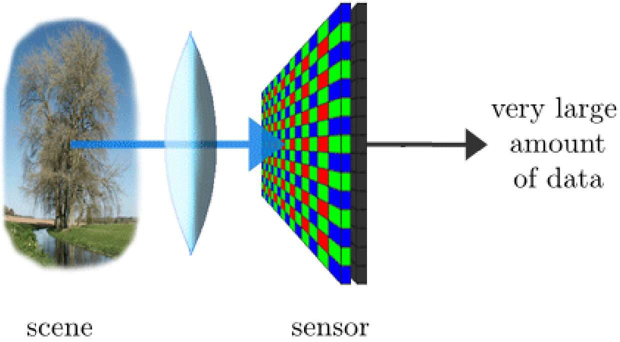
An Introduction to Compressive Sensing

Princeton, August 10, 2010

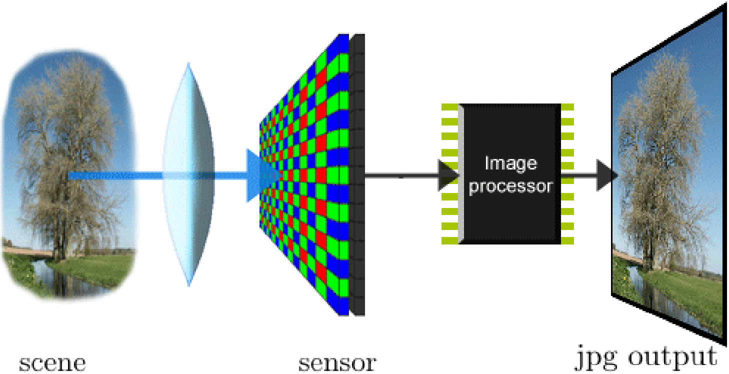
Adriana Schulz
Luiz Velho
Eduardo A. B. da Silva

Your Digital Camera



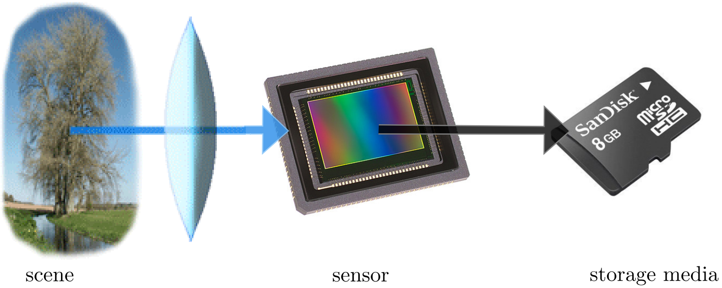
Example: 8 mega pixels - each image has 24Mbytes!

Your Digital Camera



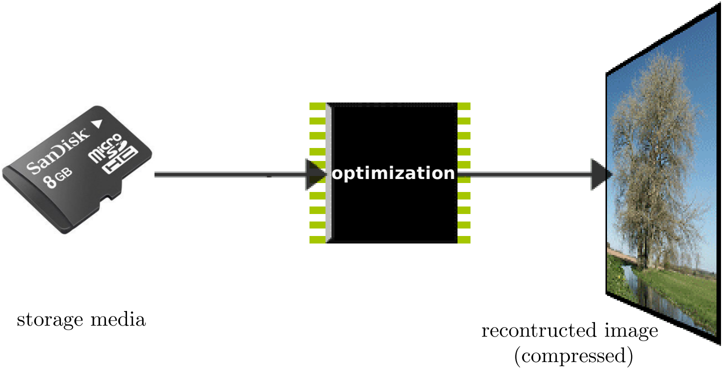
More than 90% of the data is discarded!

CS to the Rescue



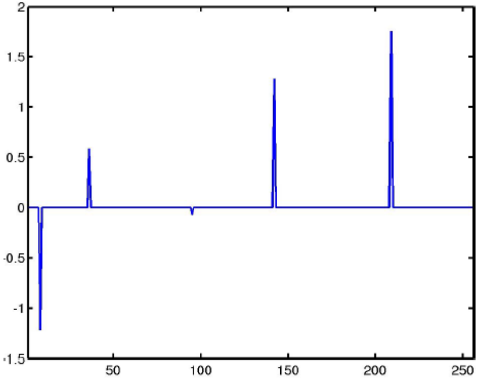
Capture only the necessary information!

CS to the Rescue

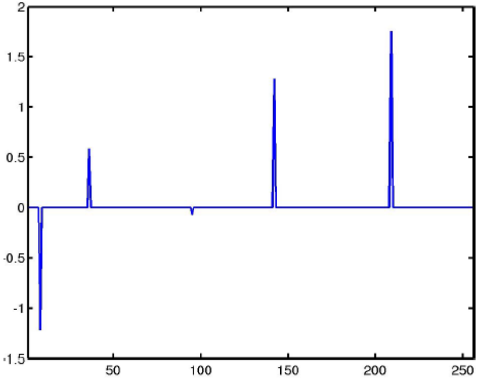


Reconstruct with convex optimization!

Sparse Signals



Sparse Signals



We don’t know the significant coefficients ...

The Twist!

point sampling $\times$ measurements

Point Sampling:

$$y_1 = \langle x, e_1 \rangle\,,\ \ y_2 = \langle x, e_2 \rangle\,,\ \ \ldots\ \ ,\ \ y_M = \langle x, e_N \rangle$$

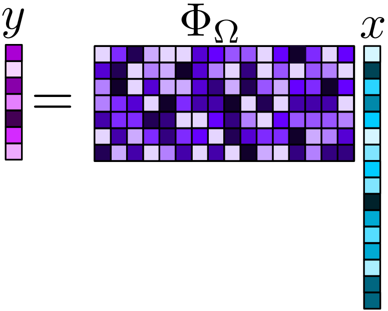
where N is size of the signal.

Now, we take different measurements:

$$y_1 = \langle x, \phi_1 \rangle\,,\ \ y_2 = \langle x, \phi_2 \rangle\,,\ \ \ldots\ \ ,\ \ y_M = \langle x, \phi_M \rangle$$

where $M \ll N$ is the number of measurements.

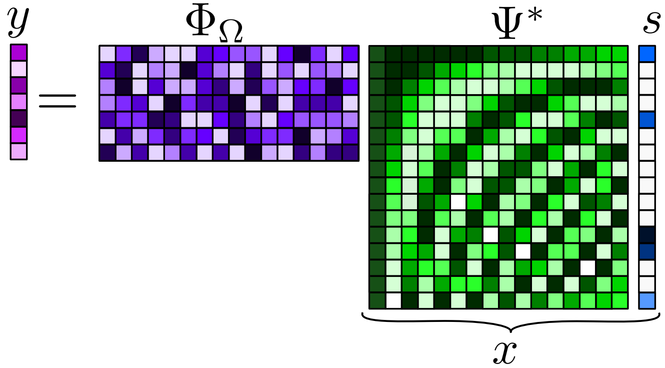
The Algebraic Problem



$$\Phi_\Omega = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_M \end{bmatrix}$$

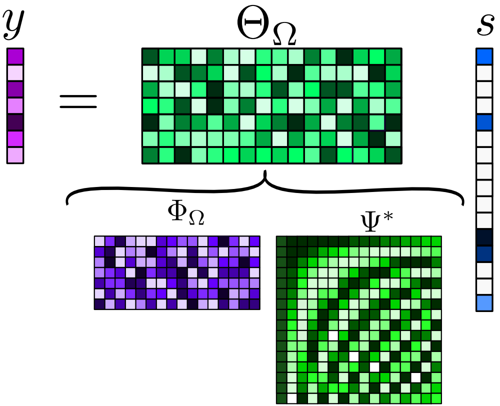
- This problem is ill-conditioned!

The Algebraic Problem



- What if there exists a domain in which x is sparse? ($s = \Psi x$)

The Algebraic Problem



- $\Theta_\Omega = \Phi_\Omega \Psi^*$
- measurements $y = \Theta_\Omega s$

The solution we want is:

$$\min_{s} \|s\|_{l_0} \quad \text{subject to} \quad \Theta_{\Omega} s = y$$

where the l_0-norm is:

$$\|\alpha\|_{l_0} = \sharp \{ i : \alpha(i) \neq 0 \}$$

The solution we want is:

$$\min_{s} \|s\|_{l_0} \quad \text{subject to} \quad \Theta_{\Omega} s = y$$

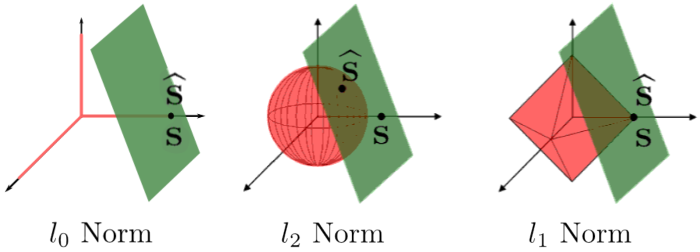
where the l_0-norm is:

$$\|\alpha\|_{l_0} = \sharp \{i : \alpha(i) \neq 0\}$$

NP-hard Problem!!!

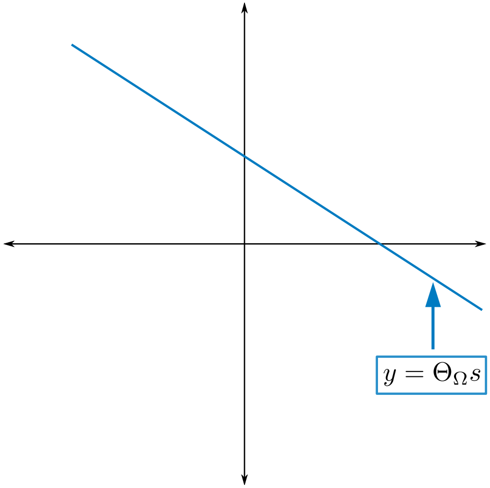
- How do we make this problem computationally tractable?

l_1 Magic!

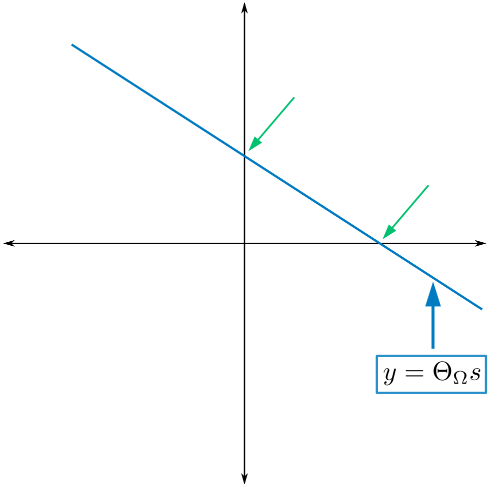


$\hat{\mathbf{s}}$
$\mathbf{s}$
l_0 Norm
$\hat{\mathbf{s}}$
$\mathbf{s}$
l_2 Norm
$\hat{\mathbf{s}}$
$\mathbf{s}$
l_1 Norm

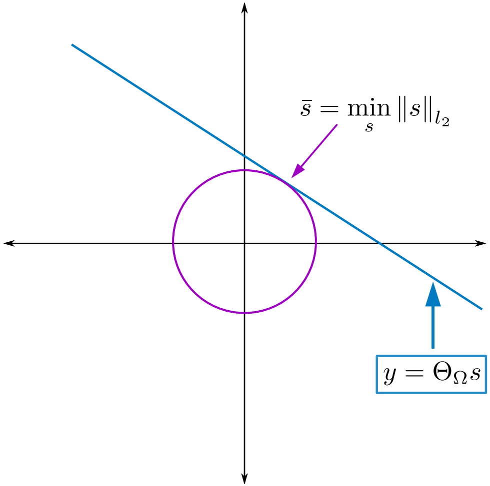
The l_1 norm and Sparsity



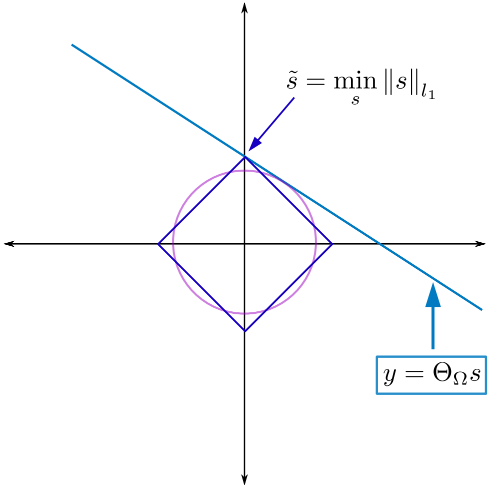
The l_1 norm and Sparsity



The l_1 norm and Sparsity



The l_1 norm and Sparsity



Reconstruction Algorithm

$$\min_{s} \|s\|_{l_1} \quad \text{subject to} \quad \Theta_\Omega s = y$$

where

$$\Theta_\Omega = \Phi_\Omega \Psi^*$$

This seems like a good procedure

- When does it work?
- What do we need to assume about the sensing matrix Θ_Ω?
- And the number of samples?
- What kind of results can we guarantee?

Fourier Sampling Theorem

On Monday:

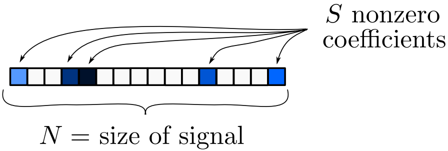
- MRI model
- Samples in the frequency domain

Theorem

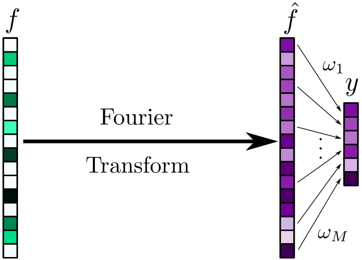
- *$s \in \mathbb{R}^N$ is S-sparse*
- *M Fourier coefficients are randomly selected*

$$M \gtrsim S \cdot \log N$$

We can reconstruct s minimizing the l_1-norm.



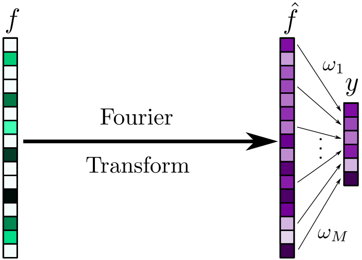
Fourier Sampling Theorem



$$min\|f\|_{l_1} \quad \text{s.t.} \quad \hat{f}(\omega_m) = y_m$$

$$\Omega = \{\omega_1, \ldots, \omega_M\}$$
$$|\Omega| = M$$

Fourier Sampling Theorem

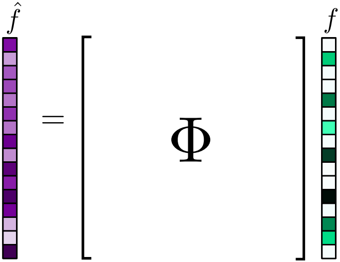


$$min\|f\|_{l_1} \quad \text{s.t.} \quad \hat{f}(\omega_m) = y_m$$

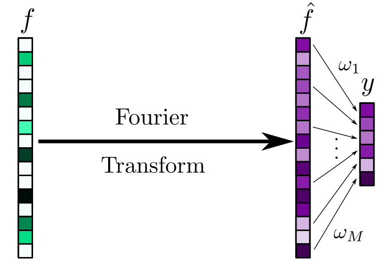
$\Omega = \{\omega_1, \ldots, \omega_M\}$

$|\Omega| = M$

New Notation: Φ = Fourier Transform Matrix



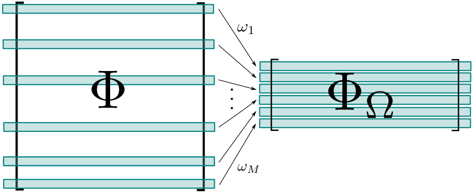
Fourier Sampling Theorem



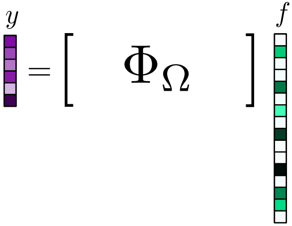
$$min\|f\|_{l_1} \quad \text{s.t.} \quad \hat{f}(\omega_m) = y_m$$

$$\Omega = \{\omega_1, \ldots, \omega_M\}$$
$$|\Omega| = M$$

New Notation: Φ = Fourier Transform Matrix



$$y = \hat{f}_{|\Omega} = (\Phi f)_{|\Omega} = \Phi_\Omega f$$



$$min\|f\|_{l_1} \quad \text{s.t.} \quad \Phi_\Omega f = y$$

Fourier Sampling Theorem

Theorem

- *$s \in \mathbb{R}^N$ is S-sparse*
- *Φ is the Fourier Transform Matrix of size $N \times N$*
- *We restrict Φ to a random set Ω of size M such that*

$$M \gtrsim S \cdot \log N$$

We can recover s by solving the convex optimization problem

$$\min_s \|s\|_{l_1} \quad \text{subject to} \quad \Phi_\Omega s = y$$

A first guarantee: if measurements are taken in the Fourier domain CS works!
A question: what is special about the Fourier domain?

Uncertainty Principles

A function and its Fourier transform cannot both be highly concentrated!

Theorem

If f is zero outside a measurable set T_f and its Fourier transform $\hat{f}$ is zero outside a measurable set Ω_f, then

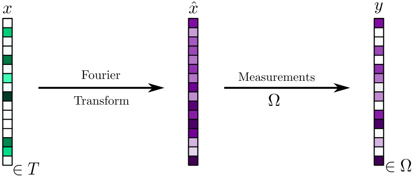
$$|T_f| \cdot |\Omega_f| \geq 1$$

Another Twist

Another Twist

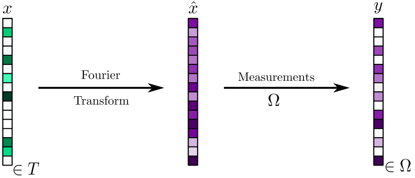
$\Longrightarrow$ allows good results!

Uncertainty Principles



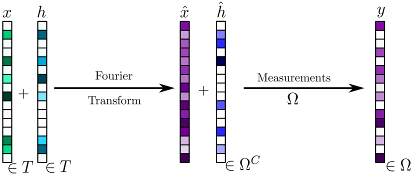
$\Rightarrow$ If $|T||\Omega^c| < 1 \Rightarrow$ we can recover x.

Uncertainty Principles

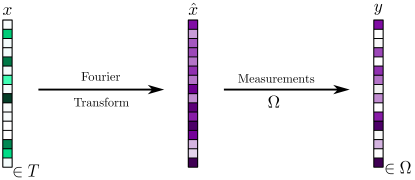


$\Rightarrow$ If $|T||\Omega^c| < 1 \Rightarrow$ we can recover x.

Problem:

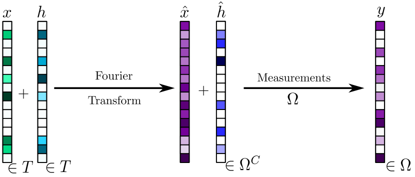


Uncertainty Principles



$\Rightarrow$ If $|T||\Omega^c| < 1 \Rightarrow$ we can recover x.

Problem:



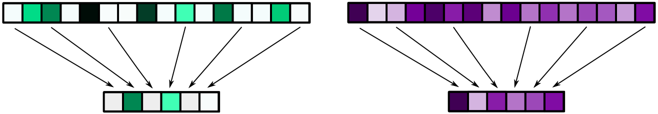
There is no such signal!

If it did the uncertainty principle would require $|T||\Omega^c| \geq 1$.

Extension of the Fourier Sampling Theorem

- It may be difficult to take samples in the Frequency domain.
- The signal may not be sparse in the time domain, but in a different Ψ domain.

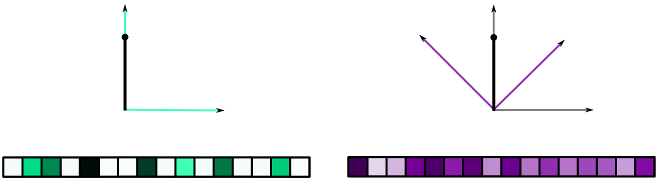
$\Rightarrow$ Other possibilities for Φ e Ψ



Coherence

Definition (Coherence between Ψ and Φ)

$$\mu(\Phi,\Psi) = \sqrt{N}\max_{i,j} |\langle \phi_i, \psi_j \rangle| \quad , \quad \|\phi_i\|_{l_2} \quad \|\psi_i\|_{l_2} = 1$$

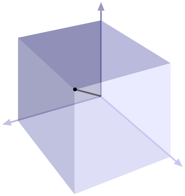


- $\mu(\Phi,\Psi)$ measures the minimum angle between ϕ_i and ψ_j
- if we look at the waveforms as vectors in R^N, then high incoherencies mean that these vectors are far apart

New Sampling Theorem

$$1 \leq \mu(\Phi, \Psi) \leq \sqrt{N}$$

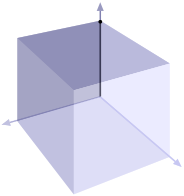
Maximally Incoherent



$\psi_1 = (1,0,0)$ $\psi_2 = (0,1,0)$ $\psi_3 = (0,0,1)$
$\phi = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$

$$\begin{cases} |\langle \psi_1, \phi \rangle| = \frac{1}{\sqrt{3}} \\ |\langle \psi_2, \phi \rangle| = \frac{1}{\sqrt{3}} \\ |\langle \psi_3, \phi \rangle| = \frac{1}{\sqrt{3}} \end{cases} \Rightarrow \mu = \frac{1}{\sqrt{3}} \cdot \sqrt{3} = 1$$

Maximally Coherent



$\psi_1 = (1,0,0)$ $\psi_2 = (0,1,0)$ $\psi_3 = (0,0,1)$
$\phi = (1,0,0)$

$$\begin{cases} |\langle \psi_1, \phi \rangle| = 1 \\ |\langle \psi_2, \phi \rangle| = 0 \\ |\langle \psi_3, \phi \rangle| = 0 \end{cases} \Rightarrow \mu = 1 \cdot \sqrt{3} = \sqrt{N}$$

New Sampling Theorem

Theorem

Reconstruction is exact if

$$M \gtrsim S \cdot \mu^2(\Phi, \Psi) \cdot \log N$$

- time and frequency are maximally incoherent (the Fourier basis $\psi_k(t) = \frac{1}{\sqrt{N}} e^{\frac{2\pi jk}{N}}$ and the canonical basis $\phi_k(t) = \delta(t-k)$ yield $\mu = 1$)
- when bases are maximally coherent **you have to see all samples**!

Restricted Isometry Property (RIP)

Definition (Restricted Isometry Constant)

For each integer $S = 1, 2, \dots, N$ we define the S-restricted isometry constant δ_S of a matrix Θ_Ω as the smallest number such that

$$(1 - \delta_S)\|s\|_{l_2}^2 \leq \|\Theta_\Omega s\|_{l_2}^2 \leq (1 + \delta_S)\|s\|_{l_2}^2$$

for all S-sparse vectors.

RIP $\Rightarrow$ a property of Θ_Ω related to the existence and limitation of δ_S

- A restricted isometry

Restricted Isometry Property (RIP)

Definition (Restricted Isometry Constant)

For each integer $S = 1, 2, \ldots, N$ we define the S-restricted isometry constant δ_S of a matrix Θ_Ω as the smallest number such that

$$(1 - \delta_S)\|s\|_{l_2}^2 \leq \|\Theta_\Omega s\|_{l_2}^2 \leq (1 + \delta_S)\|s\|_{l_2}^2$$

for all S-sparse vectors.

RIP $\Rightarrow$ a property of Θ_Ω related to the existence and limitation of δ_S

- A restricted isometry
- An Uncertainty Principle

  s is sparse $\Rightarrow$ Θs cannot be concentrated in Ω^c

Restricted Isometry Property (RIP)

Definition (Restricted Isometry Constant)

For each integer $S = 1, 2, \ldots, N$ we define the S-restricted isometry constant δ_S of a matrix Θ_Ω as the smallest number such that

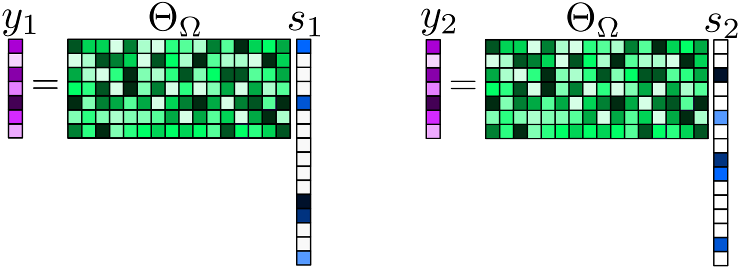
$$(1 - \delta_S)\|s\|_{l_2}^2 \leq \|\Theta_\Omega s\|_{l_2}^2 \leq (1 + \delta_S)\|s\|_{l_2}^2$$

for all S-sparse vectors.

RIP $\Rightarrow$ a property of Θ_Ω related to the existence and limitation of δ_S

- A restricted isometry
- An Uncertainty Principle
- In truth, the RIP guaranties that the sparsest solution is unique!

Unique Solution



$$y_1 \neq y_2$$

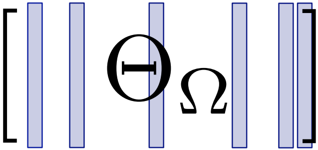
Unique Solution

We want to recover s from $y = \Theta_\Omega s$

If the columns of Θ_Ω are l.d., $\exists s_a, s_b$ such that

$$y = \Theta_\Omega s_a = \Theta_\Omega s_b$$

The columns of Θ_Ω can't be l.i. because it is a *fat* matrix!



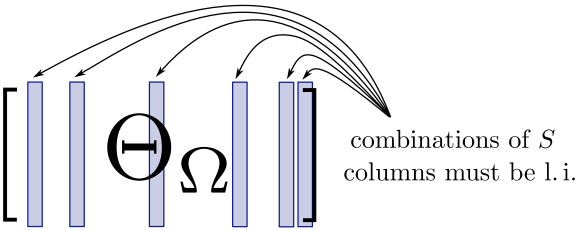
Unique Solution

We want to recover s from $y = \Theta_\Omega s$

If the columns of Θ_Ω are l.d., $\exists s_a, s_b$ such that

$$y = \Theta_\Omega s_a = \Theta_\Omega s_b$$

The columns of Θ_Ω can't be l.i. because it is a *fat* matrix!



Sparsity to the rescue: all that in necessary is that every combination of S columns of Θ be l.i.!

Theorems

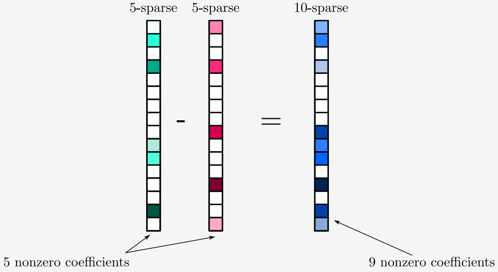
- If $\delta_{2S} < 1$ the solution that maximizes sparsity is unique.

Consider s_1 e s_2 S-sparse such that $\Theta_\Omega s_1 = \Theta_\Omega s_2 = y$.

Let $h = s_1 - s_2$.

$$\Theta_\Omega h = \Theta_\Omega (s_1 - s_2) = \Theta_\Omega s_1 - \Theta_\Omega s_2 = 0.$$

h is $2S$-sparse:



Theorems

- If $\delta_{2S} < 1$ the solution that maximizes sparsity is unique.

Consider s_1 e s_2 S-sparse such that $\Theta_\Omega s_1 = \Theta_\Omega s_2 = y$.

Let $h = s_1 - s_2$.

$$\Theta_\Omega h = \Theta_\Omega(s_1 - s_2) = \Theta_\Omega s_1 - \Theta_\Omega s_2 = 0.$$

Since h is $2S$-sparse, the RIP says that:

$$\underbrace{(1 - \delta_{2S})}_{>0} \|h\|^2 \leq \|\Theta_\Omega h\|^2 = 0$$

therefore,

$$h = 0 \rightarrow s_1 = s_2$$

Theorems

- If $\delta_{2S} < 1$ the solution that maximizes sparsity is unique.
- If $\delta_{2S} < \sqrt{2} - 1$ the solution that minimizes the l_1 norm and the one maximizes sparsity are unique and the same.

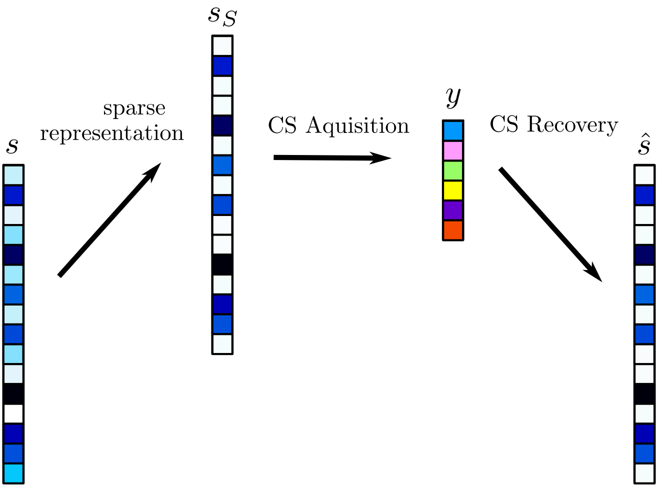
Robust CS

For CS to be suitable for real application it must be robust to two kinds of inaccuracies:

- the signal is not exactly sparse; or
- measurements are corrupted by noise.

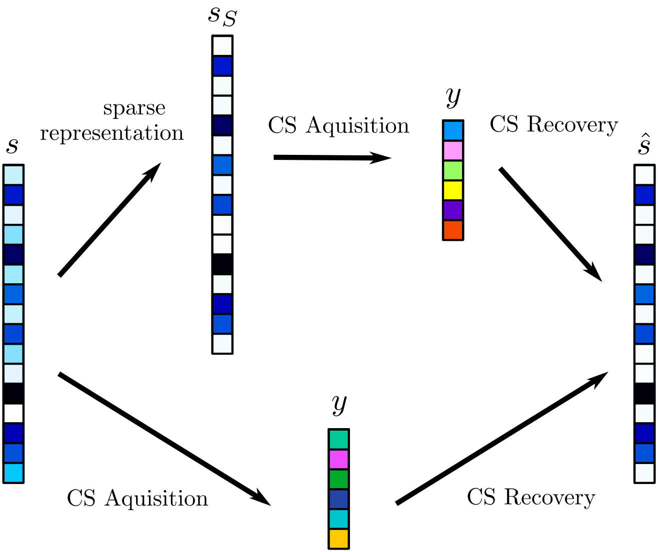
Sparsity Errors

What can we hope for?



Sparsity Errors

What can we hope for?



Result Considering the RIP

Theorem

Assume that s is approximately sparse and let s_S be as defined above. Then if $\delta_{2S} < \sqrt{2} - 1$, the solution $\tilde{s}$ to

$$\tilde{s} = \min_{s} \|s\|_{l_1} \quad \textit{subject to} \quad \Theta_\Omega s = y$$

obeys

$$\|\tilde{s} - s\|_{l_2} \lesssim \frac{1}{\sqrt{S}} \cdot \|s - s_S\|_{l_1}$$

Measurement Errors

Assume that $y = \Theta_\Omega s + n$ where $\|n\|_{l_2} \leq \epsilon$

Constraints:
$\Theta_\Omega s = y$

Measurement Errors

Assume that $y = \Theta_\Omega s + n$ where $\|n\|_{l_2} \leq \epsilon$

Constraints:

$\cancel{\Theta_\Omega s = y} \quad \rightarrow \quad \|\Theta_\Omega s - y\|_{l_2} \leq \epsilon$

Theorem

If $\delta_{2S} < \sqrt{2} - 1$, the solution $\tilde{s}$ to

$$\tilde{s} = \min_s \|s\|_{l_1} \quad \textit{subject to} \quad \|\Theta_\Omega s - y\|_{l_2} \leq \epsilon$$

obeys

$$\|\tilde{s} - s\|_{l_2} \lesssim \frac{1}{\sqrt{S}} \cdot \|s - s_S\|_{l_1} + \epsilon$$

Design of Efficient Sensing Matrices

Given a matrix Θ_Ω, the calculus of δ_S in NP–hard!

Important to determine some measurement ensembles where the RIP holds.

The actual problem: to design a fat sensing matrix Θ_Ω, so that any subset of columns of size S be approximately orthogonal.

$\rightarrow$ deterministic Θ_Ω may be a very difficult task

Randomness re-enters the Picture!

Theorems

Theorem (Gaussian Matrices)

Let the entries of Θ_Ω be i.i.d., Gaussian with mean zero and variance $1/M$. Then the RIP holds with overwhelming probability if

$$M \gtrsim S \cdot log(N/M)$$

Also valid for:

Random Projections: Θ_Ω is a random Gaussian matrix whose rows were orthonormalized.

Binary Matrices: The entries of Θ_Ω be independent taking values $\pm 1/\sqrt{M}$ with equal probability.

General Orthogonal Measurement Ensembles

Theorem

Let Θ be an orthogonal matrix and Θ_Ω be obtained by selecting M rows from Θ uniformly at random. Then the RIP holds with overwhelming probability if

$$M \gtrsim \mu^2 \cdot S \cdot (\log N)^6$$

- Relates coherence and RIP!
- Usefull if signal is sparse in a fixed Ψ: we determine Φ such that $\mu(\Phi, \Psi)$ is small