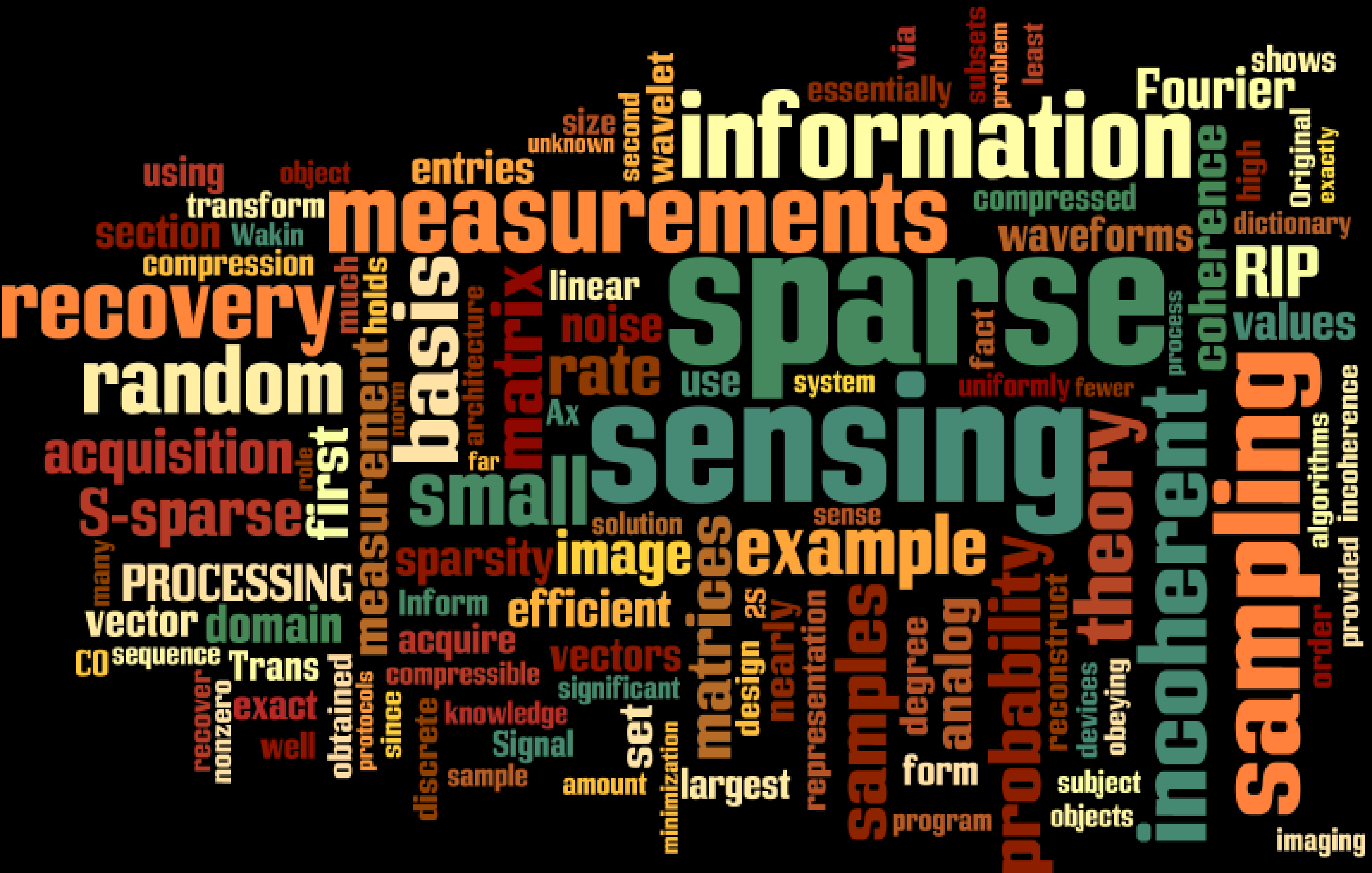
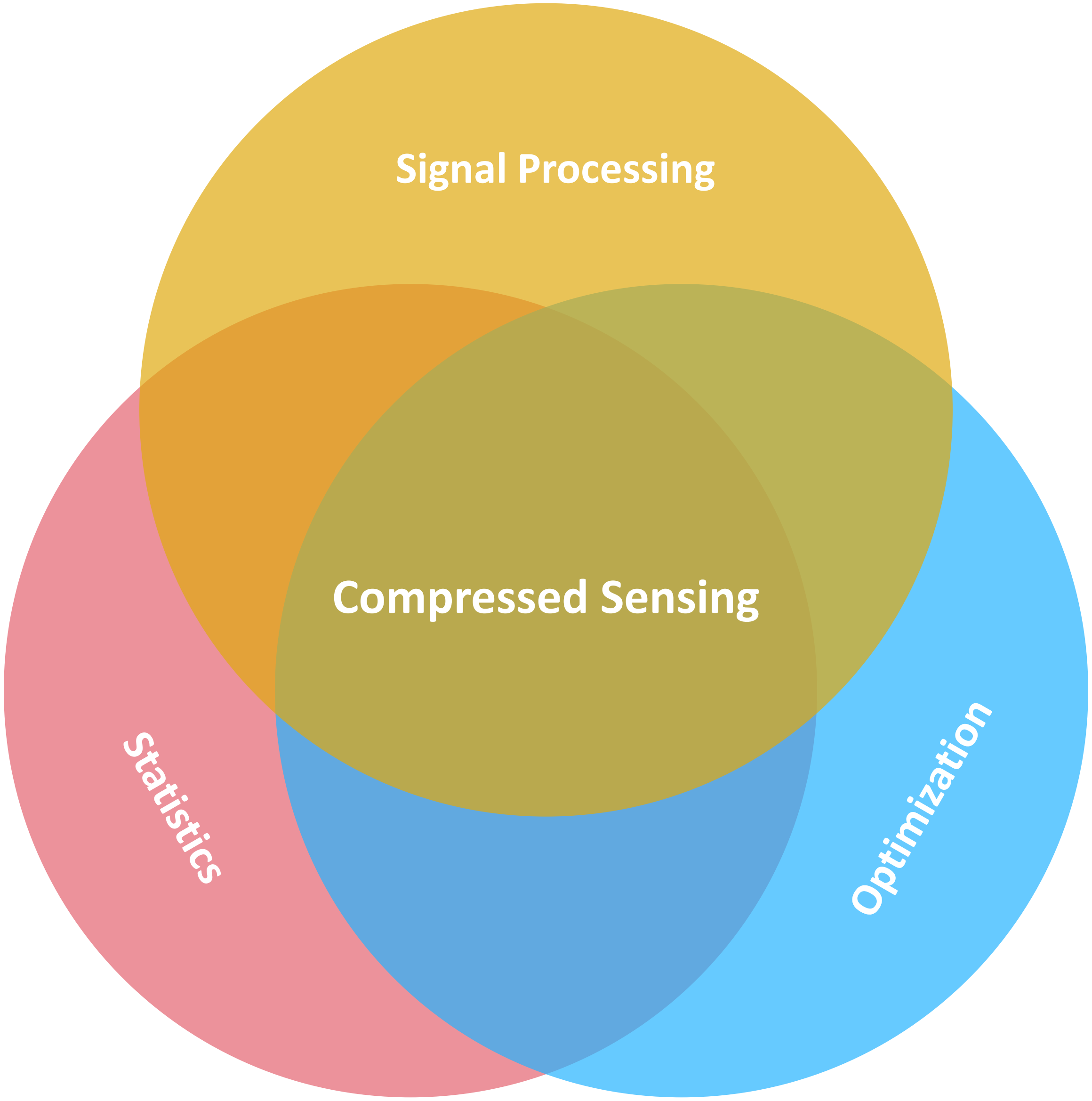
Compressive Sensing and Beyond

Sohail Bahmani

Gerorgia Tech.



shows
Fourier
information
essentially
via
subsets
problem
least
size
unknown
second
wavelet
using
object
entries
transform
measurements
compressed
Original
exactly
high
section
Wakin
waveforms
dictionary
compression
coherence
recovery
much
holds
basis
linear
sparse
RIP
noise
values
matrix
architecture
rate
fact
process
random
use
system
uniformly
fewer
norm
measurement
Ax
acquisition
sensing
role
far
first
theory
small
incoherent
sampling
S-sparse
sense
solution
algorithms
incoherence
provided
many
PROCESSING
sparsity
image
example
matrices
probability
Inform
efficient
2S
vector
domain
nearly
samples
reconstruct
order
acquire
degree
analog
CO
sequence
Trans
vectors
representation
recover
obtained
compressible
design
devices
obeying
nonzero
exact
protocols
since
significant
knowledge
set
discrete
well
Signal
minimization
form
sample
amount
largest
subject
objects
program
imaging

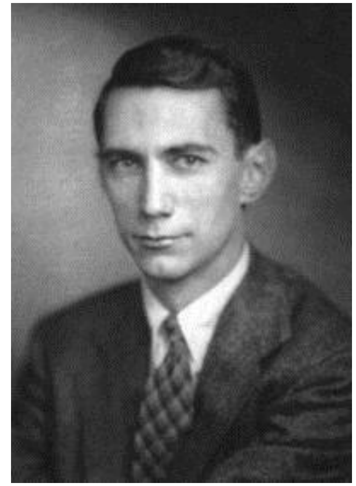

Signal Processing
Compressed Sensing
Statistics
Optimization

Signal Models

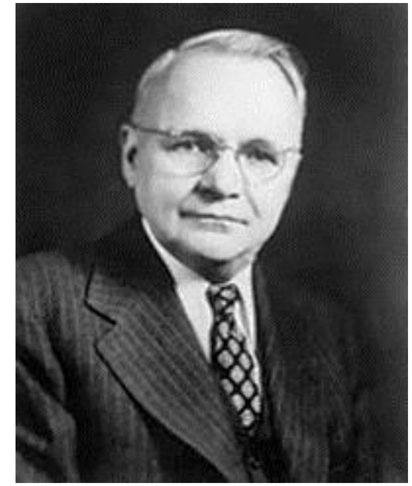
Classics: bandlimited

The Sampling Theorem

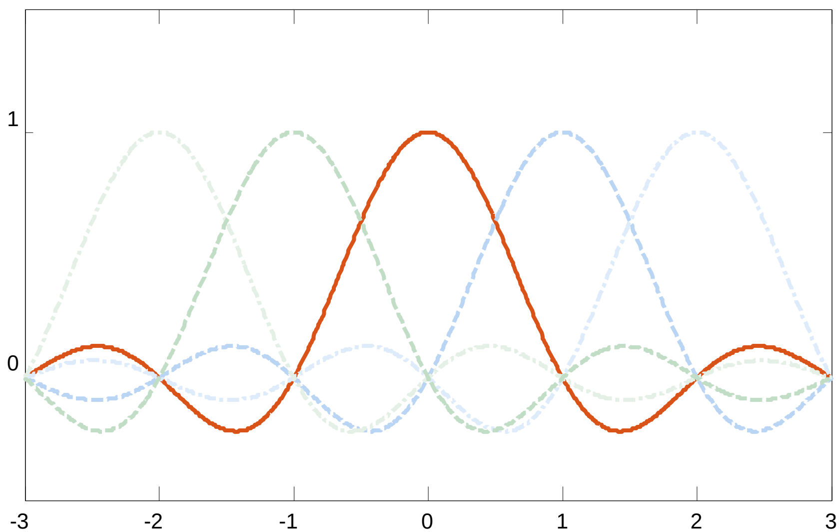
Any signal with bandwidth B can be recovered from its samples collected *uniformly* at a rate no less that $f_s = 2B$.



Claude Shannon



Harry Nyquist



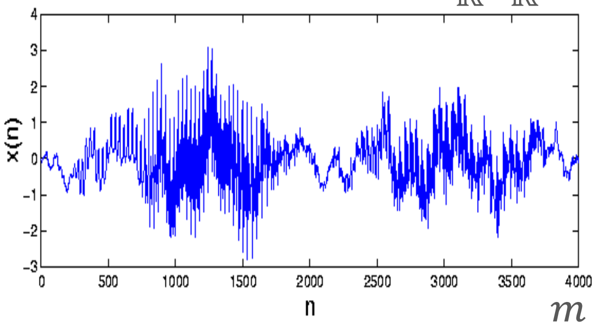
$$x(t) = \sum_{n \in \mathbb{Z}} x_n \operatorname{sinc}(f_s t - n)$$

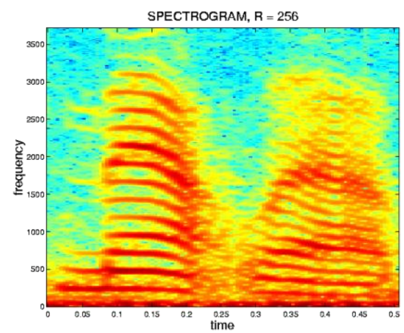
$$x_n = \langle x(t), \operatorname{sinc}(f_s t - n) \rangle$$

Classics: time-frequency

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

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





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

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

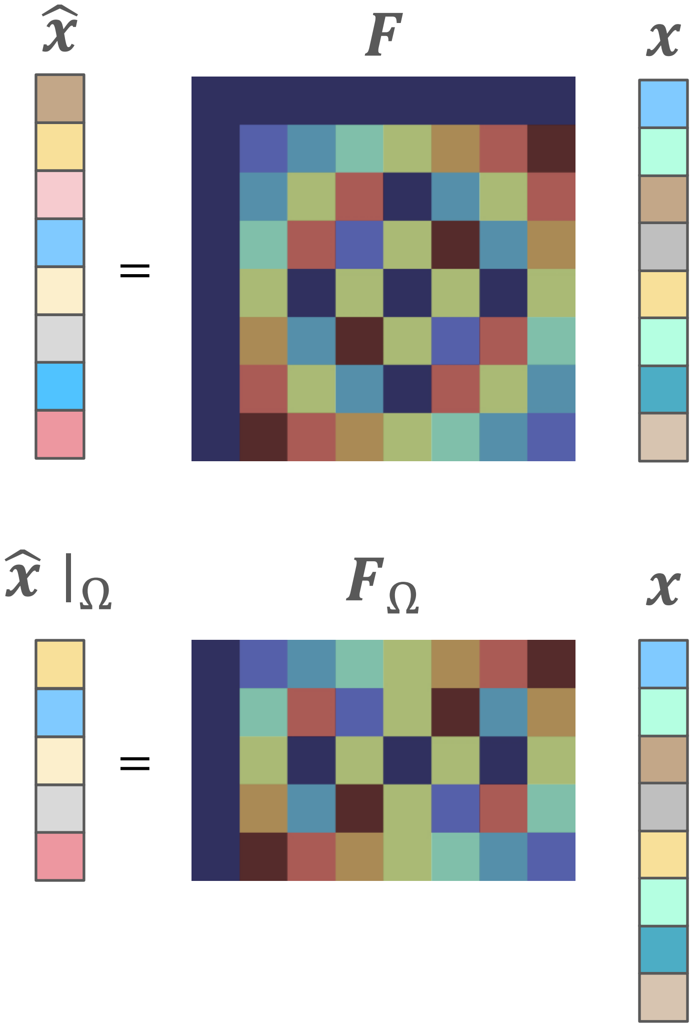
D Gabor

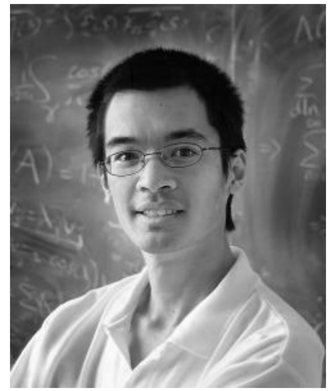
A. Haar

I. Daubechies

R. Coifman

Sparsity





T. Tao

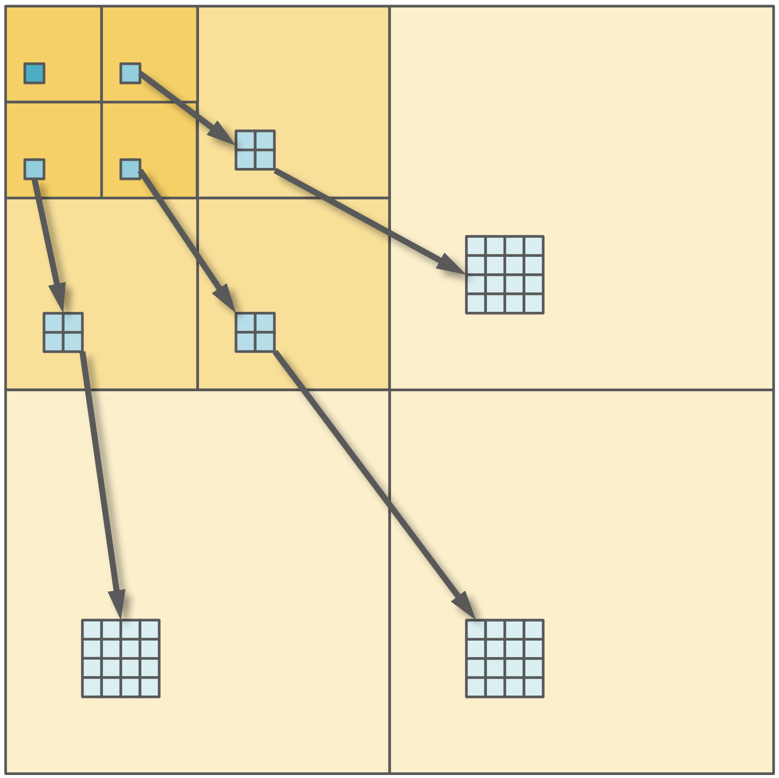
E. Candès

J. Romberg

D. Donoho

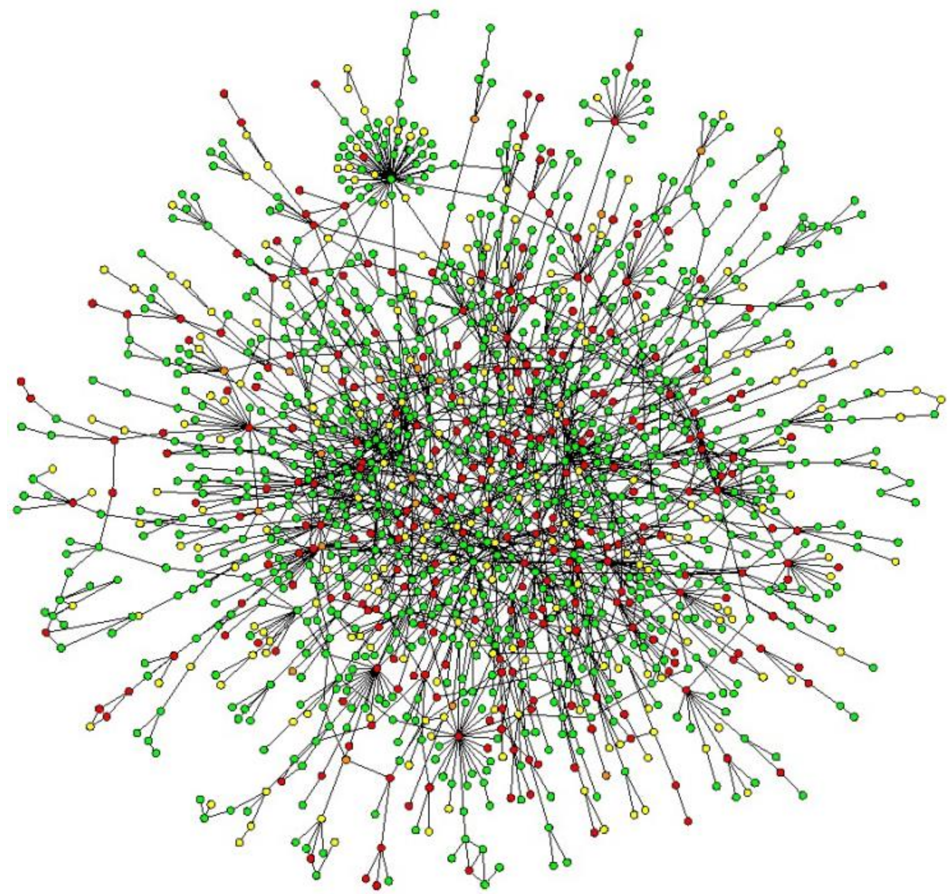
Structured sparsity

Tree Structured Models



Discrete Wavelet Transform

Graph Structured Models



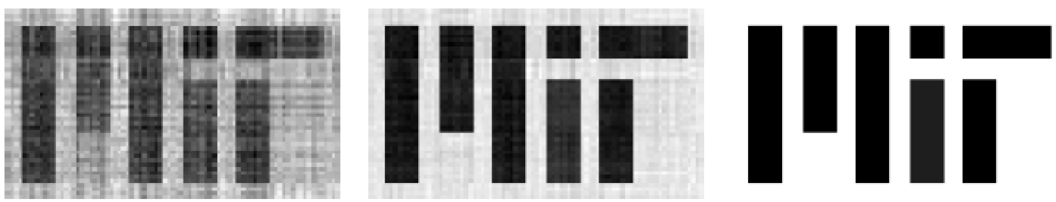
Protein-Protein Interactions
Gene Regulatory Network

Couch Lab, George Mason University, http://osf1.gmu.edu/~rcouch/protprot.jpg

Low rank

Model for Images

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



(a) (b) (c)

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.

B. Recht

M.Fazel

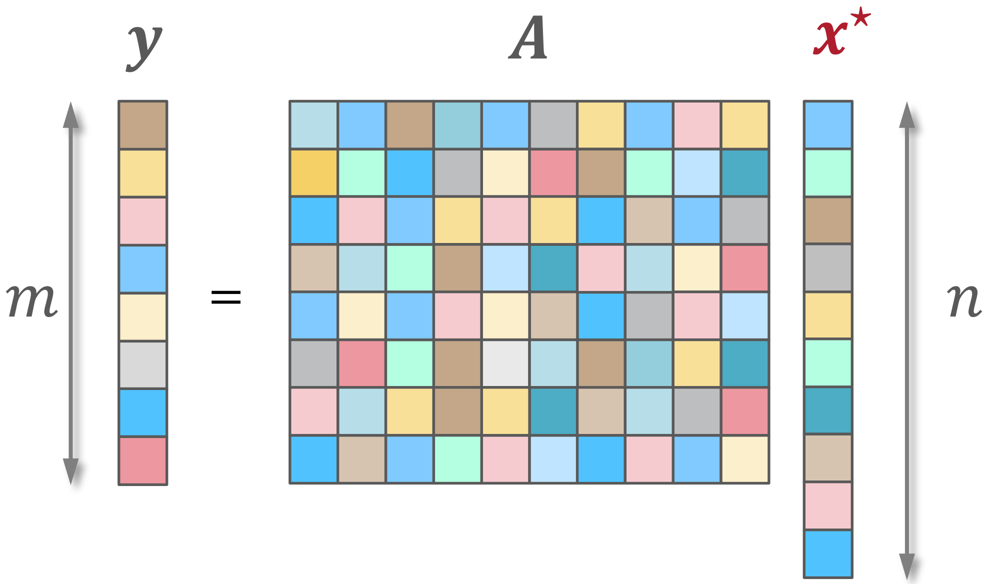
P. Parrilo

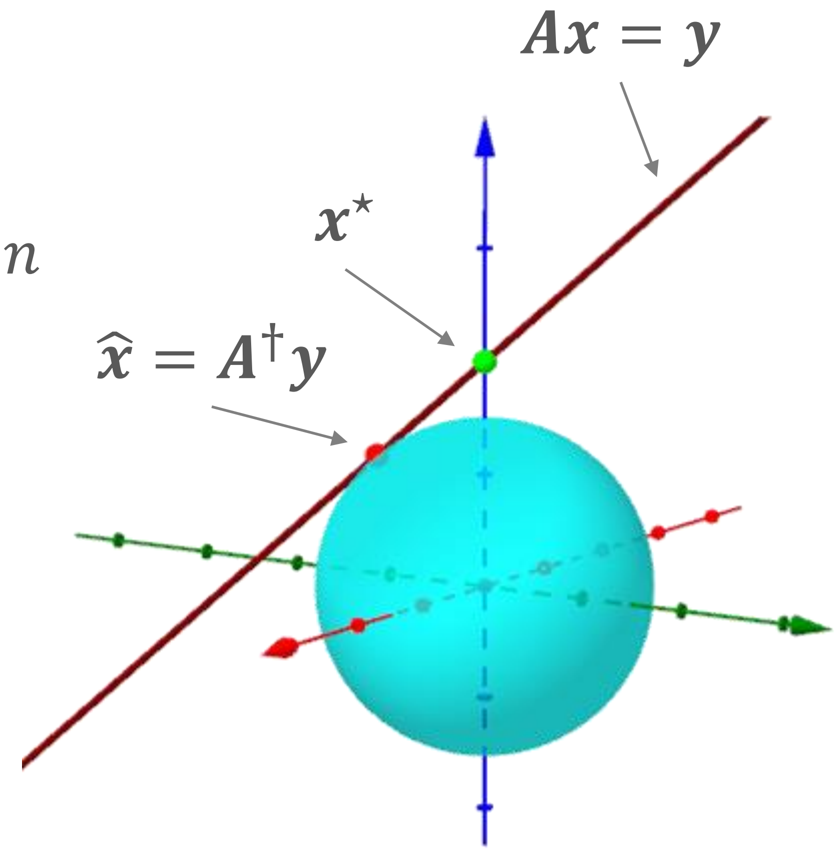
Other applications

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

Compressive Sensing

Linear inverse problems





$m < n$

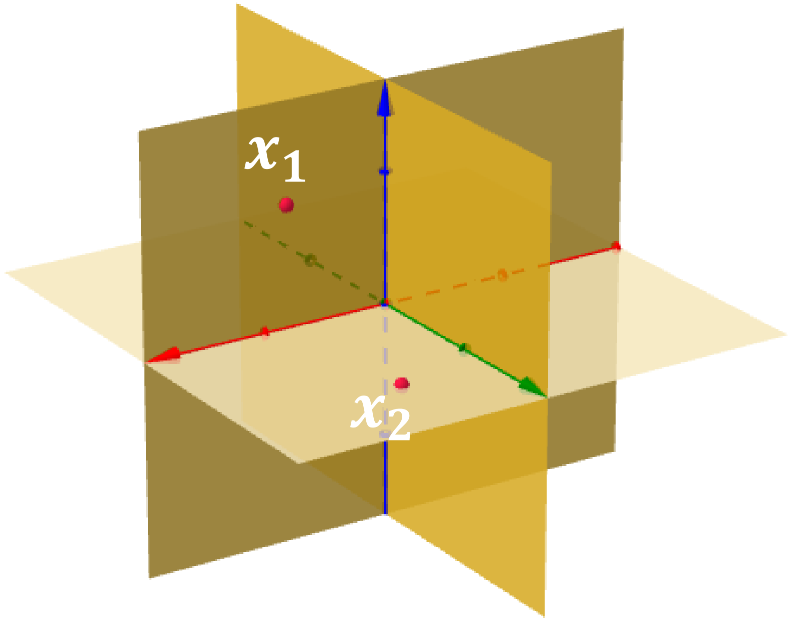
No unique inverse, but …

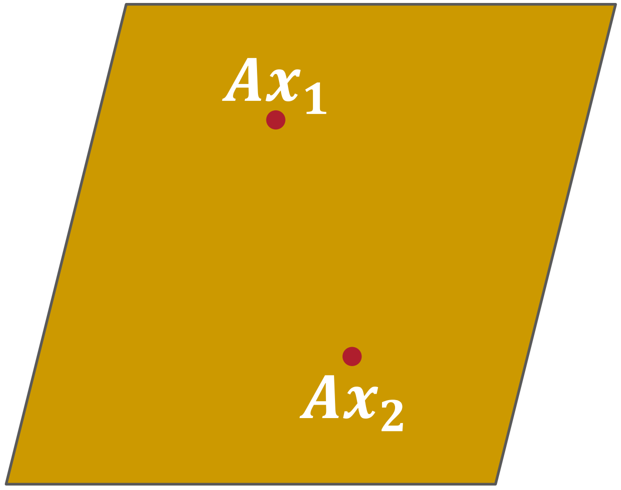
Sparse solutions

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_0$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\|\boldsymbol{x}\|_0 = |\operatorname{supp}(\boldsymbol{x})|$$
$$\operatorname{supp}(\boldsymbol{x}) = \{\, i \mid x_i \neq 0 \,\}$$

- ☹ Generally NP-hard!
- ☺ Tractable for many interesting $\boldsymbol{A}$





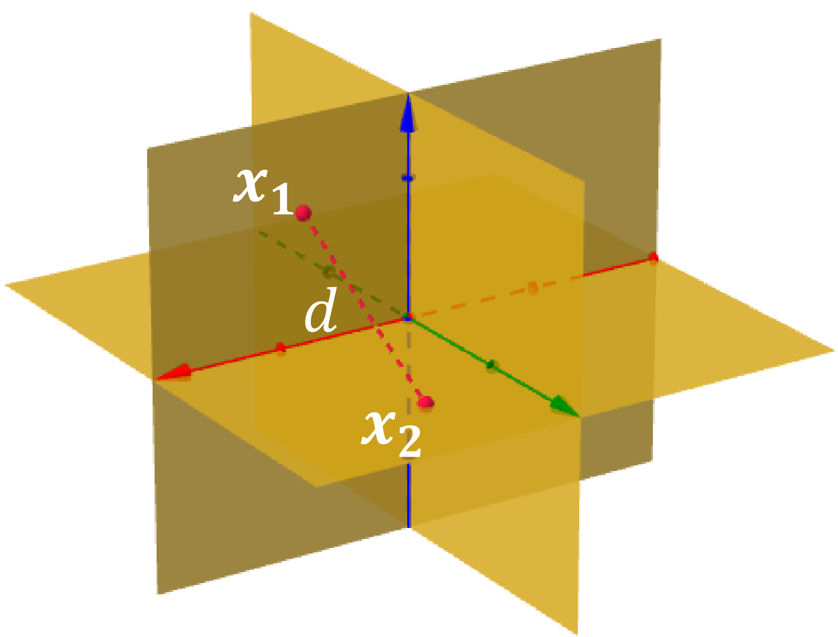
Specialize A

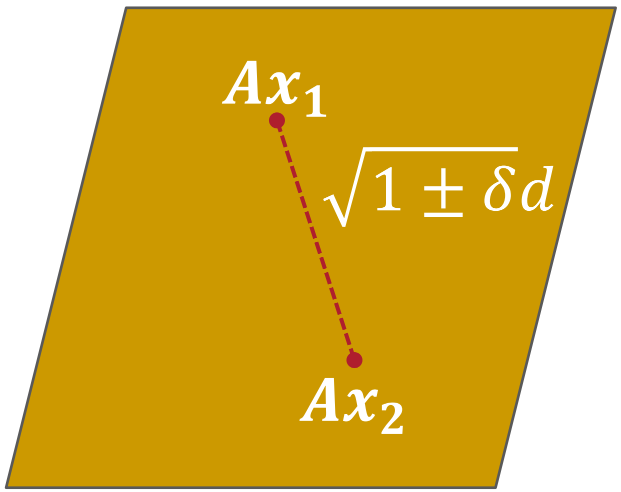
Many sufficient conditions proposed

- Nullspace property, Incoherence, Restricted Eigenvalue, …

Restricted Isometry Property

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





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

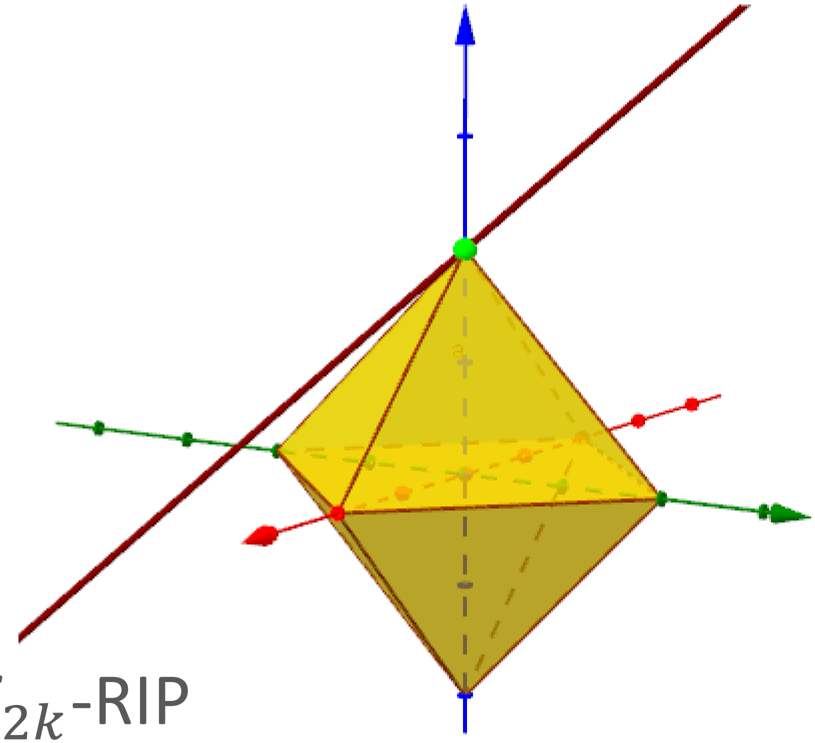
RIP holds with high probability if

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

Basis Pursuit (ℓ_1-minimization)

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_1$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

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

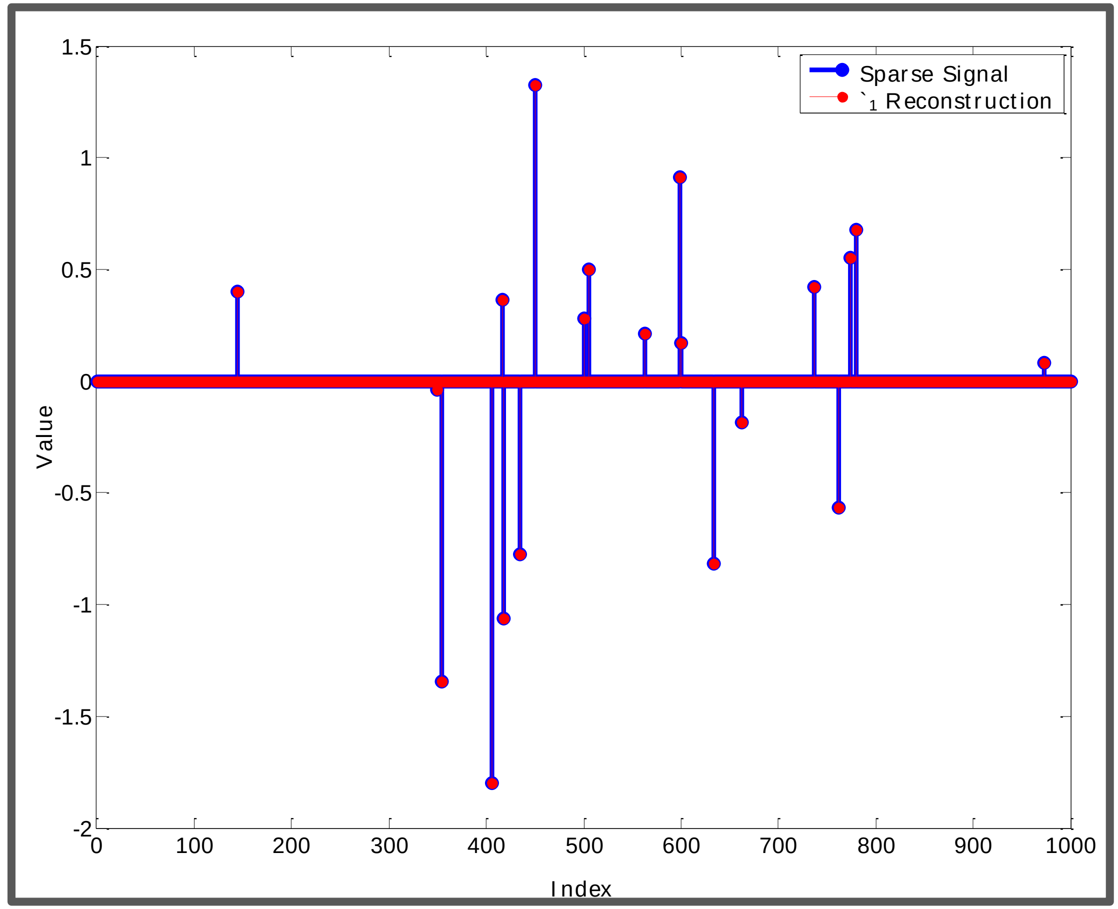


Theorem [Candès] If $\boldsymbol{A}$ satisfies δ_{2k}-RIP with $\delta_{2k} < \sqrt{2} - 1$, then BP recovers any k-sparse target exactly.

Basis Pursuit (ℓ_1-minimization)

```
S = 20;
N = 1000;
M = 200;

x = zeros(N,1);
x(randperm(N,S)) = randn(S,1);

A = randn(M,N);

y = A*x;

cvx_begin
    variable xhat(N);
    minimize(norm(xhat,1));
    subject to
            A*xhat == y;
cvx_end

stem(1:N,x,'o','LineWidth',3);
hold on;
stem(1:N,xhat,'ro','fill');
hold off;
```

CVX



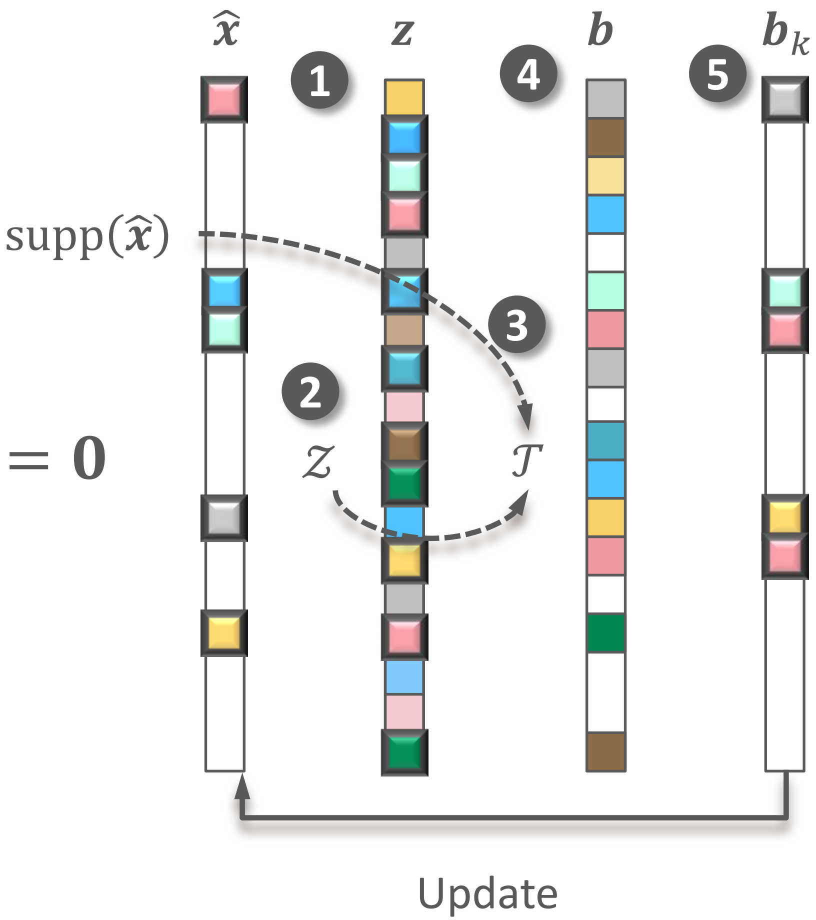
Greedy algorithms

OMP, StOMP, IHT, CoSAMP, SP, …

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

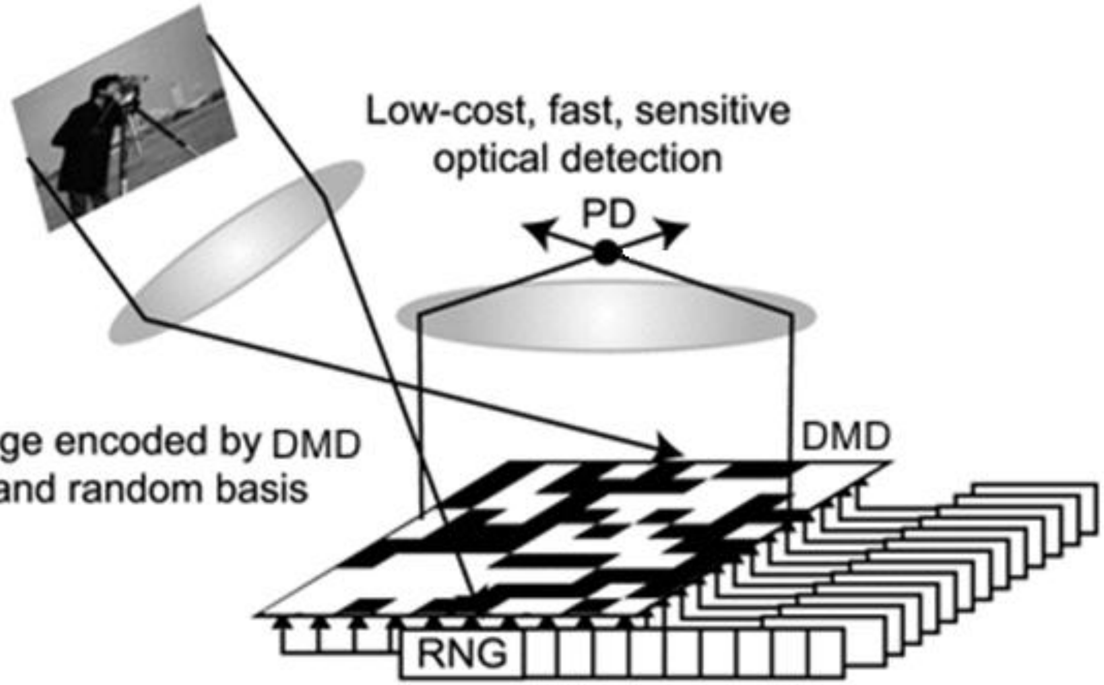
CoSaMP (Needell and Tropp)

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



Rice single-pixel camera

Wakin et al



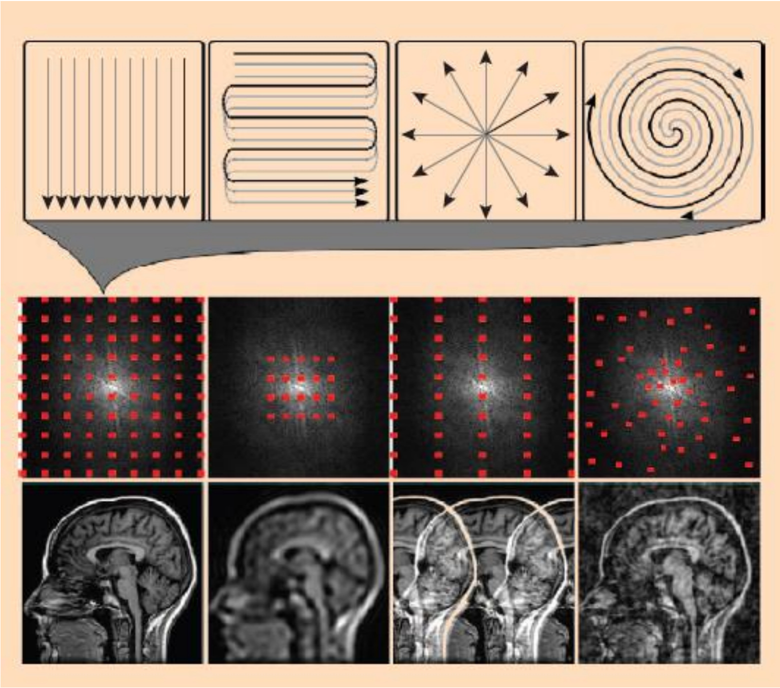
oroginal object

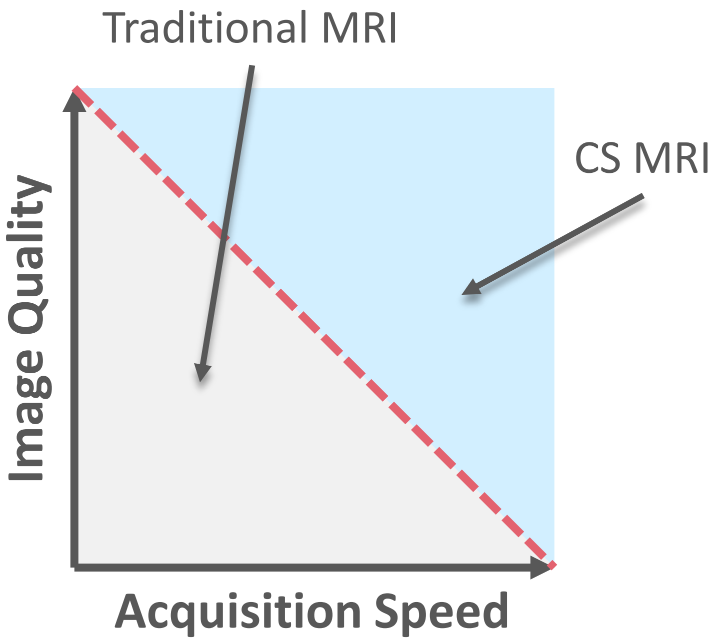
dim. reduction @ 20%

dim. reduction @ 40%

Compressive MRI

Lustig et al





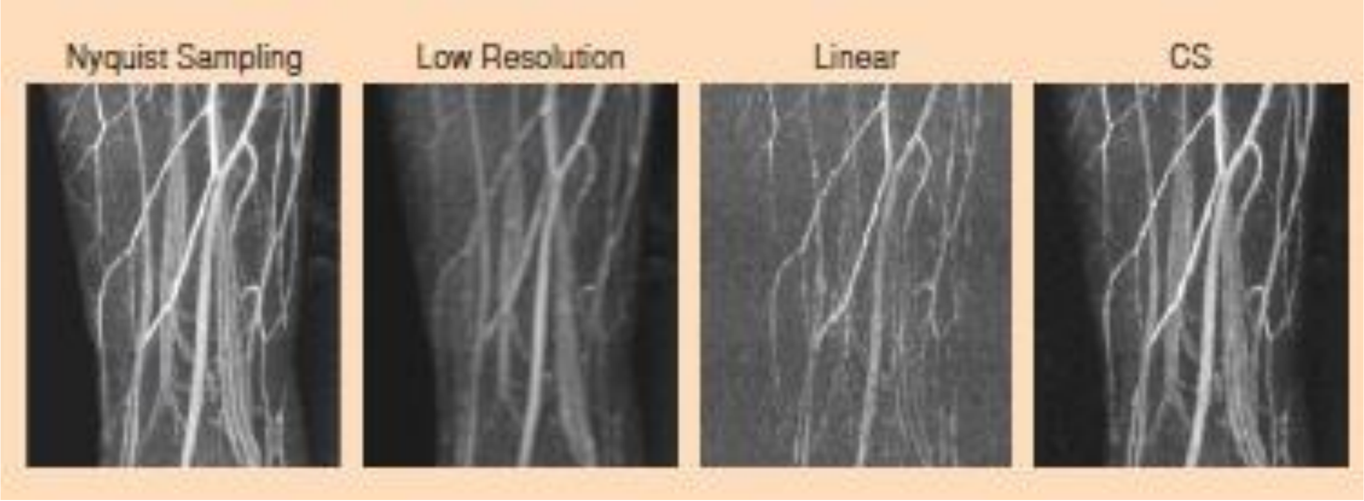


Image super-resolution

Yang et al

Sparse representation $\boldsymbol{\alpha}$

Hi-res. dictionary → $\boldsymbol{D}_{\mathrm{h}}\boldsymbol{\alpha}$

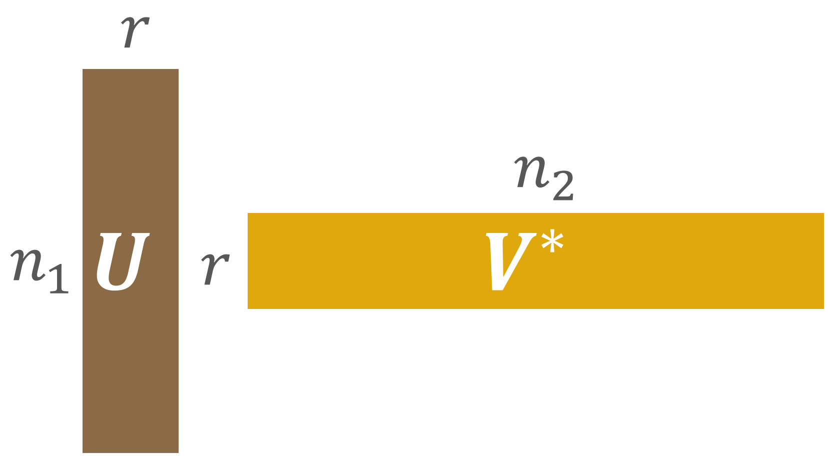
Low-res. dictionary → $\boldsymbol{D}_{\mathrm{l}}\boldsymbol{\alpha}$

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

$$y_i = \langle \boldsymbol{X}^\star = \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{A}_i \rangle$$



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

$$\underset{\boldsymbol{X}}{\operatorname{argmin}} \operatorname{rank}(\boldsymbol{X})$$
$$\text{subject to } \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$

If identifiable, it recovers the low-rank solution exactly

Just like ℓ_0-minimization, it is generally NP-hard

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

Random measurements

$\boldsymbol{A}_i$ with iid entries exhibit **low-rank RIP**

- Gaussian
- Rademacher
- Uniform

Universal

Some structured matrices

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

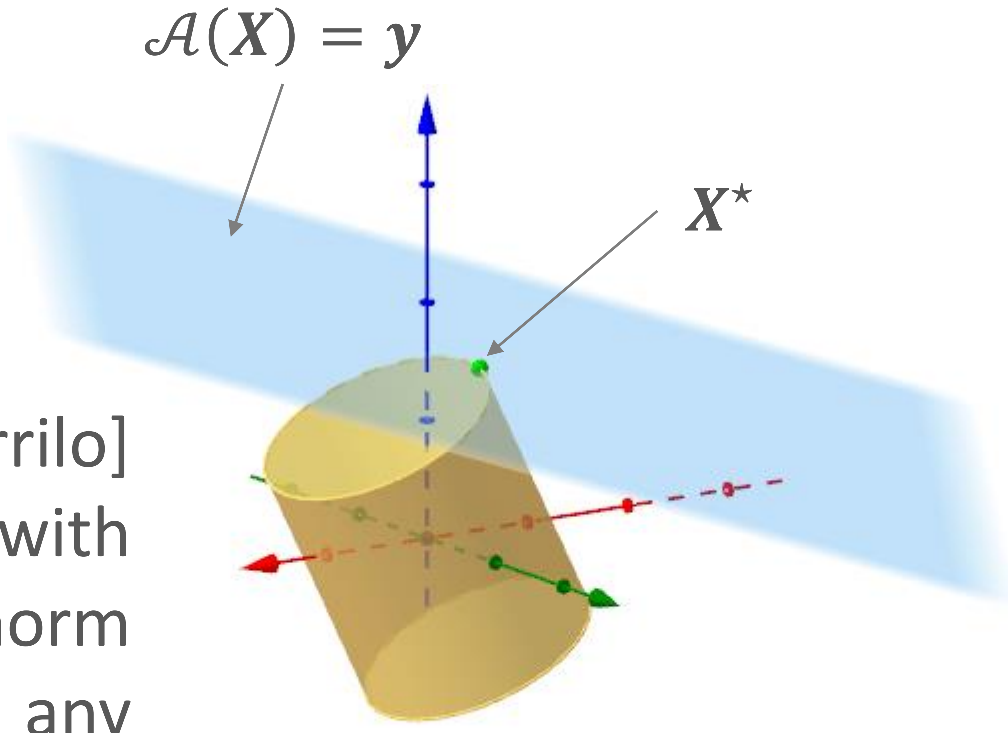
Instance optimal

Nuclear-norm minimization

$$\underset{\boldsymbol{X}}{\operatorname{argmin}} \|\boldsymbol{X}\|_*$$
$$\text{subject to } \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}$$

$$\|\boldsymbol{X}\|_* = \sum_i \sigma_i$$

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

Robust PCA

Candès et al

Ordinary PCA is sensitive to noise and outliers

low-rank component → sparse component

$$M = L + S$$

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

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

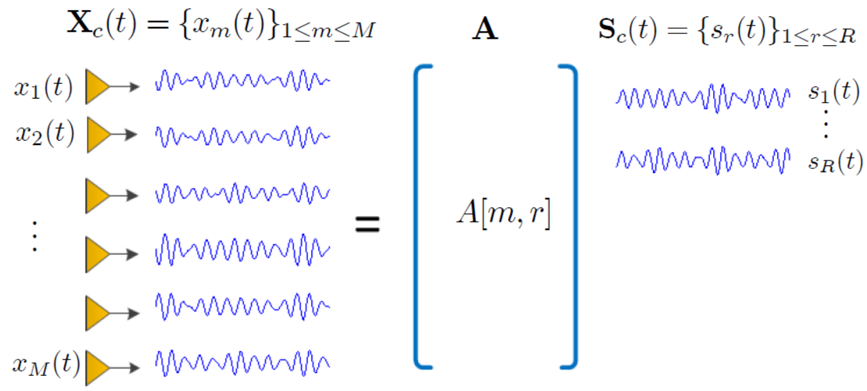
Background Modeling

(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]

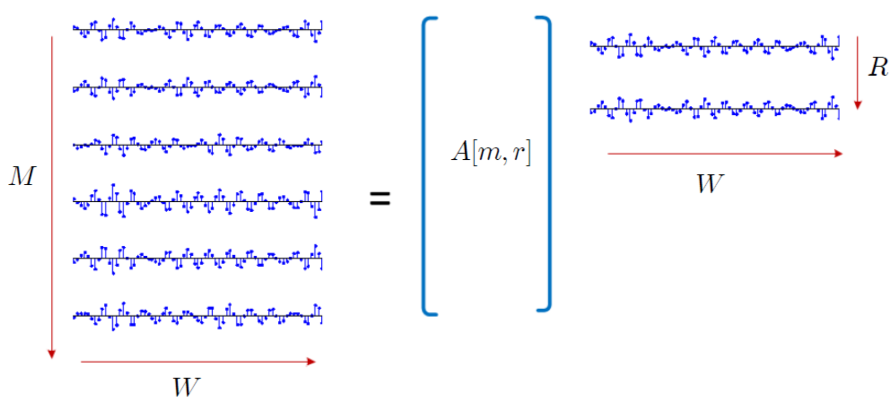
Compressive multiplexing

Ahmed and Romberg

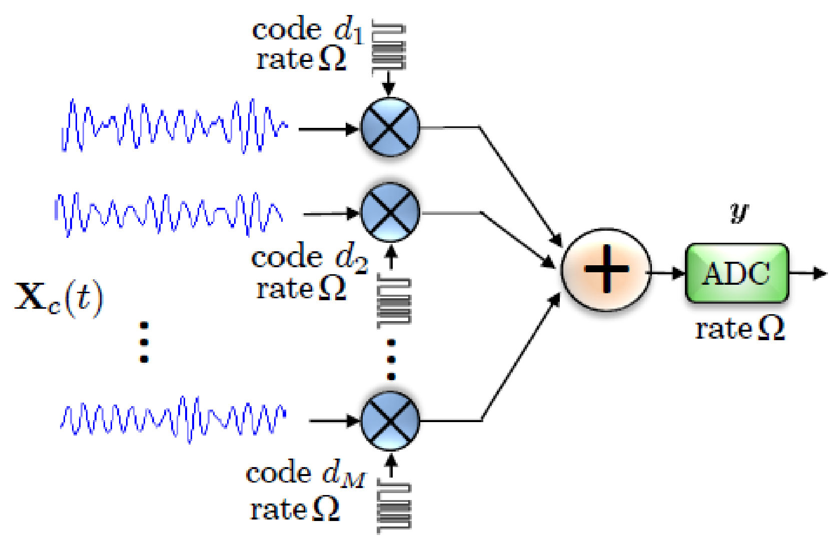


$$x_m(t) = \sum_{\omega=-B}^{B} \alpha_m[\omega] e^{j2\pi\omega t}$$

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



$$W = 2B + 1$$



Nonlinear CS

Sparsity-constrained minimization

Squared error isn't always appropriate

- example : non-Gaussian noise models

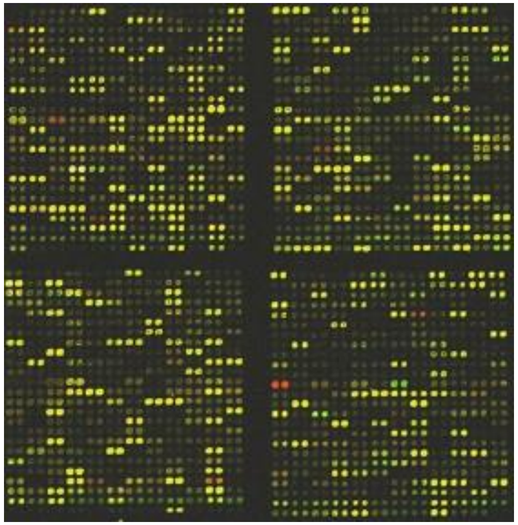
$$\operatorname{argmin}_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$\text{subject to } \|\boldsymbol{x}\|_0 \leq k$$

It's challenging in its general form

Objectives with certain properties allow approximations

- convex relaxation: ℓ_1-regularization
- greedy methods

Gene classification



DNA Microarray

$\boldsymbol{a}_i$: microarray sample

y_i : binary sample label (healthy/diseased)

model: $p(\, y \mid \boldsymbol{a}; \boldsymbol{x}\,)$

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

Sparse Logistic Regression

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

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

Imaging under photon noise

Harmany et al

Photon noise follows a Poisson distribution

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

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

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

$R(\boldsymbol{x})$: $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_{\mathrm{TV}}$, …

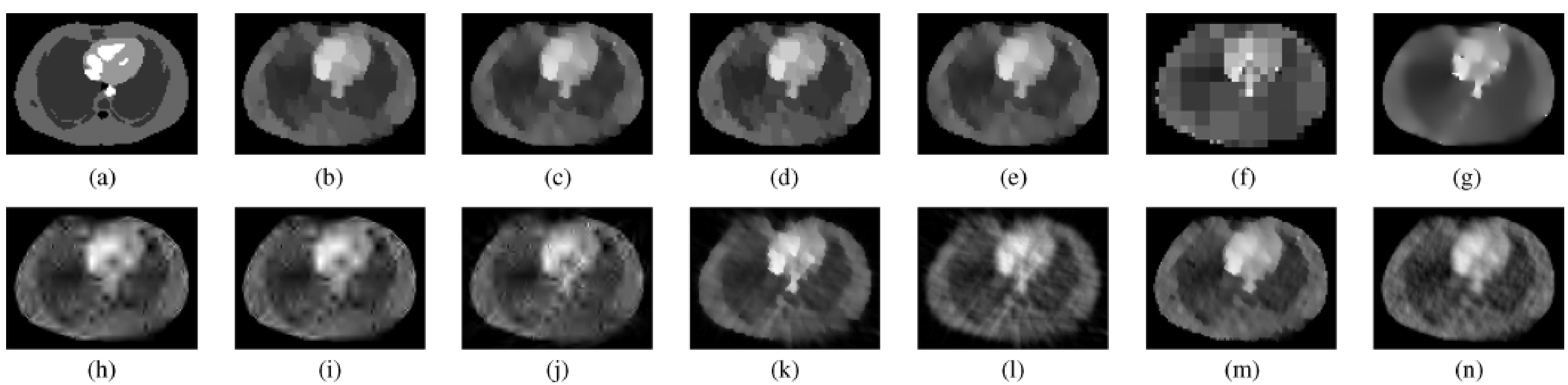
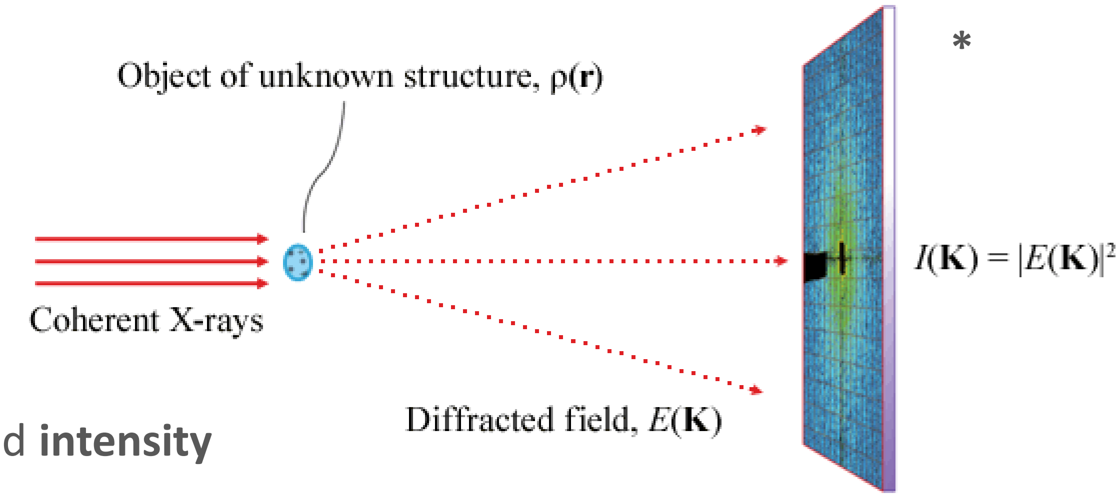


Fig. 3. Single-trial reconstructed images for all methods considered. Note RMSE $(\%) = 100 \cdot \|\hat{f} - f^\star\|_2 / \|f^\star\|_2$. (a) Ground truth. (b) SPIRAL-TV loose monotonic (RMSE $= 24.404\%$). (c) SPIRAL-TV loose nonmonotonic (RMSE $= 24.962\%$). (d) SPIRAL-TV tight monotonic (RMSE $= 24.526\%$). (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-ℓ_1 Loose (DB-6) (RMSE $= 28.626\%$). (i) SPIRAL-ℓ_1 Tight (DB-6) (RMSE $= 28.665\%$). (j) SpaRSA ℓ_2-ℓ_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

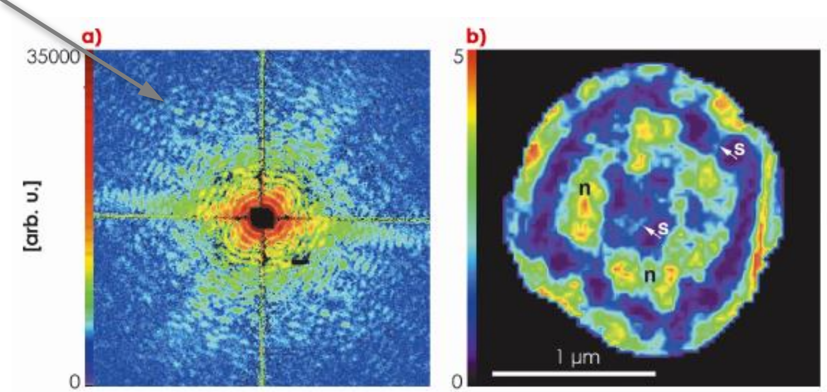
Szameit et al



*

I : diffracted field **intensity**

*



Cx **:** real image

C : Shifts of **generating function**

L **:** low-pass filter

F : Fourier transform

$$f(x) = \| |\boldsymbol{LFCx}|^2 - \boldsymbol{I} \|_2^2$$

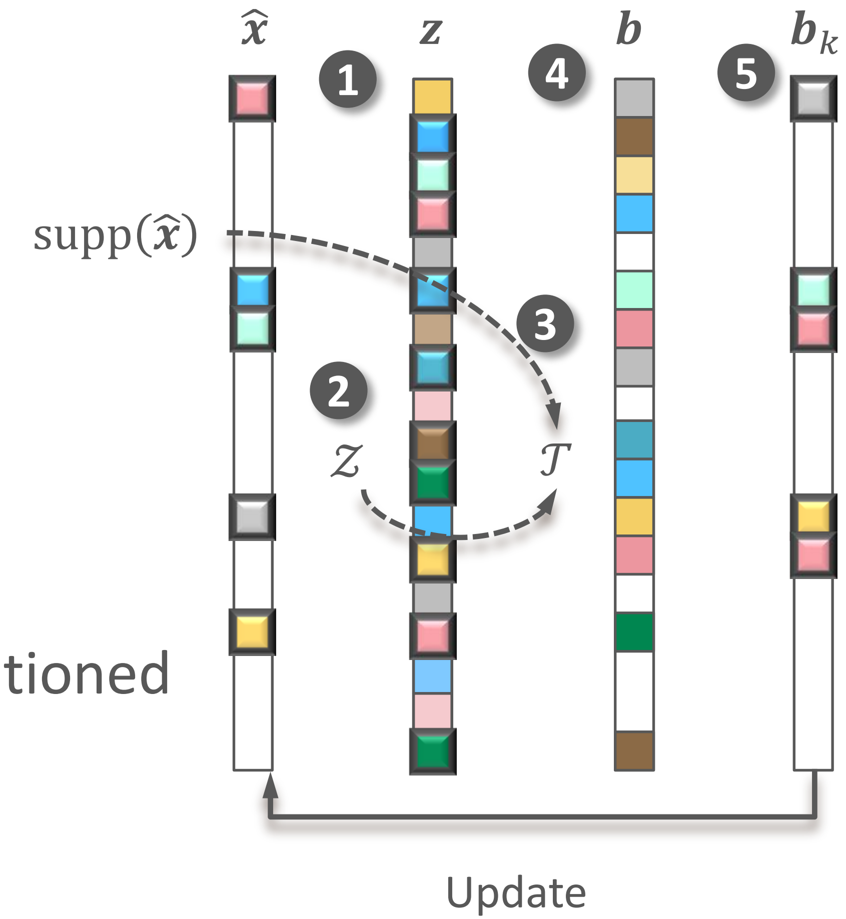
* *Lima et al.*, ESRF (www.esrf.eu)

Gradient Support Pursuit

w/ Raj and Boufounos

Generalizes CoSaMP

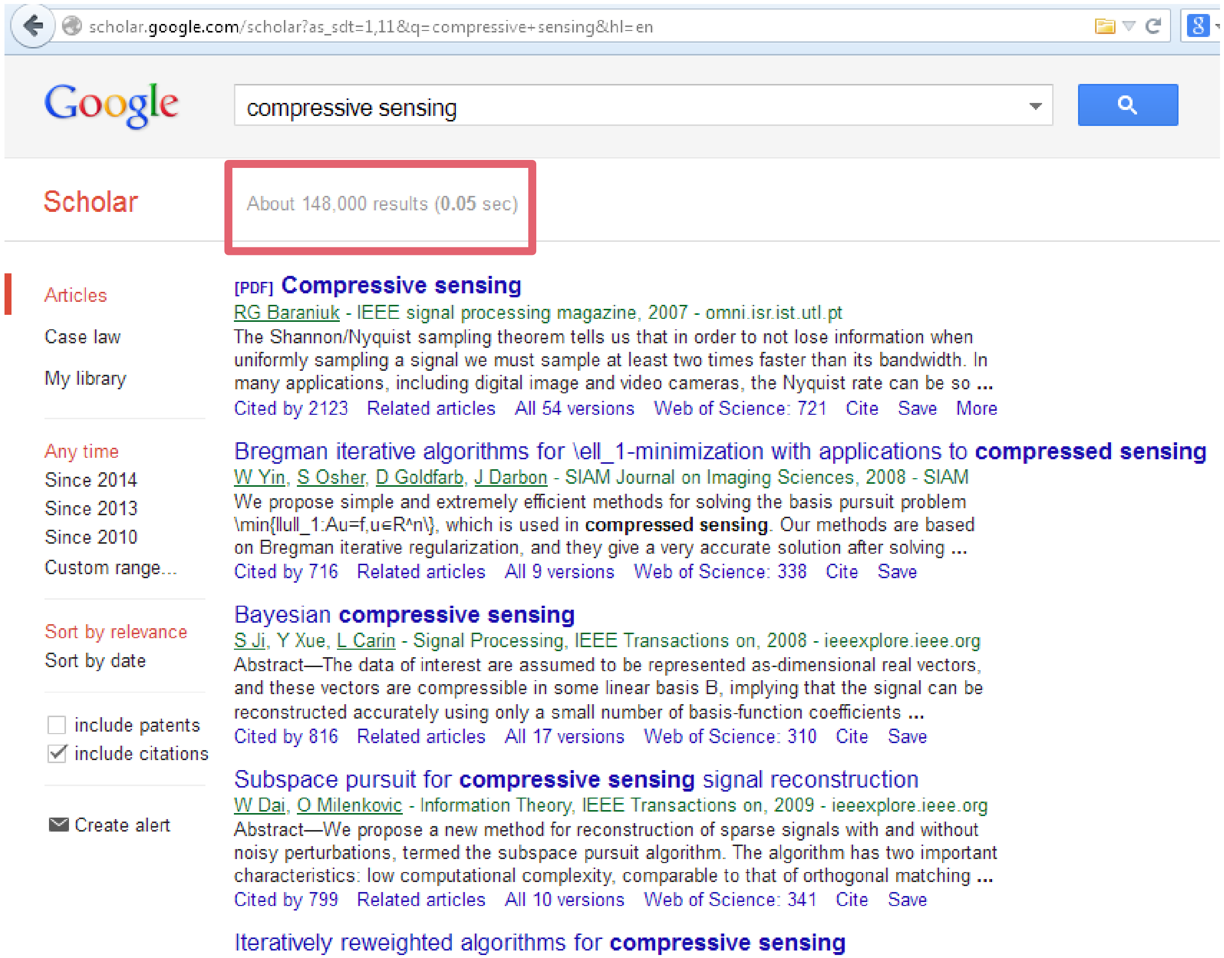
- Proxy vector $\boldsymbol{z} = \nabla f(\boldsymbol{x})$
- $\mathcal{Z} = \mathrm{supp}(\boldsymbol{z}_{2k})$
- $\boldsymbol{b} = \mathrm{argmin}_{\boldsymbol{x}}\, f(\boldsymbol{x})\;$ s.t. $\boldsymbol{x}\,|_{\mathcal{T}^{\mathrm{c}}} = \boldsymbol{0}$



Converges under SRH

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

There's much more …



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

a weblog focusing on CS and broader computational areas

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