A Mathematical Introduction to Compressed Sensing

Stefania Petra
Mathematical Imaging Group
Heidelberg University

Lecture WT 2019/20

Contact

Stefania Petra, MIG - lecture
Room A4 209
Email petra@math.uni-heidelberg.de

Jan Plier, MIG - tutorial
Room A4 212
Email jan.plier@iwr.uni-heidelberg.de

Organizational Matters

Time/Location

Lecture SR 6 - Mathematikon INF 205

Wed. 9:00 - 11:00

Tutorial SR 2 - Mathematikon INF 205

Thu. 16:00 - 18:00

Evaluation Oral exam ← 50% programming exercises

Creditpoints 6CP **+** 2CP optional programming project

Media forms blackboard / lecture notes / slides

Previous knowledge Linear algebra, analysis I + II
(+ basic tools from probability theory,
convex analysis & optimization)

Web

Check your email and

`http://ipa.iwr.uni-heidelberg.de/dokuwiki/doku.php?id=teaching`

for announcements and updates of

- exercise sheets
- handouts
- lecture notes

Please also sign up using MÜSLI
`https://muesli.mathi.uni-heidelberg.de`

Literature

S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013

S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004

M. Ledoux, *The Concentration of Measure Phenomenon* American Mathematical Society, 2005

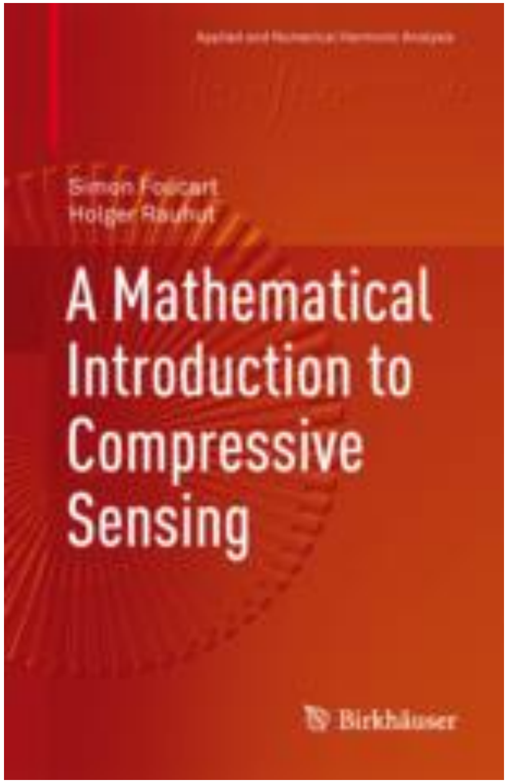
R. Schneider, W. Weil, *Stochastic and Integral Geometry*, Springer, 2008

J.-L. Starck, F. Mutagh, J.M. Fadili, *Sparse Image and Signal Processing*, Cambridge University Press, 2010

Literature

main reference

S. Foucart, H. Rauhut
Birkhäuser, 2013



Simon Foucart
Holger Rauhut

A Mathematical Introduction to Compressive Sensing

Birkhäuser

covers: 2000 - 2012

Content

Theory

sparse reconstruction via l0/l1-minimization;
basic properties: coherence, nullspace property, restricted isometry property; random sensors; phase transitions;
basic tools from convex analysis, probabilities and integral geometry

Algorithms

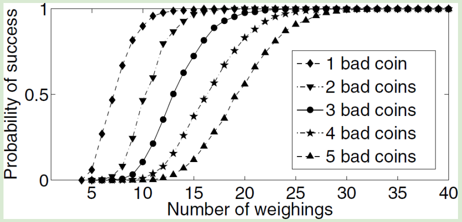
orthogonal matching pursuit; thresholding based methods;
primal-dual methods

Applications

sparse approximation; image processing (tomographic inversion, deblurring, etc.); low-rank completion

Overview

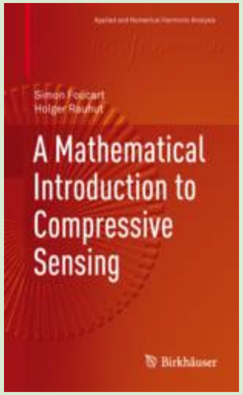
CS principles



idealized scenarios

enable

established theory



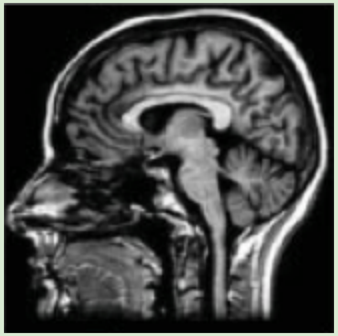
Simon Foucart
Holger Rauhut

A Mathematical Introduction to Compressive Sensing

Birkhäuser

applies

applications

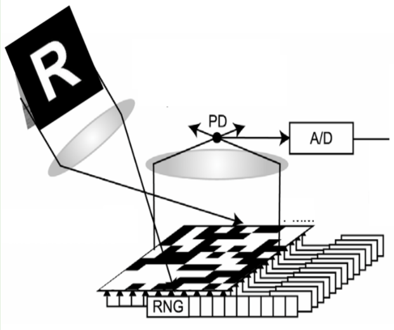


classical MRI

applies

applications

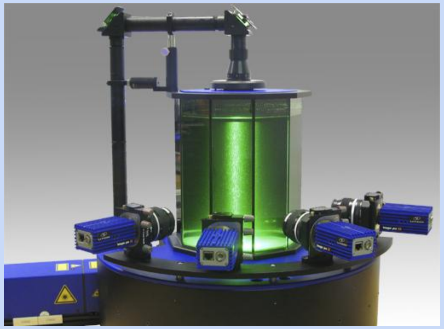
surveillance



single pixel camera

RIP fails

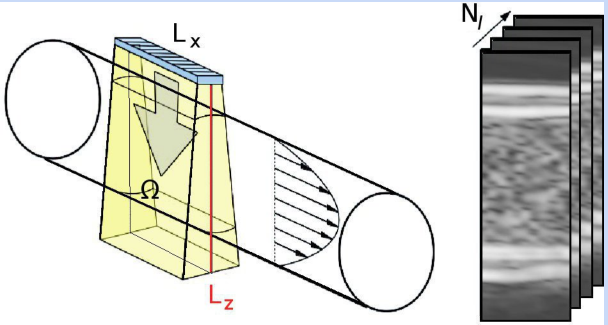
applications



non-standard tomography

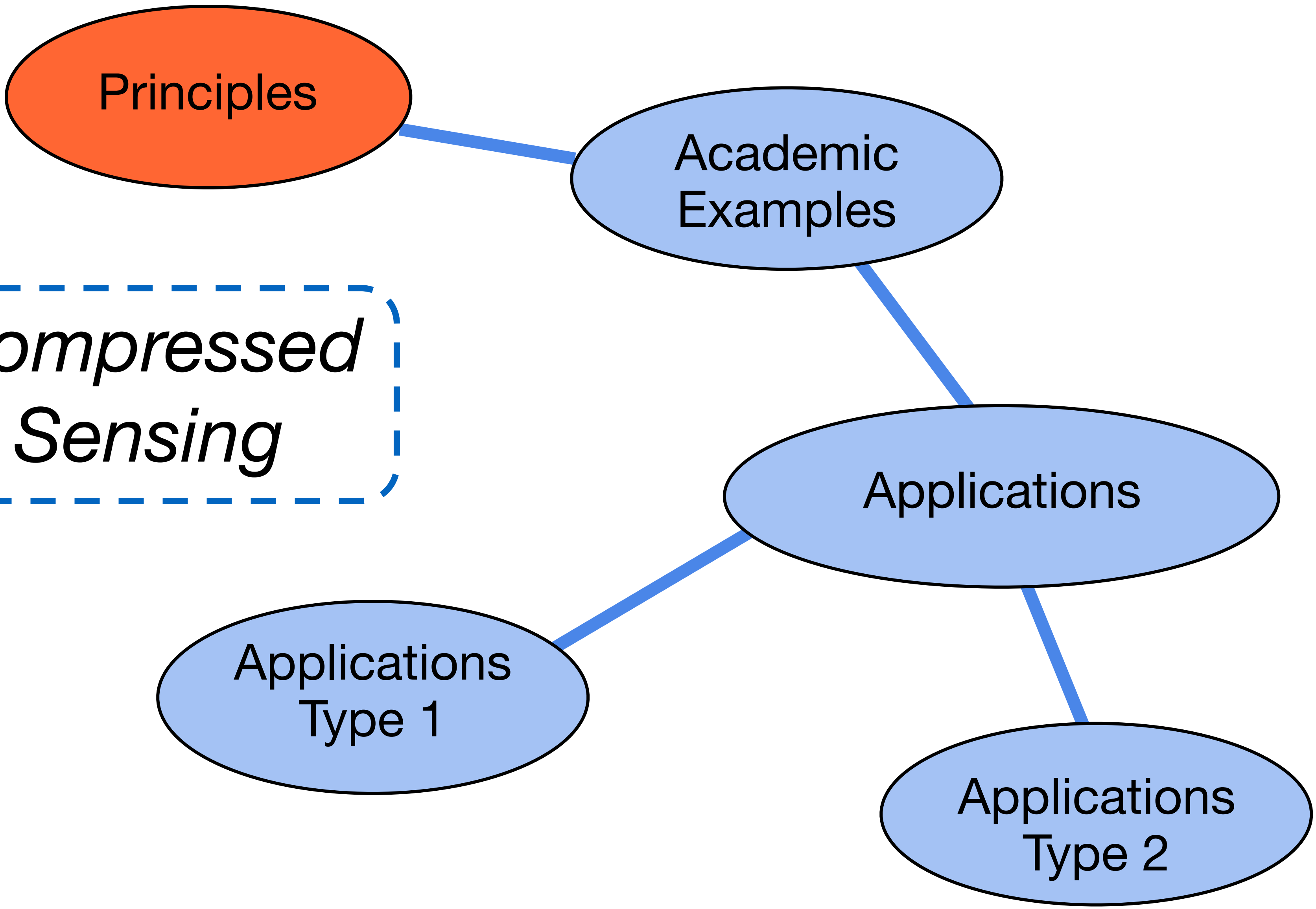
RIP fails

applications



ultrasound PIV

Compressed Sensing



Classical Sampling vs Compressed Sensing

Claude
Shannon

Emmanuel
Candés

Harry
Nyquist

Terence
Tao

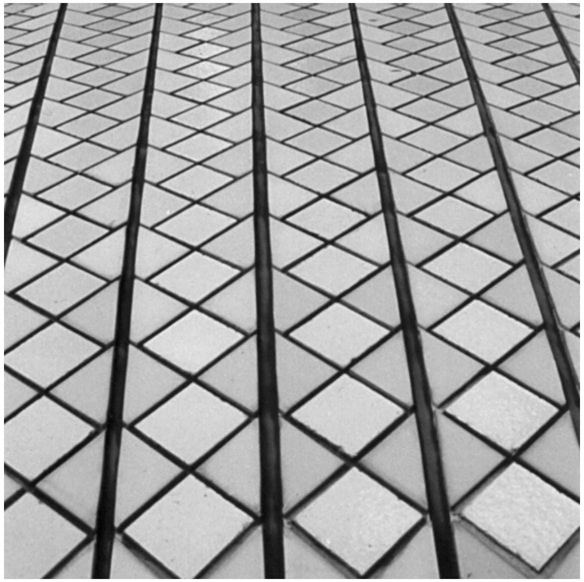
David
Donoho

Classical Sampling

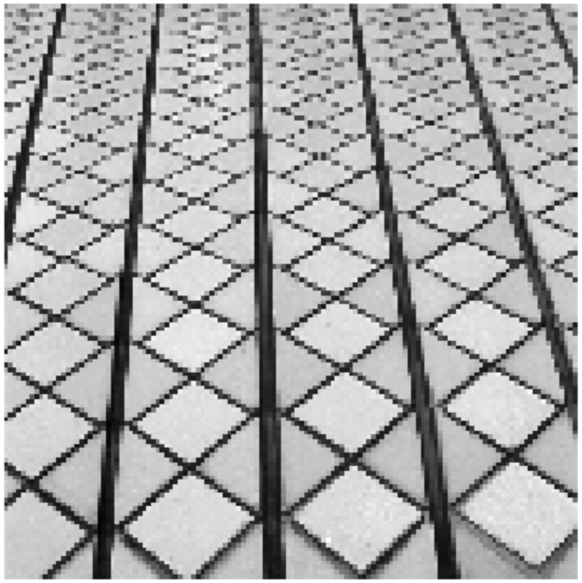
Shannon-Nyquist theorem

finite support in
the Fourier domain

correct sampling rate
correct (continuous) recovery



incorrect sampling rate
severe artefacts



… but “big data”

Compressed Sensing

encoding below the Nyquist rate: $m \ll n$

$$b = Ax \qquad x \in \mathbb{R}^n \, , \, b \in \mathbb{R}^m$$

observation measurement matrix (“sensor”)

Compressed Sensing

encoding below the Nyquist rate: $m \ll n$

$$b = Ax \qquad x \in \mathbb{R}^n \, , \, b \in \mathbb{R}^m$$

observation — measurement matrix (“sensor”)

nonlinear decoding by convex programming, e.g.

$$\Delta(b) = \underset{x:\, Ax=b}{\operatorname{argmin}} \, \|x\|_1$$

performance of en-/decoding pair: $\|x - \Delta(Ax)\|$

Compressed Sensing: Basic Requirements

When does it (provably) work?!

small support in *some* *transformed* domain

(1)Signals are **sparse** ...

$$\mathcal{X}_k := \{x \in \mathbb{R}^n : \|x\|_0 = |\mathrm{supp}(x)| \leq k\}$$

… or have a **sparse representation** $x = Dz$
(basis, dictionary; henceforth for simplicity: $D = I$)

Compressed Sensing: Basic Requirements

When does it (provably) work?!

small support in *some transformed* domain

(1)Signals are **sparse** ...

$$\mathcal{X}_k := \{x \in \mathbb{R}^n : \|x\|_0 = |\mathrm{supp}(x)| \leq k\}$$

… or have a **sparse representation** $x = Dz$
(basis, dictionary; henceforth for simplicity: $D = I$)

(2)Sensor matrix A is an isometry on $\mathcal{X}_k$

$$\exists \delta_k \in (0,1), \quad \forall x \in \mathcal{X}_k$$

$$(1-\delta_k)\|x\|^2_{\mathbb{R}^n} \leq \|Ax\|^2_{\mathbb{R}^m} \leq (1+\delta_k)\|x\|^2_{\mathbb{R}^n}$$

“restricted isometry property (RIP)”

Some Basic Questions

- Can we trust our model to return an intended sparse signal?
- Does the model have a unique solution? (otherwise, different algorithms may return different answers)
- Is the solution exactly equal to the original sparse signal?
- If not (due to noise), is the solution a faithful approximation of it?
- How much effort is needed to numerically solve the model?

Compressed Sensing: Stable Recovery

Candés, Romberg, Tao 2006

$$\delta_{2k} < \sqrt{2} - 1$$ (RIP)

$$b = Ax^* + \xi,\ \|\xi\|_2 \le \varepsilon$$ (bounded noise)

$$\boxed{\Delta(b) = \operatorname{argmin}_{x:\ \|Ax-b\|_2 \le \varepsilon} \|x\|_1}$$

decoding

Stable recovery guarantee

$$\|x^* - \Delta(b)\|_2 \le C\left(\frac{\|x^* - x^*_k\|_1}{\sqrt{k}} + \varepsilon\right)$$

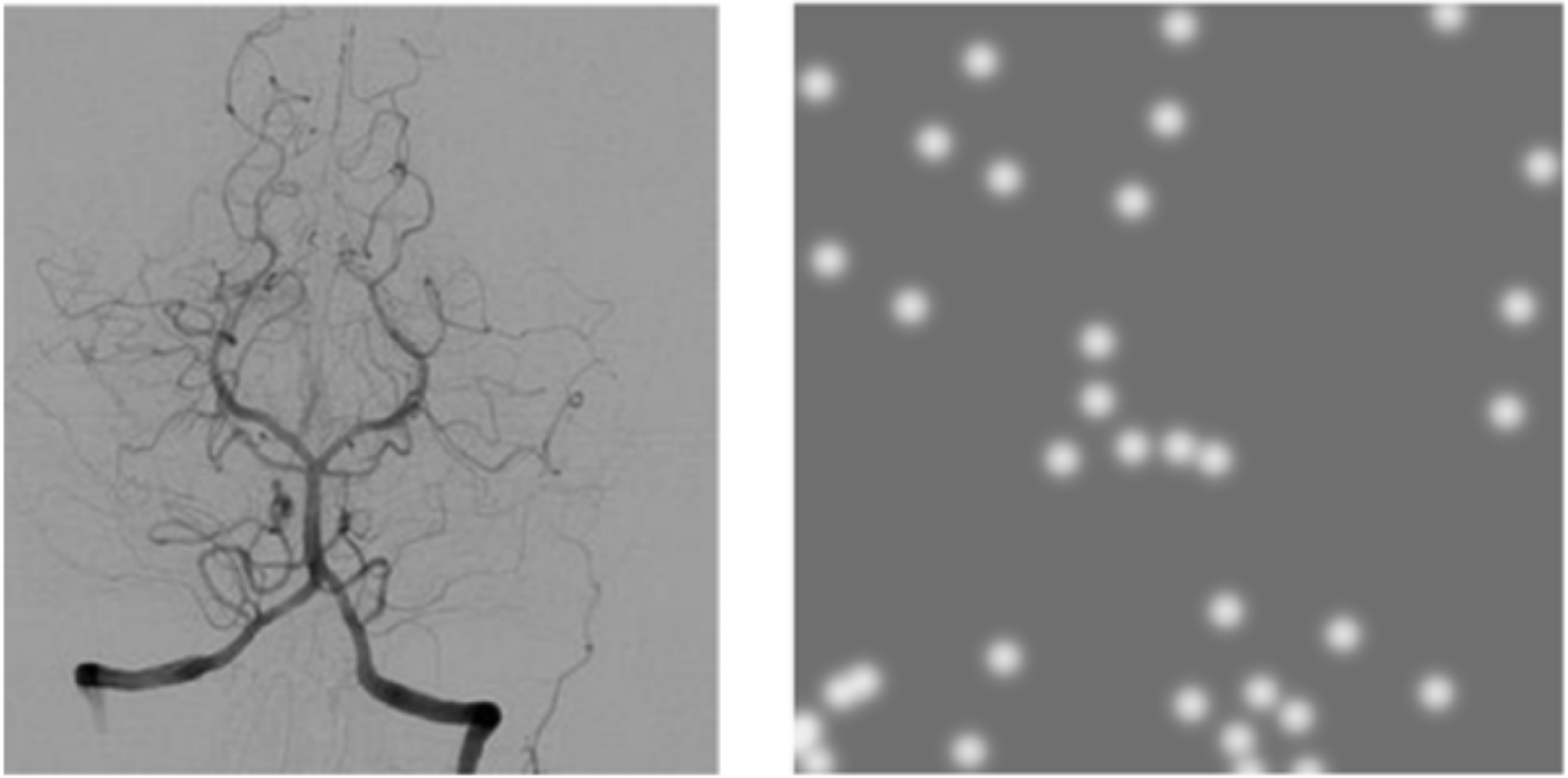
$\varepsilon \to$ *0* and *k*-sparsity ⇒ perfect reconstruction

How to Read Recovery Guarantees

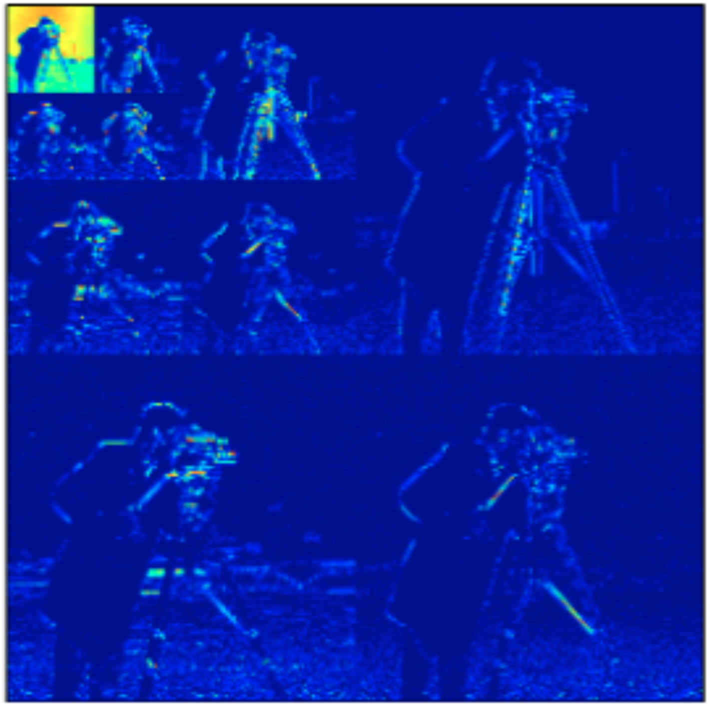
Some basic aspects that distinguish different types of guarantees:

- Recoverability (exact) vs stability (inexact)
- General A or special A?
- Universal (all sparse vectors) or instance (certain sparse vector(s))? Uniform/non uniform recovery guarantees?
- General optimality? or specific to model/algorithm?
- Required property of A: spark, NSP, coherence, RIP, dual certificate?
- If randomness is involved, what is its role?

Sparsity in Coordinate Basis



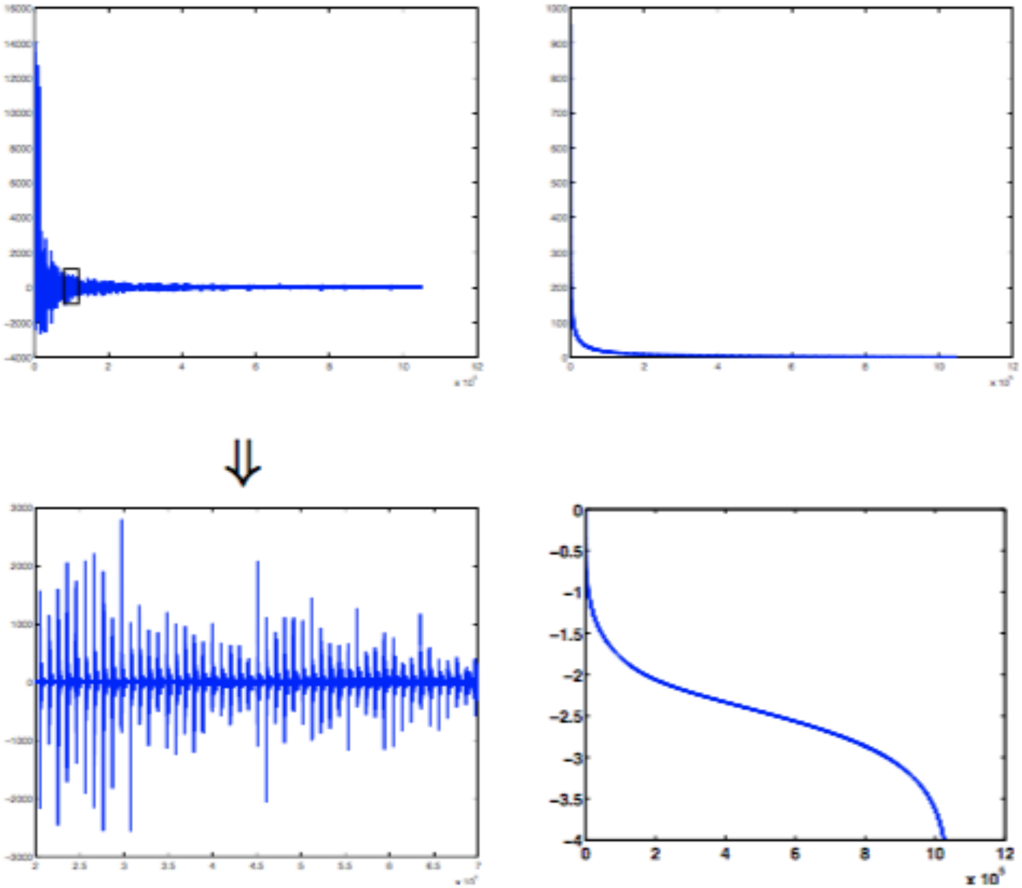
Sparsity in Orthonormal Basis



$H^e_{j,k}(x) = 2^j H^e(2^j x - k), \quad e \in V = \{\{0,1\},\{1,0\},\{1,1\}\}$ and
$j \geq 0, \quad k \in \mathbb{Z}^2 \cap 2^j[0,1)^2$

Sparsity vs Compressibility

1 megapixel image



wavelet coeffs

(sorted)

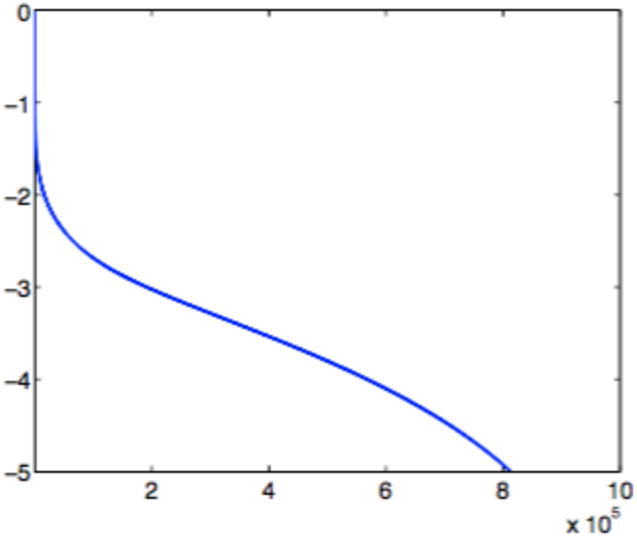
zoom in

($\log_{10}$ sorted)

Sparsity vs Compressibility

1 megapixel image

25k term approx

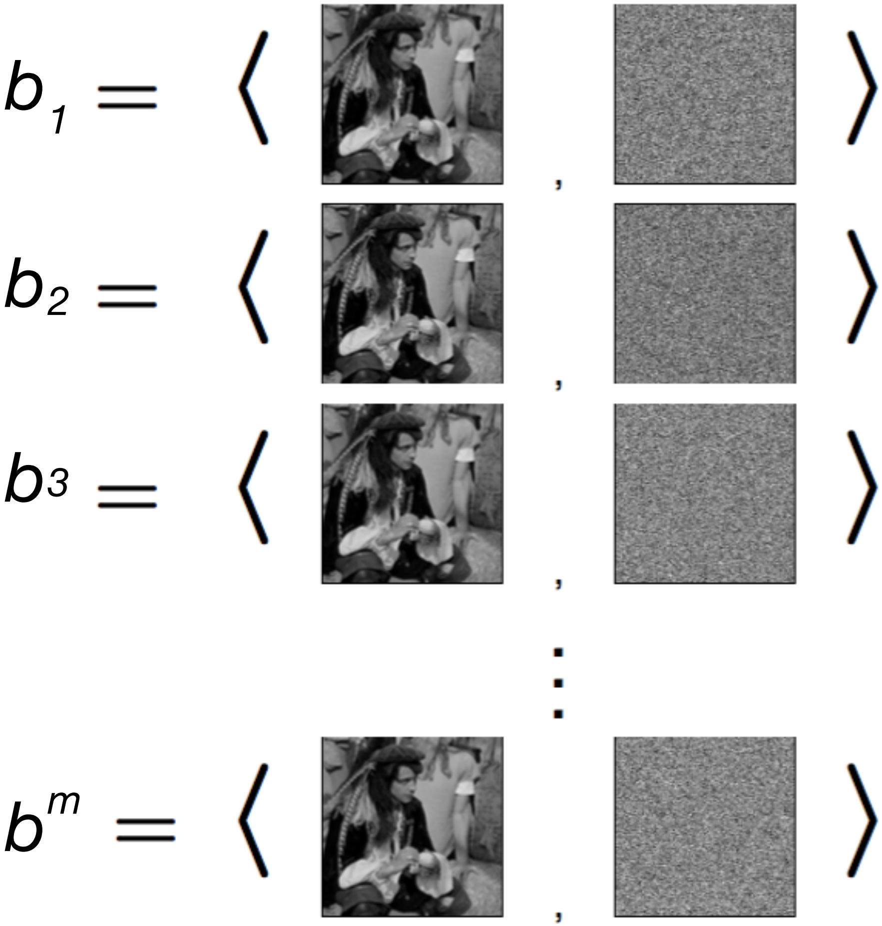


B-term approx error

- Sparse structure in Wavelet domain: few large coeffs, many small coeffs
- Basis for JPEG2000 image compression standard
- Do not confuse image compression with CS

Good Sensors?

- preserve structure and information in sparse/ compressible signals models with high probability
- the number of samples should be minimal
- each measurement carries the same amount of information



Surprise

- measurements do not match image structure at all
- measurements look like random noise
- same measurements can be used for **any** compressible signal class (universal)

Random Sensors: Sufficient RIP Conditions

Mathematical ways to design good sensor matrices A enjoying RIP include

- Gaussian i.i.d. entries $A_{ij} \sim \mathcal{N}(0, \frac{1}{m})$, Bernoulli, …
- random partial Fourier (DFT) matrices

Random Sensors: Sufficient RIP Conditions

Mathematical ways to design good sensor matrices A enjoying RIP include

- Gaussian i.i.d. entries $A_{ij} \sim \mathcal{N}(0, \frac{1}{m})$, Bernoulli, …
- random partial Fourier (DFT) matrices

\# measurements

Then $\exists c_1 = c_1(\delta), c_2 = c_2(\delta)$ such that

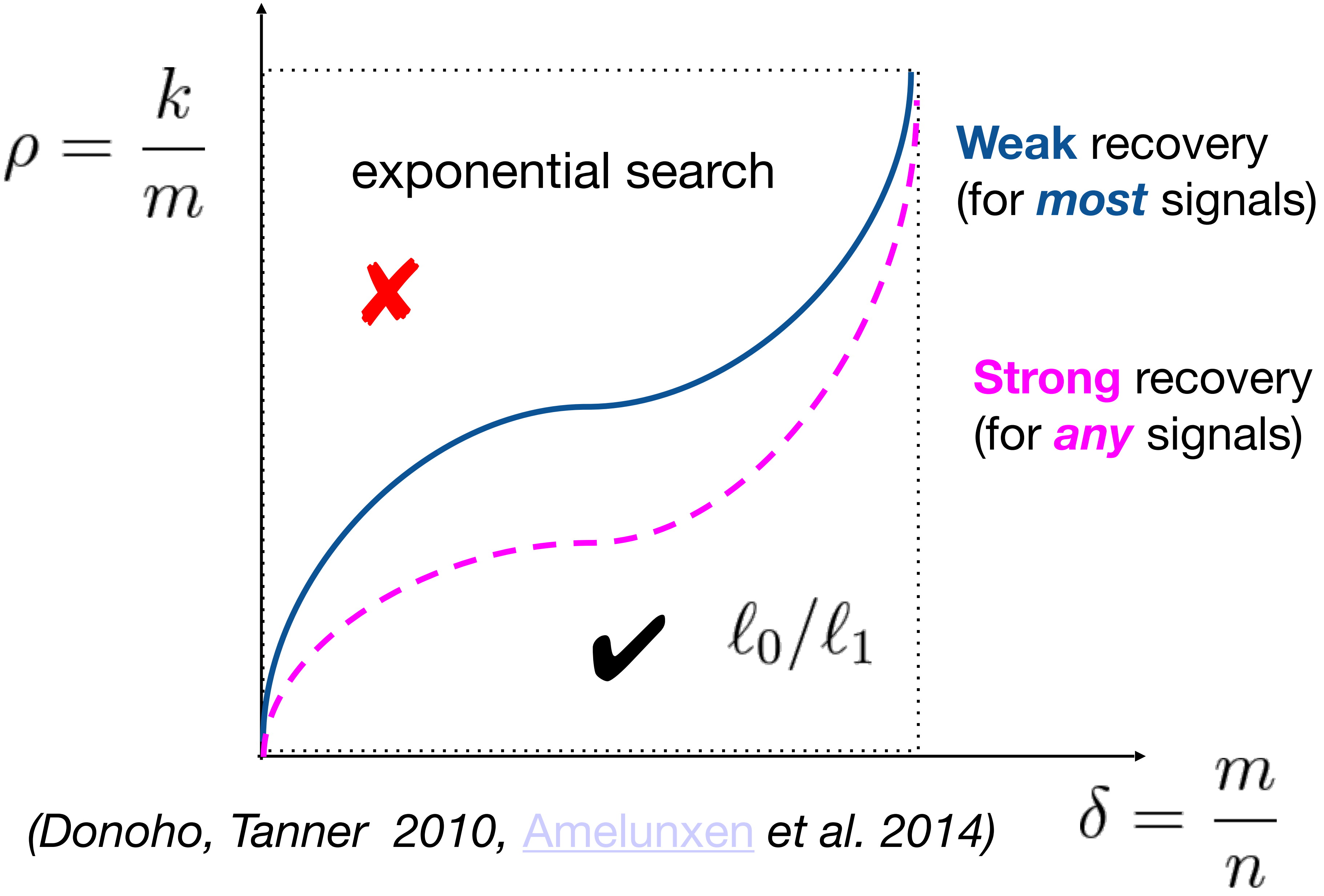
$$\Pr\left(A \,\text{has}\, \text{RIP}_{\text{k}}\right) \geq 1 - 2e^{-c_1 m}$$

provided the number of measurements satisfies

$$\underset{\uparrow}{m} \geq c_2 \underset{\uparrow}{k} \log(n/k)$$

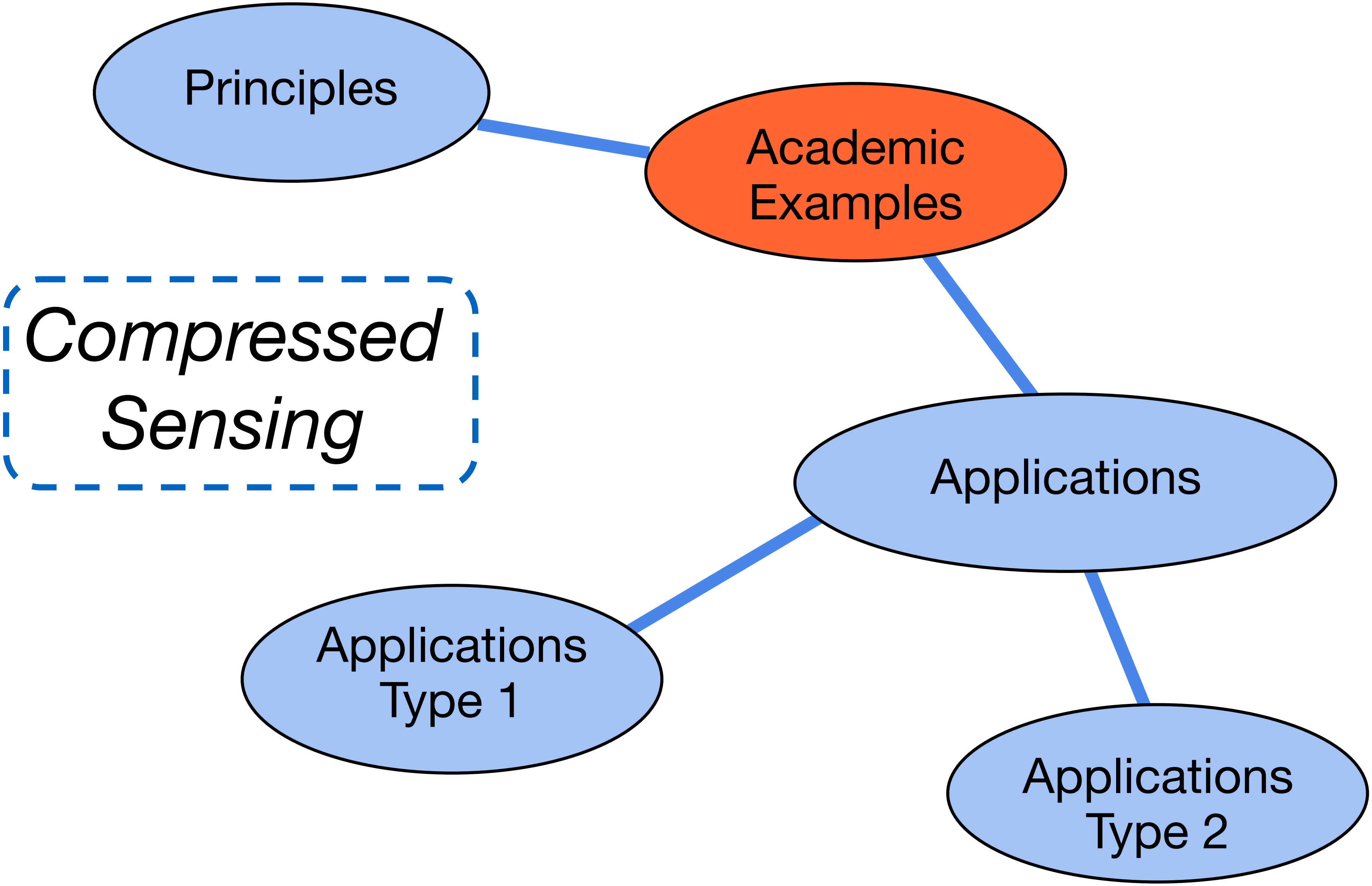
($m \propto$ sparsity !)

Known Phase Transitions: Gaussian *A*



(Donoho, Tanner 2010, Amelunxen et al. 2014)

Compressed Sensing



Academic Example

$$\underbrace{\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}}^{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

max. 3 weighings!!!

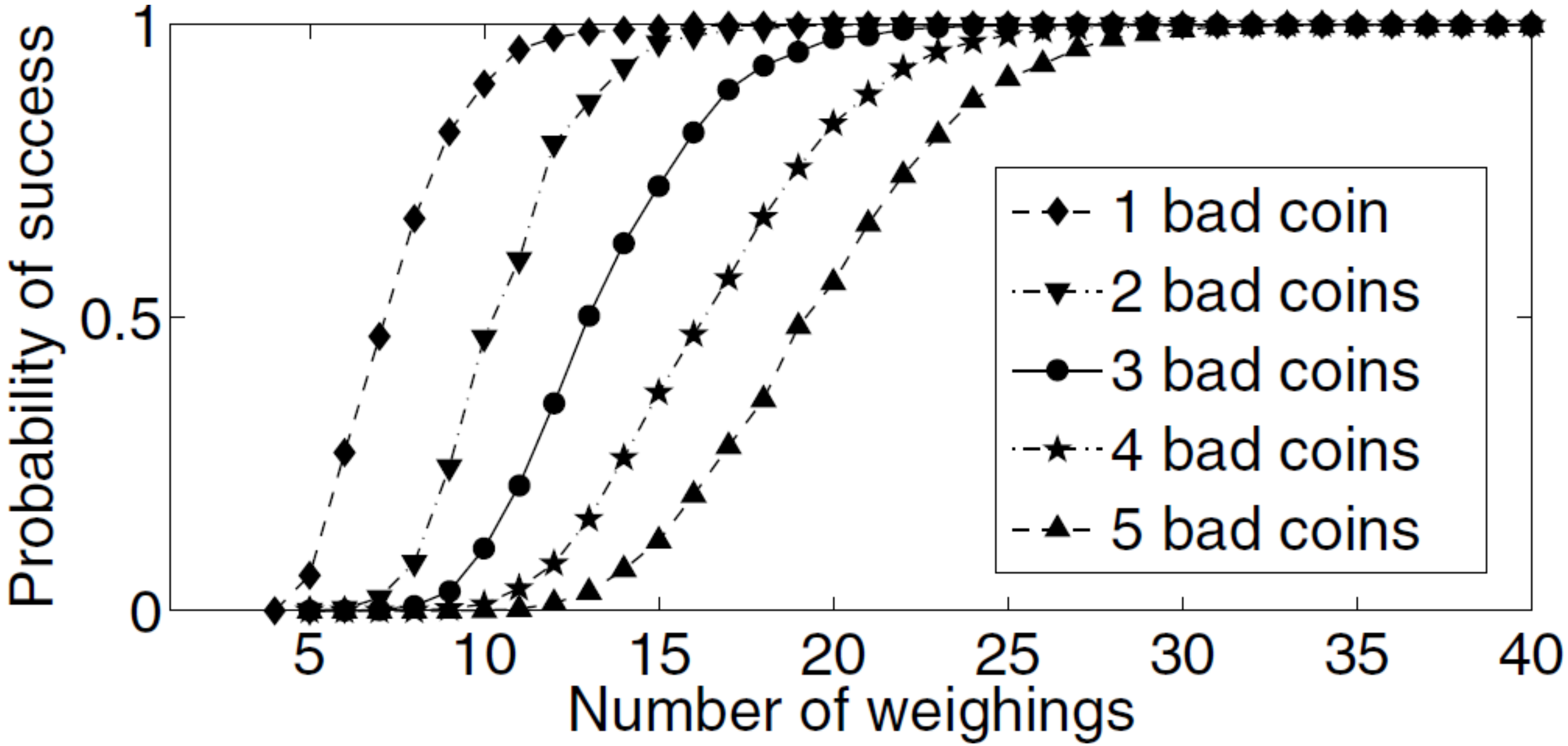
x_j	deviation of the j th coin from the correct mass
b_i	i th weighting

One coin is false. Which one?

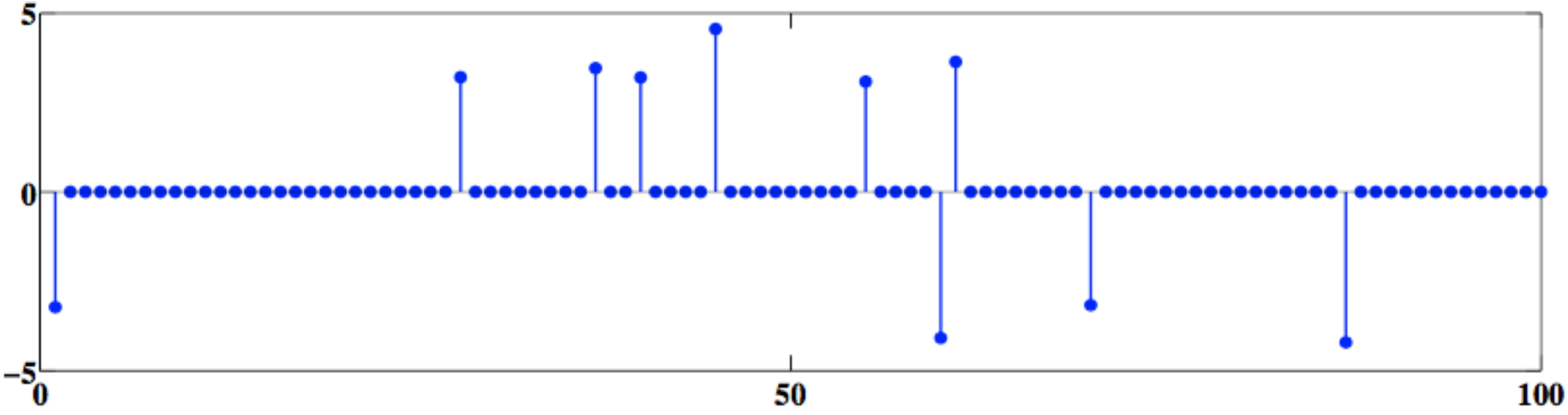
Weigh Random Subsets

Example: $n = 100$ coins, m weighings

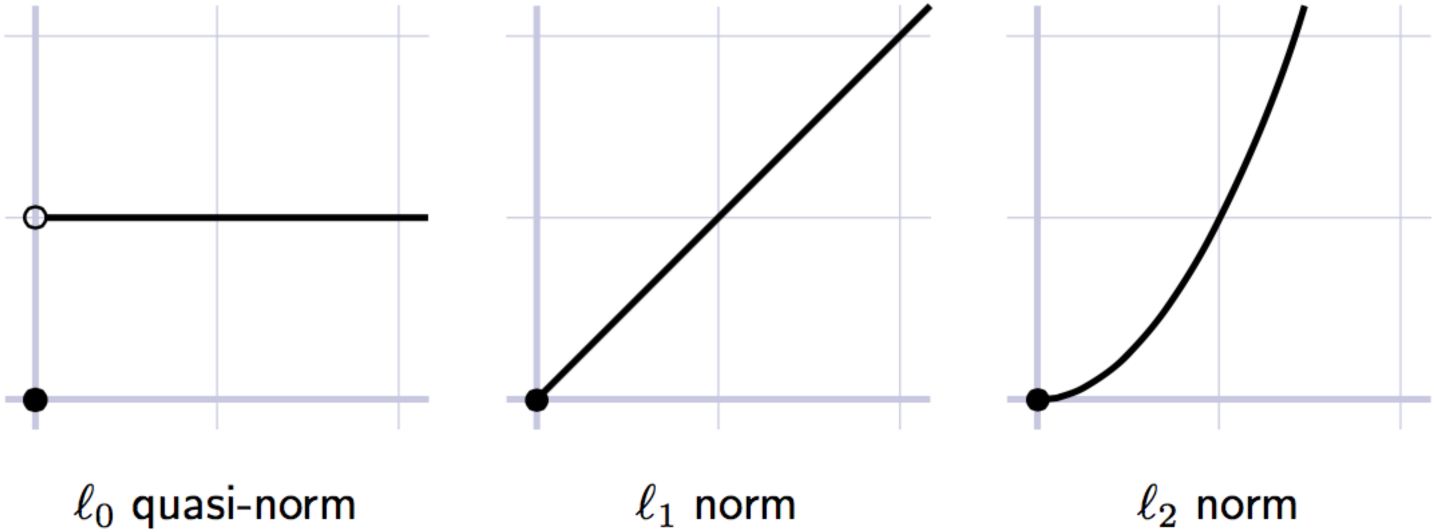
A *Bernoulli* $m \times n$ underdetermined !



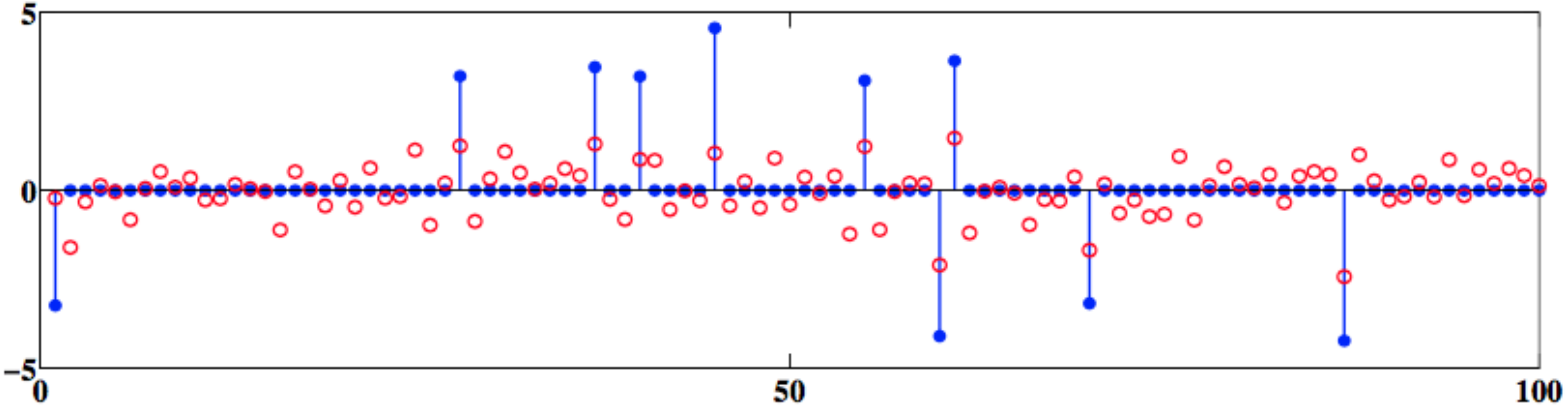
Academic Example

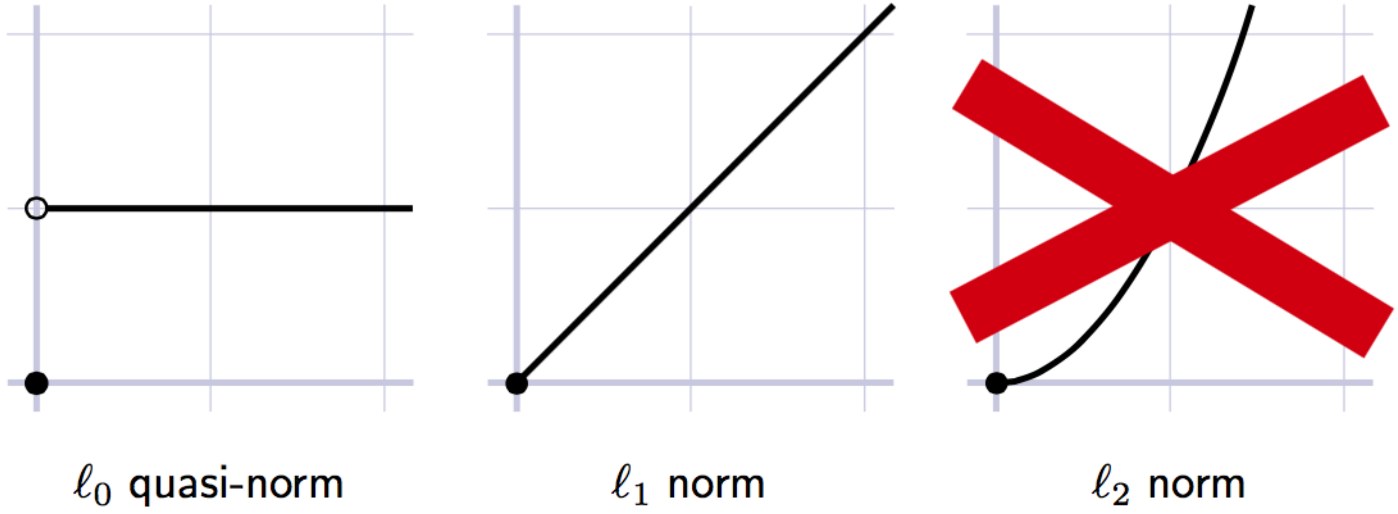


$x \in \mathbb{R}^{100}, \|x\|_0 = 10, b = Ax, A \in \mathbb{R}^{30\times 100}$ Gaussian

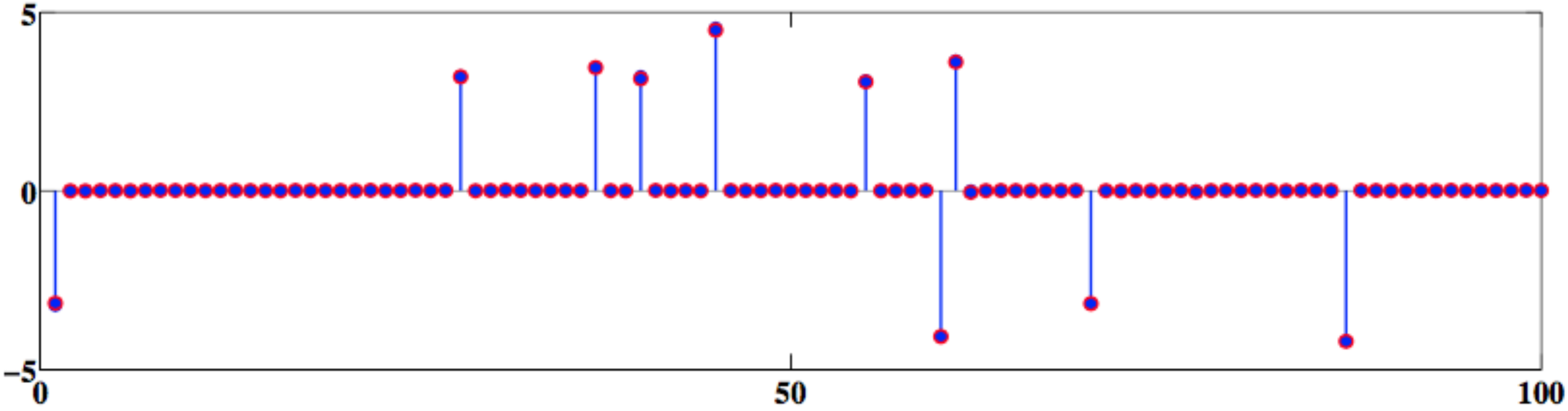


Academic Example

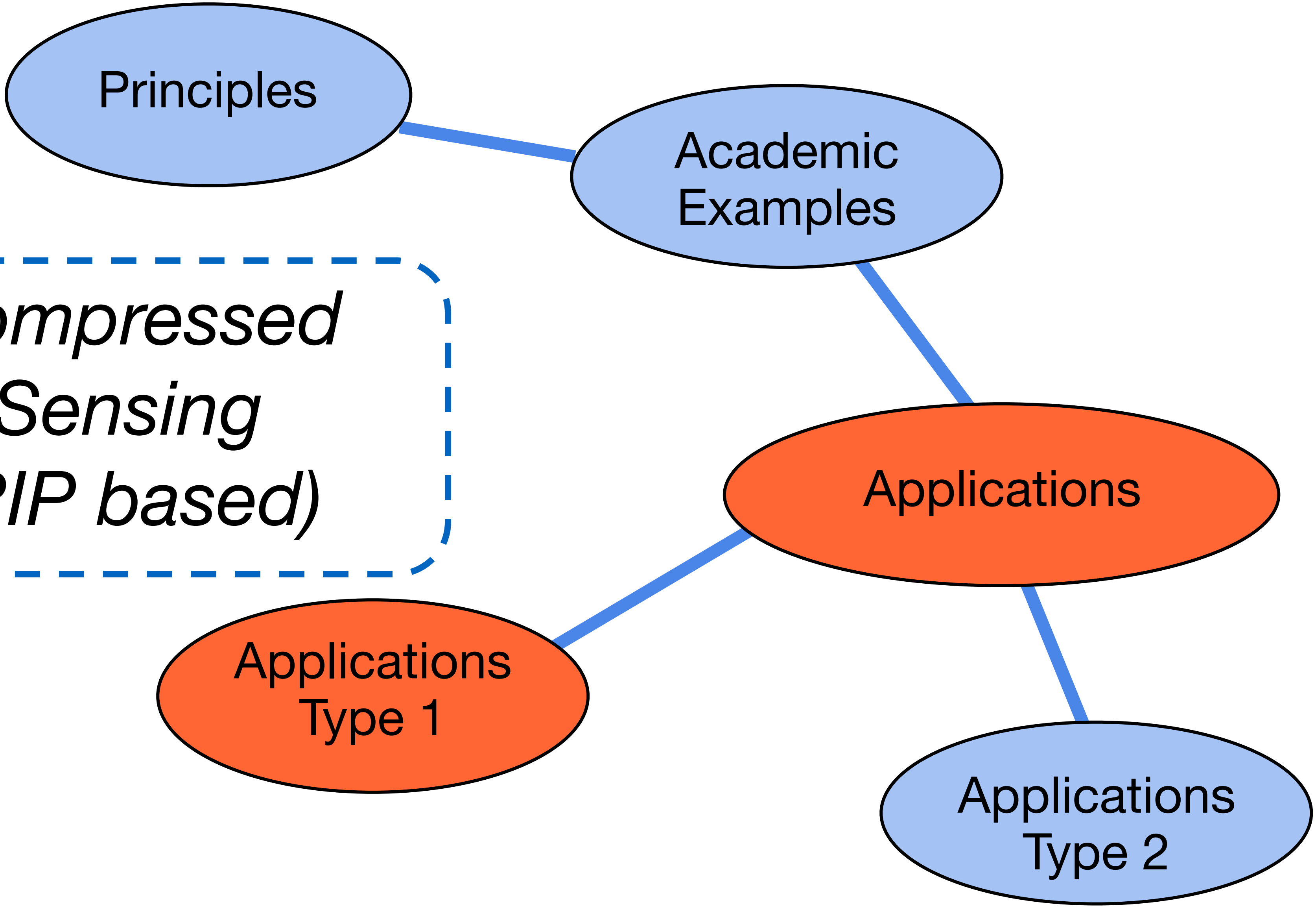




Academic Example



- ℓ_1 can be used as a proxy for ℓ_0
- This is a convex program and can be solved in polynomial time


Principles

Academic Examples

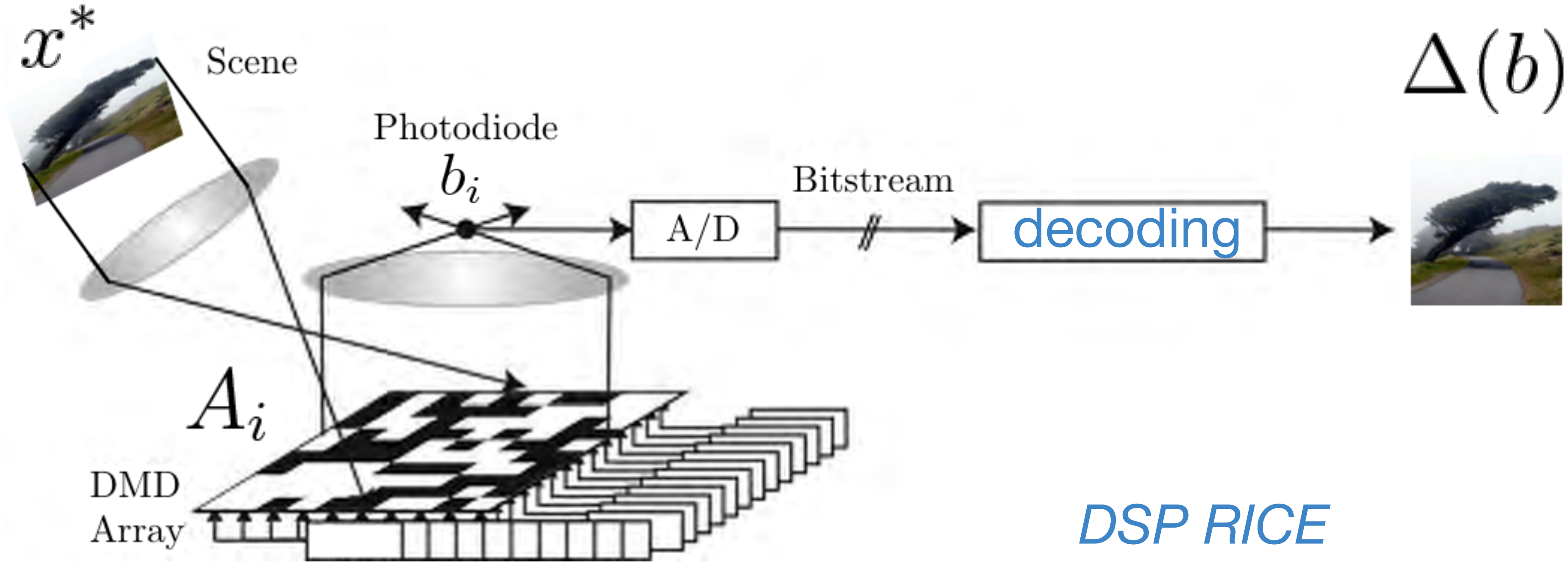
Applications

Applications Type 1

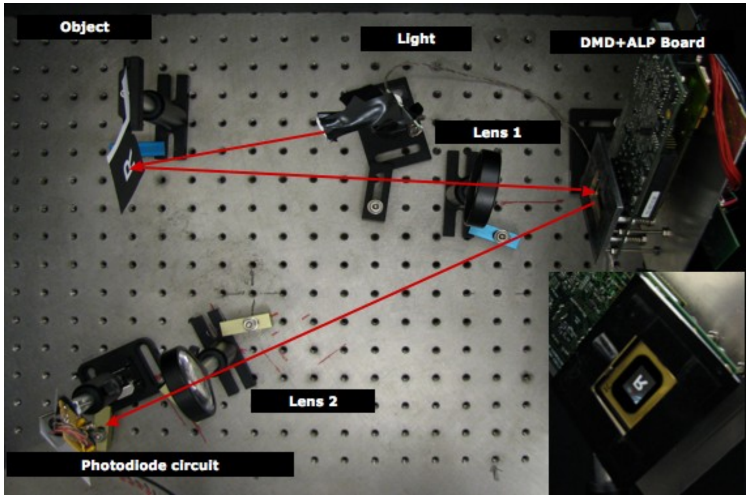
Applications Type 2

Compressed Sensing (RIP based)

Single Pixel Camera



DSP RICE



$$m = O(n^{1/4}\log^5(n)) \ll n$$

Single Pixel Camera

target
65536 pixels

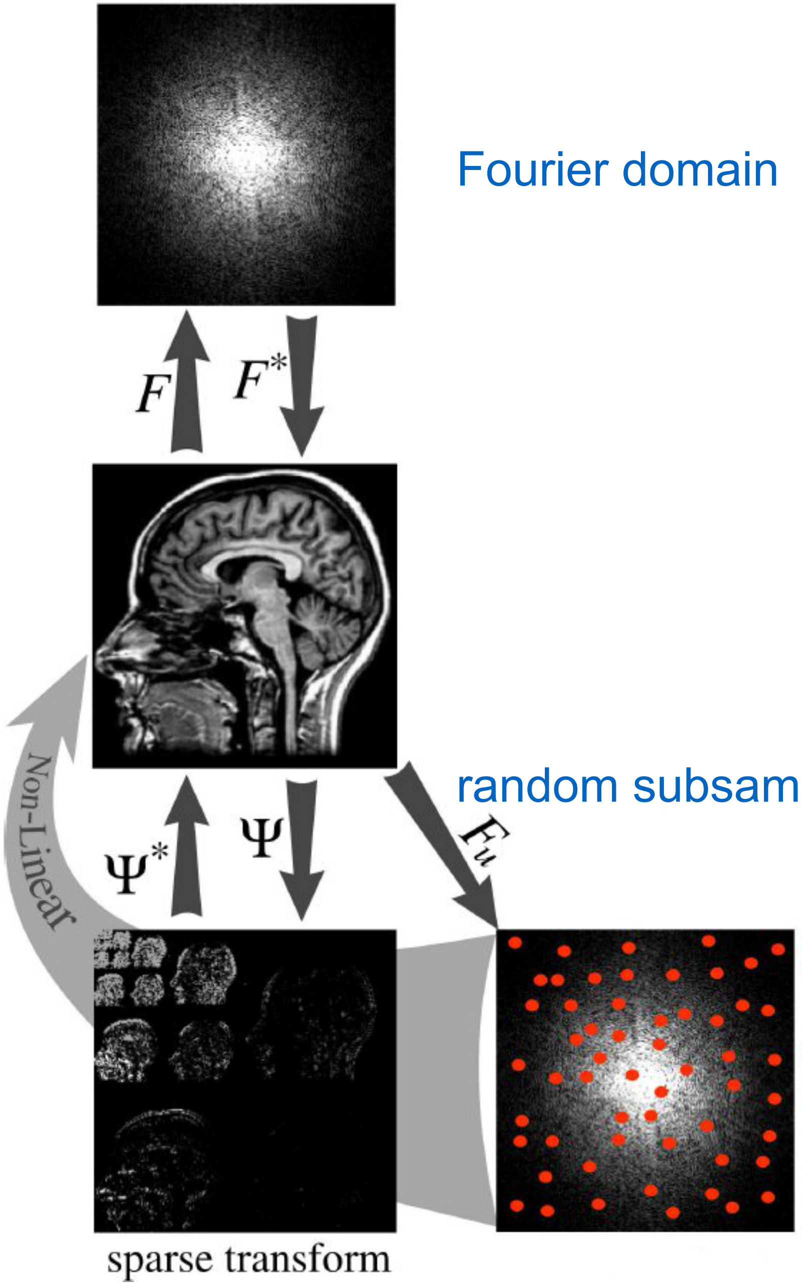
11000 measurements
(16%)

1300 measurements
(2%)

Rapid Sparse Magnetic Resonance Imaging

Lustig, Donoho, Pauly 2007
(Google Scholar: 3105 citations
October 2016)

compression &
speed-up factor: 8



Fourier domain

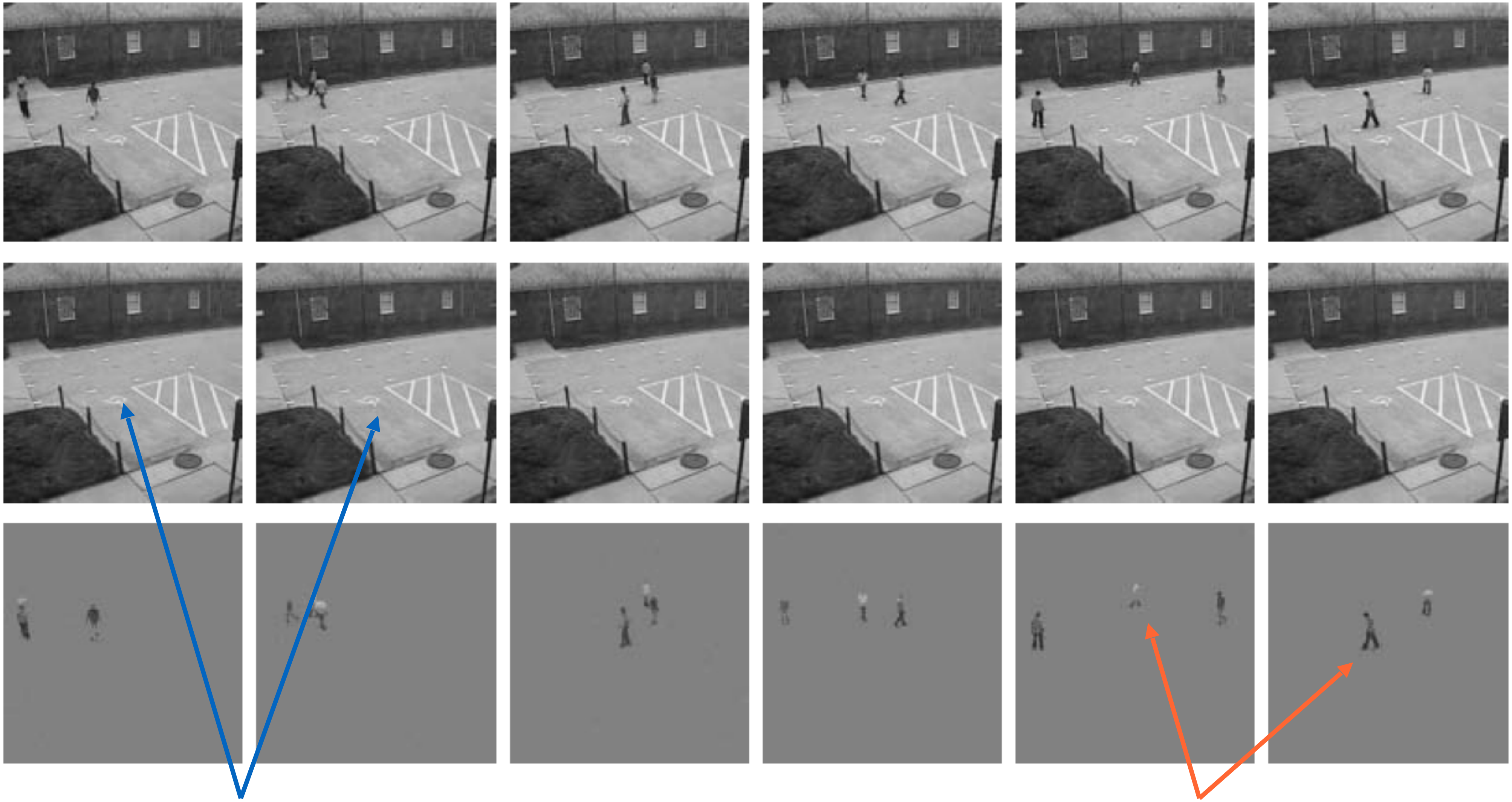
signal

random subsampling

nonlinear (convex)
decoding & recovery

Compressed Video Sensing

© Baraniuk et al.,
NIPS'11



illum. changes

$$F = L + S$$

motion

video = low-rank + sparse signal

Compressed Video Sensing

$$F = L + S$$

video = low-rank + sparse signal

Recovery by *convex* programming

$$\min_{L,S} \|L\|_* + \lambda \|\operatorname{vec}(S)\|_1$$

subject to

$$\|f - \mathcal{A}(L + S)\|_2 \leq \varepsilon$$

low-rank sparse sampling

Scientific Imaging: 2007 - 2018

Medical Imaging

- Magnetic Resonance Imaging
- Computerized Tomography

Imaging in other Sciences

- Thermoacoustic Tomography
- Photoacoustic Tomography
- Electrical Impedance Tomography
- Electron Tomography
- Seismic Tomography
- Fluorescence Microscopy
- Radio Interferometry

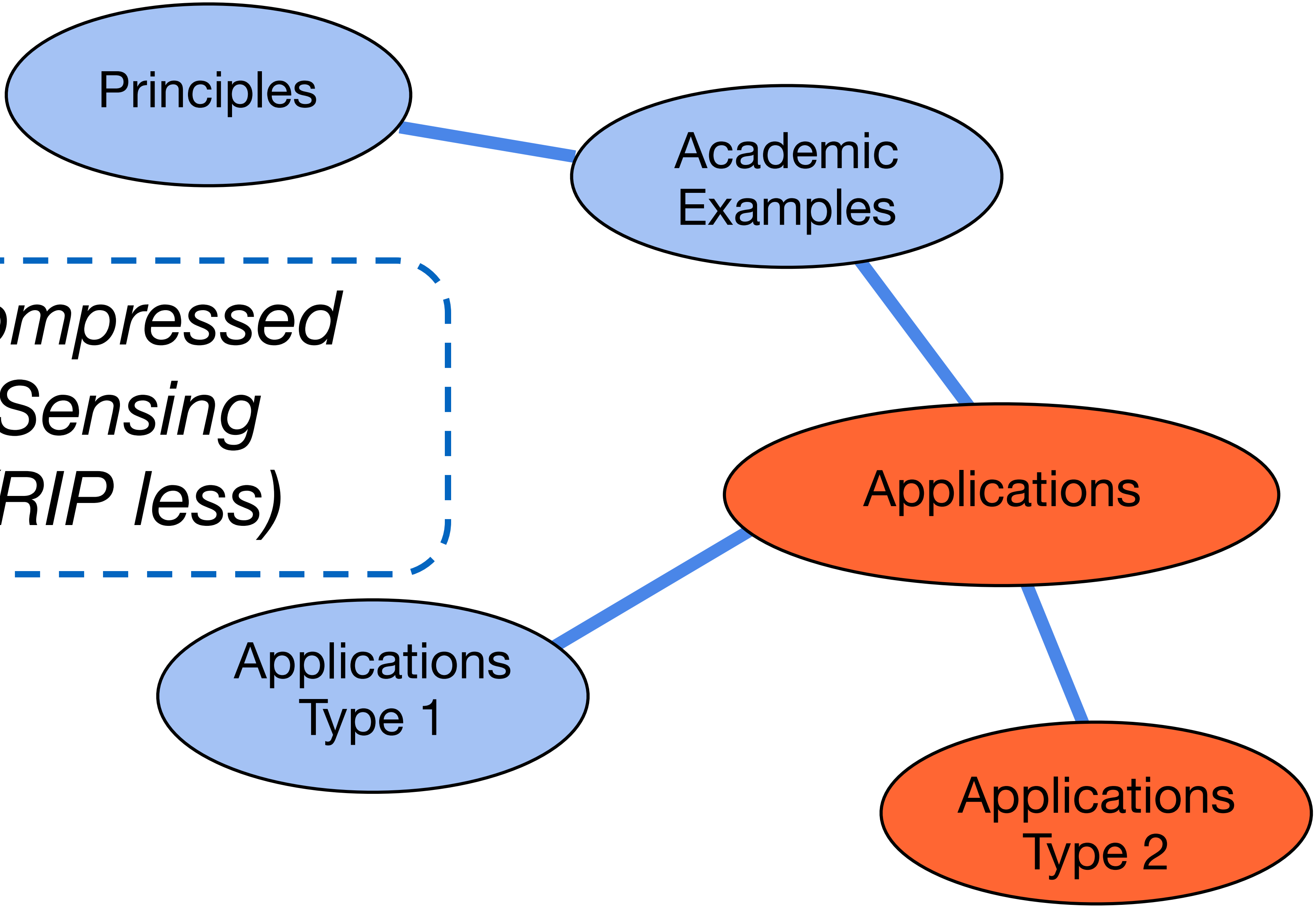
Established RIP-based CS-theory cannot explain much better *empirical* performance

Significant gap between *mathematical* theory and applied fields

Major recent trend (mathematics)

$\Longrightarrow$ dispense with RIP and universal sensing operators

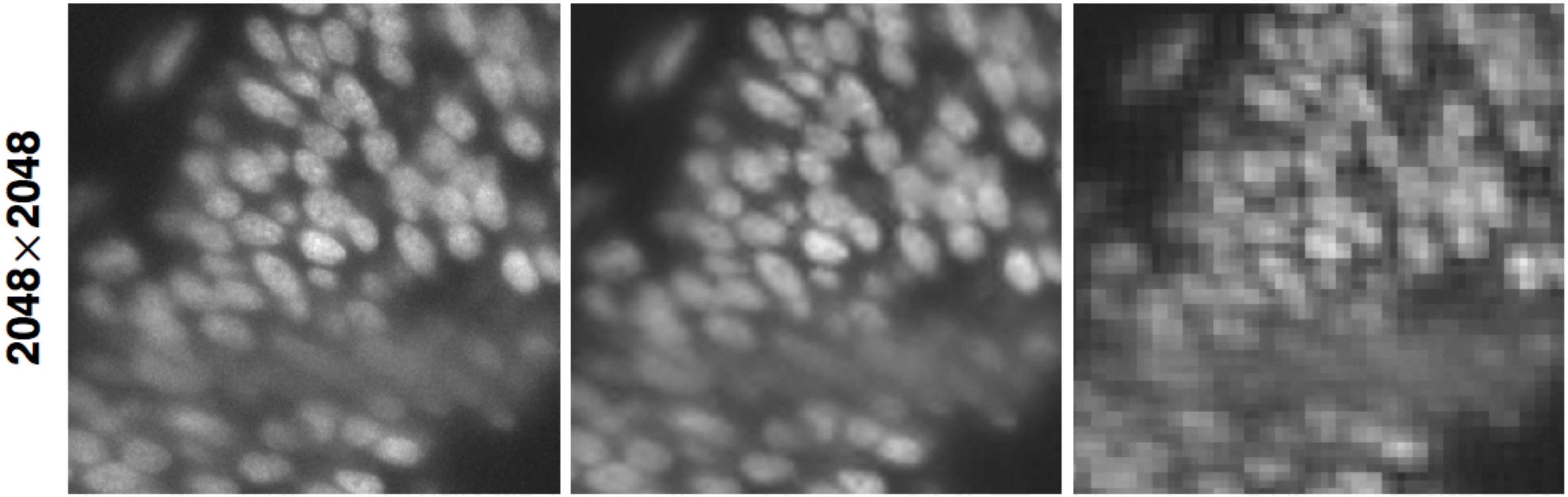
$\Longrightarrow$ exploit *structured sparsity*



Compressed Sensing (RIP less)

Example (© A. Hansen, 2014)

Cambridge Advanced Imaging Center (CAIC)
Fluorescence Microscopy: zebra fish cells



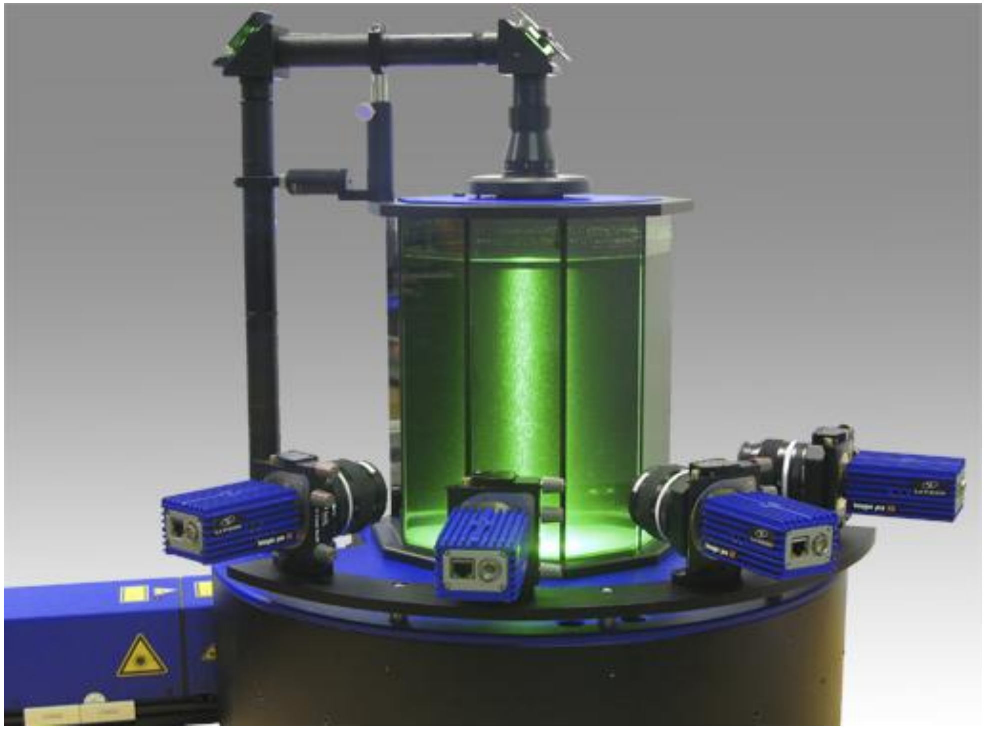
raster scan

recovery from 6.25 %

established scheme

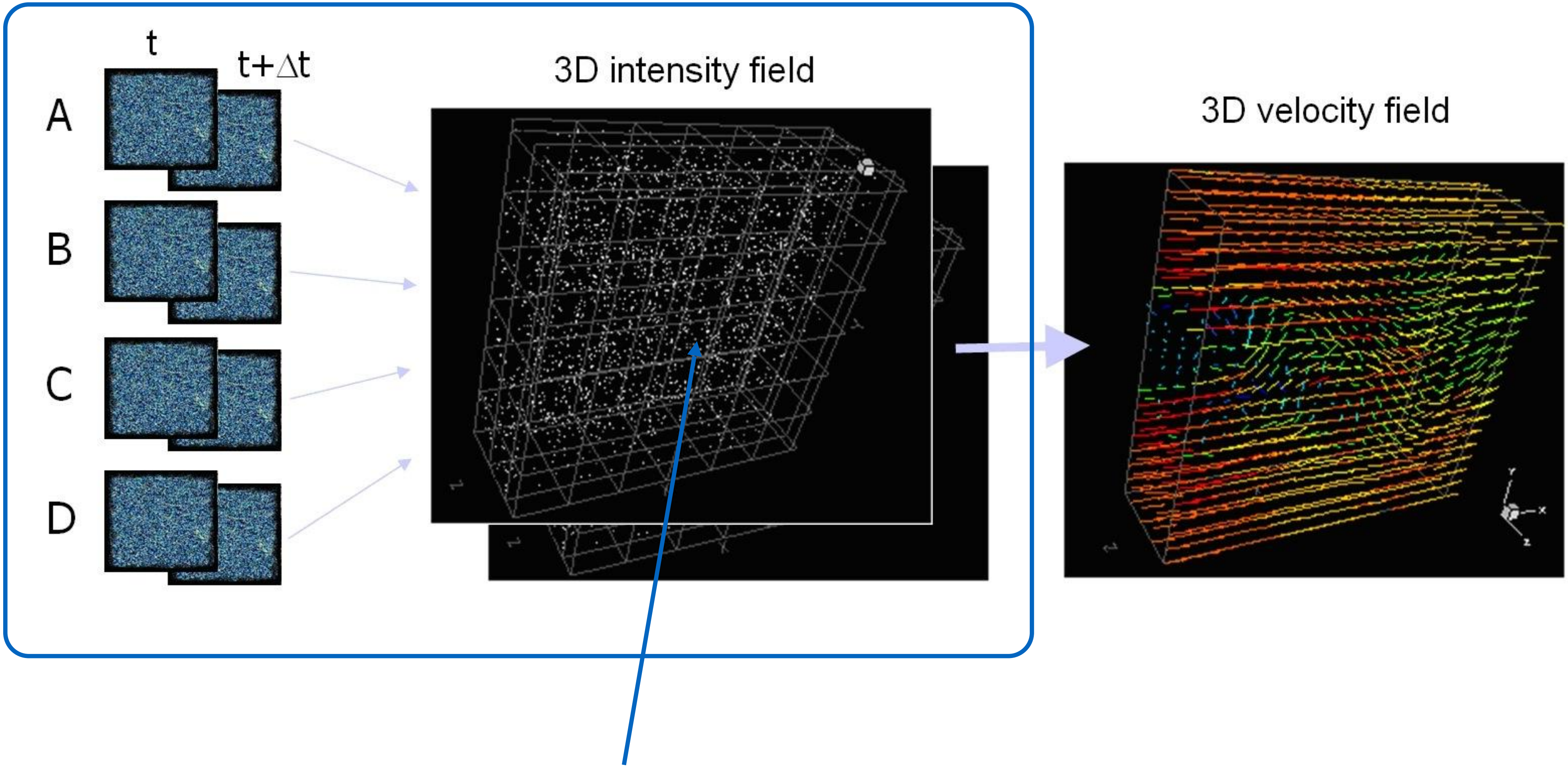
CS and Non Standard Tomography

Experiments in Fluids
DFG SPP 1147



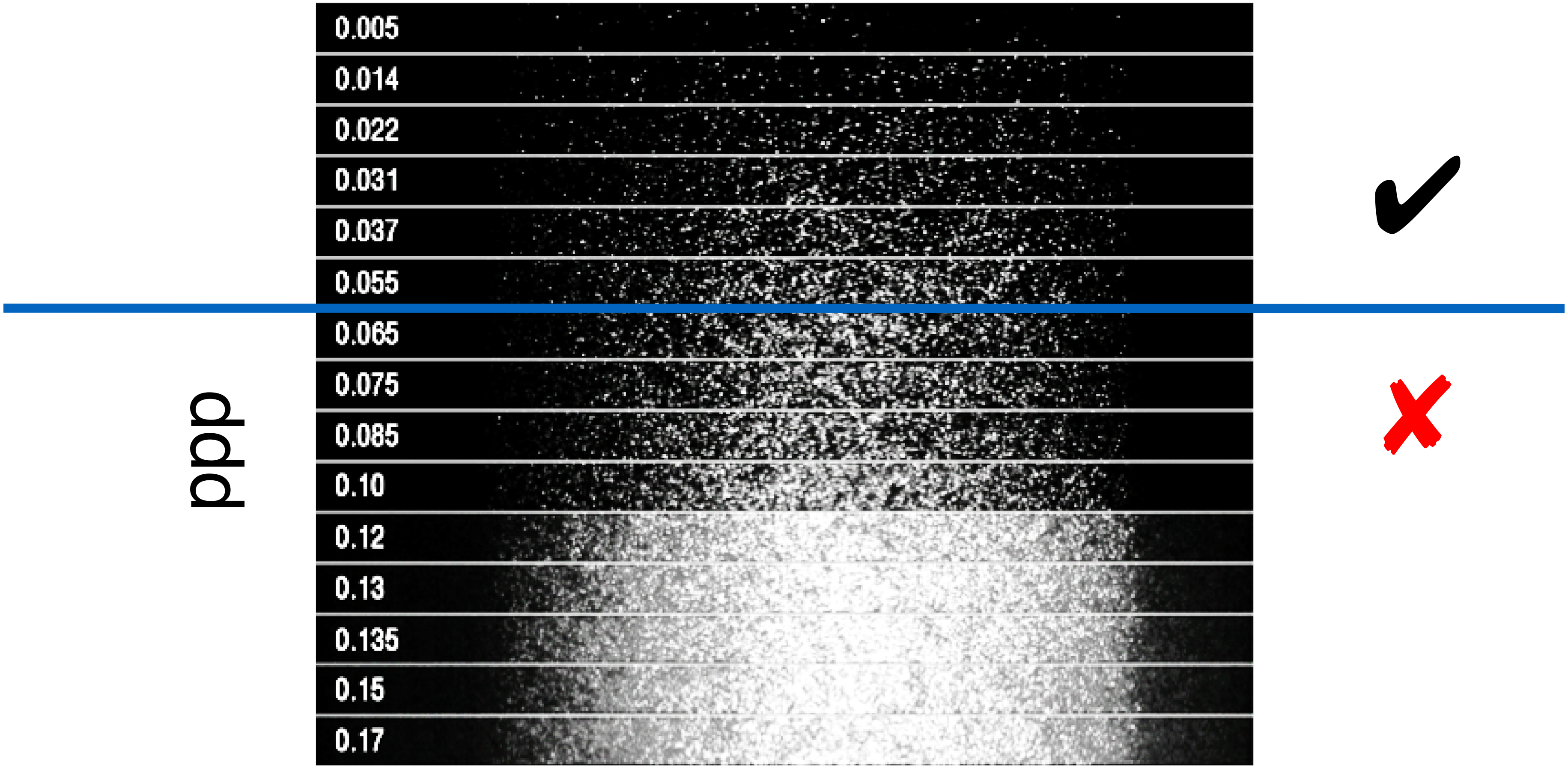
Tomo PIV

Sparsity in Tomo PIV



key parameter: seeding density

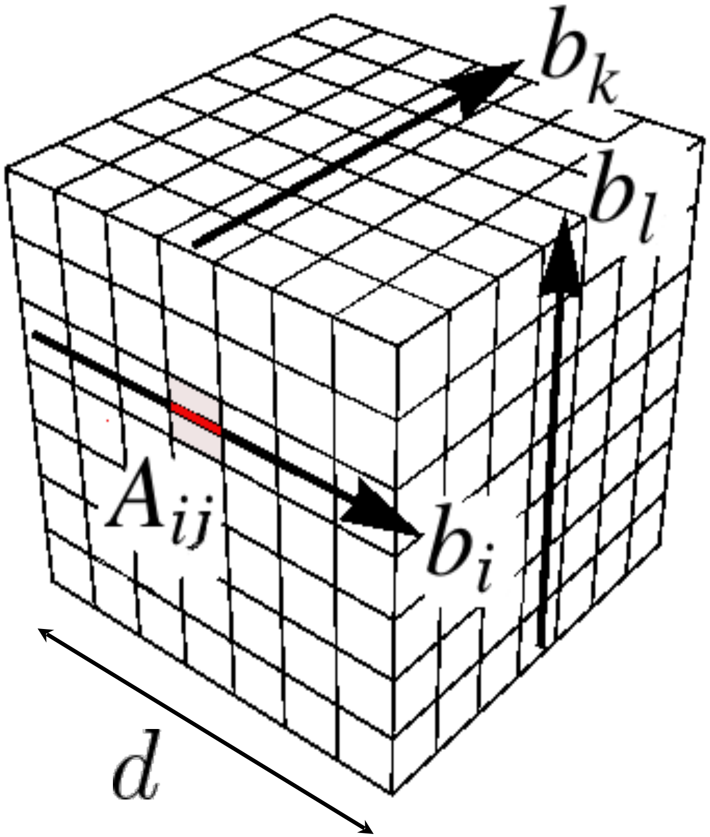
Motivation



key parameter: seeding density

Imaging Set-Up

Few measurements



d problem size, resolution

$n = d^3$

$A \in \mathbb{R}_+^{m \times n}$

$\boxed{m} = p\, d^2 < \boxed{n}$

p projections

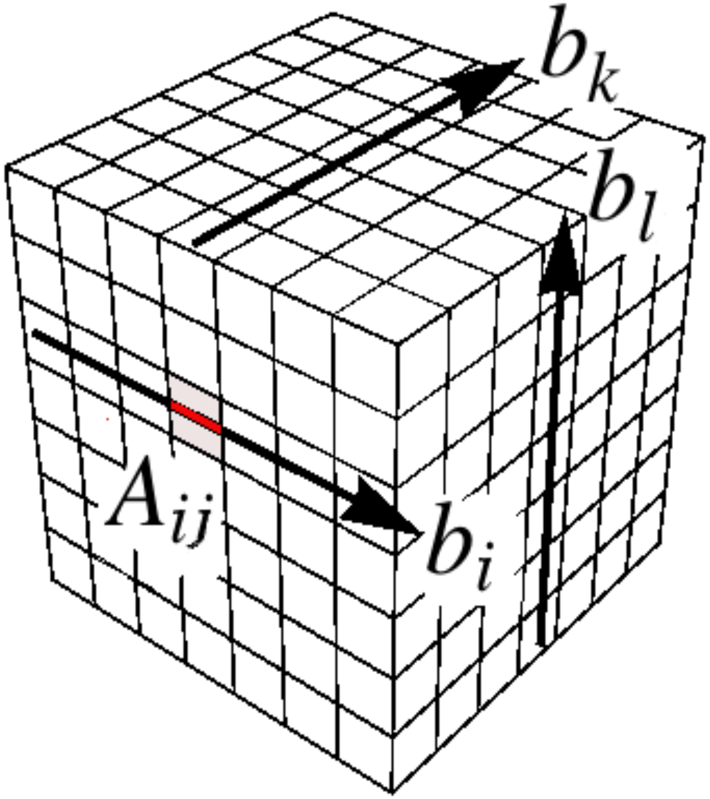
$\boxed{k \text{ particles}}$

$$Ax = b, \quad x \in \{0,1\}^n, \; x \in \mathbb{R}_+^n$$

Poor CS-Sensor

Our sensors have ***poor strong*** recovery properties.

(P. & Schnörr, PMA 2009)



RIP, neighborly polytopes,
mutual coherence, …

Recovery of *any* vector from

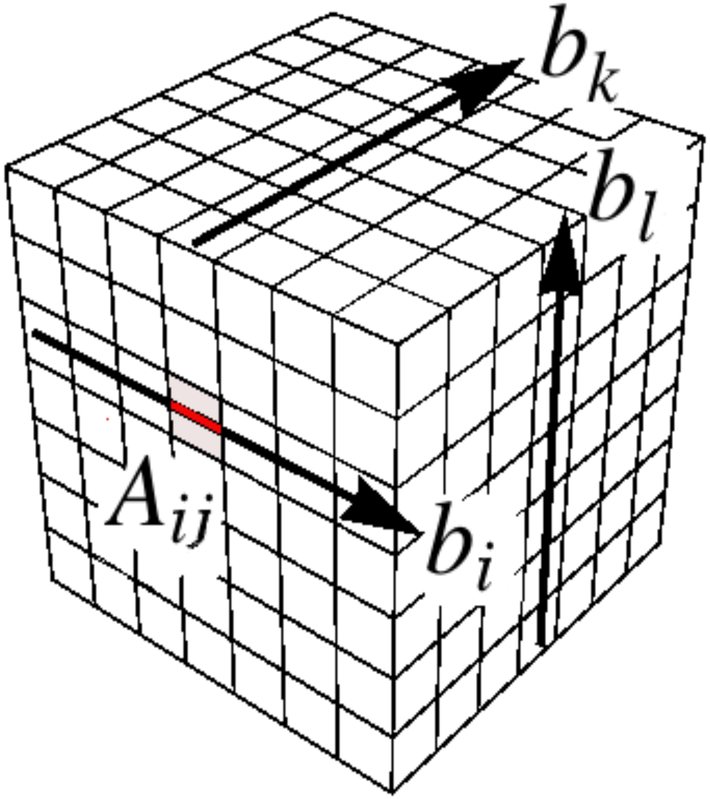
$$\mathcal{X}^n(k)$$

$\Rightarrow k$ ***very small***

CS Theory Does Not Apply

Our sensors have ***poor strong*** recovery properties.

(P. & Schnörr, PMA 2009)



RIP, neighborly polytopes, mutual coherence, …

Recovery of *any* vector from

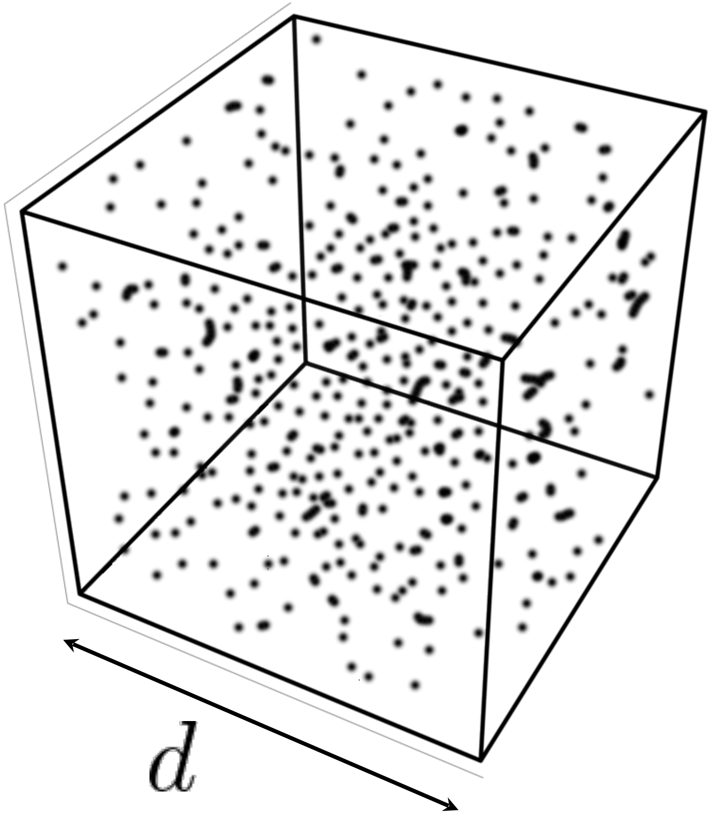
$$\mathcal{X}^n(k)$$

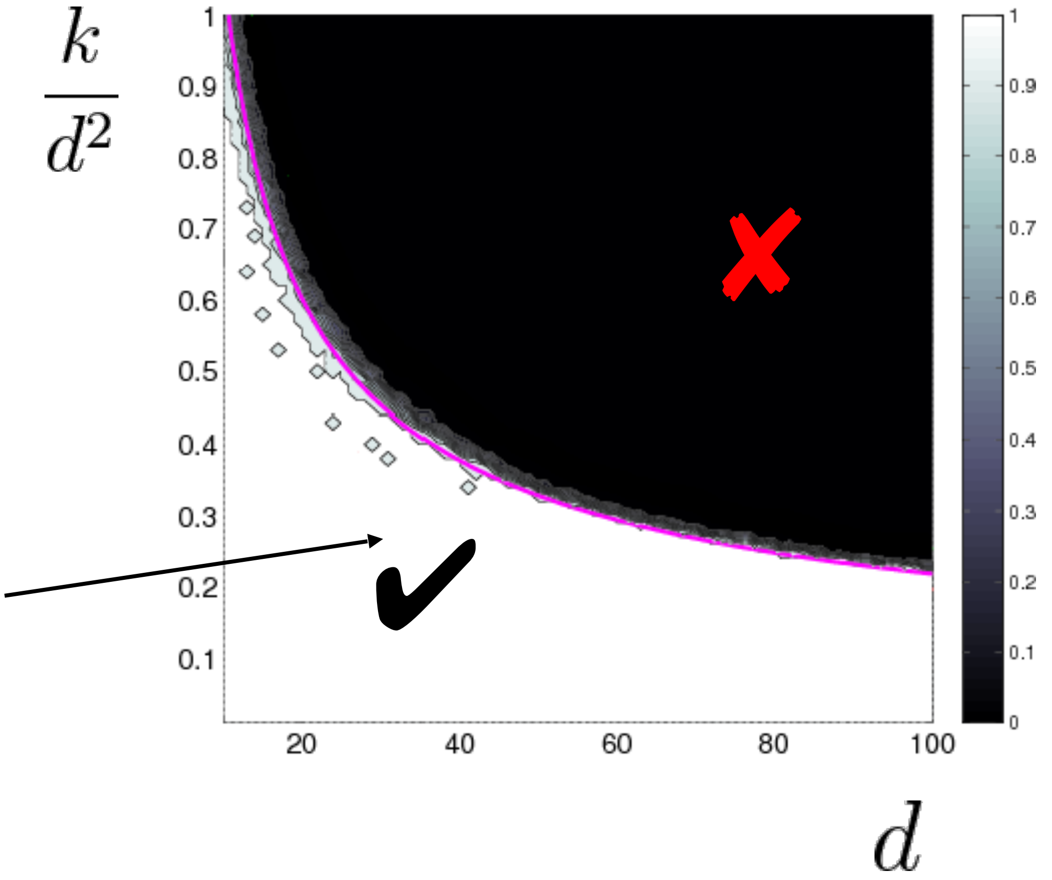
$\Rightarrow k$ ***very small***

☞ average case analysis

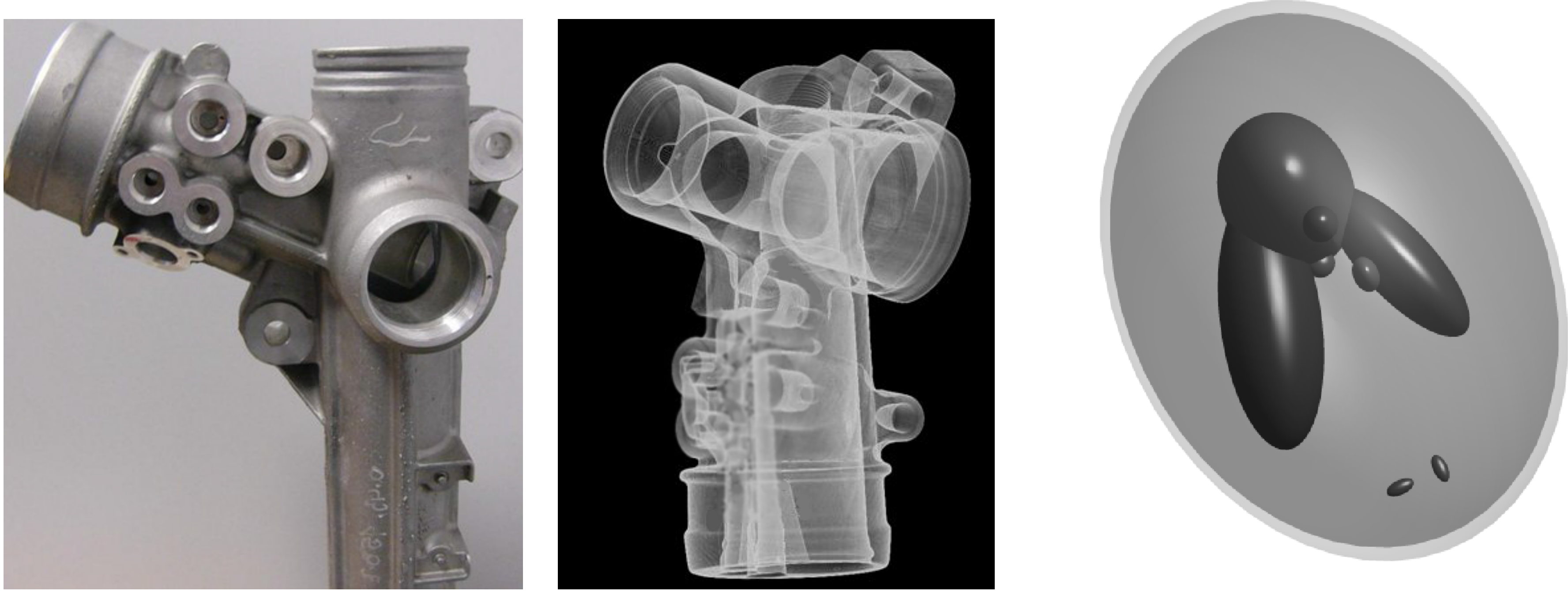
Practice Matches Theory

of k particles





Non-Destructive Testing (NDT)



key parameter: cosparsity,

sparse ∇x

“object complexity” $\longleftarrow$ material

interfaces

Recovery Performance

(Needell, Ward 2013)

$$\boxed{\Delta(b) = \mathrm{argmin}_{x:\ \|Ax-b\|_2 \le \varepsilon} \|\nabla x\|_1}$$

decoding

Recovery Performance

(Needell, Ward 2013)

$b = Ax^* + \xi, \; \|\xi\|_2 \leq \varepsilon$ (bounded noise)

and $AH^\top \leftarrow RIP_{\delta_{5k}}, \delta_{5k} < \frac{1}{3}$ H Haar transform

$$\boxed{\Delta(b) = \operatorname{argmin}_{x:\, \|Ax-b\|_2 \leq \varepsilon} \|\nabla x\|_1}$$

decoding

Stable recovery guarantee

$$\|x^* - \Delta(b)\|_2 \leq C\left(\frac{\|\nabla x^* - (\nabla x^*)_k\|_1}{\sqrt{k}} + \varepsilon\right)$$

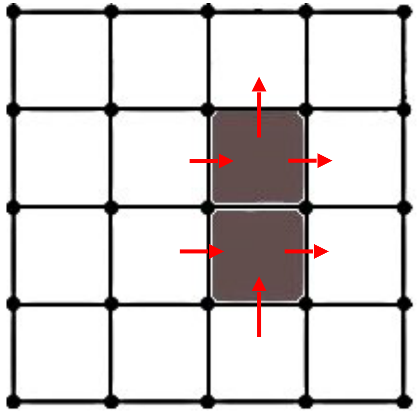
ε → *0* and *k*-sparsity of ∇x^* ⇒ perfect reconstruction

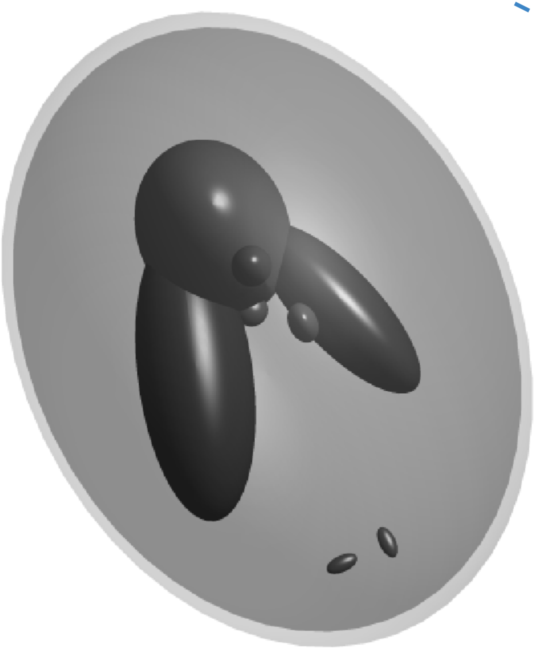
Worst Case vs Practical Recovery

Thm. (Needell, Ward 2013)
our scenario, 8 views

$\Rightarrow$

Image gradient ∇x^*
can be at most 6-sparse





But: perfect recovery from
8 views, $\|\nabla x^*\|_0 = 109930$

Cosparsity & Recovery

(Nam et al, 2013) $\Lambda := \{i : (\nabla x)_i = 0\}$ cosupport

cosparsity $\ell := |\Lambda| = \dim(\nabla x) - \|\nabla x\|_0$

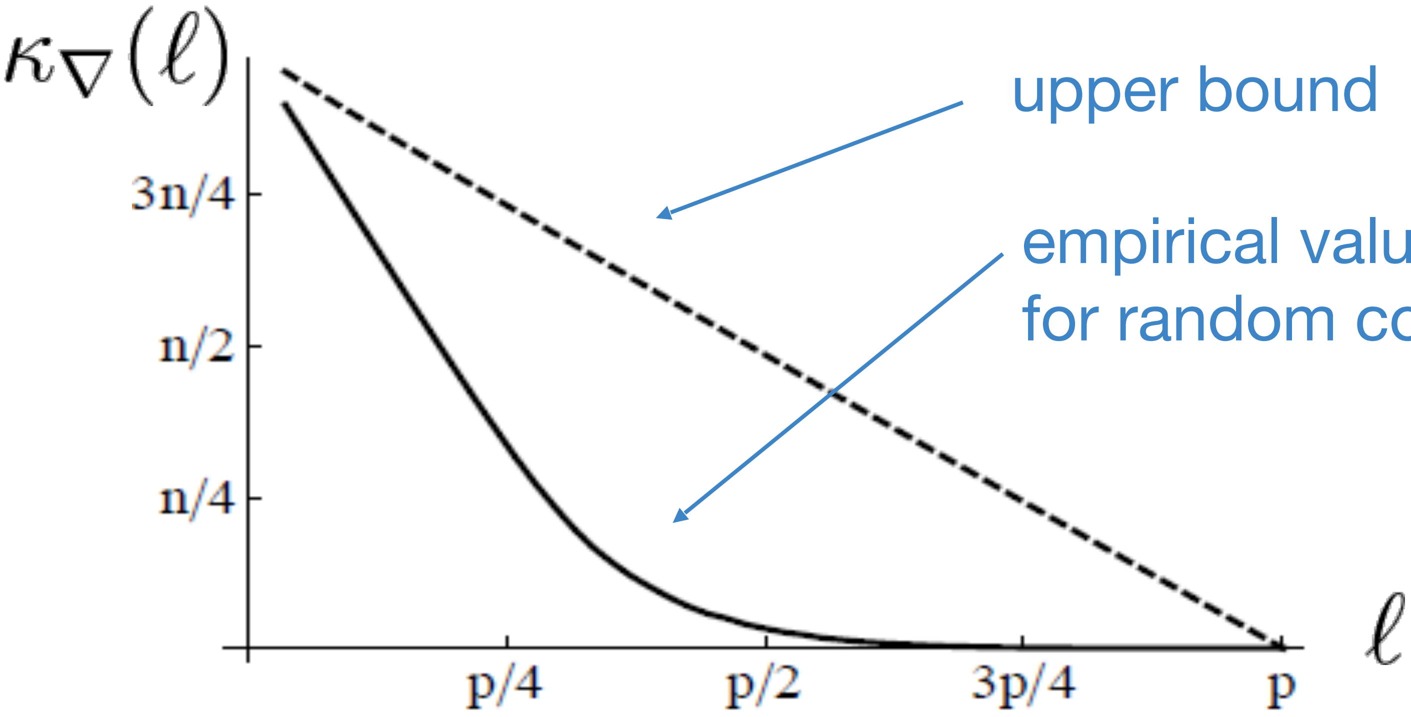
key quantity $\boxed{\kappa_\nabla(\ell)} := \max_{|\Lambda|=\ell} \dim\ \ker \nabla_\Lambda$

Cosparsity & Recovery

(Denitiu et al, 2014)

$$m \geq 2n - (\ell + \sqrt{2\ell + 1} - 1) \qquad 2D$$

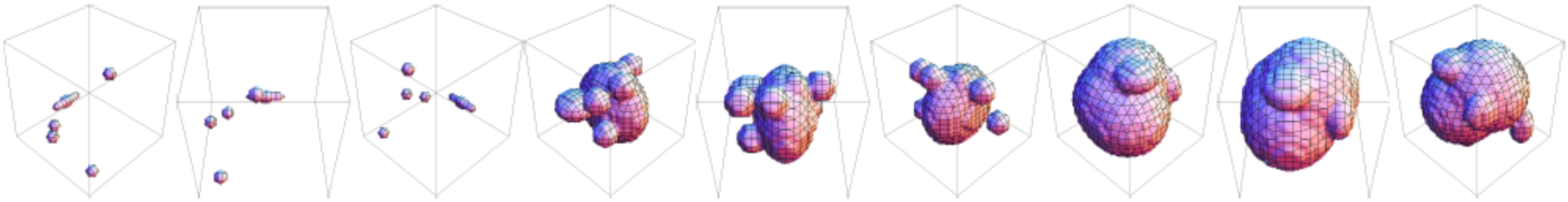
$$m \geq 2n - \frac{2}{3}\left(\ell + \sqrt[3]{3\ell^2} + 2\sqrt[3]{\frac{\ell}{3}} - 2\right) \qquad 3D$$

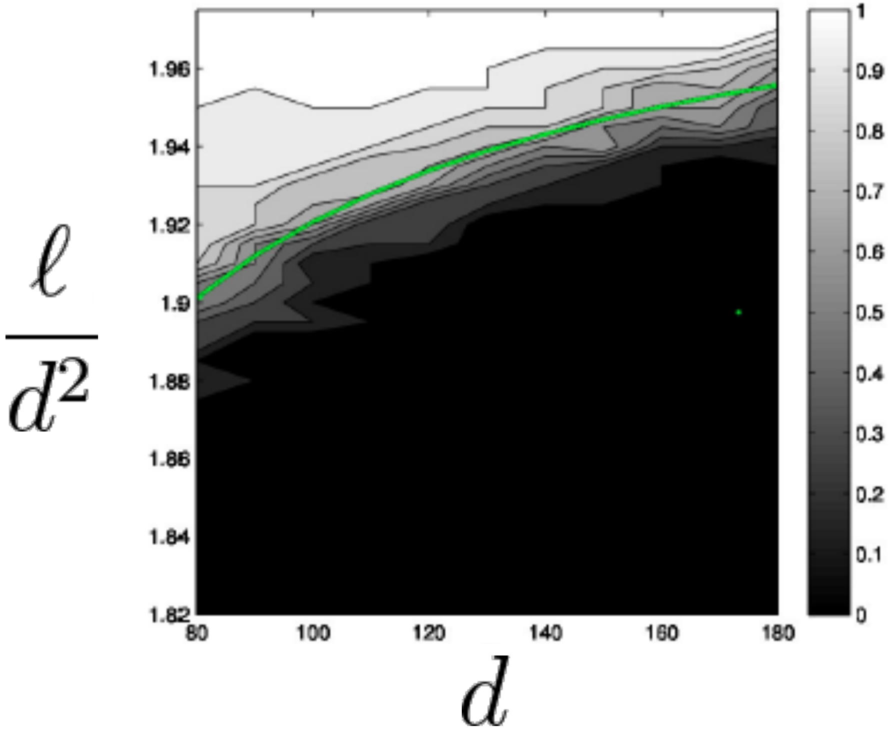


upper bound

empirical value
for random cosupport

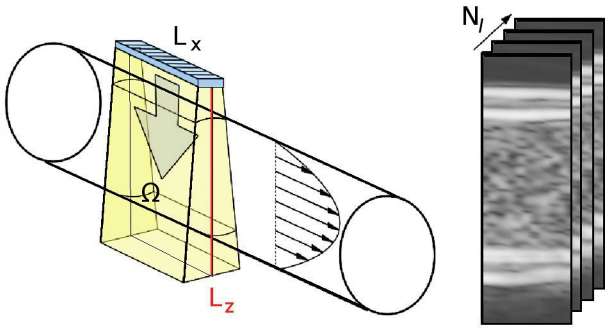
Recovery for Solid Objects

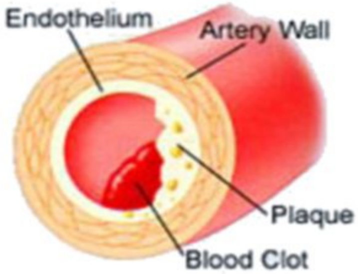




4 views

Compressed Motion Sensing



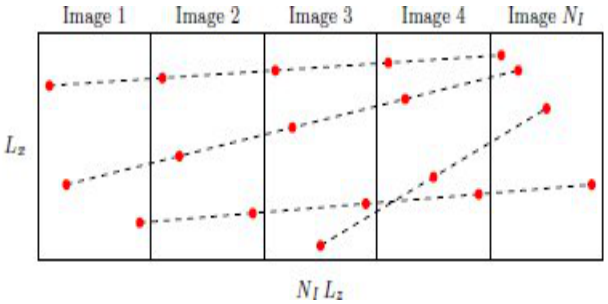


Diseased artery

entire
video

ultrasound imaging

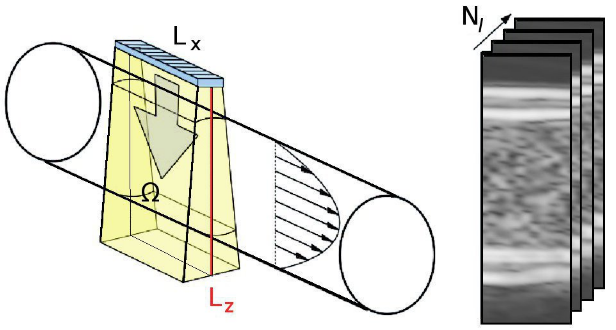
Echo PIV

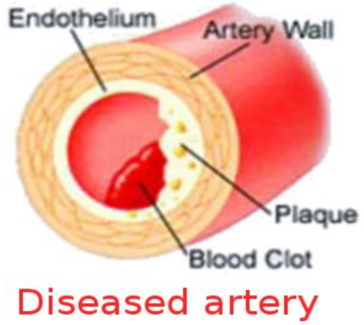


Few trajectories

(Bodnariuc et al, EMMCVPR 2014)

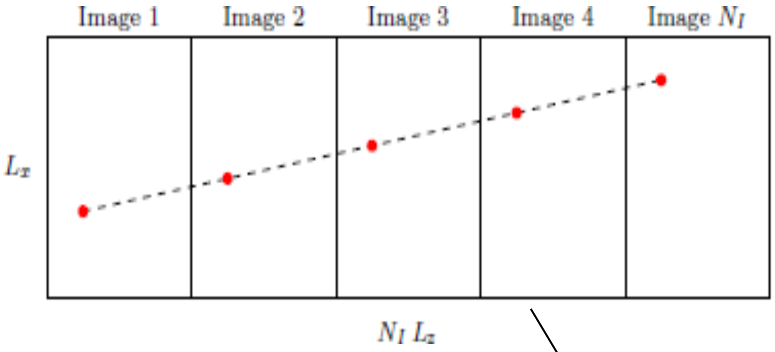
Compressed Motion Sensing

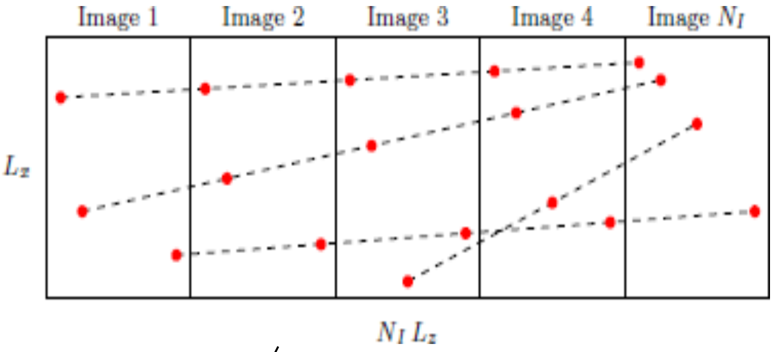




Diseased artery

entire video





column in adaptive dictionary

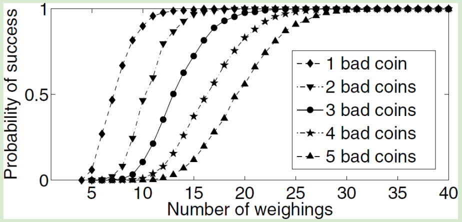
sparsity

$$\min_{v} \min_{u \in [0,1]^n} \|D(v)u - F\|_1 + \lambda \|u\|_1$$

(Bodnariuc et al, EMMCVPR 2014)

Overview

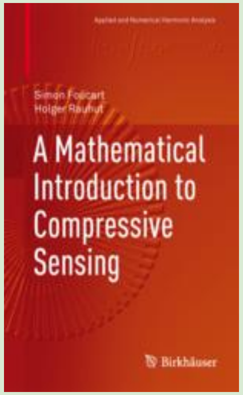
CS principles



idealized scenarios

enable

established theory



Applied and Numerical Harmonic Analysis

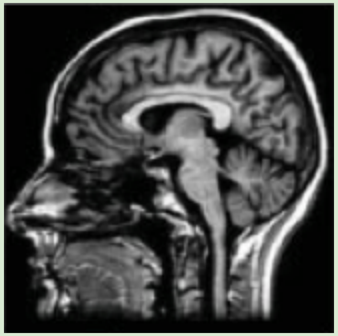
Simon Foucart
Holger Rauhut

A Mathematical Introduction to Compressive Sensing

Birkhäuser

applies

applications

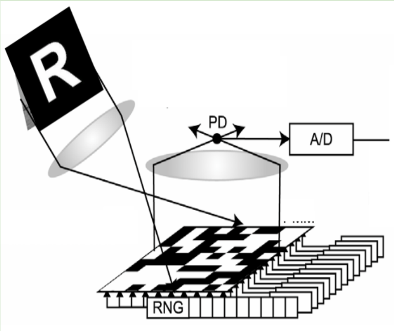


classical MRI

applies

applications

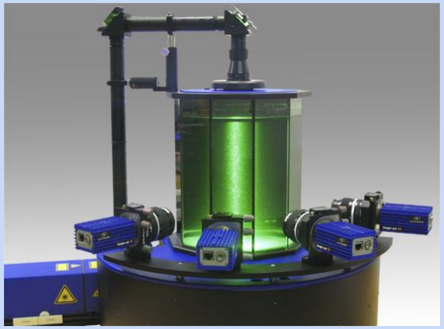
surveillance



single pixel camera

RIP fails

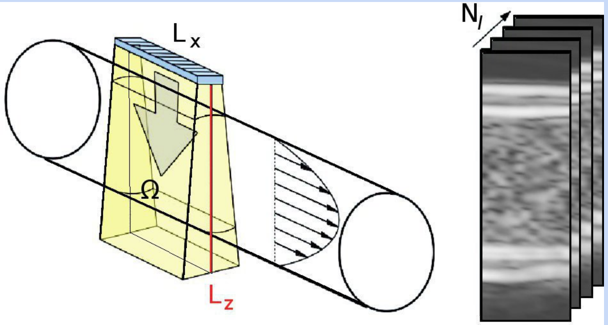
applications



non-standard tomography

RIP fails

applications



ultrasound PIV

A Mathematical Introduction to Compressed Sensing

Stefania Petra
Mathematical Imaging Group
Heidelberg University

Lecture WT 2018/19

Content

Content

Content

Content