

# Compressive Sensing and Beyond

Sohail Bahmani

Georgia Tech.



information

measurements

sparse

sensing

basis

matrix

small

image

example

matrices

theory

incoherent

sampling

recovery

random

acquisition

S-sparse

first

sparsity

image

example

efficient

PROCESSING

vector domain

sequence Trans

exact well

obtain

protocols

since

discrete

sample knowledge

Signal

set

minimization

largest

representation

samples

degree

form

probability

reconstruct

devices

obeying

subject

objects

imaging

Fourier

shows

Original exactly

RIP

values

coherence

high

dictionary

waveforms

fact

uniformly fewer

system

use

linear noise rate

Ax

architecture

much holds

norm

far

sense

solution

sense

design as

nearly

representation

degree

form

reconstruct

devices

obeying

subject

objects

imaging

wavelet

size unknown

second

essentially

subsets problem least

via

entries

transform

compression

using

object

section Wakin

acquisition

S-sparse

first

sparsity

image

example

efficient

PROCESSING

vector domain

sequence Trans

exact well

obtain

protocols

since

discrete

sample knowledge

Signal

set

minimization

largest

representation

samples

degree

form

probability

reconstruct

devices

obeying

subject

objects

imaging

wavelet

size unknown

second

essentially

subsets problem least

via

entries

transform

compression

using

object

section Wakin

acquisition

S-sparse

first

sparsity

image

example

efficient

PROCESSING

vector domain

sequence Trans

exact well

obtain

protocols

since

discrete

sample knowledge

Signal

set

minimization

largest

representation

samples

degree

form

probability

reconstruct

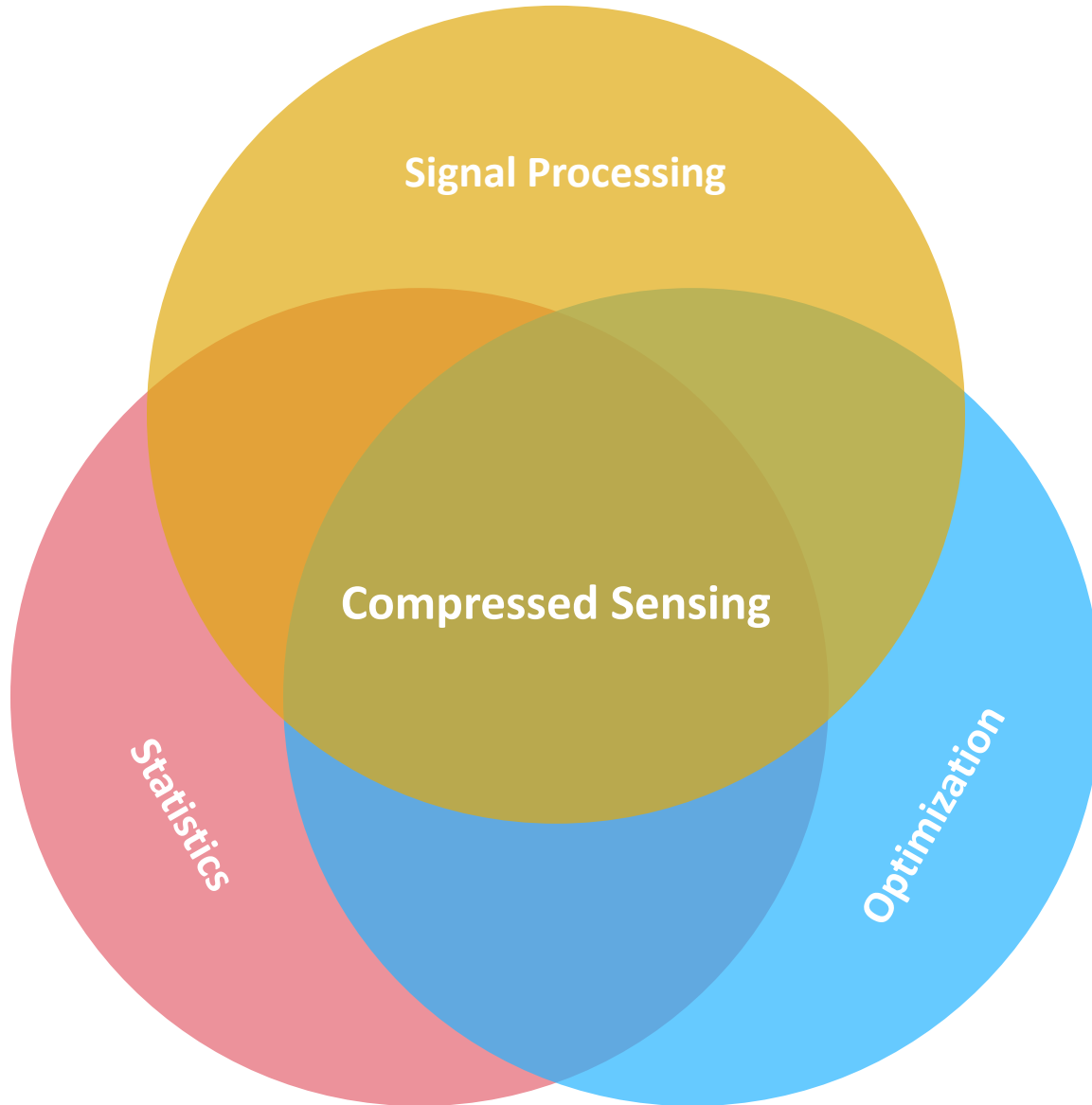
devices

obeying

subject

objects

imaging



Signal Processing

Compressed Sensing

Statistics

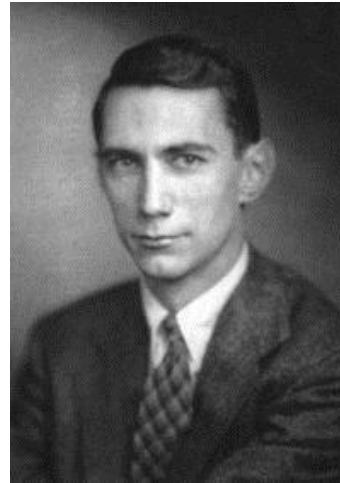
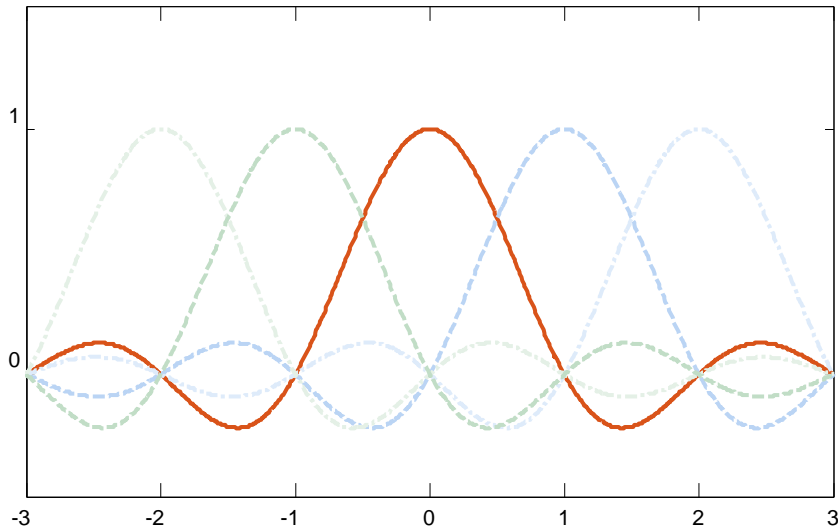
Optimization

# Signal Models

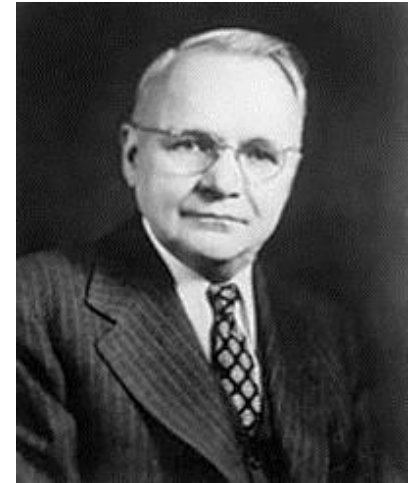
# Classics: bandlimited

## *The Sampling Theorem*

Any signal with bandwidth  $B$  can be recovered from its samples collected *uniformly* at a rate no less than  $f_s = 2B$ .



Claude Shannon



Harry Nyquist

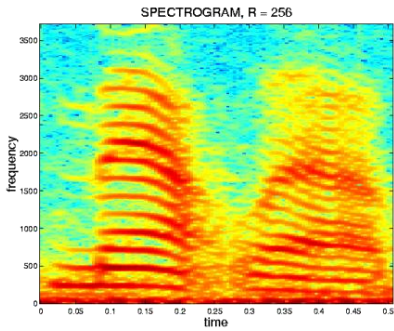
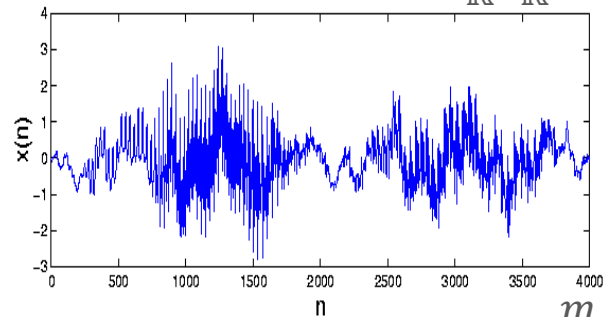
$$x(t) = \sum_{n \in \mathbb{Z}} x_n \text{sinc}(f_s t - n)$$

$$x_n = \langle x(t), \text{sinc}(f_s t - n) \rangle$$

# Classics: time-frequency

$$X(\tau, \omega) = \langle x(t), w(t - \tau)e^{j\omega t} \rangle$$

$$x(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} X(\tau, \omega) e^{j\omega t} d\omega d\tau$$



$$\psi_{m,n}(t) = a^{-\frac{m}{2}} \psi(a^{-m}t - nb)$$

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle x(t), \psi_{m,n}(t) \rangle \psi_{m,n}(t)$$



D Gabor



A. Haar

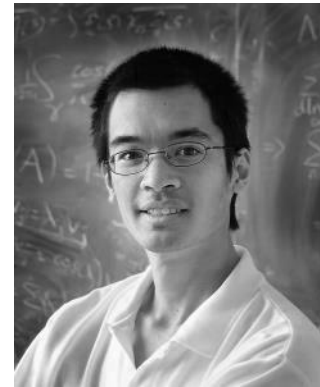
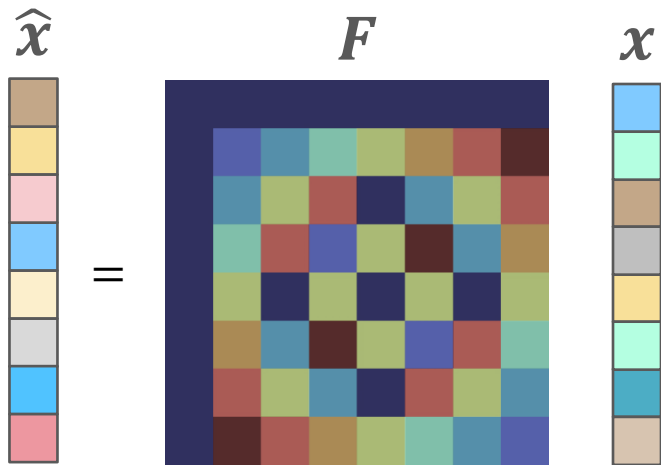


I. Daubechies



R. Coifman

# Sparsity



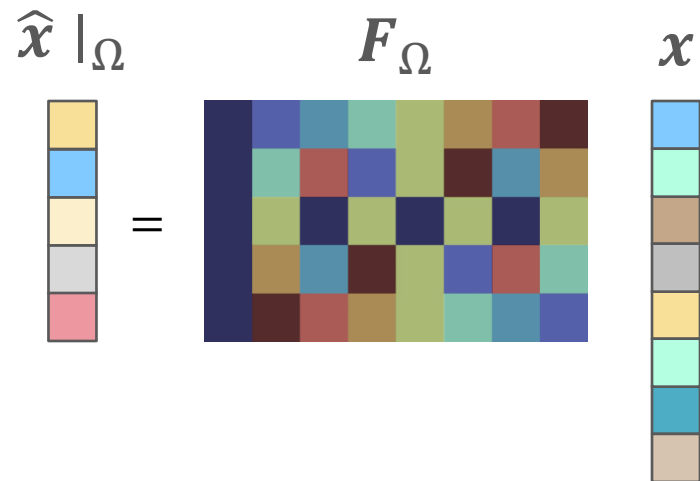
T. Tao



E. Candès



J. Romberg

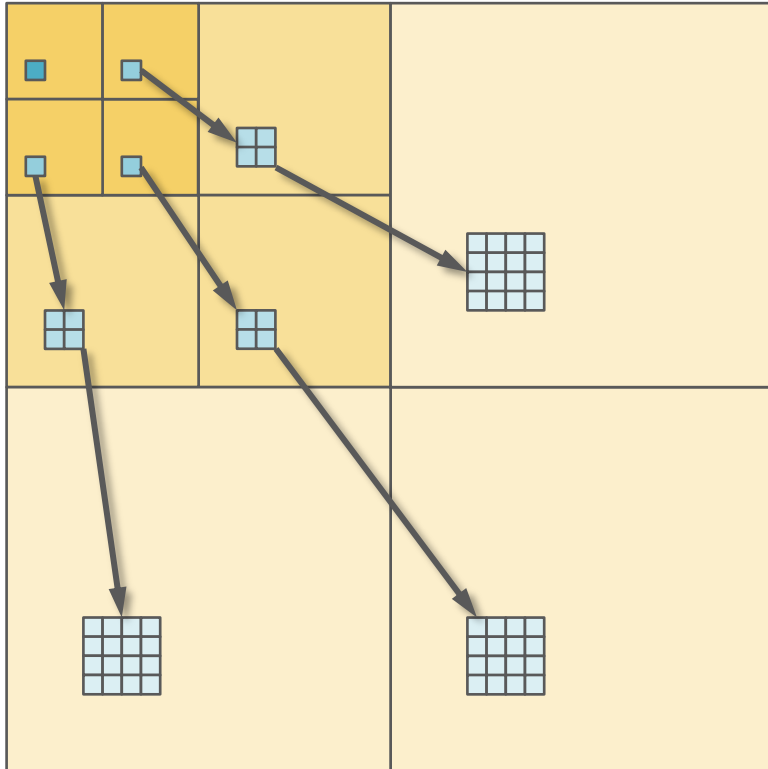


D. Donoho



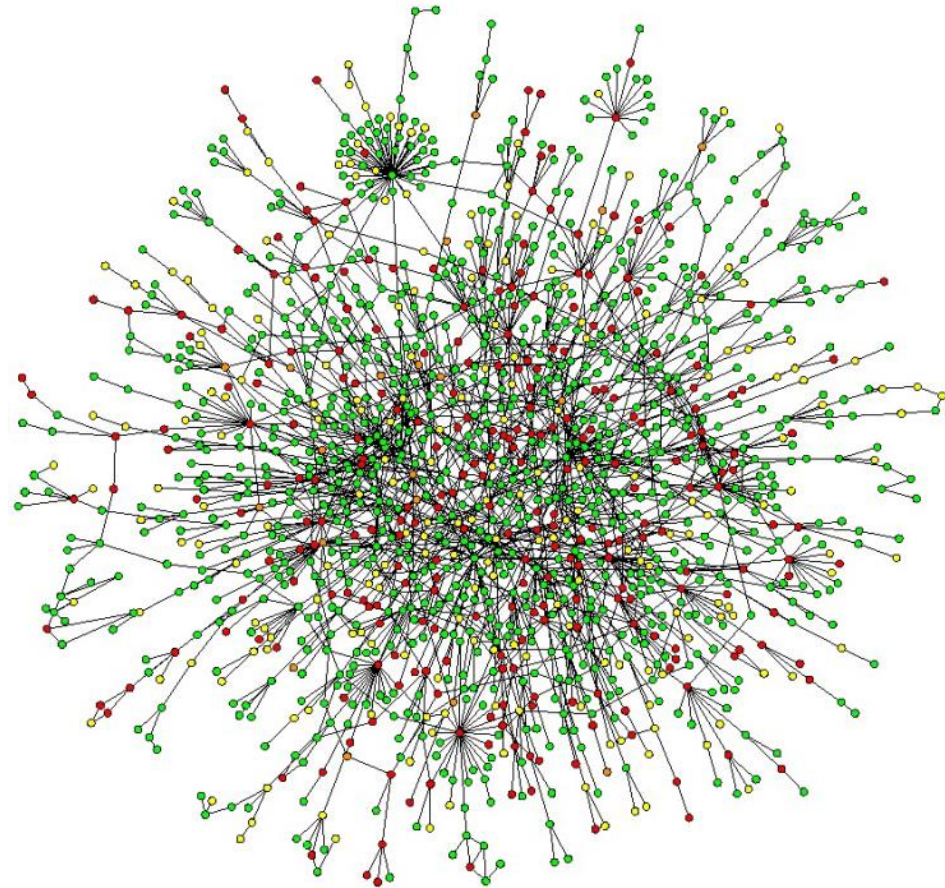
# Structured sparsity

## Tree Structured Models



Discrete Wavelet Transform

## Graph Structured Models



Protein-Protein Interactions  
Gene Regulatory Network



# Low rank

## Model for Images



Figure 1: The MIT logo image. The associated matrix has dimensions  $46 \times 81$  and has rank 5.

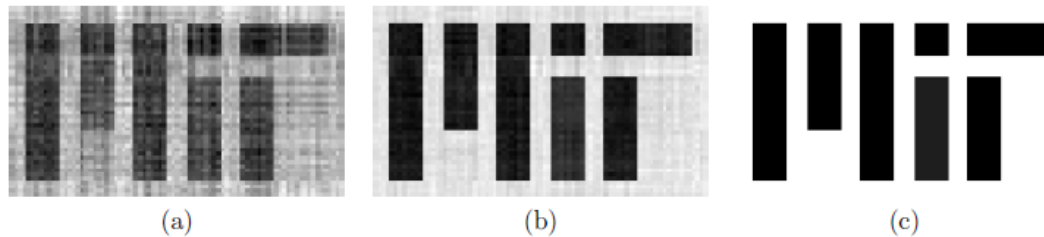


Figure 3: Example recovered images using the Gaussian ensemble. (a) 700 measurements. (b) 1100 measurements (c) 1250 measurements. The total number of pixels is  $46 \times 81 = 3726$ . Note that the error is plotted on a logarithmic scale.

## Other applications

- minimum order linear system realization
- low-rank matrix completion
- low-dimensional Euclidean embedding



B. Recht



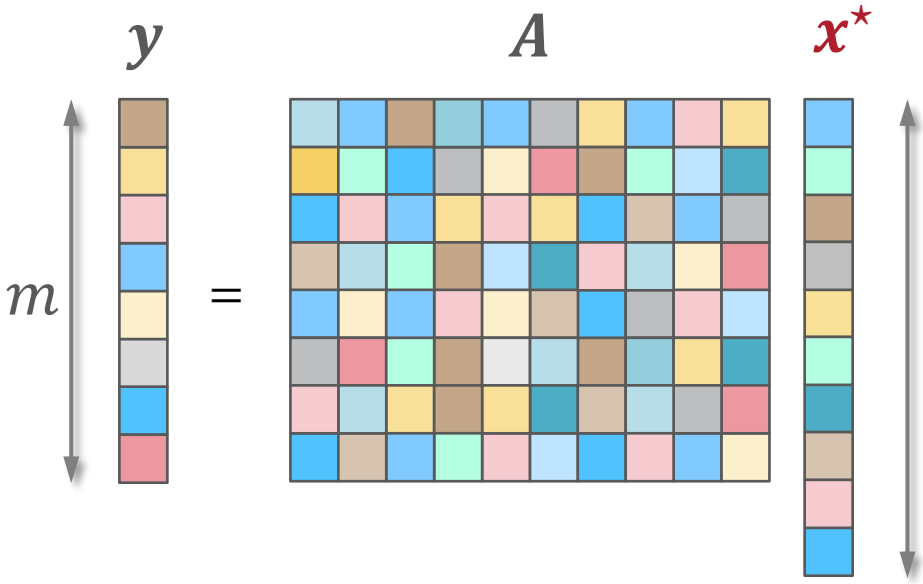
M. Fazel



P. Parrilo

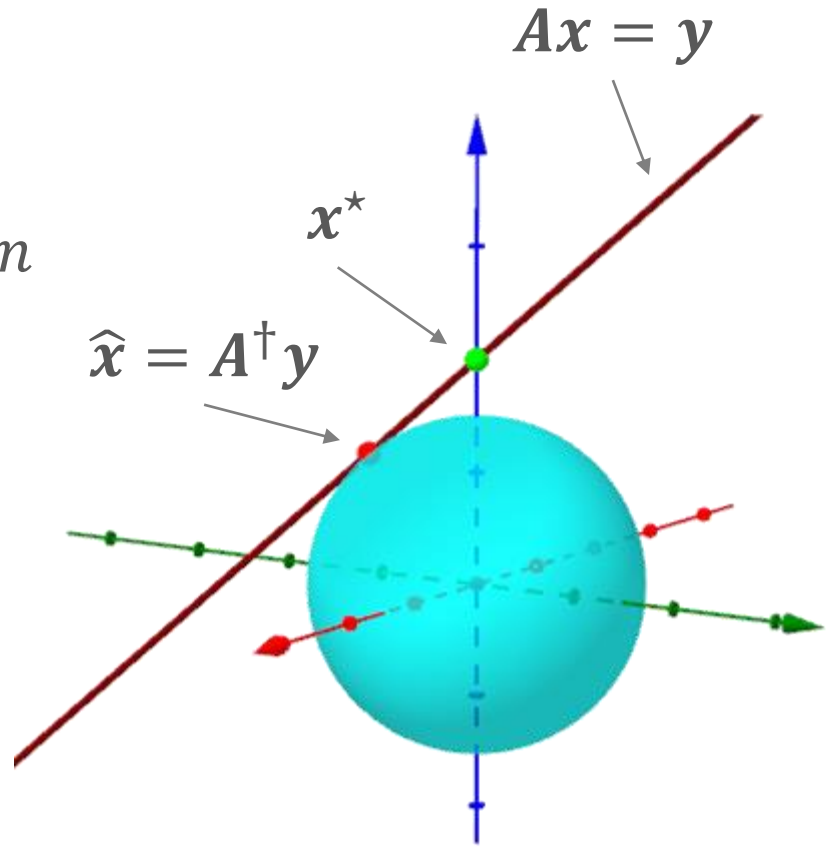
# Compressive Sensing

# Linear inverse problems



$$m < n$$

No unique inverse, but ...



# Sparse solutions

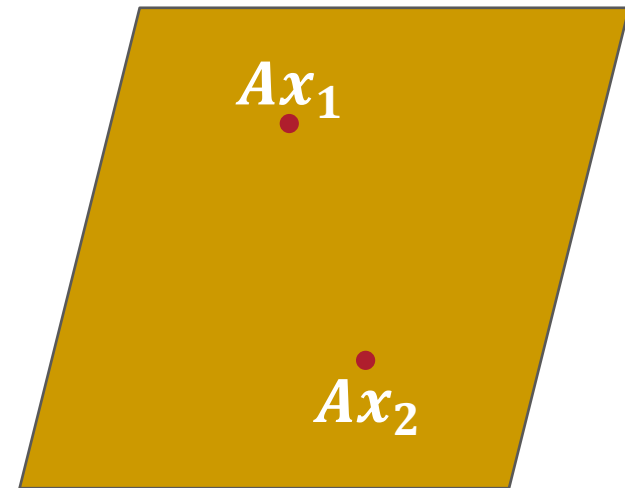
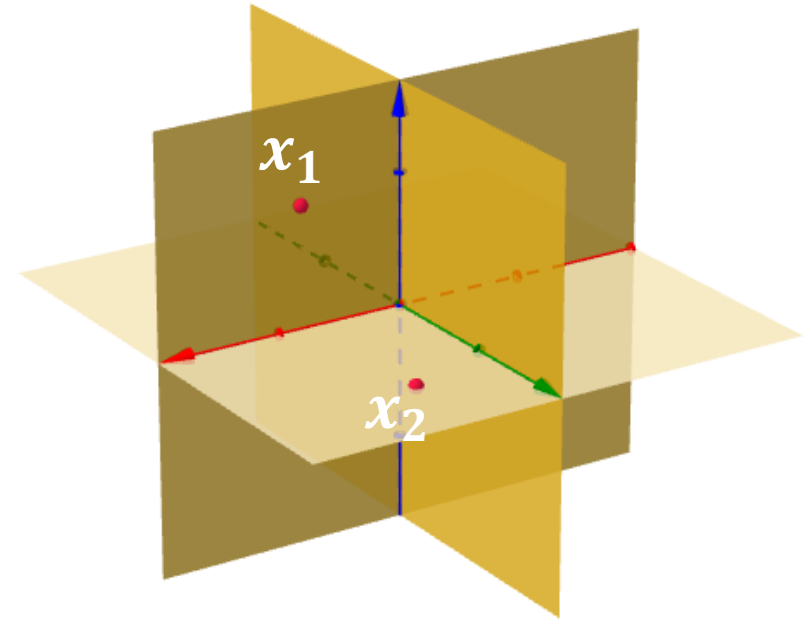
$$\begin{aligned} & \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \\ & \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y} \end{aligned}$$

$$\|\mathbf{x}\|_0 = |\operatorname{supp}(\mathbf{x})|$$

$$\operatorname{supp}(\mathbf{x}) = \{i \mid x_i \neq 0\}$$

☹ Generally NP-hard!

😊 Tractable for many interesting  $\mathbf{A}$



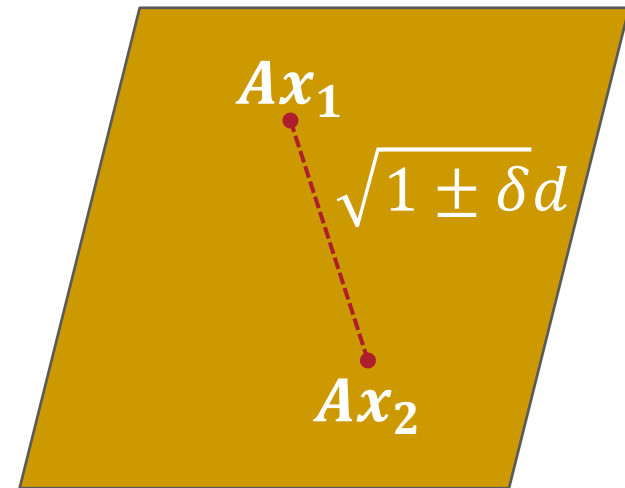
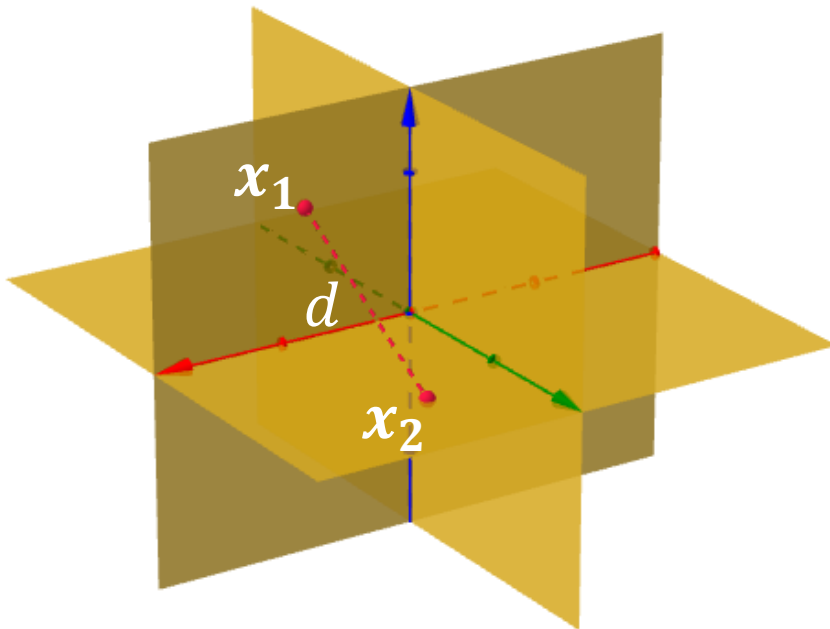
# Specialize $A$

Many sufficient conditions proposed

- Nullspace property, Incoherence, Restricted Eigenvalue, ...

## Restricted Isometry Property

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2, \quad \forall \mathbf{x}: \|\mathbf{x}\|_0 \leq k$$



# Randomness comes to rescue

## Random matrices can exhibit RIP:

Random matrices with iid entries

- Gaussian, Rademacher (symmetric Bernoulli), Uniform

Structured random matrices

- Random partial Fourier matrices
- Random circulant matrices

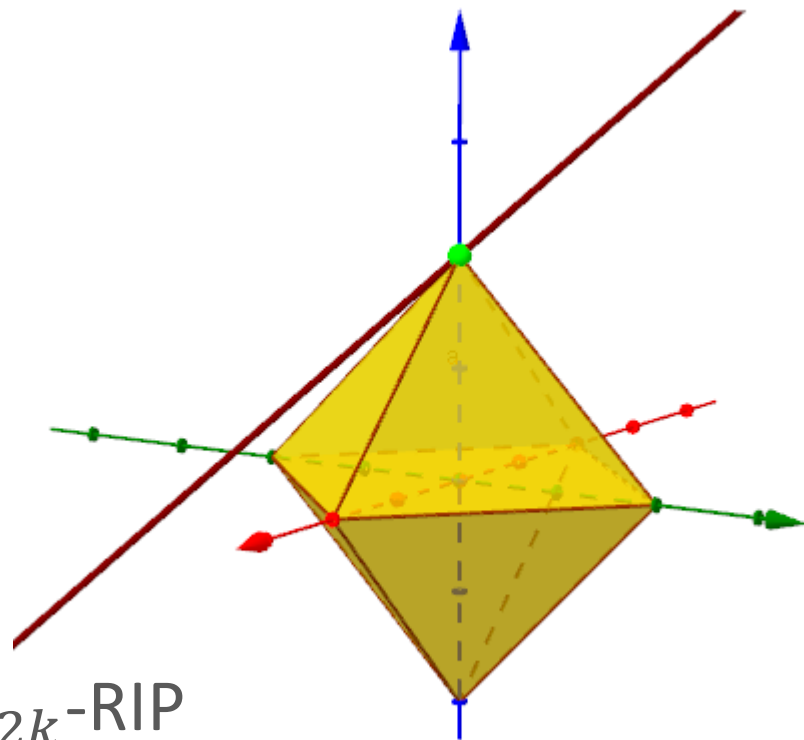
RIP holds with high probability if

$$m \gtrsim k \log^\gamma n$$

# Basis Pursuit ( $\ell_1$ -minimization)

$$\begin{aligned} & \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \\ & \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y} \end{aligned}$$

$$\|\mathbf{x}\|_1 = \sum_i |x_i|$$



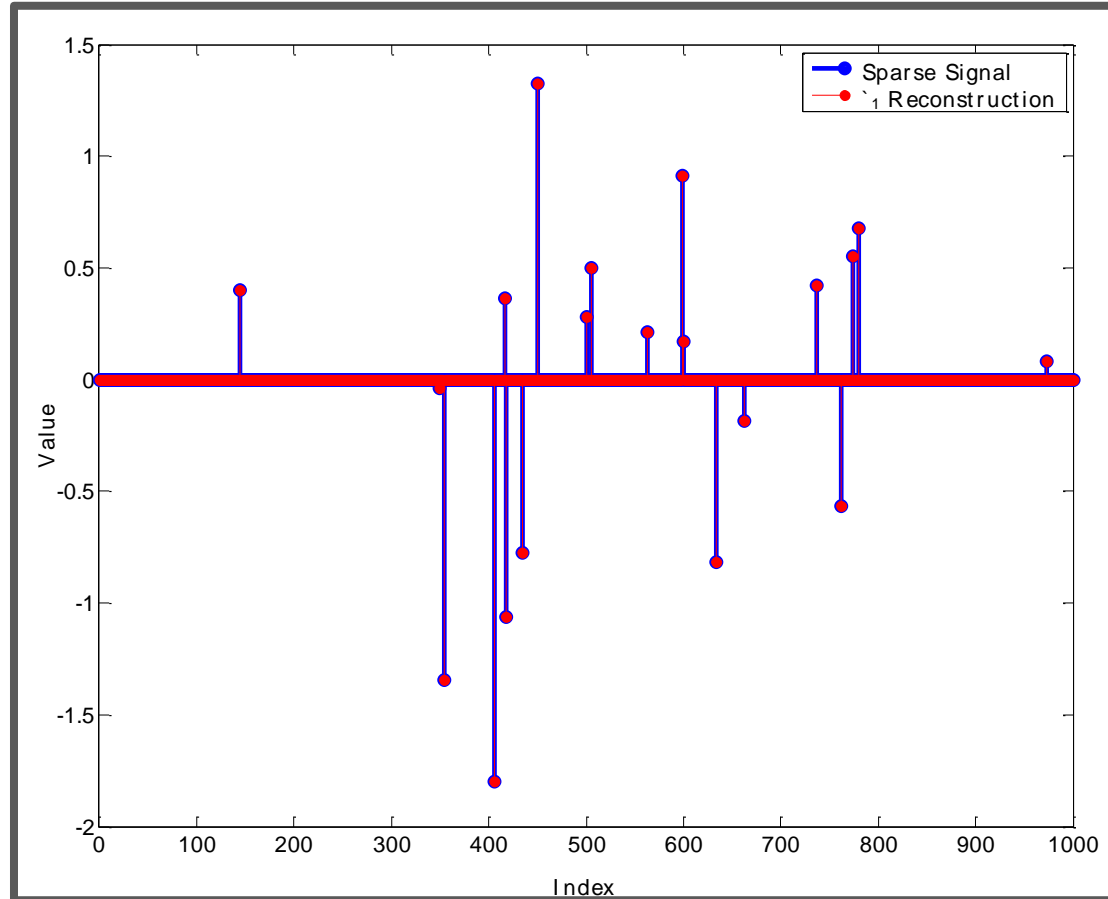
**Theorem** [Candès] If  $\mathbf{A}$  satisfies  $\delta_{2k}$ -RIP with  $\delta_{2k} < \sqrt{2} - 1$ , then BP recovers any  $k$ -sparse target exactly.



# Basis Pursuit ( $\ell_1$ -minimization)

```
1  S = 20;  
2  N = 1000;  
3  M = 200;  
4  
5  x = zeros(N,1);  
6  x(randperm(N,S)) = randn(S,1);  
7  
8  A = randn(M,N);  
9  
10 y = A*x;  
11  
12 cvx_begin  
13     variable xhat(N);  
14     minimize(norm(xhat,1));  
15     subject to  
16         A*xhat == y;  
17 cvx_end  
18  
19 stem(1:N,x,'o','LineWidth',3);  
20 hold on;  
21 stem(1:N,xhat,'ro','fill');  
22 hold off;
```

CVX



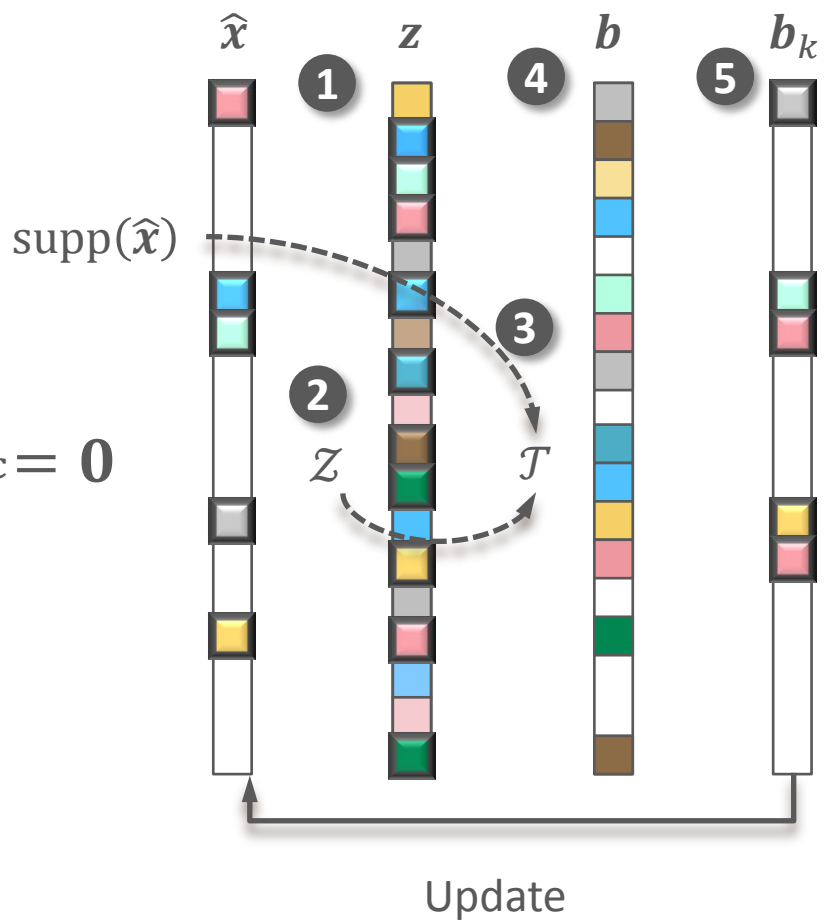
# Greedy algorithms

OMP, StOMP, IHT, CoSaMP, SP, ...

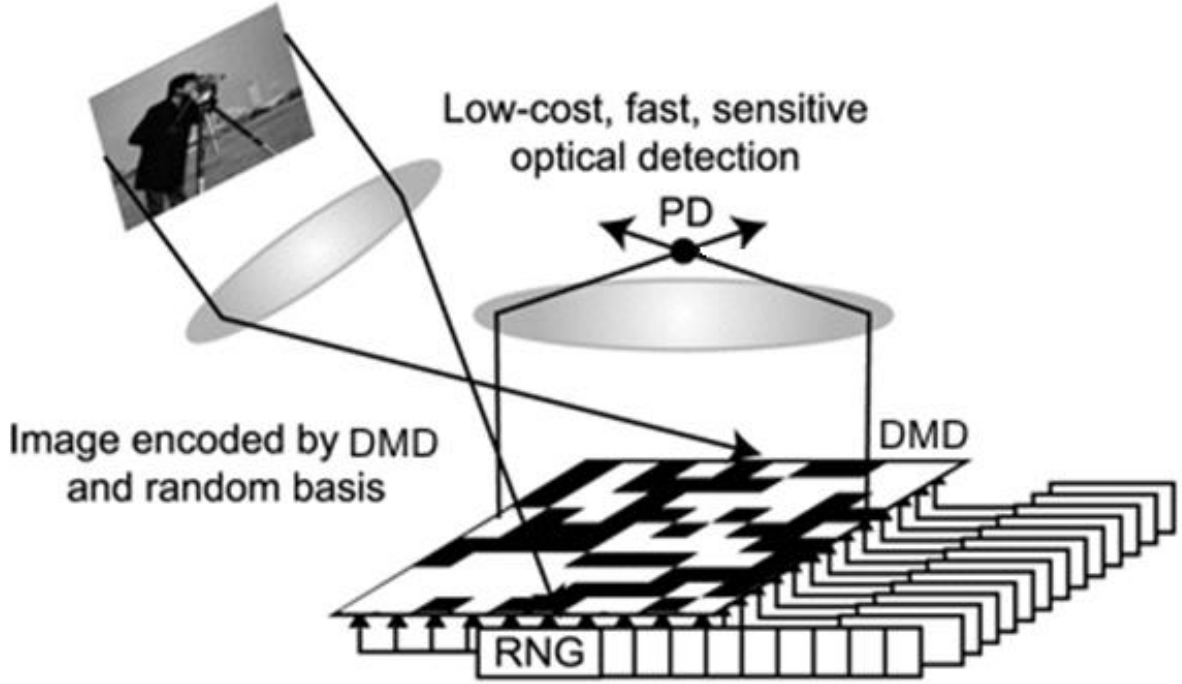
- Iteratively estimate the support and the values on the support
- Many of them have RIP-based convergence guarantees

## CoSaMP (Needell and Tropp)

- Proxy vector  $\mathbf{z} = \mathbf{A}^T(\mathbf{A}\hat{\mathbf{x}} - \mathbf{y})$
- $\mathcal{Z} = \text{supp}(\mathbf{z}_{2k})$
- $\mathbf{b} = \text{argmin}_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \text{ s.t. } \mathbf{x} |_{\mathcal{J}^c} = \mathbf{0}$
- Converges with  $\delta_{4k} \leq 0.1$



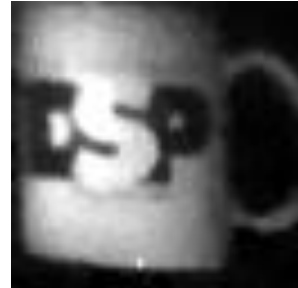
# Rice single-pixel camera



original object



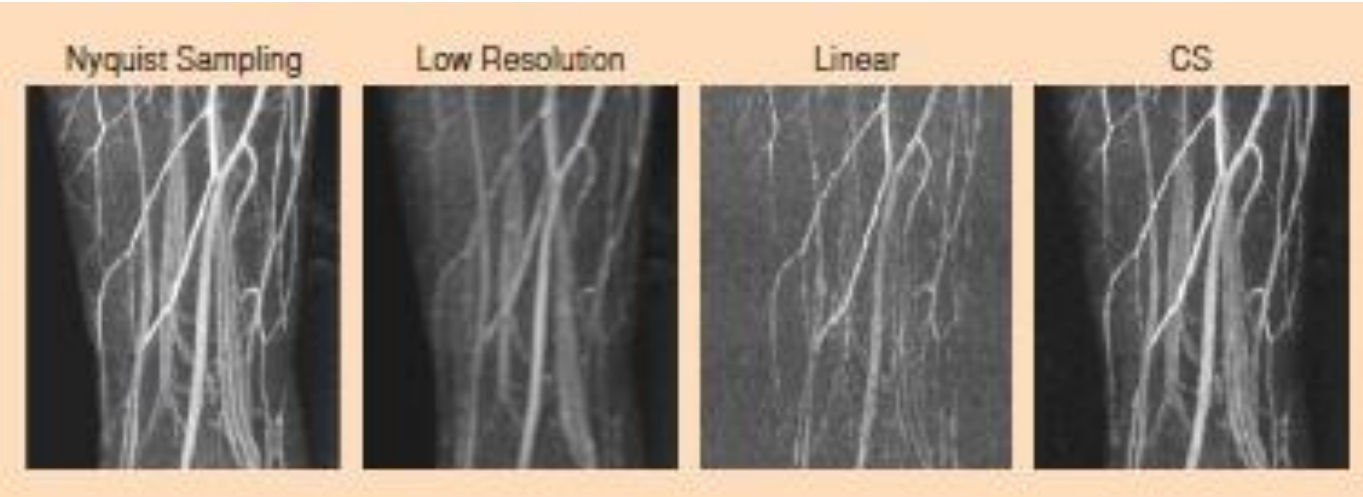
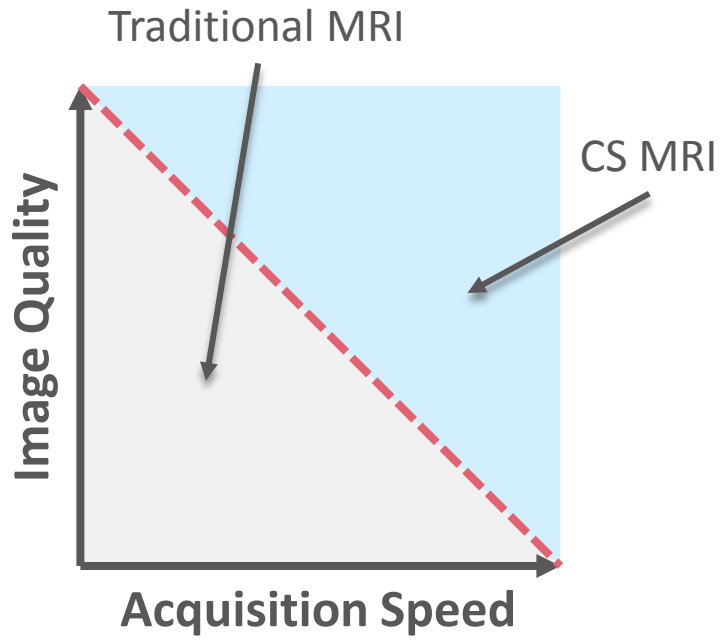
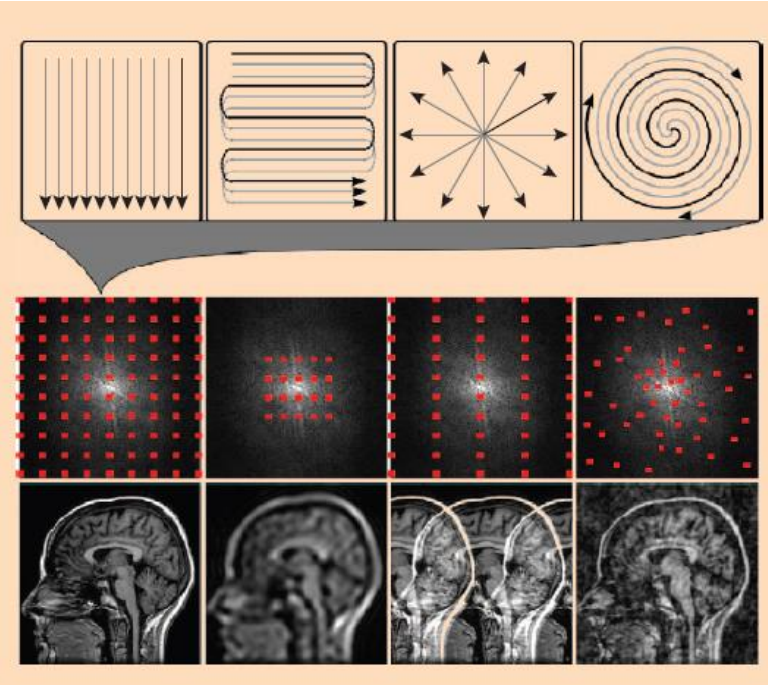
dim. reduction @ 20%



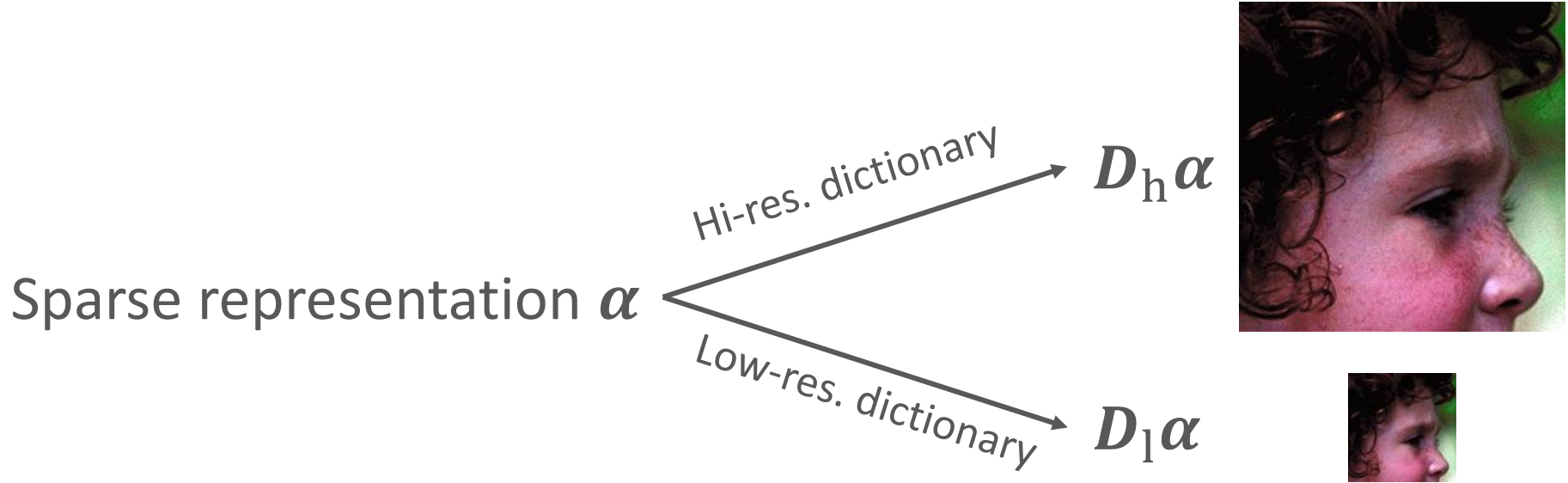
dim. reduction @ 40%

# Compressive MRI

Lustig et al



# Image super-resolution



CS-based reconstruction Original

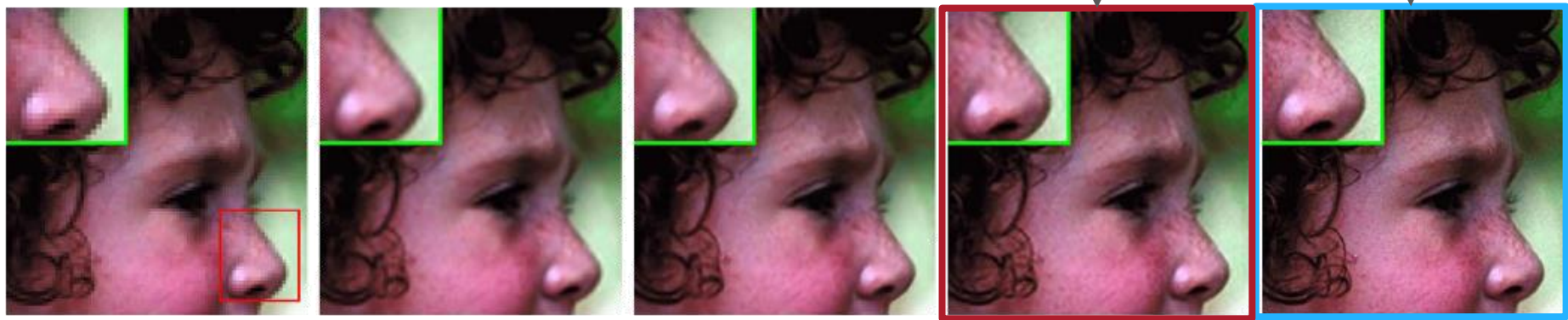


Fig. 4. Results of the girl image magnified by a factor of 3 and the corresponding RMSEs. Left to right: input, bicubic interpolation (RMSE: 6.843), NE [11] (RMSE: 7.740), our method (RMSE: 6.525), and the original.

# Low-Rank Matrix Recovery

# Linear systems w/ low-rank solutions

$$y_i = \langle \mathbf{X}^* = n_1 \begin{matrix} r \\ \mathbf{U} \end{matrix} \begin{matrix} n_2 \\ \mathbf{V}^* \end{matrix}, \mathbf{A}_i \rangle$$

Simple least squares requires  $n_1 n_2$  measurements

Degrees of freedom is  $r(n_1 + n_2 - 1)$

Can we close the gap?



# Rank minimization

$$\begin{aligned} & \underset{\mathbf{X}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{X}) \\ & \text{subject to } \mathcal{A}(\mathbf{X}) = \mathbf{y} \end{aligned}$$

If identifiable, it recovers the low-rank solution exactly

Just like  $\ell_0$ -minimization, it is generally NP-hard

Special measurement operators  $\mathcal{A}$  admit efficient solvers

# Random measurements

$A_i$  with iid entries exhibit **low-rank RIP**

- Gaussian
- Rademacher
- Uniform

Universal



Some structured matrices

- Rank-one  $A_i = \mathbf{a}_i \mathbf{b}_i^*$
- $A_i$  with dependent entries
- Standard RIP usually doesn't hold, but other approaches exist

Instance optimal

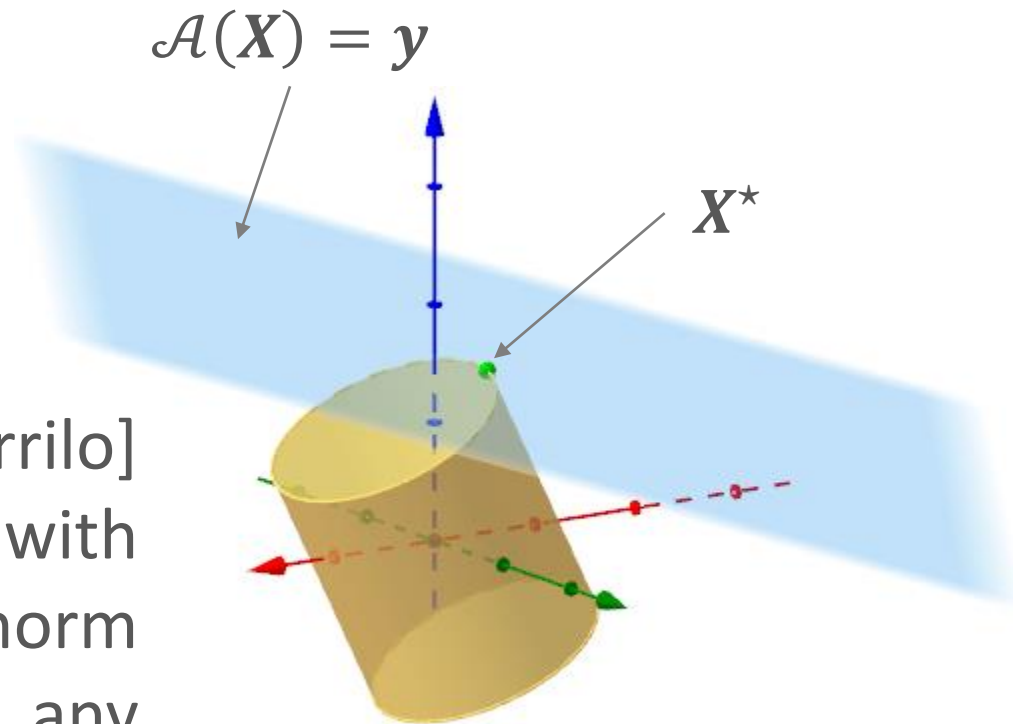


# Nuclear-norm minimization

$$\begin{aligned} & \operatorname{argmin}_X \|X\|_* \\ & \text{subject to } \mathcal{A}(X) = \mathbf{y} \end{aligned}$$

$$\|X\|_* = \sum_i \sigma_i$$

**Theorem** [Recht, Fazel, Parrilo]  
If  $\mathcal{A}$  obeys  $\delta_{5r}$ -RIP with  $\delta_{5r} < 0.1$  then nuclear-norm minimization recovers any rank- $r$  target exactly.



# Matrix completion

						...		
Alice	4	5	?	3	?	...	?	1
Bob	?	2	?	4	5	...	?	?
.								
.								
.								
Yvon	5	?	4	3	2	...	2	?
Zelda	3	2	?	?	5	...	2	2

**Theorem** [Candès and Recht] If  $M$  is a rank- $r$  matrix from the “random orthogonal model”, then w.h.p. the nuclear-norm minimization recovers  $M$  exactly from  $\mathcal{O}(rn^{5/4} \log n)$  uniformly observed entries, where  $n = n_1 \vee n_2$ .

Ordinary PCA is sensitive to noise and outliers

low-rank component  $\swarrow$   $\searrow$  sparse component

$$M = L + S$$

$$\operatorname{argmin}_{L,S} \|L\|_* + \lambda \|S\|_1$$

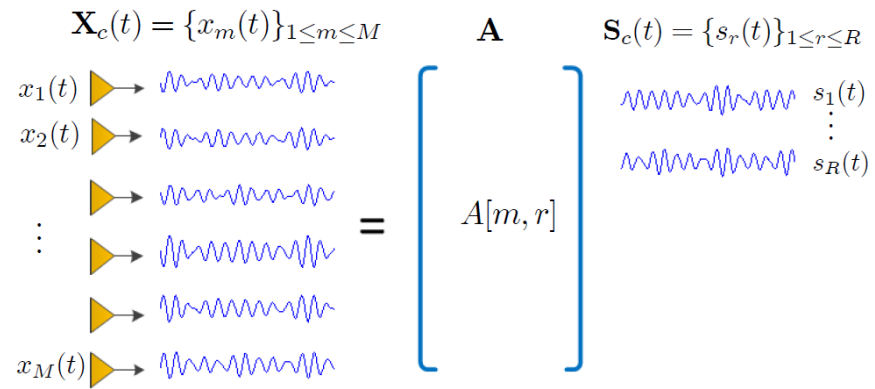
$$\text{subject to } P_{\Omega_{\text{obs}}} (L + S) = Y$$

Background Modeling



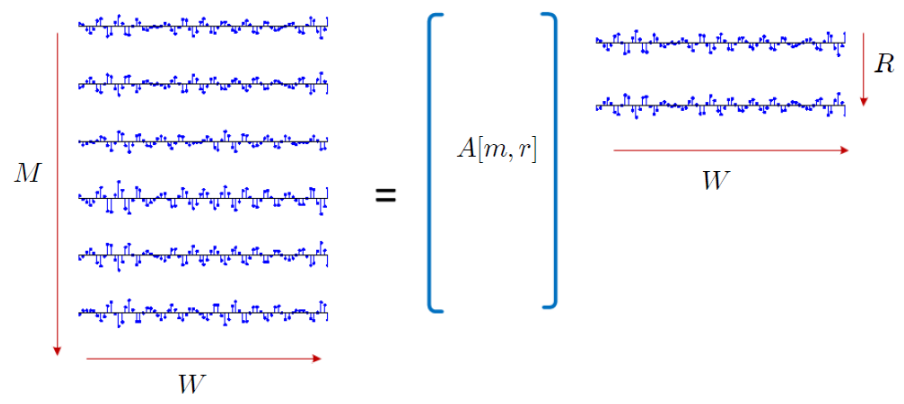
(a) Original frames      (b) Low-rank  $\hat{L}$       (c) Sparse  $\hat{S}$       (d) Low-rank  $\hat{L}$       (e) Sparse  $\hat{S}$   
Convex optimization (this work)      Alternating minimization [47]

# Compressive multiplexing

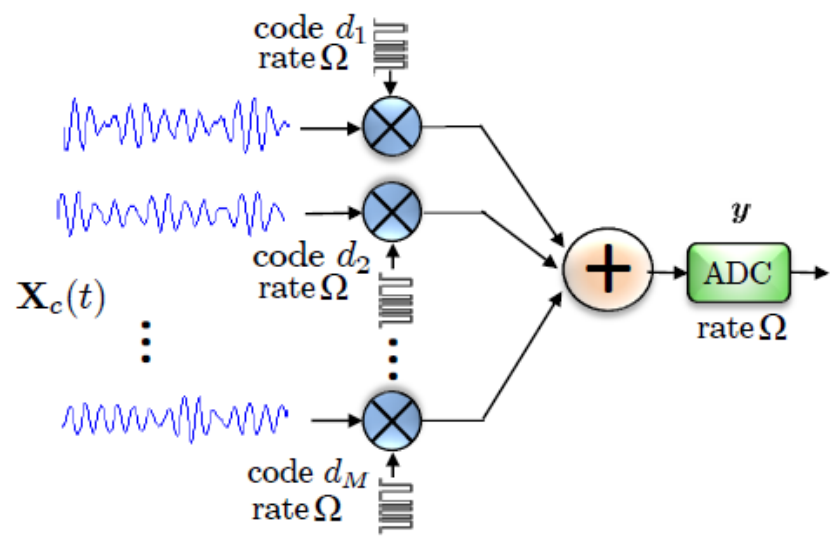


$$x_m(t) = \sum_{\omega=-B}^B \alpha_m[\omega] e^{j2\pi\omega t}$$

$$\Omega \sim R(M + W) \log^3(MW)$$



$$W = 2B + 1$$



# Nonlinear CS



# Sparsity-constrained minimization

Squared error isn't always appropriate

- example : non-Gaussian noise models

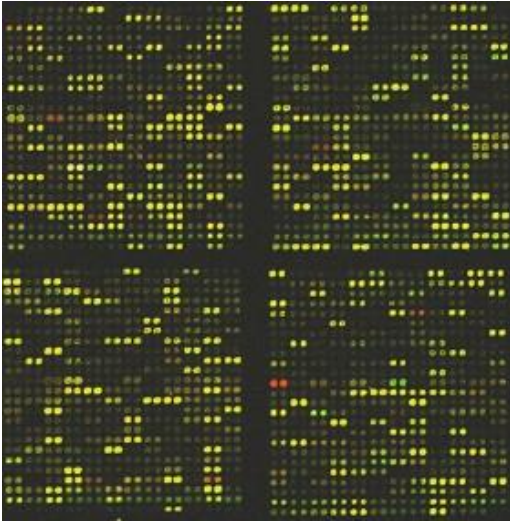
$$\begin{aligned} & \operatorname{argmin}_{\mathbf{x}} f(\mathbf{x}) \\ & \text{subject to } \|\mathbf{x}\|_0 \leq k \end{aligned}$$

It's challenging in its general form

Objectives with certain properties allow approximations

- convex relaxation:  $\ell_1$ -regularization
- greedy methods

# Gene classification



DNA Microarray

$\mathbf{a}_i$  : microarray sample

$y_i$  : binary sample label (healthy/diseased)

model:  $p(y | \mathbf{a}; \mathbf{x})$

$\mathbf{x}$  : sparse weights for genes

## Sparse Logistic Regression

$$\operatorname{argmin}_{\mathbf{x}} \sum_{i=1}^m \log \left( 1 + e^{\mathbf{a}_i^T \mathbf{x}} \right) - y_i \mathbf{a}_i^T \mathbf{x}$$

$$\text{subject to } \|\mathbf{x}\|_0 \leq k, \|\mathbf{x}\|_2 \leq 1$$

# Imaging under photon noise

Photon noise follows a Poisson distribution

$$p(\mathbf{y} \mid \mathbf{A}; \mathbf{x}) = \prod_i \frac{(\mathbf{Ax})_i^{y_i}}{y_i!} e^{-(\mathbf{Ax})_i}$$

- Physical constraints  $\mathbf{A} \geq 0, \mathbf{Ax} \geq 0, \mathbf{x} \geq 0$

## SPIRAL-TAP

$$\underset{\mathbf{x} \geq 0}{\operatorname{argmin}} -\log p(\mathbf{y} \mid \mathbf{A}; \mathbf{x}) + \tau R(\mathbf{x})$$

- $\|\mathbf{x}\|_1$
- $\|\mathbf{W}^T \mathbf{x}\|_1$
- $\|\mathbf{x}\|_{\text{TV}}$
- ...

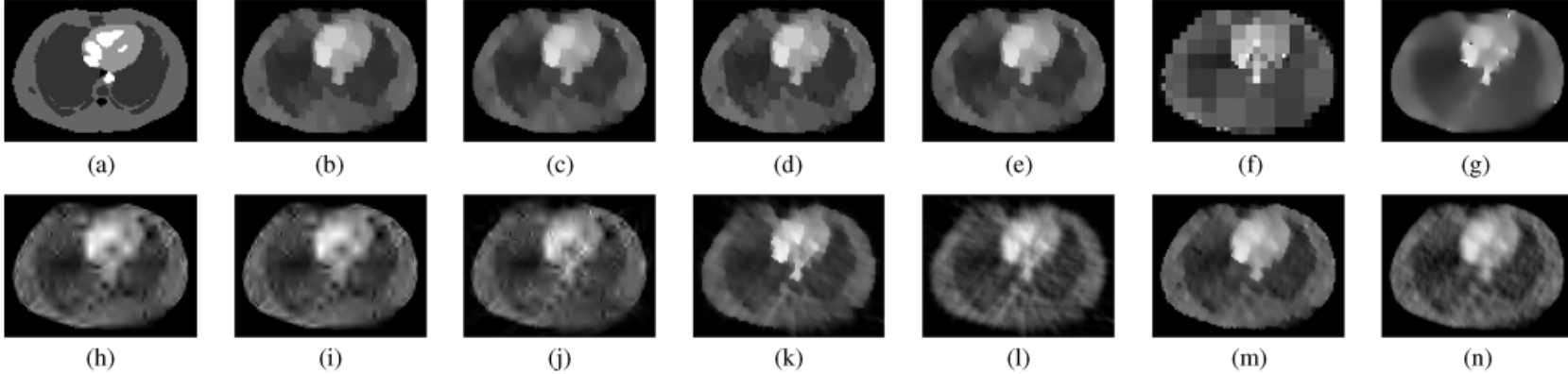
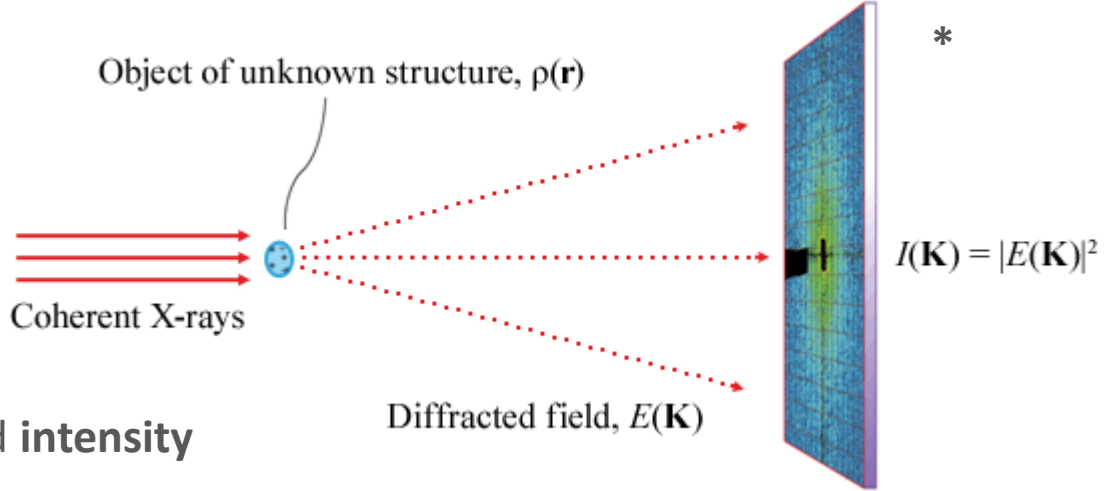
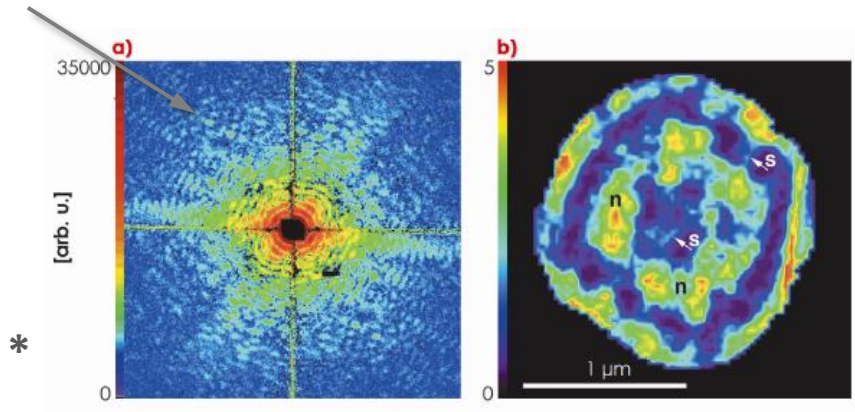


Fig. 3. Single-trial reconstructed images for all methods considered. Note  $\text{RMSE}(\%) = 100 \cdot \|\hat{f} - f^*\|_2 / \|f^*\|_2$ . (a) Ground truth. (b) SPIRAL-TV loose monotonic (RMSE = 24.404%). (c) SPIRAL-TV loose nonmonotonic (RMSE = 24.962%). (d) SPIRAL-TV tight monotonic (RMSE = 24.526%). (e) SPIRAL-TV tight nonmonotonic (RMSE = 24.467%). (f) SPIRAL-RDP (translation variant) (RMSE = 33.959%). (g) SPIRAL-RDP-TI (Cycle-Spun) (RMSE = 27.557%). (h) SPIRAL- $\ell_1$  Loose (DB-6) (RMSE = 28.626%). (i) SPIRAL- $\ell_1$  Tight (DB-6) (RMSE = 28.665%). (j) SpaRSA  $\ell_2$ - $\ell_1$  (DB-6) (RMSE = 31.172%). (k) SPS-OS (Huber potential) (RMSE = 27.555%). (l) SPS-OS (quadratic potential) (RMSE = 29.420%). (m) EPL-INC-3 (Huber potential) (RMSE = 24.748%). (n) EPL-INC-3 (quadratic potential) (RMSE = 26.474%).

# Coherent diffractive imaging



$I$  : diffracted field **intensity**



$Cx$  : real image

- $C$  : Shifts of **generating function**
- $L$  : low-pass filter
- $F$  : Fourier transform

$$f(x) = |||LFCx|^2 - I||_2^2$$

\* Lima et al., ESRF (www.esrf.eu)

# Gradient Support Pursuit

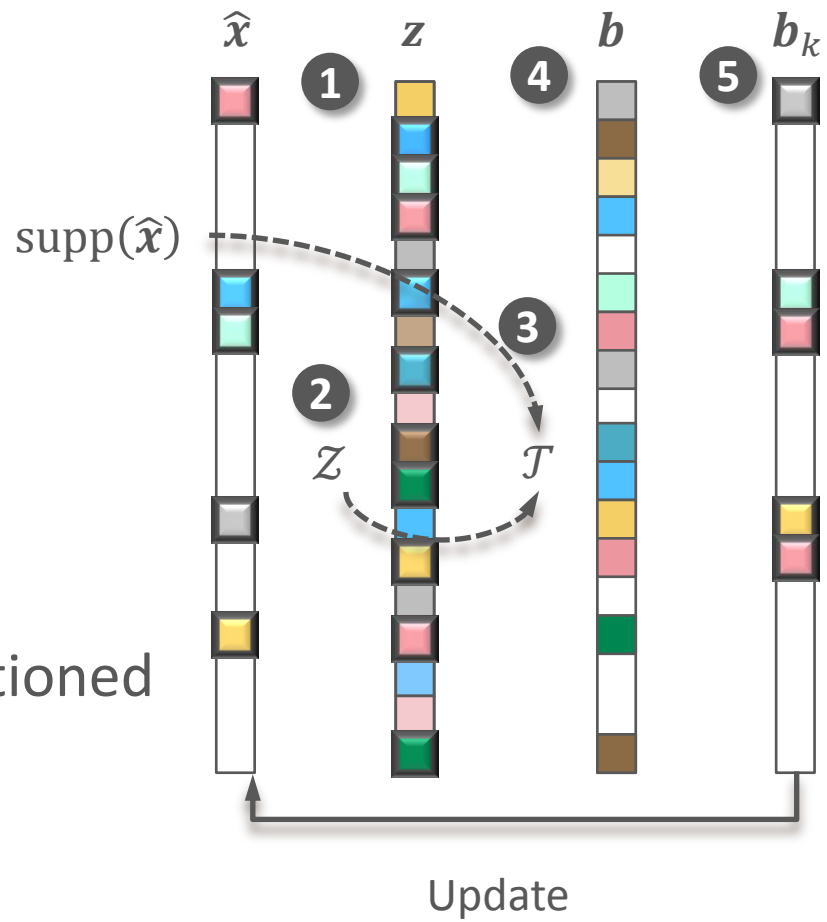
w/ Raj and Boufounos

## Generalizes CoSaMP

- Proxy vector  $\mathbf{z} = \nabla f(\mathbf{x})$
- $\mathcal{Z} = \text{supp}(\mathbf{z}_{2k})$
- $\mathbf{b} = \text{argmin}_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } \mathbf{x} |_{\mathcal{J}^c} = \mathbf{0}$

## Converges under SRH

- the Hessian  $\nabla^2 f(\mathbf{x})$  is well-conditioned when restricted to sparse subspaces
- a generalization of the RIP



# There's much more ...

The screenshot shows a Google Scholar search interface. The search bar contains the text "compressive sensing" and a search button. Below the search bar, the word "Scholar" is displayed, followed by a red-bordered box containing the text "About 148,000 results (0.05 sec)".

On the left side, there are several filter options:

- Articles** (selected)
- Case law
- My library
- Any time**
  - Since 2014
  - Since 2013
  - Since 2010
  - Custom range...
- Sort by relevance** (selected)
- Sort by date
- include patents
- include citations
- Create alert

The search results are listed on the right side:

- [PDF] Compressive sensing**  
[RG Baraniuk](#) - IEEE signal processing magazine, 2007 - omni.isr.ist.utl.pt  
The Shannon/Nyquist sampling theorem tells us that in order to not lose information when uniformly sampling a signal we must sample at least two times faster than its bandwidth. In many applications, including digital image and video cameras, the Nyquist rate can be so ...  
Cited by 2123 Related articles All 54 versions Web of Science: 721 Cite Save More
- Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing**  
[W Yin](#), [S Osher](#), [D Goldfarb](#), [J Darbon](#) - SIAM Journal on Imaging Sciences, 2008 - SIAM  
We propose simple and extremely efficient methods for solving the basis pursuit problem  $\min\{\|u\|_1 : Au=f, u \in \mathbb{R}^n\}$ , which is used in **compressed sensing**. Our methods are based on Bregman iterative regularization, and they give a very accurate solution after solving ...  
Cited by 716 Related articles All 9 versions Web of Science: 338 Cite Save
- Bayesian compressive sensing**  
[S Ji](#), [Y Xue](#), [L Carin](#) - Signal Processing, IEEE Transactions on, 2008 - ieeexplore.ieee.org  
Abstract—The data of interest are assumed to be represented as-dimensional real vectors, and these vectors are compressible in some linear basis  $B$ , implying that the signal can be reconstructed accurately using only a small number of basis-function coefficients ...  
Cited by 816 Related articles All 17 versions Web of Science: 310 Cite Save
- Subspace pursuit for compressive sensing signal reconstruction**  
[W Dai](#), [O Milenkovic](#) - Information Theory, IEEE Transactions on, 2009 - ieeexplore.ieee.org  
Abstract—We propose a new method for reconstruction of sparse signals with and without noisy perturbations, termed the subspace pursuit algorithm. The algorithm has two important characteristics: low computational complexity, comparable to that of orthogonal matching ...  
Cited by 799 Related articles All 10 versions Web of Science: 341 Cite Save
- Iteratively reweighted algorithms for compressive sensing**

## Resources

a collection of CS papers, tutorials, and softwares

- <http://dsp.rice.edu/cs>

sparse and low-rank algorithms wiki

- [http://ugcs.caltech.edu/~srbecker/wiki/Main\\_Page](http://ugcs.caltech.edu/~srbecker/wiki/Main_Page)

a weblog focusing on CS and broader computational areas

- <http://nuit-blanche.blogspot.com/>