THE FUTURE NOW
THE CONTROL CENTERS
OFFICER ALGORITHM
NEW WAYS OF SEEING
PLUS
SPECIAL ISSUE
DATA IS POWER
HOW INFORMATION IS DRIVING THE FUTURE

So much data…

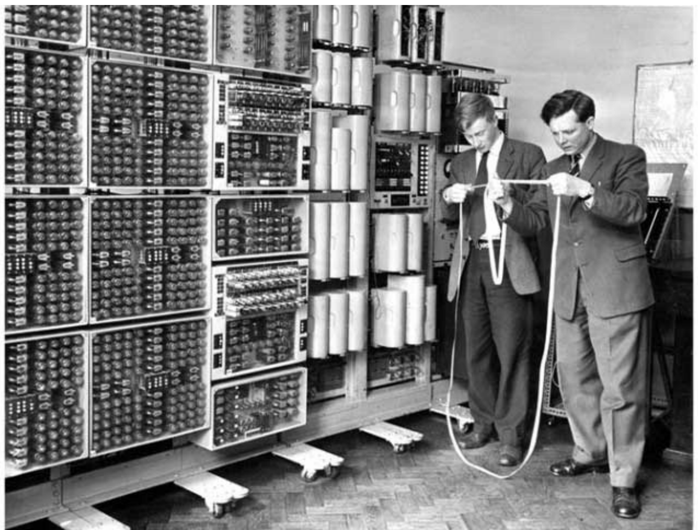

METU siam. Student Chapter Presents
16-24 May 2012
WHAT WOULD YOU DO WITH ALL THIS DATA?
Gerhard Wilhelm Weber
Estimation of Dynamics under Uncertainty
May 16, Wednesday, 11:40-12:30 *
Annette Hohenberger
Fractals in Cognitive Science
May 17, Thursday, 11:40-12:30 *
Özlem İlk
Introduction to R and GGobi
May 17, Thursday, 14:00-17:30 **
Aybar Acar
MapReduce and Hadoop: Mining Big Data in the Cloud
May 23, Wednesday, 14:00-16:30 *
Cem İyigün
Introduction to Clustering
May 23, Wednesday, 11:40-12:30 *
Fatma Yerlikaya-Özkurt
Modeling with MARS and CMARS
May 24, Thursday, 14:00-15:30 *
* Institute of Applied Math, S 209
** Department of Mathematics, Computer Lab (M 202)
For more info visit http://siam.metu.edu.tr
Mathematics, Statistics, and the Data Deluge
MATHEMATICS AWARENESS MONTH

So much data…


Systems to handle big data might be this generation's moon landing
by Stacey Higginbotham Apr. 1, 2012 - 9:00 PM PST
5 Comments

How can we handle all this data?

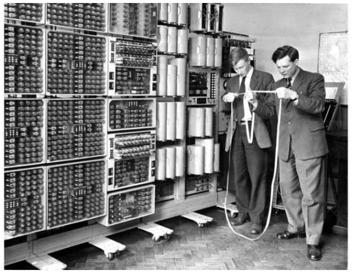
Option 1 : Build bigger computing systems



How can we handle all this data?

Option 1 : Build bigger computing systems

- We need the resources
- Fundamental limitations
- Wasteful (resources, energy, cost, …)



How can we handle all this data?

Option 1 : Build bigger computing systems

- We need the resources
- Fundamental limitations
- Wasteful (resources, energy, cost, …)

3 MB of internet data transfer = boiling one cup of water

(https://www.katescomment.com/energy-of-downloads/)

How can we handle all this data?

Option 2 : Design more efficient compression methods

Enter the world of : Compressed sensing

Compressed sensing: motivation

Applications are numerous :

- Data storage
- Reliable data transmission
- Collaborative filtering (e.g. Netflix predictions)
- Radar
- DNA array sequencing
- Neuroscience
- Predicting earthquakes
- Restoring damaged artwork
- Crime prediction
- Image compression
- Medical imaging
- Many, many, many more…

Representations of High Dimensional Data

Key Idea :

Modern data is too large-scale. Big data ≠ Big understanding

→ Mathematical tools like Compressed Sensing provide rigorous means for representing large data in efficient ways.

→ This allows for efficient data acquisition, storage, and analysis.

Representations of High Dimensional Data

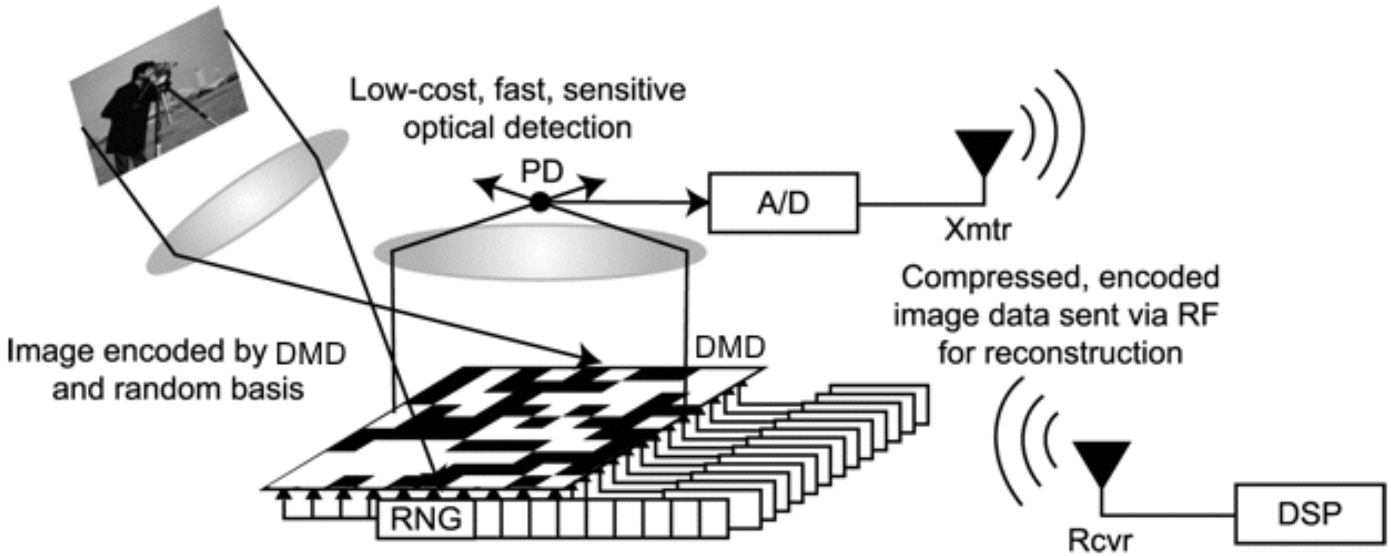
Key Topics :

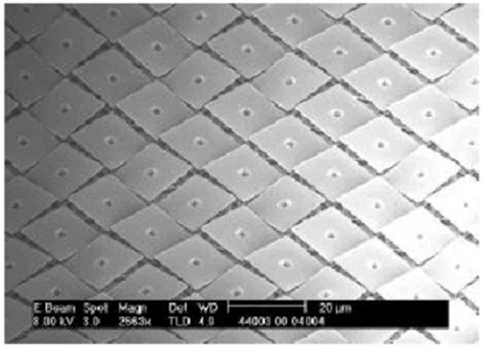
- Mathematics of sparsity and compressed sensing
  - Sampling designs
  - Reconstruction methods
  - Quantization issues

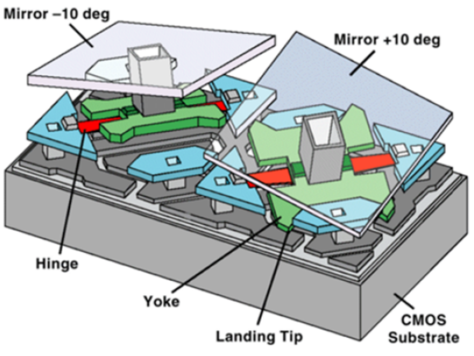
- Inferential tasks
  - Topic modeling
  - Clustering and classification methods
  - Numerical optimization

Applications

- Digital Camera (Rice Univ.)

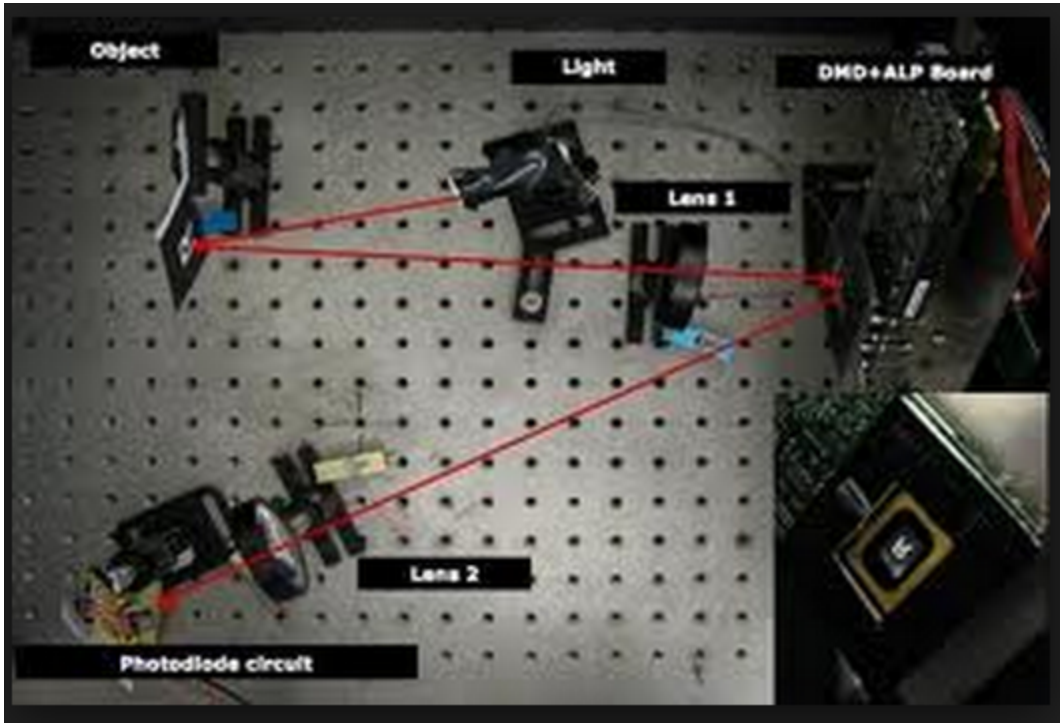






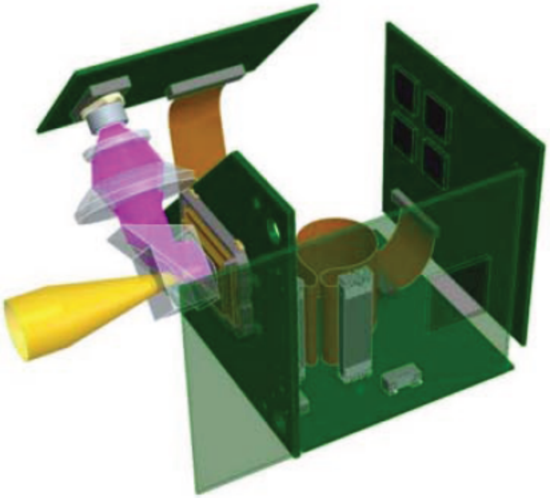
Applications

- Digital Camera (Rice Univ.)



Applications

❖ Hyperspectral camera (InView Corp.)



Applications

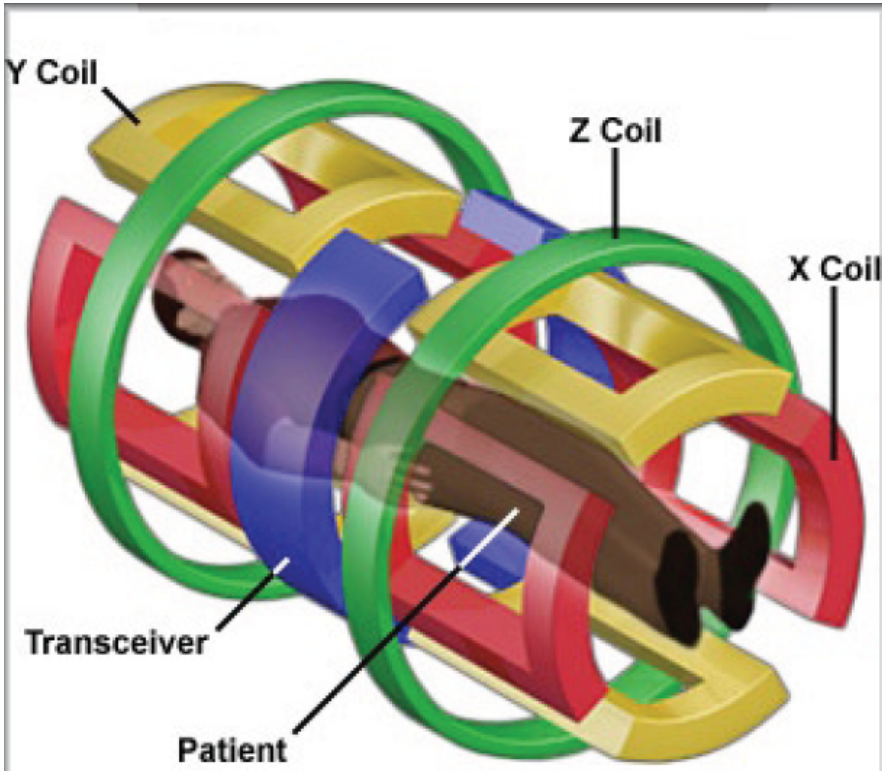
- Magnetic Resonance Imaging (MRI)

Applications

❖ Magnetic Resonance Imaging (MRI)

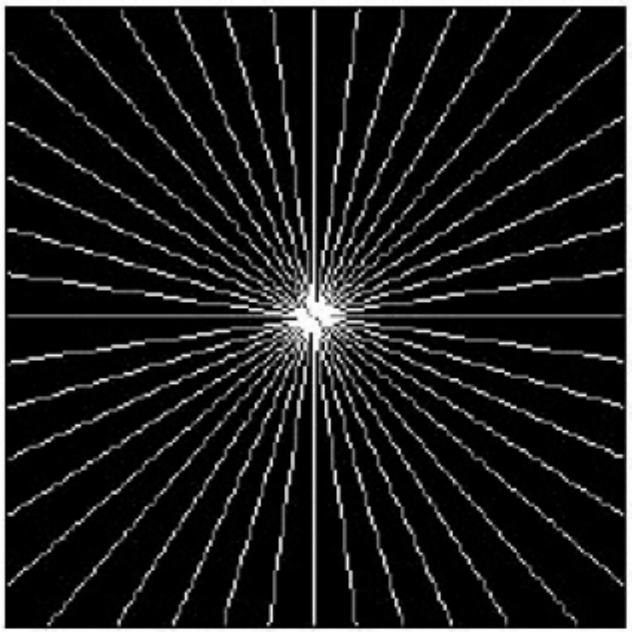
Applications

- Magnetic Resonance Imaging (MRI)



Applications

❖ Magnetic Resonance Imaging (MRI)

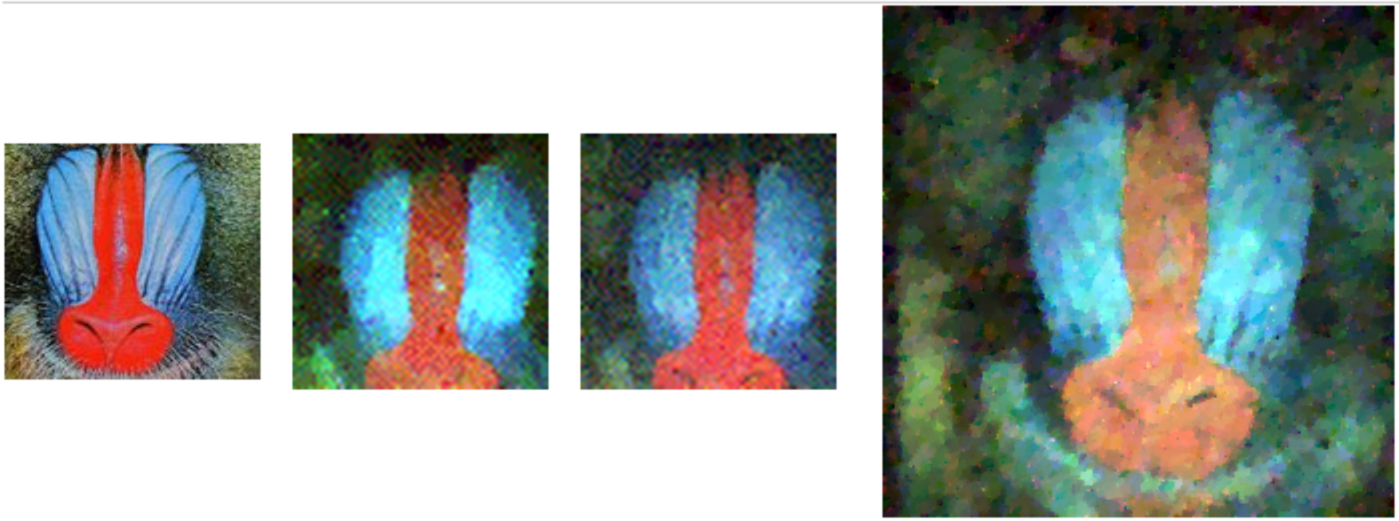


Less measurements = less time

Applications

❖ Magnetic Resonance Imaging (MRI)

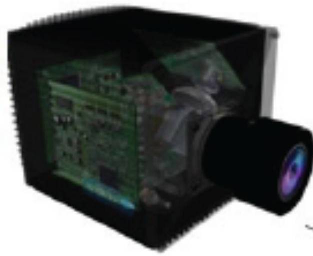
Results of Compressed Sensing

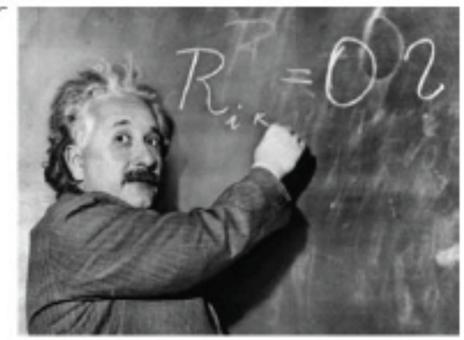


Original	4096 Pixels 800 Measurements (20%)	4096 Pixels 1600 Measurements (40%)	65536 Pixels 6600 Measurements (10%)

Results of Compressed Sensing

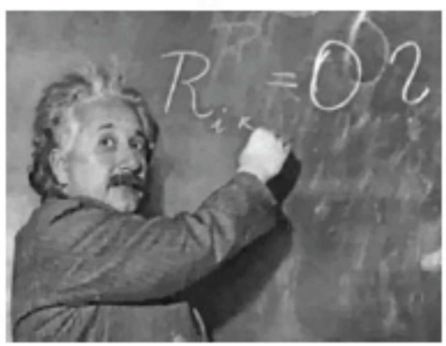
InView
"single-pixel"
Compressive
Sensing
SWIR
Camera
(1024x768)



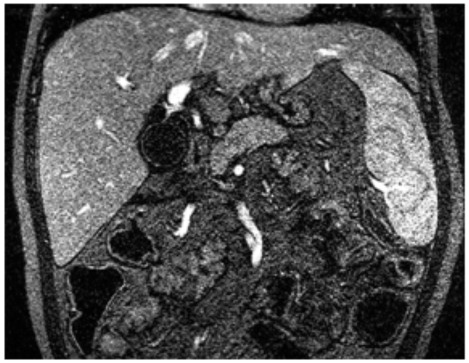


- Target: black and white print-out
- Illumination: SWIR source (no visual source)

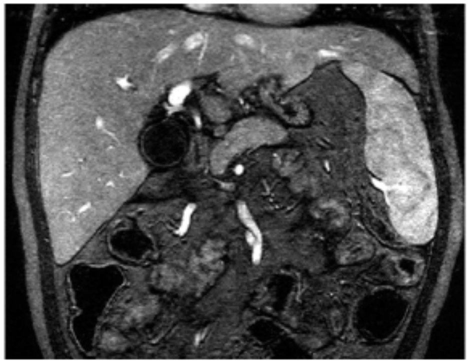
InView
SWIR
Result



Results of Compressed Sensing



(a)



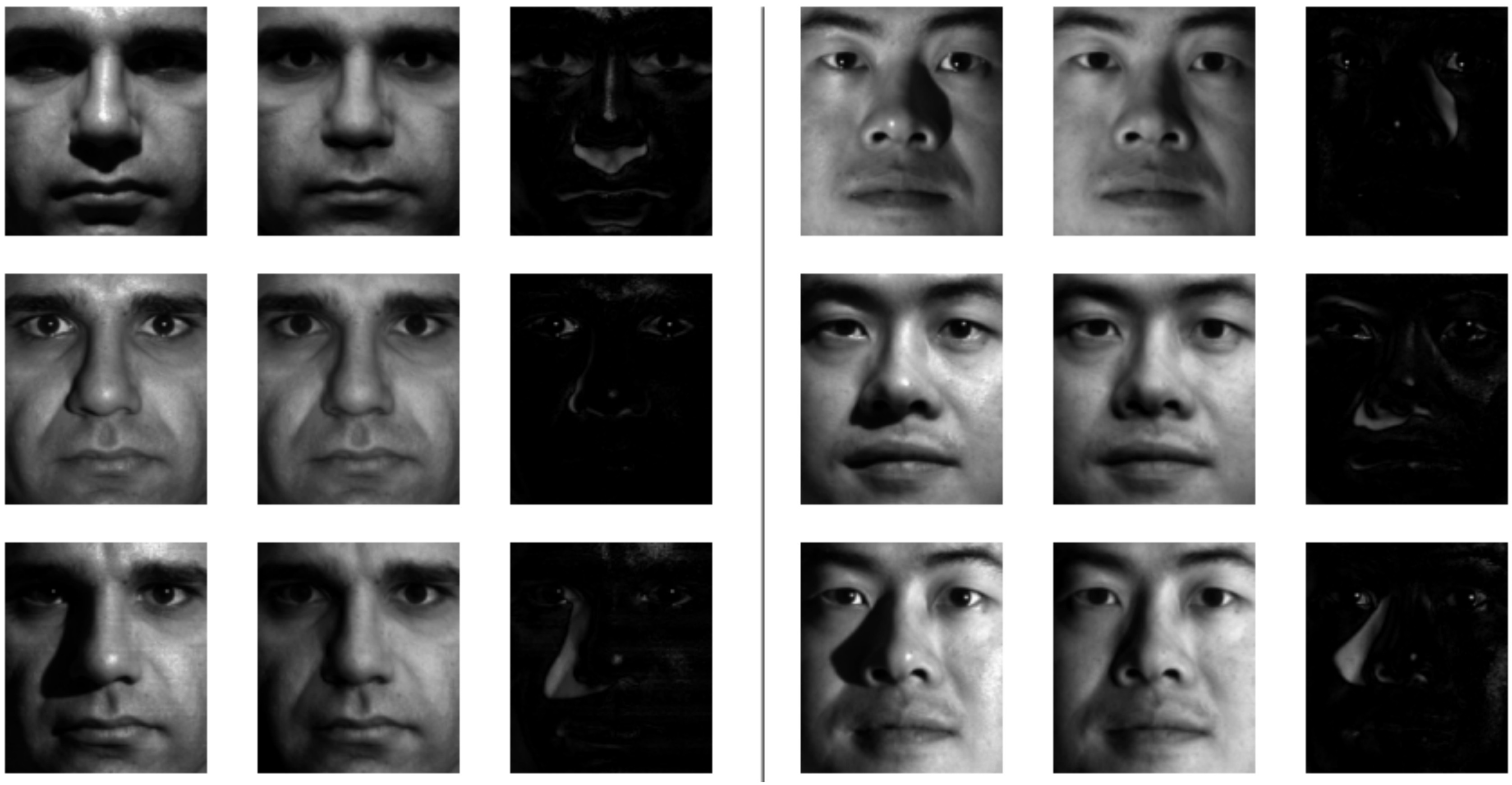
(b)

(c)

(d)

Results of Compressed Sensing

Results of Compressed Sensing



Results of Compressed Sensing

Original D

Repaired A

Corruptions

Frame 1

480×620 pixels

Why is compression possible?

Why is compression possible?

Because most practical signals, such as images, contain much less information than their dimension (e.g. 256x256 = 65,536 pixels) would suggest.

Why is compression possible?

Because most practical signals, such as images, contain much less information than their dimension (e.g. 256x256 = 65,536 pixels) would suggest.

How to quantify this?

A believable example

This image is sparse.

A believable example

This image is sparse.

In a computer, images are represented by an array of numbers (0=black, 2555=white). Sparse images are those which are mostly zeros (black).

A little bit harder…

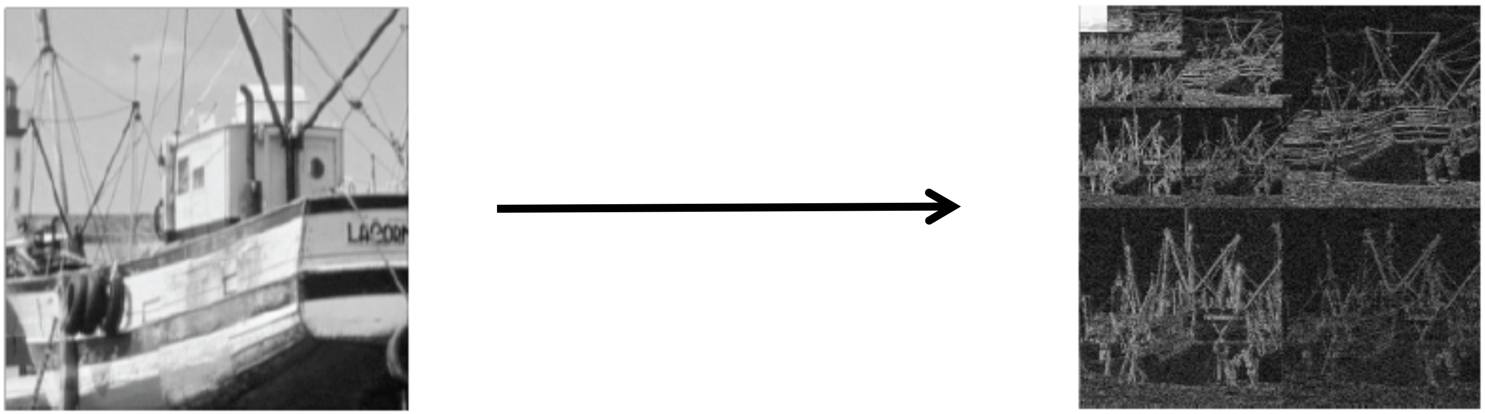
This image is NOT sparse…uh oh.

We call an image "compressible" if it is well approximated by a sparse image.

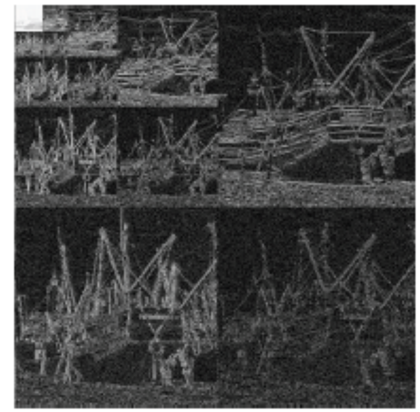
Ok, this one is really hard…

This image is NOT EVEN CLOSE to sparse…uh oh.

Sparsifying transformations

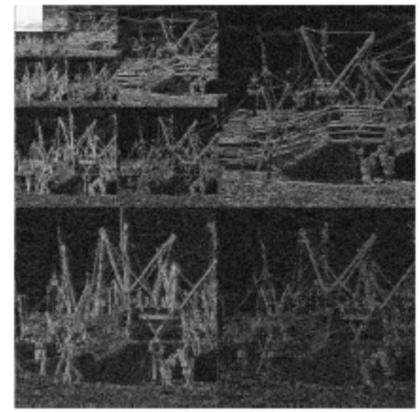


Sparsifying transformations



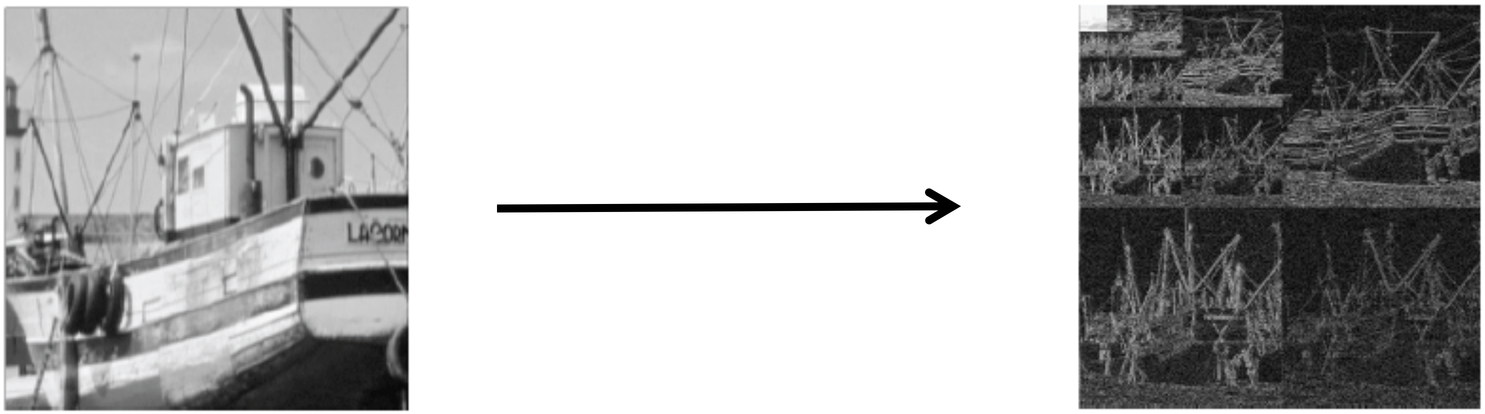
❖ Haar wavelet transformation (Haar, 1909)

Sparsifying transformations



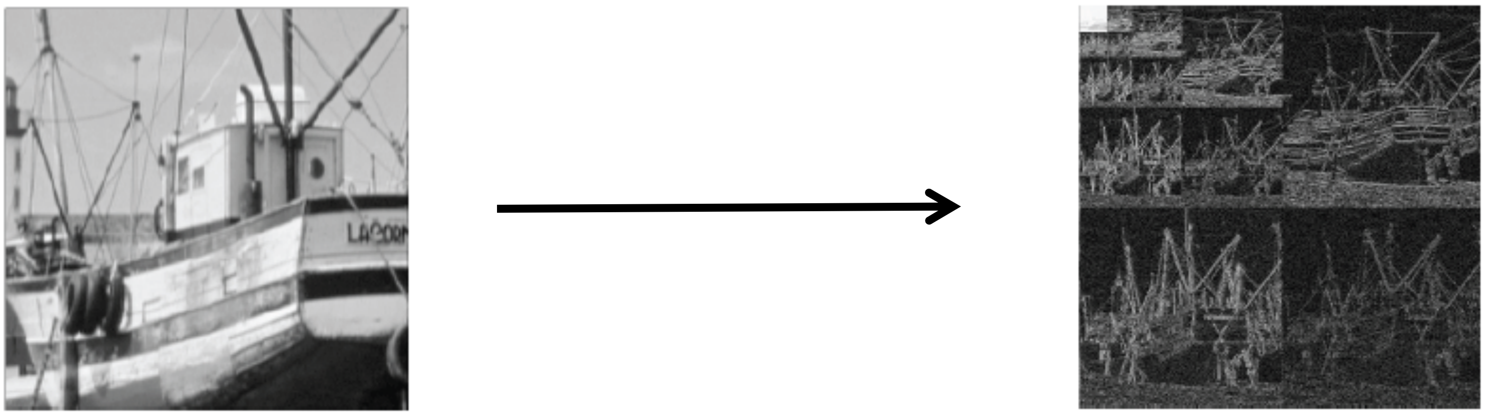
- ❖ Haar wavelet transformation (Haar, 1909)
- ❖ Daubechies wavelet transformation (Daubechies, 1988)

Sparsifying transformations



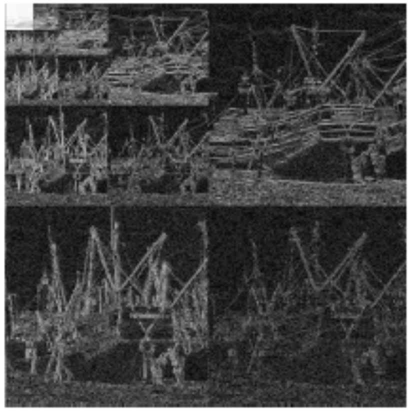
- ❖ Haar wavelet transformation (Haar, 1909)
- ❖ Daubechies wavelet transformation (Daubechies, 1988)
- ❖ Curvelets (Candes et.al., 2002)

Sparsifying transformations

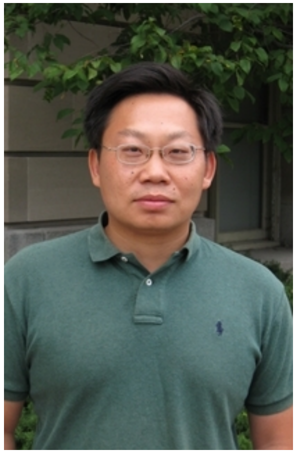


- Haar wavelet transformation (Haar, 1909)
- Daubechies wavelet transformation (Daubechies, 1988)
- Curvelets (Candes et.al., 2002)
- Shearlets (Kutyniok et.al., 2005)

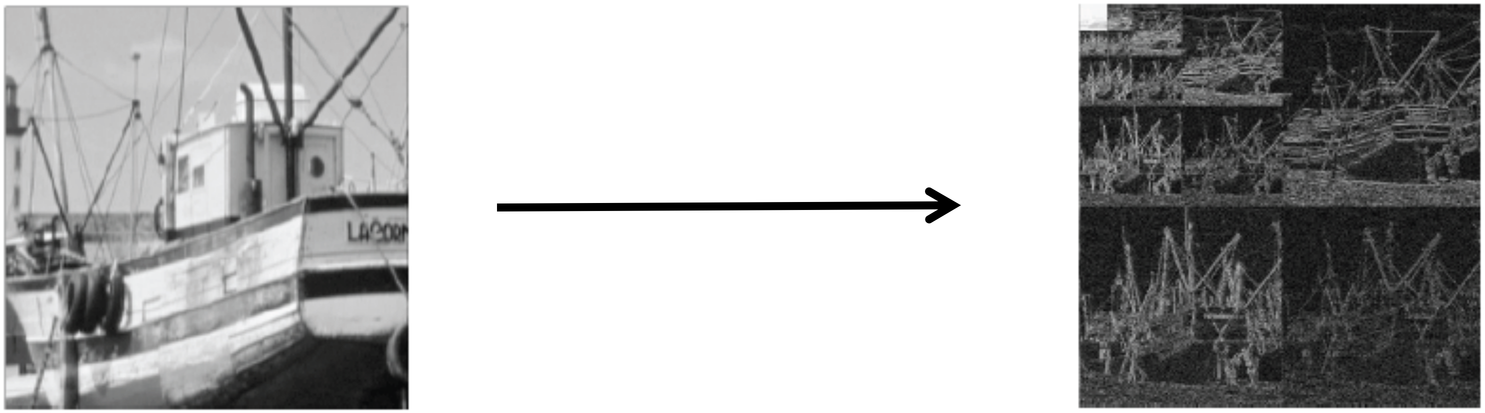
Sparsifying transformations



- Haar wavelet transformation (Haar, 1909)
- Daubechies wavelet transformation (Daubechies, 1988)
- Curvelets (Candes et.al., 2002)
- Shearlets (Kutyniok et.al., 2005)
- Framelets (Cai et.al., 2008)

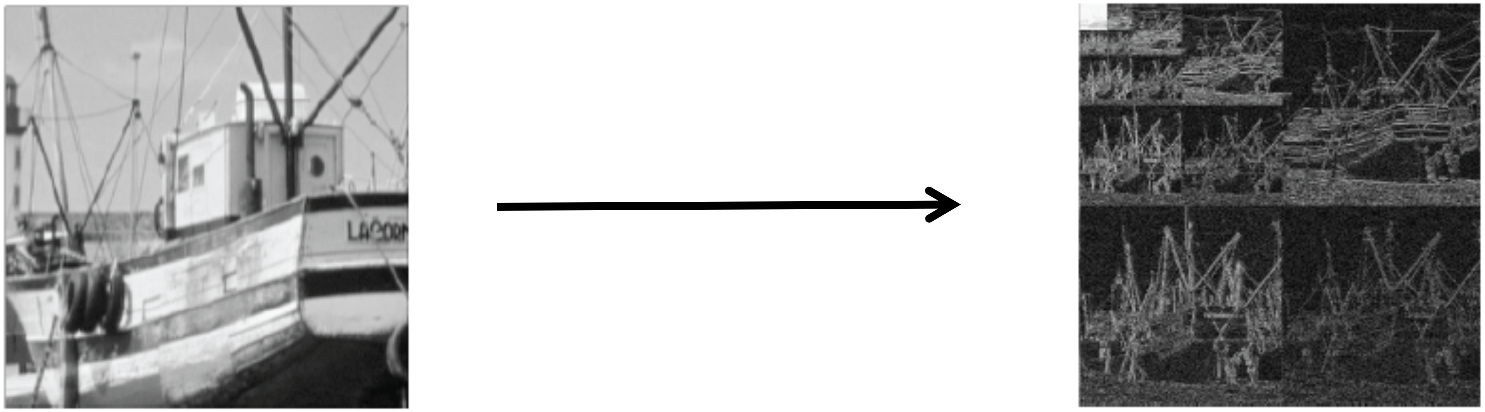


Sparsifying transformations



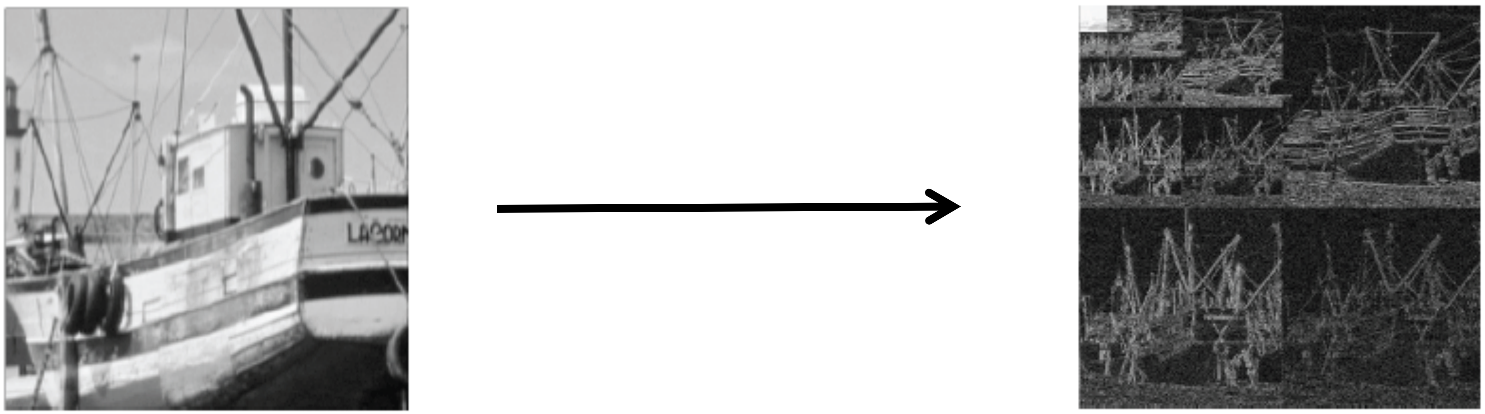
- Haar wavelet transformation (Haar, 1909)
- Daubechies wavelet transformation (Daubechies, 1988)
- Curvelets (Candes et.al., 2002)
- Shearlets (Kutyniok et.al., 2005)
- Framelets (Cai et.al., 2008)
- Omelets

Sparsifying transformations



- Haar wavelet transformation (Haar, 1909)
- Daubechies wavelet transformation (Daubechies, 1988)
- Curvelets (Candes et.al., 2002)
- Shearlets (Kutyniok et.al., 2005)
- Framelets (Cai et.al., 2008)
- Omelets  (jk)

Sparsifying transformations



- Mathematically, a basis or redundant frame B such that:

$$x = Bz, \quad z \text{ is } s\text{-sparse } (s << d)$$

Sparsifying transformations

- We can thus assume the images of interest are sparse
- *How* do we actually compress them and then *how* do we reconstruct them from that compression?
- Simple ad-hoc methods not feasible for practice. Need sophisticated robust machinery, motivated by applications.

Mathematical formulation

1. Signal of interest $f \in \mathbb{C}^n$ (or $\mathbb{C}^{N \times N}$)
2. Sampling operator $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$.
3. Samples $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

Mathematical formulation

1. Signal of interest $f \in \mathbb{C}^n$ (or $\mathbb{C}^{N \times N}$)
2. Sampling operator $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$.
3. Samples $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

Mathematical formulation

1. Signal of interest $f \in \mathbb{C}^n$ (or $\mathbb{C}^{N \times N}$)
2. Sampling operator $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$.
3. Samples $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

Mathematical formulation

1. Signal of interest $f \in \mathbb{C}^n$ (or $\mathbb{C}^{N \times N}$)
2. Sampling operator $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$.
3. Samples $y = \mathcal{A}f + \xi$.

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

4. Problem: Reconstruct signal f from measurements y

Mathematical formulation

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

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad \mathcal{A} \quad \end{bmatrix} \begin{bmatrix} f \end{bmatrix} + \begin{bmatrix} \xi \end{bmatrix}$$

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\mathrm{supp}(f)| \leq s \ll n$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \leq s \ll n$
- In other dictionary: $f = Dx$ where $\|x\|_0 \leq s \ll n$

In practice, we encounter compressible signals.

Mathematical formulation

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

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

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll n$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \leq s \ll n$
- In other dictionary: $f = Dx$ where $\|x\|_0 \leq s \ll n$

In practice, we encounter compressible signals.

Mathematical formulation

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

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

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\text{supp}(f)| \leq s \ll n$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \leq s \ll n$
- In other dictionary: $f = Dx$ where $\|x\|_0 \leq s \ll n$

In practice, we encounter compressible signals.

Mathematical formulation

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

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

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\mathrm{supp}(f)| \leq s \ll n$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \leq s \ll n$
- In other dictionary: $f = Dx$ where $\|x\|_0 \leq s \ll n$

In practice, we encounter compressible signals.

Mathematical formulation

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

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

Assume f is sparse:

- In the coordinate basis: $\|f\|_0 \stackrel{\text{def}}{=} |\mathrm{supp}(f)| \leq s \ll n$
- In orthonormal basis: $f = Bx$ where $\|x\|_0 \leq s \ll n$
- In other dictionary: $f = Dx$ where $\|x\|_0 \leq s \ll n$

In practice, we encounter compressible signals.

Restricted Isometry Property

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

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

- Sub-gaussian measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log n.$$

- Subsampled bounded orthogonal (e.g. Fourier) matrices have similar property: $m \gtrsim s \log^4 n$.

Restricted Isometry Property

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

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

- Sub-gaussian measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log n.$$

- Subsampled bounded orthogonal (e.g. Fourier) matrices have similar property: $m \gtrsim s \log^4 n$.

Restricted Isometry Property

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

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

- Sub-gaussian measurement matrices satisfy the RIP with high probability when

$$m \gtrsim s \log n.$$

- Subsampled bounded orthogonal (e.g. Fourier) matrices have similar property: $m \gtrsim s \log^4 n$.

Recovery guarantees via ℓ_1-minimization

ℓ_1-minimization Candès-Romberg-Tao '06

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

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

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

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

Recovery guarantees via ℓ_1-minimization

ℓ_1-minimization Candès-Romberg-Tao '06

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

$$\hat{f} = \mathrm{argmin}_g \|g\|_1 \quad \text{such that} \quad \|\mathcal{A}f - y\|_2 \leq \varepsilon,$$

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

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

Greedy methods

(Jeff Blanchard)

1. OMP
2. CoSaMP
3. IHT
4. ...

Extensions of CS

Some non-trivial branches

1. Non-orthonormal bases
2. Quantization
3. Matrix completion (Mark Davenport)

Non-orthonormal sparsifying bases

Many (most) signals are sparse in highly redundant tight frames.

1. Oversampled DFT
2. Gabor frames
3. Curvelet frames
4. Undecimated wavelet frames
5. ONB concatenations
6. ...
7. Gradient

Non-orthonormal sparsifying bases

ℓ_1-analysis

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

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

ℓ_1-analysis Candès-Eldar-N-Randall '10

Let D be an arbitrary tight frame and let $\mathcal{A}$ satisfy (a variant of the) RIP. Then the solution $\hat{f}$ to ℓ_1-analysis satisfies

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

Non-orthonormal sparsifying bases

ℓ_1-analysis

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

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

ℓ_1-analysis Candès-Eldar-N-Randall '10

Let D be an arbitrary tight frame and let $\mathcal{A}$ satisfy (a variant of the) RIP. Then the solution $\hat{f}$ to ℓ_1-analysis satisfies

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

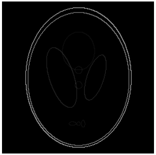
Non-orthonormal sparsifying bases

Many (most) signals are sparse in highly redundant tight frames.

1. Oversampled DFT
2. Gabor frames
3. Curvelet frames
4. Undecimated wavelet frames
5. ONB concatenations
6. ...
7. Gradient

Gradient sparsity

Natural images and smoothly varying signals are compressible in the *discrete gradient*.



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

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

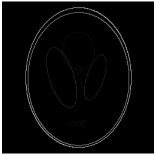
and the discrete gradient operator is

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

$\|\nabla[f]\|_1 := \|f\|_{TV}$ is the *total variation* (TV).

Gradient sparsity

Natural images and smoothly varying signals are compressible in the *discrete gradient*.

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

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

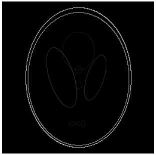
and the discrete gradient operator is

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

$\|\nabla[f]\|_1 := \|f\|_{\mathrm{TV}}$ is the *total variation* (TV).

Gradient sparsity

Natural images and smoothly varying signals are compressible in the *discrete gradient*.

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

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

and the discrete gradient operator is

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

$\|\nabla[f]\|_1 := \|f\|_{\mathrm{TV}}$ is the *total variation* (TV).

Stable signal recovery using total-variation minimization

Theorem N-Ward '13

From $m \gtrsim s \log(N^d)$ RIP measurements, for any $f \in \mathbb{C}^{N^d}$ $(d \geq 2)$,

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

satisfies

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

and

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

This error guarantee is optimal up to log factors.

Stable signal recovery using total-variation minimization

Theorem N-Ward '13

From $m \gtrsim s \log(N^d)$ RIP measurements, for any $f \in \mathbb{C}^{N^d}$ $(d \geq 2)$,

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

satisfies

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

and

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

This error guarantee is optimal up to log factors.

Stable signal recovery using total-variation minimization

Theorem N-Ward '13

From $m \gtrsim s \log(N^d)$ RIP measurements, for any $f \in \mathbb{C}^{N^d}$ $(d \geq 2)$,

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

satisfies

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

and

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

This error guarantee is optimal up to log factors.

Stable signal recovery using total-variation minimization

Theorem N-Ward '13

From $m \gtrsim s \log(N^d)$ RIP measurements, for any $f \in \mathbb{C}^{N^d}$ $(d \geq 2)$,

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

satisfies

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

and

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

This error guarantee is optimal up to log factors.

Stable signal recovery using total-variation minimization

Theorem N-Ward '13

From $m \gtrsim s\log(N^d)$ RIP measurements, for any $f \in \mathbb{C}^{N^d}$ $(d \geq 2)$,

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

satisfies

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

and

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

This error guarantee is optimal up to log factors.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \leq s$ acquired via nonadaptive linear measurements $\langle a_i, f \rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \leq h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \leq s$ acquired via nonadaptive linear measurements $\langle a_i, f \rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \leq h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \leq s$ acquired via nonadaptive linear measurements $\langle a_i, f\rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f\rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \leq h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \le s$ acquired via nonadaptive linear measurements $\langle a_i, f\rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f\rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \le h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \leq s$ acquired via nonadaptive linear measurements $\langle a_i, f \rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \leq h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \leq s$ acquired via nonadaptive linear measurements $\langle a_i, f\rangle + e_i$, $i = 1, \ldots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f\rangle + e_i), \qquad i = 1, \ldots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \leq h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

The One-Bit Sparse reconstruction problem

- Standard: $f \in \mathbb{R}^n$ with $\|f\|_0 \le s$ acquired via nonadaptive linear measurements $\langle a_i, f \rangle + e_i$, $i = 1, \dots, m$.
- In practice, measurements need to be quantized.
- One-Bit: extreme quantization as $y = \mathrm{sign}(\mathcal{A}f + e)$, i.e.,

$$y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i), \qquad i = 1, \dots, m.$$

- Goal: find reconstruction maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^n$ such that, assuming the ℓ_2-normalization of f,

$$\|f - \Delta(y)\| \le h(\lambda)$$

where the oversampling factor is denoted

$$\lambda := \frac{m}{s \ln(n/s)}$$

and h is rapidly decreasing to zero when λ increases.

Limitations of the Framework

- Power decay is optimal since

$$\|f - \Delta_{\mathrm{opt}}(y)\|_2 \gtrsim \lambda^{-1}$$

even if $\mathrm{supp}(f)$ known in advance [Goyal-Vetterli-Thao '98].

- Geometric intuition

Limitations of the Framework

- Power decay is optimal since

$$\|f - \Delta_{\mathrm{opt}}(y)\|_2 \gtrsim \lambda^{-1}$$

even if $\mathrm{supp}(f)$ known in advance [Goyal-Vetterli-Thao '98].

- Geometric intuition

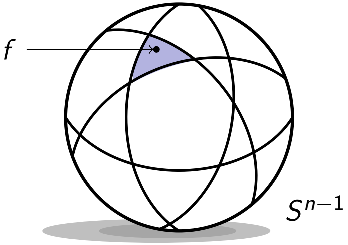
Limitations of the Framework

- Power decay is optimal since

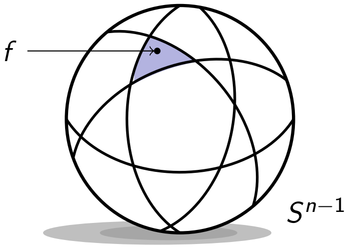
$$\|f - \Delta_{\mathrm{opt}}(y)\|_2 \gtrsim \lambda^{-1}$$

even if $\mathrm{supp}(f)$ known in advance [Goyal-Vetterli-Thao '98].

- Geometric intuition



Adaptivity



- Remedy: adaptive choice of dithers $\tau_1, \ldots, \tau_m$ in

$$y_i = \operatorname{sign}(\langle a_i, f \rangle - \tau_i), \qquad i = 1, \ldots, m.$$

Main results

Theorem Baraniuk-Foucart-N-Plan-Wootters '16

- Pre-quantization error, $y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i - \tau_i)$:
  if $\|e\|_\infty \le \varepsilon R\, 2^{-T}$ (or $\|e^t\|_2 \le \varepsilon \sqrt{q} \|f - f^t\|_2$ throughout), then
  $$\|f - f^T\|_2 \le R\, 2^{-T} = R \exp(-c\lambda)$$
  for the convex-optimization and hard-thresholding schemes.
- Post-quantization error, $y_i = f_i \, \mathrm{sign}(\langle a_i, f \rangle + e_i - \tau_i)$:
  if $|\{i : f_i^t = -1\}| \le \eta q$ throughout, then
  $$\|f - f^T\|_2 \le R\, 2^{-T} = R \exp(-c\lambda)$$
  for the hard-thresholding scheme.

Main results

Theorem Baraniuk-Foucart-N-Plan-Wootters '16

- Pre-quantization error, $y_i = \mathrm{sign}(\langle a_i, f \rangle + e_i - \tau_i)$:
  if $\|e\|_\infty \leq \varepsilon R\, 2^{-T}$ (or $\|e^t\|_2 \leq \varepsilon \sqrt{q} \|f - f^t\|_2$ throughout), then
  $$\|f - f^T\|_2 \leq R\, 2^{-T} = R \exp(-c\lambda)$$
  for the convex-optimization and hard-thresholding schemes.
- Post-quantization error, $y_i = f_i \,\mathrm{sign}(\langle a_i, f \rangle + e_i - \tau_i)$:
  if $|\{i : f_i^t = -1\}| \leq \eta q$ throughout, then
  $$\|f - f^T\|_2 \leq R\, 2^{-T} = R \exp(-c\lambda)$$
  for the hard-thresholding scheme.

Thank you!

E-mail:

- deanna@math.ucla.edu

Web:

- www.math.ucla.edu/~deanna/

References:

- 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.
- D. Needell and R. Ward. Stable image reconstruction using total variation minimization. SIAM Journal on Imaging Sciences, 6(2):1035-1058.
- D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization. IEEE Transactions on Image Processing, 22(10):3941-3949.
- R. Baraniuk, S. Foucart, D. Needell, Y. Plan, M. Wootters. Exponential decay of reconstruction error from binary measurements of sparse signals. IEEE Trans. Information Theory, vol. 63, num. 6, 3368 - 3385, 2017.