Introduction to Compressed Sensing

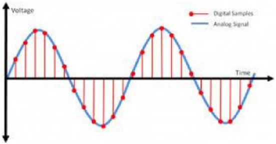
Alejandro Parada, Gonzalo Arce

University of Delaware

August 25, 2016

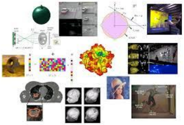
Motivation: Classical Sampling

Shannon-Nyquist sampling theorem



Continuous Signals

Discrete Sequences



Applications

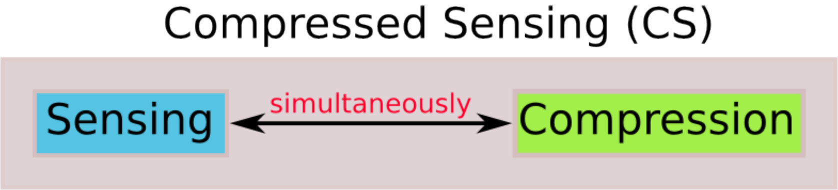
Motivation: Classical Sampling Issues

Some applications

- Radar
- Spectral Imaging
- Medical Imaging
- Remote surveillance

Issue

Sampling rate is too high!



$\Updownarrow$

Reduction in the computation costs for measuring signals that have an **sparse** representation.

Compressed Sensing (CS) vs Classical Sampling

CS

- $\mathbf{x} \in \mathbb{R}^n$
- **Random** measurements
- Measurements as inner products $\langle \mathbf{x}, \boldsymbol{\phi} \rangle$

Classical Sampling

- Continuous signals
- Infinite-length signals

Compressed Sensing (CS) vs Classical Sampling

CS

- **Recovery**: Non linear.

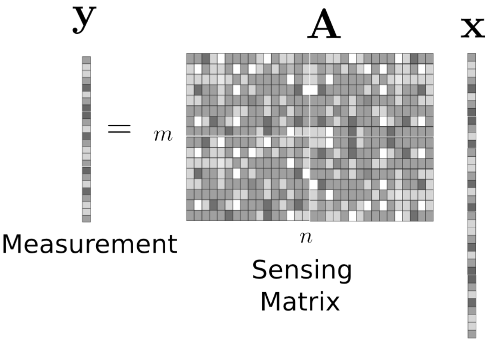
Classical Sampling

- Recovery: linear processing.

Compressed sensing basics

$\mathbf{x} \in \mathbb{R}^n$ is acquired taking $m < n$ measurements

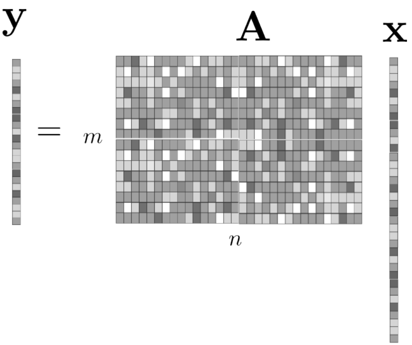
$$\boxed{\mathbf{y} = \mathbf{A}\mathbf{x}} \quad (1)$$



Compressed Sensing Basics

$\mathbf{x} \in \mathbb{R}^n$ is acquired taking $m < n$ measurements

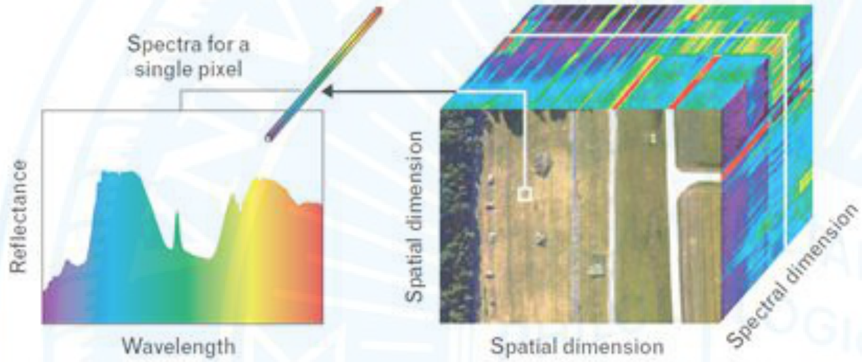
$$\boxed{\mathbf{y} = \mathbf{A}\mathbf{x}} \qquad (1)$$

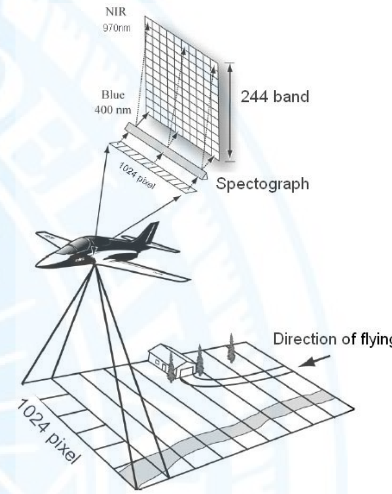


- $\mathbf{y}$: Measurement vector
- $\mathbf{A}$: CS sensing matrix
- $m \ll n$

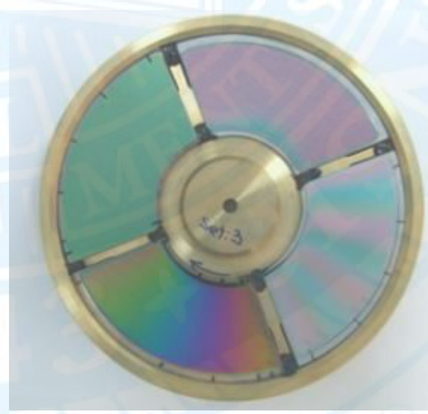
The Spectral Imaging Problem

- Push broom spectral imaging: Expensive, low sensing speed, senses $N \times N \times L$ voxels





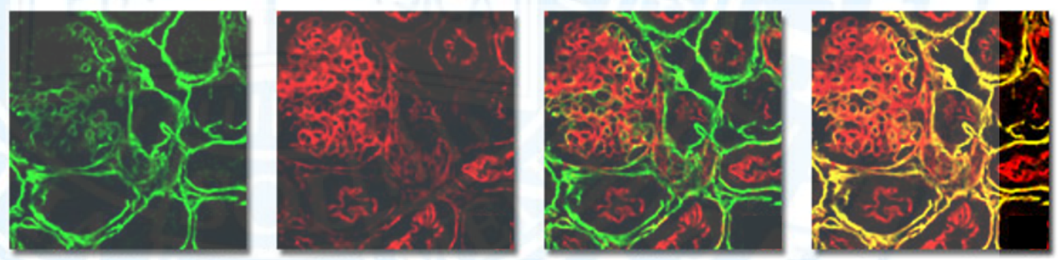
- Optical Filters; Sequential sensing of $N \times N \times L$ voxels; limited by number of colors



Why is this Important?

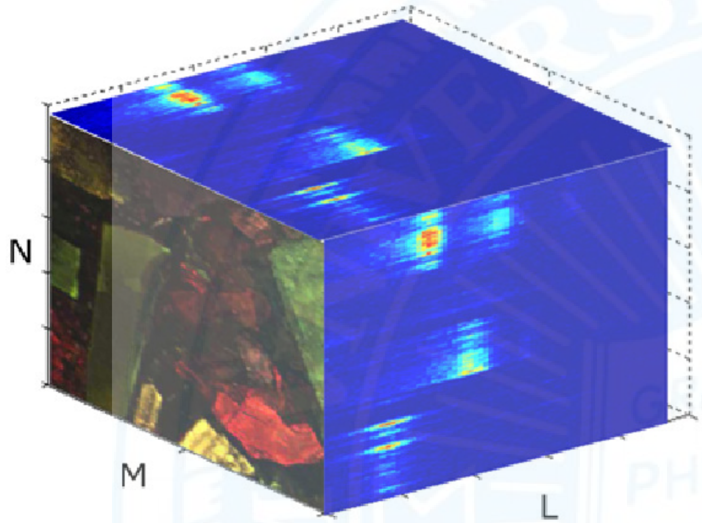
- Remote sensing and surveillance in the Visible, NIR, SWIR

- Devices are challenging in NIR and SWIR due to SWaP



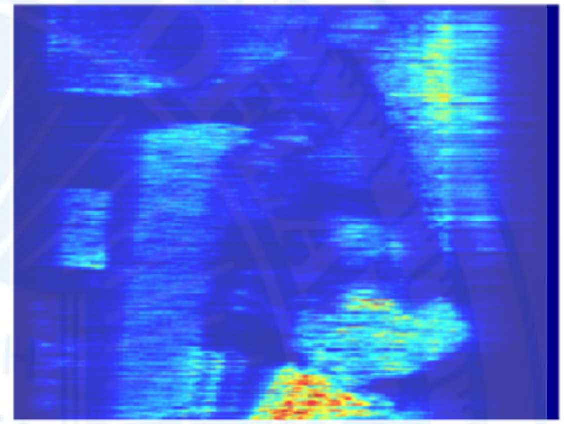
- Medical imaging and other applications

Introduction



Datacube

$$\mathbf{f} = \boldsymbol{\Psi}\boldsymbol{\theta}$$



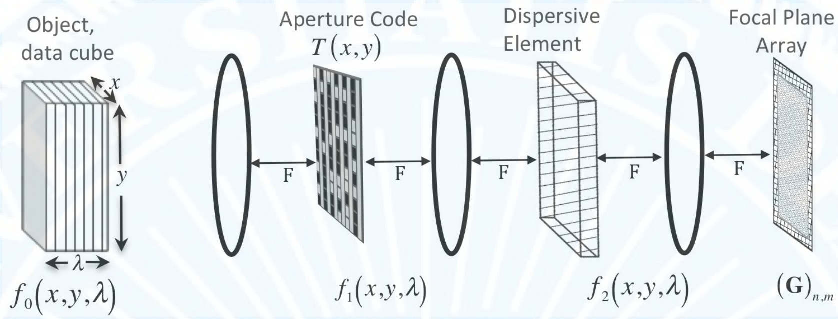
Compressive Measurements

$$\mathbf{g} = \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta} + \mathbf{w}$$

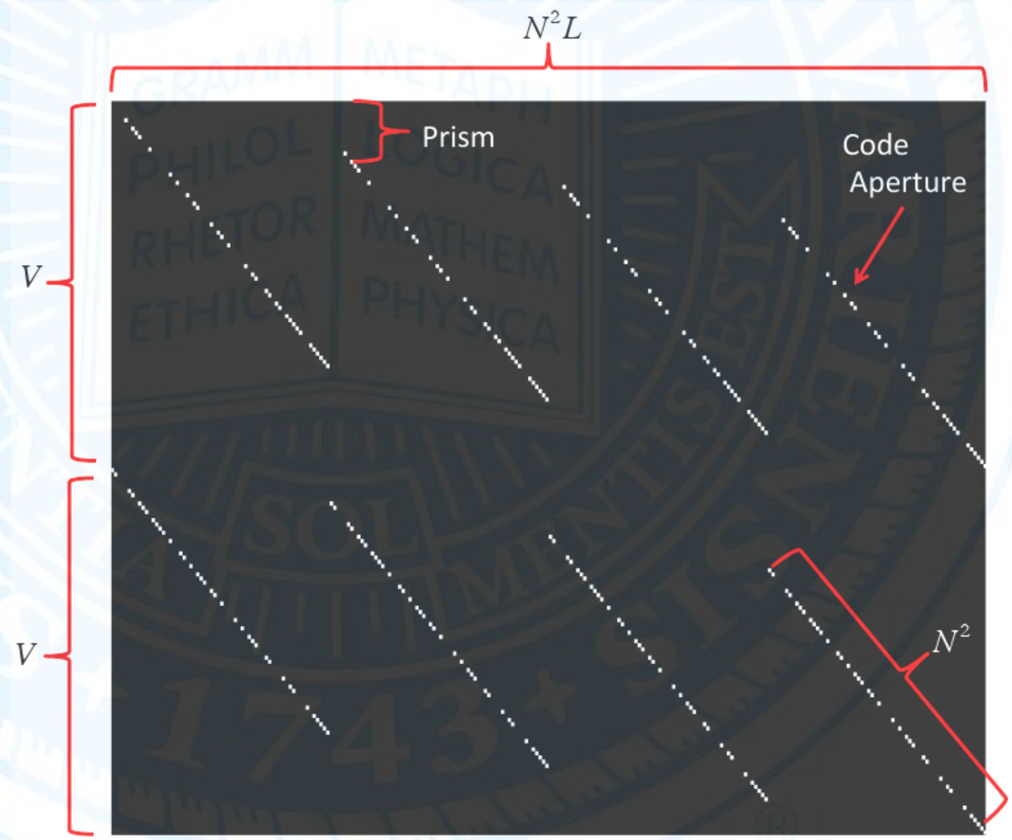
Underdetermined system of equations

$$\hat{\mathbf{f}} = \boldsymbol{\Psi}\{\min_{\boldsymbol{\theta}} \|\mathbf{g} - \mathbf{H}\boldsymbol{\Psi}\boldsymbol{\theta}\|_2 + \tau\|\boldsymbol{\theta}\|_1\}$$

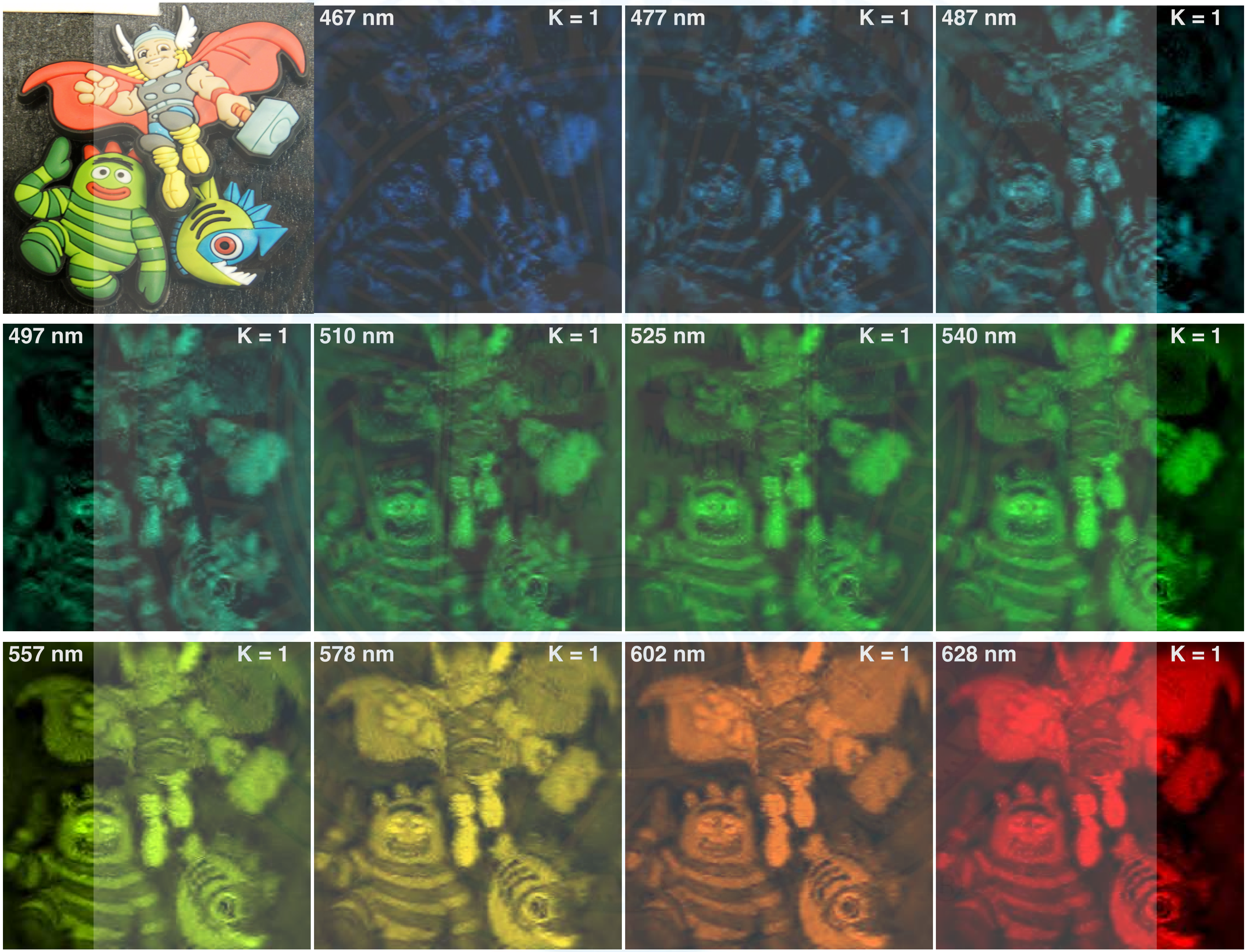
Matrix CASSI representation $\mathbf{g} = \mathbf{H}\mathbf{f}$



- Data cube: $N \times N \times L$
- Spectral bands: L
- Spatial resolution: $N \times N$
- Sensor size $N \times (N + L - 1)$
- $V = N(N + L - 1)$

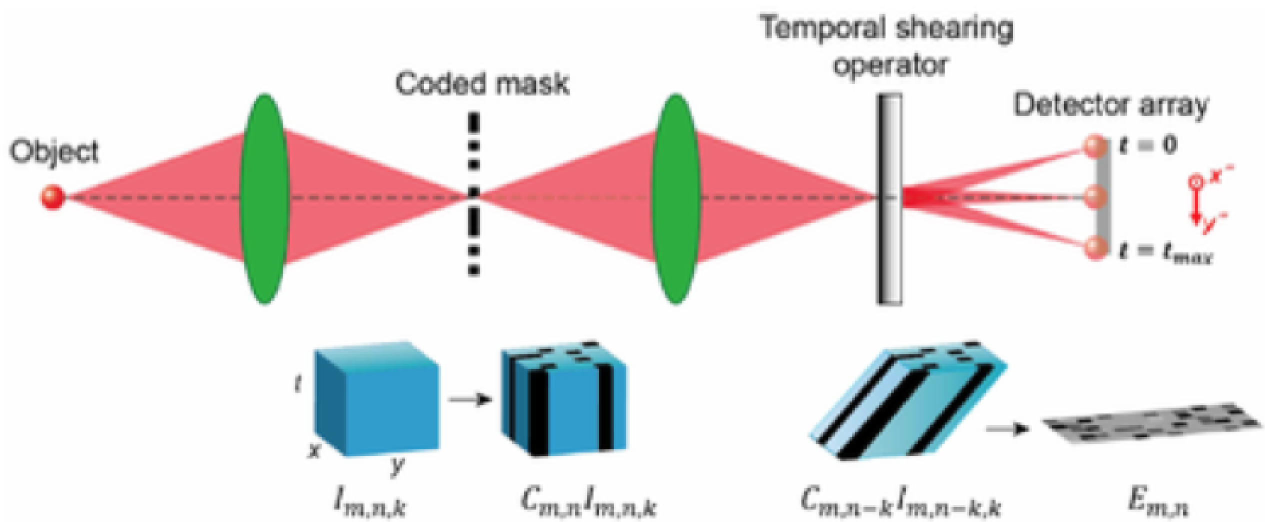


Preliminar Results: $K = 1$ Random Snapshots



Ultrafast Photography

Single-shot compressed ultrafast photography at one hundred billion frames per second



High dimensional signal

Low number of linear measurements

Since $m < n$, it is possible to have

$$\mathbf{y} = \begin{cases} \mathbf{A}\mathbf{x} \\ \\ \mathbf{A}\mathbf{x}' \end{cases}$$

with $\mathbf{x} \neq \mathbf{x}'$

Motivation

- Design $\mathbf{A}$ such that $\mathbf{x}$ can be uniquely identifiable from $\mathbf{y}$, with $\mathbf{x}$ in an **specific** space of signals

Even when $m \ll n$

Characteristics of this space?

$$\mathbf{x} = \sum_{i=1}^{n} \boldsymbol{\phi}_i \theta_i \qquad (2)$$

$$\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\theta} \qquad (3)$$

with

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n]$$

$$\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T$$

Definition

A signal $\mathbf{x}$ is k sparse in the basis frame $\boldsymbol{\Phi}$ if exists $\boldsymbol{\theta} \in \mathbb{R}^n$ with $k = |\text{supp}(\boldsymbol{\theta})| \ll n$ such that $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\theta}$

Characteristics of this space ?

Definition

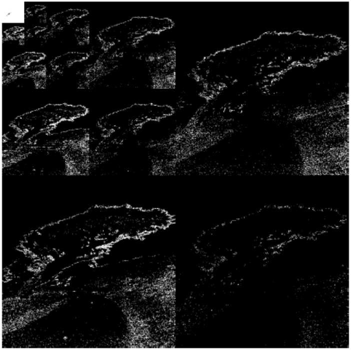
The space of k sparse signals Σ_k is defined as

$$\Sigma_k = \{\mathbf{x} : \|\mathbf{x}\|_0 \leq k\} \quad (4)$$

$\|\mathbf{x}\|_0$: Number of nonzero elements in $\mathbf{x}$ (It is called the ℓ_0-norm)

Examples of Sparse Signals

(a)



(b)

Figure: (a): Original Image. (b) Wavelet Representation.[1]

[1] Compressed Sensing: Theory and Applications, Eldar

Compressible Signals

Real signals

Non exactly sparse.

Real signals

Good approximations on Σ_k

Compressible Signals

Compressible Signals

(a) (b)

Figure: (a): Original Image. (b) Wavelet Representation (Keeping 10%).

Compressible Signals

Real signals

Non exactly sparse.

Real signals

Good approximations on Σ_k

Compressible Signals

$$\sigma_k(\mathbf{x})_p \triangleq \min_{\hat{x} \in \Sigma_k} \|\mathbf{x} - \hat{\mathbf{x}}\|_p \quad (5)$$

where $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

Sensing Matrices $\mathbf{A}$

Sensing matrix $\mathbf{A}$

1

Identify uniquely $\mathbf{x} \in \Sigma_k$ given

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

2

How to get $\mathbf{x}$ given $\mathbf{y}$?

Sensing Matrices

Let $\Lambda \subset \{1, 2, \ldots, n\}$

$\mathbf{A}_\Lambda$: The matrix that contains all columns of $\mathbf{A}$ indexed by Λ.

Example If $\Lambda = \{1, 3\}$

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \end{bmatrix}$$

Then

$$\mathbf{A}_\Lambda = \begin{bmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}$$

Sensing Matrices: When supp(x) is known

$$\Lambda = \text{supp}(\mathbf{x})$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}_\Lambda \mathbf{x}_\Lambda \quad (6)$$

If $\mathbf{A}_\Lambda$ is full column rank

$$\mathbf{x}_\Lambda = \mathbf{A}_\Lambda^\dagger \mathbf{y}$$

where

$$\mathbf{A}_\Lambda^\dagger = \left(\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda\right)^{-1} \mathbf{A}_\Lambda^*$$

- If $\mathbf{A}_\Lambda$ is full rank

$$m \geq k$$

If Λ is known

Recover $\mathbf{x}$ from a subspace.

Sensing matrices: When supp(x) is unknown

CS central idea

How to choose $\mathbf{A}$?

- Information of $\mathbf{x}$ is preserved.
- Recover uniquely $\mathbf{x}$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$

Null Space Conditions

Null space of the matrix $\mathbf{A}$

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{z} : \mathbf{A}\mathbf{z} = \mathbf{0}\}$$

Uniqueness in recovery

$$\mathbf{A}\mathbf{x} \neq \mathbf{A}\mathbf{x}^{'}, \quad \mathbf{x} \neq \mathbf{x}^{'}$$

$$\mathbf{A}(\mathbf{x} - \mathbf{x}^{'}) \neq \mathbf{0} \quad \mathbf{x}, \mathbf{x}^{'} \in \Sigma_k$$

so $(\mathbf{x} - \mathbf{x}^{'}) \notin \mathcal{N}(\mathbf{A})$

but $(\mathbf{x} - \mathbf{x}^{'}) \in \Sigma_{2k}$

Desired

$$\mathcal{N}(\mathbf{A}) \cap \Sigma_{2k} = \emptyset$$

The **spark** of a given matrix $\mathbf{A}$ is the smallest number of columns of $\mathbf{A}$ that are linearly dependent.

The procedure

- Look for all combinations of r columns, $r = 2, 3, \ldots, n$.
- If for any of the combinations we get linear dependency, then the spark is given for the number of vectors in that combination.

The **spark** of a given matrix $\mathbf{A}$ is the smallest number of columns of $\mathbf{A}$ that are linearly dependent.

Example

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Rank

$$\text{Rank}(\mathbf{A}) = 4$$

Spark

$$\text{Spark}(\mathbf{A}) = 2$$

Example

$\mathbf{A}$ of size $m \times n$ with $m < n$

- All entries of $\mathbf{A}$ represented by i.i.d random variables.

[3]

Rank

$$\text{Rank}(\mathbf{A}) = m$$

Spark

$$\text{Spark}(\mathbf{A}) = m + 1$$

Any submatrix of size $m \times m$ is non singular

[3] "The rank of a random matrix, X. Feng"

Uniqueness of sparse recovery

Theorem

For any vector $\mathbf{y} \in \mathbb{R}^m$, *there exists at most one signal* $\mathbf{x} \in \Sigma_k$ *such that* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *if and only if* $spark(\mathbf{A}) > 2k$. *For uniqueness we must have that* $m \geq 2k$

Example

$\mathbf{A}$ of size $m \times n$ with $m < n$

- All entries of $\mathbf{A}$ are i.i.d random variables

4

Spark

$$\text{Spark}(\mathbf{A}) = m + 1$$

Unique recovery of $\mathbf{x} \in \Sigma_k$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$ if

$$\text{spark}(\mathbf{A}) > 2k$$

$$m + 1 > 2k$$

$$k < \frac{m+1}{2}$$

[4] "The rank of a random matrix, X. Feng"

Robust Signal Recovery

CS

Where we are?

- Spark Condition
- NSP

Exactly sparse signals

$\Downarrow$

Real signals?

The restricted isometry property (RIP)

CS

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad (20)$$

- $\mathbf{x}$: Exactly sparse.
- $\mathbf{x}$: approximately sparse.

Noise?

What happen if:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \text{Noise} \qquad (21)$$

The restricted isometry property (RIP)

Definition

A matrix $\mathbf{A}$ satisfies the restricted isometry property (RIP) of order k if there exists a δ_k such that

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2 \qquad (22)$$

for all $\mathbf{x} \in \Sigma_k$

Coherence

RIP

$\Downarrow$

NSP

$\Downarrow$

Spark condition

Disadvantages

- RIP: NP-hard to calculate
- NSP: NP-hard to calculate

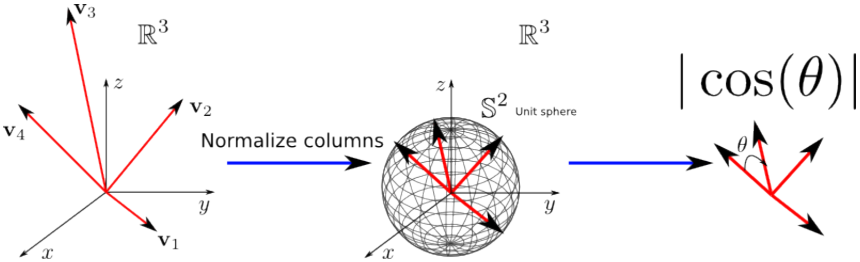
Coherence

Definition

The coherence of a matrix $\mathbf{A}$, denoted $\mu(\mathbf{A})$, is the largest absolute inner product between any two columns $\mathbf{a}_i, \mathbf{a}_j$ of $\mathbf{A}$:

$$\mu(\mathbf{A}) = \max_{1 \le i < j \le n} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2} \tag{30}$$

Coherence


$\mathbf{v}_3$
$\mathbb{R}^3$
z
$\mathbf{v}_2$
$\mathbf{v}_4$
y
x
$\mathbf{v}_1$
Normalize columns
$\mathbb{R}^3$
z
$\mathbb{S}^2$
Unit sphere
y
x
$|\cos(\theta)|$
θ

Properties of the coherence

Theorem

Let $\mathbf{A}$ *be a matrix of size* $m \times n$ *with* $m \leq n$, $(n \geq 2)$ *whose columns are normalized so that* $\|\mathbf{a}_i\| = 1$ *for all* i. *Then the coherence of* $\mathbf{A}$ *satisfies*

$$\sqrt{\frac{n-m}{m(n-1)}} \leq \mu(\mathbf{A}) \leq 1 \tag{31}$$

lower bound: **Welch bound**

Coherence and the Spark

Lemma

For any matrix $\mathbf{A}$

$$spark(\mathbf{A}) \geq 1 + \frac{1}{\mu(\mathbf{A})} \tag{35}$$

Unique recovery on Σ_k

$$\text{spark}(\mathbf{A}) > 2k$$

Uniqueness via coherence

Theorem (Uniqueness via coherence)

If

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{A})}\right) \tag{36}$$

then for each measurement vector $\mathbf{y} \in \mathbb{R}^m$ *there exists at most one signal* $\mathbf{x} \in \Sigma_k$ *such that* $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Theorem

Let $\mathbf{A}$ *be an* $m \times n$ *matrix that satisfies the RIP of order* $2k$ *with constant* $\delta \in (0, 1/2]$. *Then*

$$m \geq Ck \log\left(\frac{n}{k}\right) \quad (37)$$

where $C = (1/2)\log(\sqrt{24} + 1) \approx 0.28$

Johnson-Lindenstrauss lemma

CS Where we are?

Given $\mathbf{y}$ and $\mathbf{A}$ find $\mathbf{x} \in \Sigma_k$ such that

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad (43)$$

What we have?

- Conditions = Unique $\mathbf{x}$.
  - Spark
  - NSP
  - RIP
  - Coherence
- How to design $\mathbf{A}$

How to get $\mathbf{x}$?

Outline

Recovery algorithms

The problem can be formulated as

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \quad \text{s.t} \quad \mathbf{x} \in \mathcal{B}(\mathbf{y}) \qquad (2)$$

- Noise free recovery: $\mathcal{B}(\mathbf{y}) = \{\mathbf{x} : \mathbf{Ax} = \mathbf{y}\}$
- Noise: $\mathcal{B}(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{Ax} - \mathbf{y}\| \leq \varepsilon\}$

If $\mathbf{x}$ is represented in a basis $\mathbf{\Phi}$ such as $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\theta}$, the the problem is written as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\theta\|_0 \quad \text{s.t.} \quad \boldsymbol{\theta} \in \mathcal{B}(\mathbf{y}) \tag{3}$$

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0$$

$$\text{s.t} \quad \mathbf{x} \in \mathcal{B}(\mathbf{y})$$

- Check for different values of k
- Solve the problem for all $|\Lambda| = k$

$$\mathbf{y} - \mathbf{A}_\Lambda \mathbf{x}_\Lambda$$

Expose extremely high computational cost.

ℓ_1−recovery

Relaxation

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{s.t} \quad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{4}$$

Computationally feasible.

Can formulated as a (Linear programming) LP problem: *Basis Pursuit*

In the presence of noise

$$\mathcal{B}(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon\} \tag{5}$$

It is possible to get a Lagrangian relaxation as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \tag{6}$$

known as basis pursuit denoising

Example(BP)

$$\mathbf{A} = \frac{1}{\sqrt{4.44}} \begin{bmatrix} 1.2 & -1 & -1.2 & 1 & -1 \\ -1 & 1 & -1 & 1.2 & 1.2 \\ 1 & 1.2 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 0.2 \\ 0 \\ 2.2 \\ 0 \end{bmatrix}, \quad \text{with} \quad \mathbf{x} = \begin{bmatrix} \sqrt{4.44} \\ \sqrt{4.44} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Note: $\sqrt{4.44} \approx 2.1071307$

Solution

Using CVX solve ℓ_1 relaxed version.

$$\hat{\mathbf{x}} = \begin{bmatrix} 2.1071 \\ 2.1071 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Because the computational cost

It is clear why, replacing

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0$$

$$\text{s.t} \quad \mathbf{x} \in \mathcal{B}(\mathbf{y})$$

by

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \quad \text{s.t} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

is convenient

Not so trivial to see how the solution of ℓ_1-relaxation problem is an approximate solution to the original problem $\ell_0 - problem$.

The ℓ_1 Norm and Sparsity

- The ℓ_0 norm is defined by: $\|x\|_0 = \#\{i : x(i) \neq 0\}$
  Sparsity of x is measured by its number of non-zero elements.

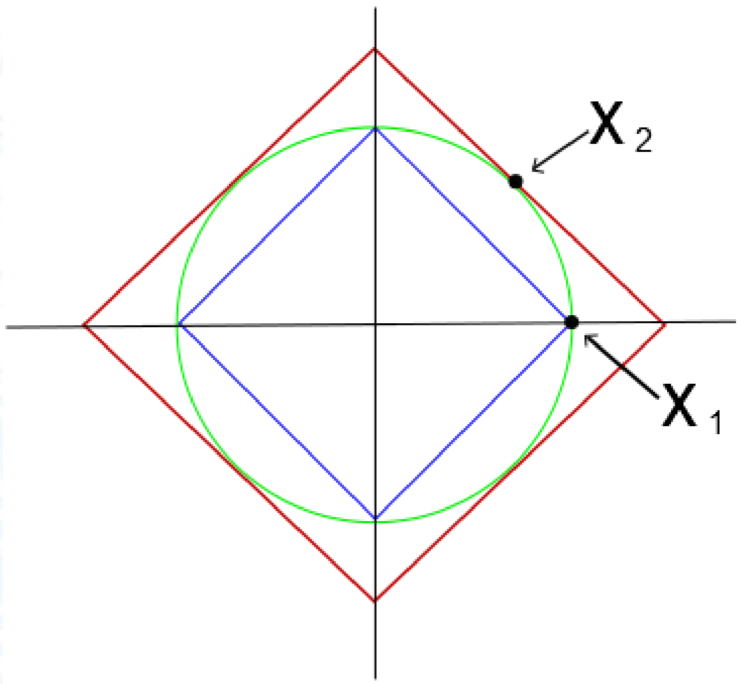
- The ℓ_1 norm is defined by: $\|x\|_1 = \sum_i |x(i)|$
  ℓ_1 norm has two key properties:
  - Robust data fitting
  - Sparsity inducing norm

- The ℓ_2 norm is defined by: $\|x\|_2 = (\sum_i |x(i)|^2)^{1/2}$
  ℓ_2 norm is not effective in measuring *sparsity* of x

Why ℓ_1 Norm Promotes Sparsity?

Given two N-dimensional signals:

- $x_1 = (1, 0, ..., 0) \rightarrow$ "Spike" signal
- $x_2 = (1/\sqrt{N}, 1/\sqrt{N}, ..., 1/\sqrt{N}) \rightarrow$ "Comb" signal



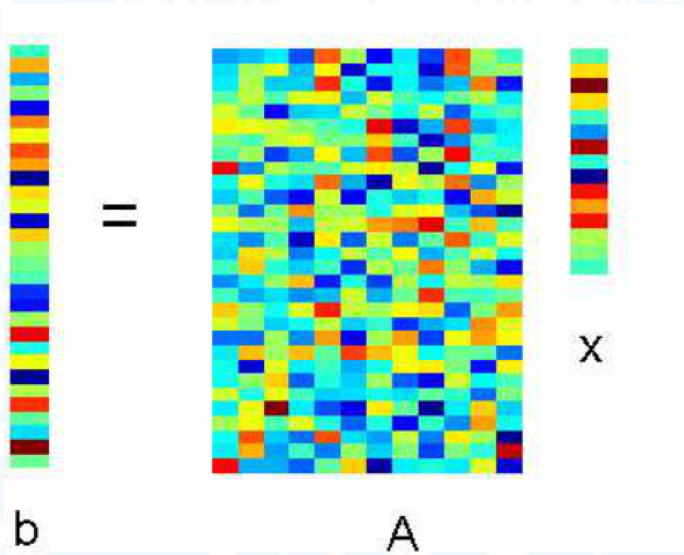
- x_1 and x_2 have the same ℓ_2 norm: $\|x_1\|_2 = 1$ and $\|x_2\|_2 = 1$.
- However, $\|x_1\|_1 = 1$ and $\|x_2\|_1 = \sqrt{N}$.

ℓ_1 Norm in Regression

- Linear regression is widely used in science and engineering.

$$\text{Given} \quad A \in R^{m \times n} \quad \text{and} \quad b \in R^m; \;\; m > n$$

$$\text{Find} \quad x \;\; \text{s.t.} \;\; b = Ax \qquad \text{(overdetermined)}$$



ℓ_1 Norm Regression

Two approaches:

- Minimize the ℓ_2 norm of the residuals

$$\min_{x \in R^n} \|b - Ax\|_2$$

The ℓ_2 norm penalizes large residuals

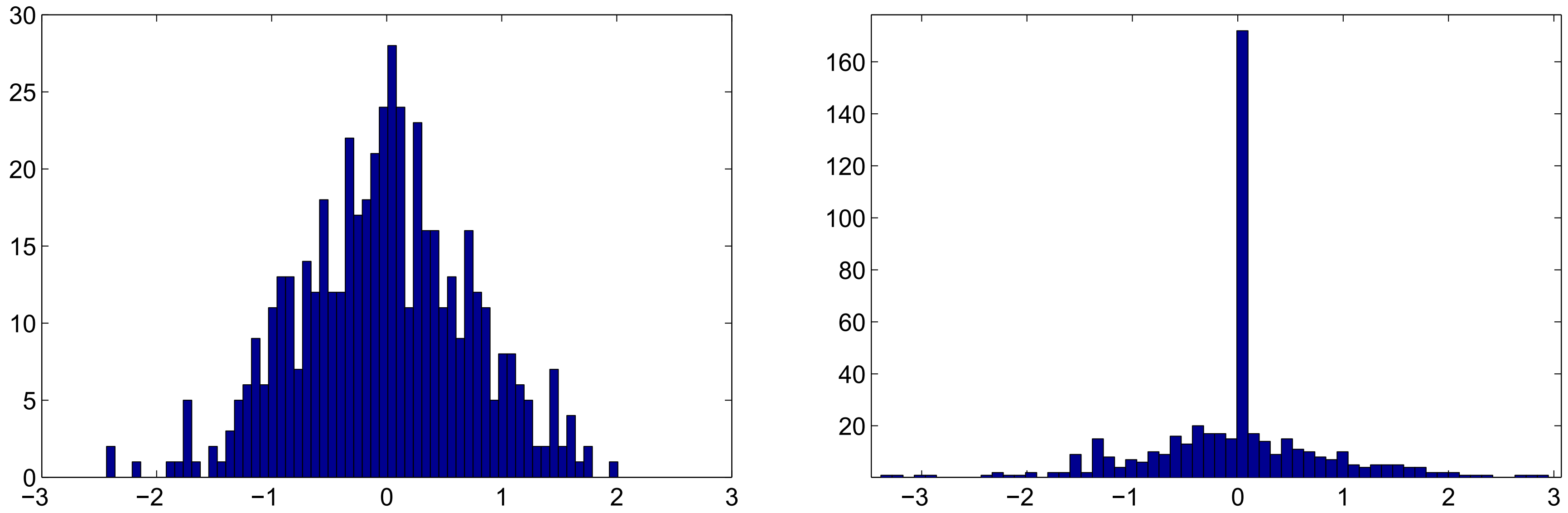
- Minimizes the ℓ_1 norm of the residuals

$$\min_{x \in R^n} \|b - Ax\|_1$$

The ℓ_1 norm puts much more weight on small residuals

ℓ_1 Norm Regression

$m = 500$, $n = 150$. $A = randn(m, n)$ and $b = randn(m, 1)$



ℓ_2 Residuals

ℓ_1 Residuals

Greedy Algorithms

Greedy Pursuits

Built iteratively an estimate of $\mathbf{x}$ starting by $\mathbf{0}$ and iteratively add new components. For each iteration the nonzero components of $\mathbf{x}$ are optimized.

Thresholding algorithms

Built iteratively an estimate of $\mathbf{x}$. For each iteration a subset of nonzero components of $\mathbf{x}$ are selected, while other components are removed (make their valus 0).

Greedy Algorithms

Greedy Pursuits

- Matching Pursuit (MP).
- Orthogonal Matching Pursuit (OMP).

Thresholding algorithms

- Compressive Sampling Matching Pursuit (CoSaMP).
- Iterative Hard Thresholding (IHT).

Intuition

OMP

- Find the column of $\mathbf{A}$ most correlated with $\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$
- Determine the support
- **Update all coefficients over the support.**

MP

- Find the column of $\mathbf{A}$ most correlated with $\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$
- Determine the support
- **Update the coefficient with the related column.**

Matching Pursuit (MP)

$$\min_{i,x} \|\mathbf{y} - \mathbf{a}_i x\|^2$$

$$i = \arg\max_j \frac{(\mathbf{a}_j^T \mathbf{y})^2}{\|\mathbf{a}_j\|_2^2}$$

$$x = \frac{\mathbf{a}_i^T \mathbf{y}}{\|\mathbf{a}_i\|_2^2}$$

MP update

with $\mathbf{r}_0 = \mathbf{y}$

$$\mathbf{r}_\ell = \mathbf{r}_{\ell-1} - \frac{\mathbf{a}_i^T \mathbf{r}_{\ell-1}}{\|\mathbf{a}_i\|_2^2} \mathbf{a}_i$$

$$\hat{\mathbf{x}}_\ell \leftarrow \hat{\mathbf{x}}_{\ell-1}$$

$$\hat{\mathbf{x}}_\ell\big|_i \leftarrow \frac{\mathbf{a}_i^T \mathbf{r}_{\ell-1}}{\|\mathbf{a}_i\|_2^2}$$

Iterative Hard Thresholding (IHT) Variant of CoSaMP

IHT update

$$\hat{\mathbf{x}}_i = \mathcal{T}(\hat{\mathbf{x}}_{i-1} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{i-1}), k)$$

Example

Comparison

- $\ell_1 - optimization$
- OMP
- MP
- IHT
- CoSaMP

Error versus sparsity level

- $\mathbf{A}$ is of dimension 512×1024
- $\mathbf{A}$: Entries taken from $\mathcal{N}(0, 1)$
- $\mathbf{A}$: Random partial Fourier matrix.
- For each value of sparsity k, a $k-$sparse vector of dimension $n \times 1$ is built.
- The nonzero locations in $\mathbf{x}$ are selected at random, and the values are taken from $\mathcal{N}(0, 1)$.

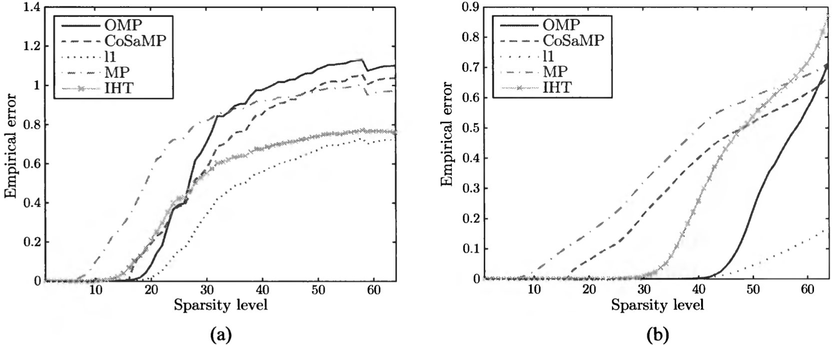


Figure: Comparison of the performance of different reconstruction algorithms in terms of the sparsity level. (a) Gaussian matrix (b) Random partial Fourier matrix

Recovery guarantees

Guarantees

- RIP-based.
- Coherence based.

Pessimistic

Recovery is possible for much more relaxed versions than those stated by some Theoretic results.

Signal recovery in noise

Theorem (RIP-based noisy ℓ_1 recovery)

Suppose that $\mathbf{A}$ *satisfies the RIP of order* $2k$ *with* $\delta_{2k} < \sqrt{2} - 1$*, and let* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *where* $\|\mathbf{e}\|_2 \leq \varepsilon$*. Then, when* $\mathcal{B}(\mathbf{y}) = \{\mathbf{z} : \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2\}$*, the solution* $\hat{\mathbf{x}}$ *to*

$$\hat{\mathbf{x}} = arg\min \|\mathbf{x}\|_1 \quad s.t \quad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{11}$$

obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}} + C_2\varepsilon \tag{12}$$

RIP Guarantees

- Difficult to calculate the RIP for large size matrices

Coherence Guarantees

Exploit advantages of using coherence for structured matrices.

Theorem (Coherence-based ℓ_1 recovery with bounded noise)

Suppose that $\mathbf{A}$ *has coherence* μ *and that* $\mathbf{x} \in \Sigma_k$ *with* $k < (1/\mu + 1)/4$. *Furthermore, suppose that we obtain measurements of th form* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *with* $\gamma = \|\mathbf{e}\|_2$. *Then when* $\mathcal{B}(\mathbf{y}) = \{\mathbf{z} : \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2\}$ *with* $\varepsilon > \gamma$, *the solution* $\hat{\mathbf{x}}$ *to*

$$\hat{\mathbf{x}} = arg \min \|\mathbf{x}\|_1 \quad s.t \quad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{14}$$

obeys

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{\gamma + \varepsilon}{\sqrt{1 - \mu(4k-1)}} \tag{15}$$

Guarantees on greedy methods

Theorem (RIP-based OMP recovery)

Suppose that $\mathbf{A}$ *satisfies the RIP of order* $k+1$ *with* $\delta_{k+1} < 1/(3\sqrt{k})$ *and let* $\mathbf{y} = \mathbf{A}\mathbf{x}$. *Then OMP can recover a* $k-$*sparse signal exactly in* k *iterations.*

Theorem (RIP-based thresholding recovery)

Suppose that $\mathbf{A}$ *satisfies the RIP or order* ck *with constant* δ *and let* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *where* $\|\mathbf{e}\|_2 \leq \varepsilon$*. Then the outputs* $\hat{\mathbf{x}}$ *of the CoSaMP, subspace pursuit, and IHT algorithms obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_1 \sigma_k(\mathbf{x})_2 + C_2 \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}} + C_3 \varepsilon \qquad (19)$$

The requirements on the parameters c, δ *of the RIP and the values of* C_1, C_2 *and* C_3 *are specific to each algorithm.*