Sparsity & Co.:
An Overview of Analysis vs Synthesis in Low-Dimensional Signal Models

Rémi Gribonval, INRIA Rennes - Bretagne Atlantique

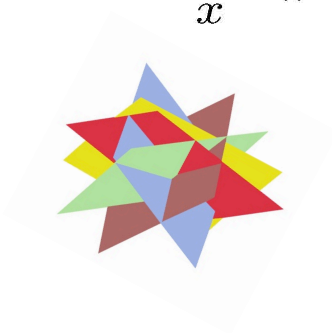
remi.gribonval@inria.fr

http://www.irisa.fr/metiss/members/remi/talks

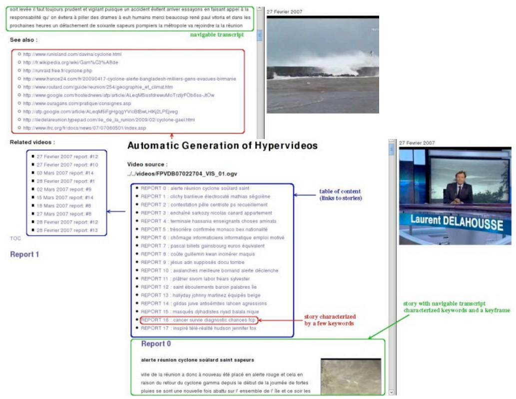
METISS Rennes

Speech and audio modeling and processing

$$\min_{x} \|x\|_1 \text{ s.t. } \|y - \mathbf{A}x\| \leq \epsilon$$



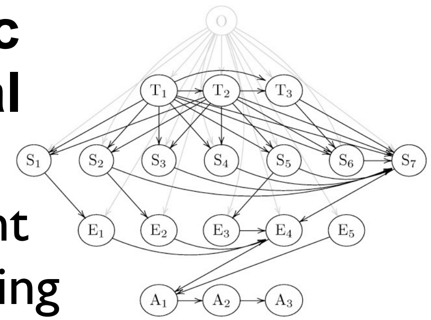
Sparse representations
models and algorithms



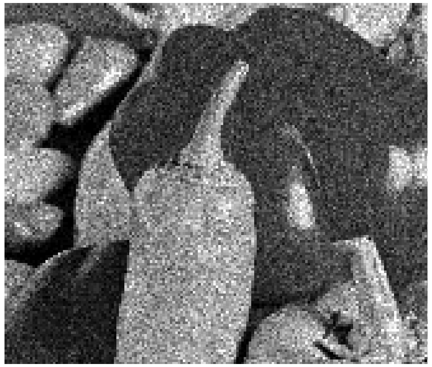
Audio and **spoken content** processing and description

Audio **source localization** and **separation**

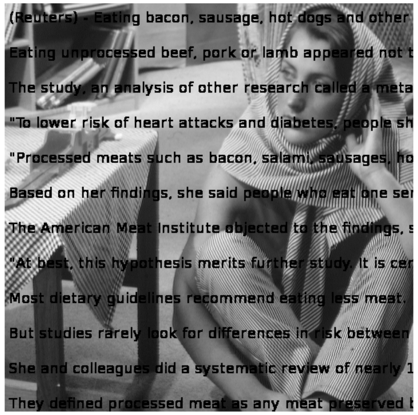
Music signal and content processing



Sparse Signal / Image Processing



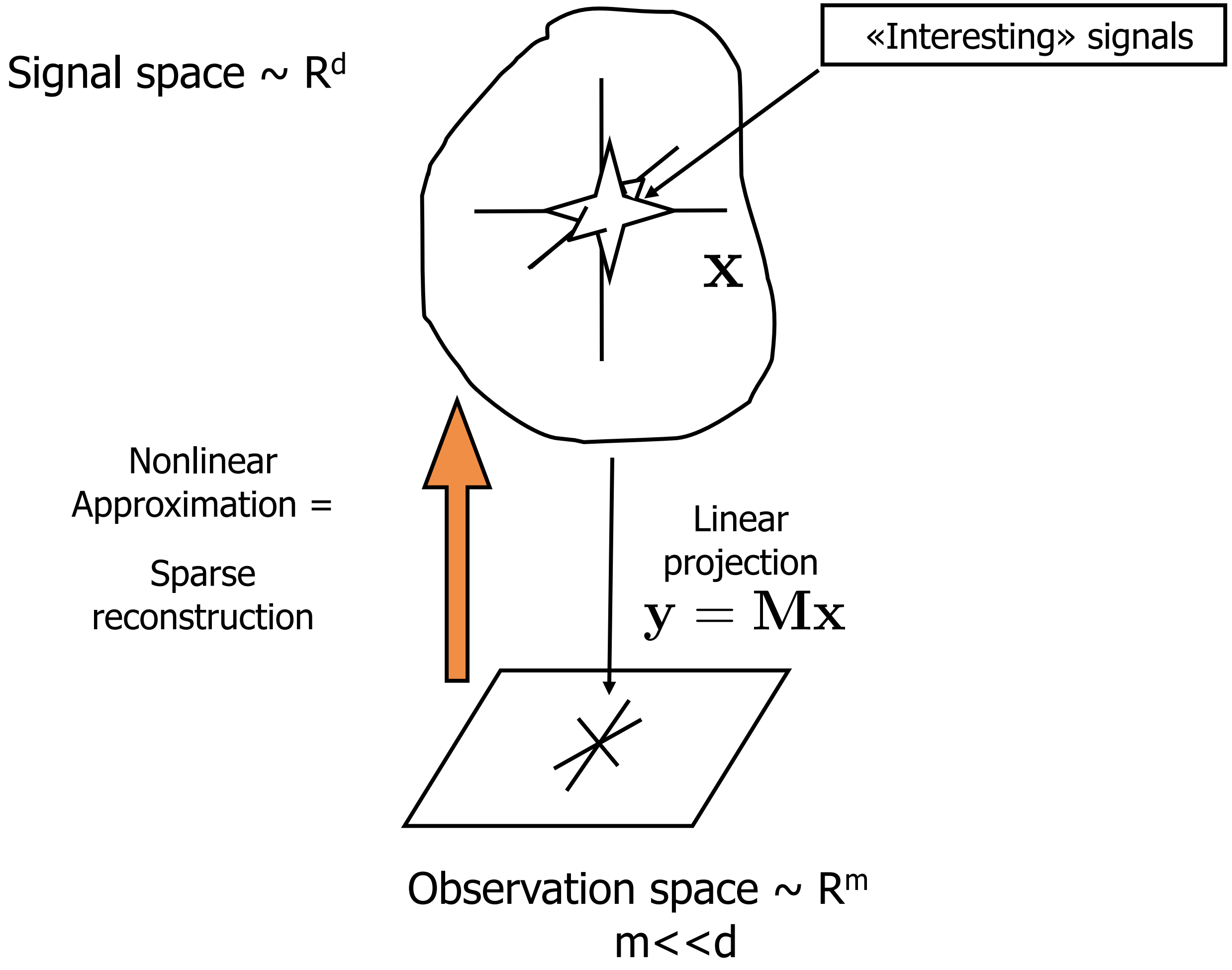
denoising



inpainting

+ Compression,
Source Localization, Separation,
Compressed Sensing ...

Geometry of Inverse Problems



Thanks to M. Davies, U. Edinburgh

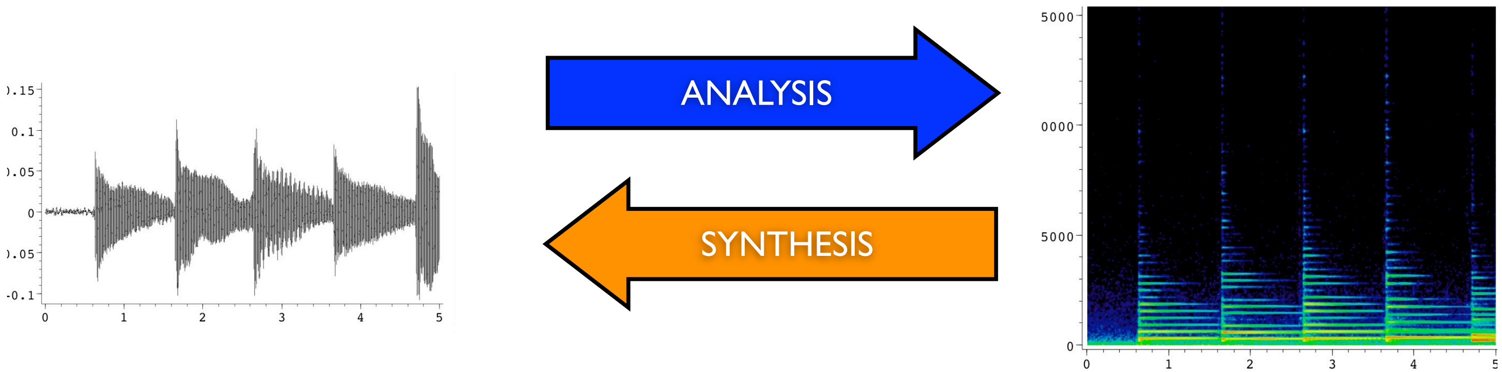
Sparse Atomic Decompositions

$$\mathbf{x} \approx \mathbf{D}\mathbf{z}$$

Signal Image	(Overcomplete) dictionary of atoms	Representation Coefficients

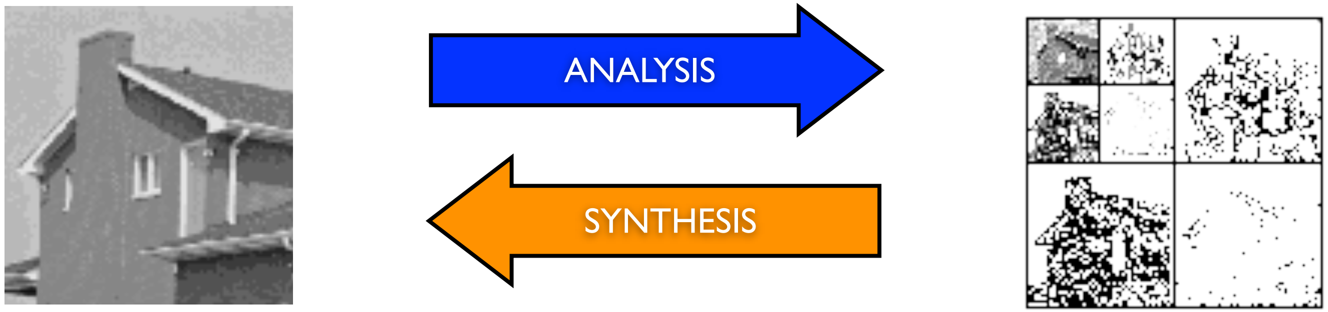
Supporting Evidence: Sparsifying *Transforms*

- Audio : time-frequency representations (MP3)



Black = zero

- Images : wavelet transform (JPEG2000)



White = zero

Sparsity and inverse problems
Uniqueness guarantees

Mathematical foundations

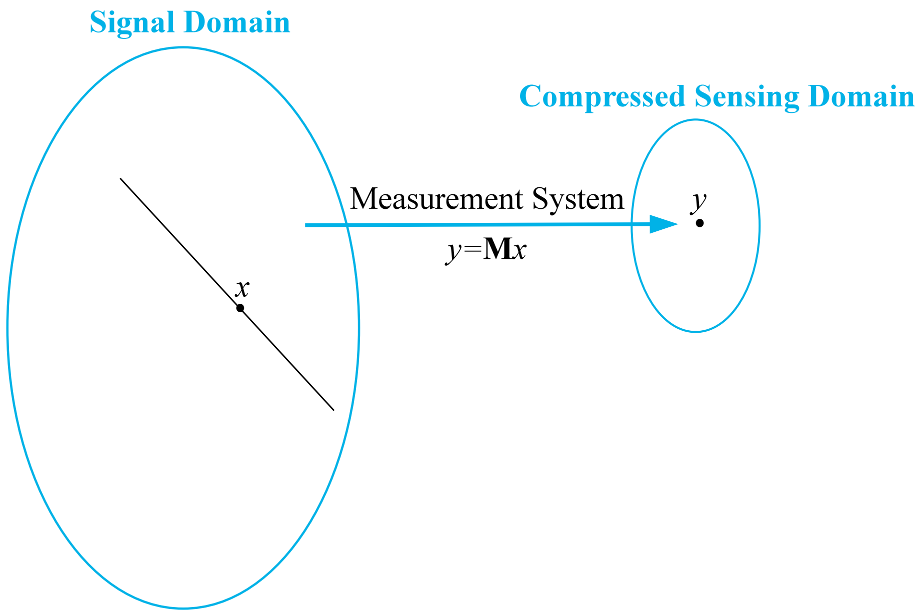
- Bottleneck (1990-2000) :
  - ✓ fewer equations than unknowns = ill-posed ...

$$\mathbf{A}x_0 = \mathbf{A}x_1 \not\Rightarrow x_0 = x_1$$

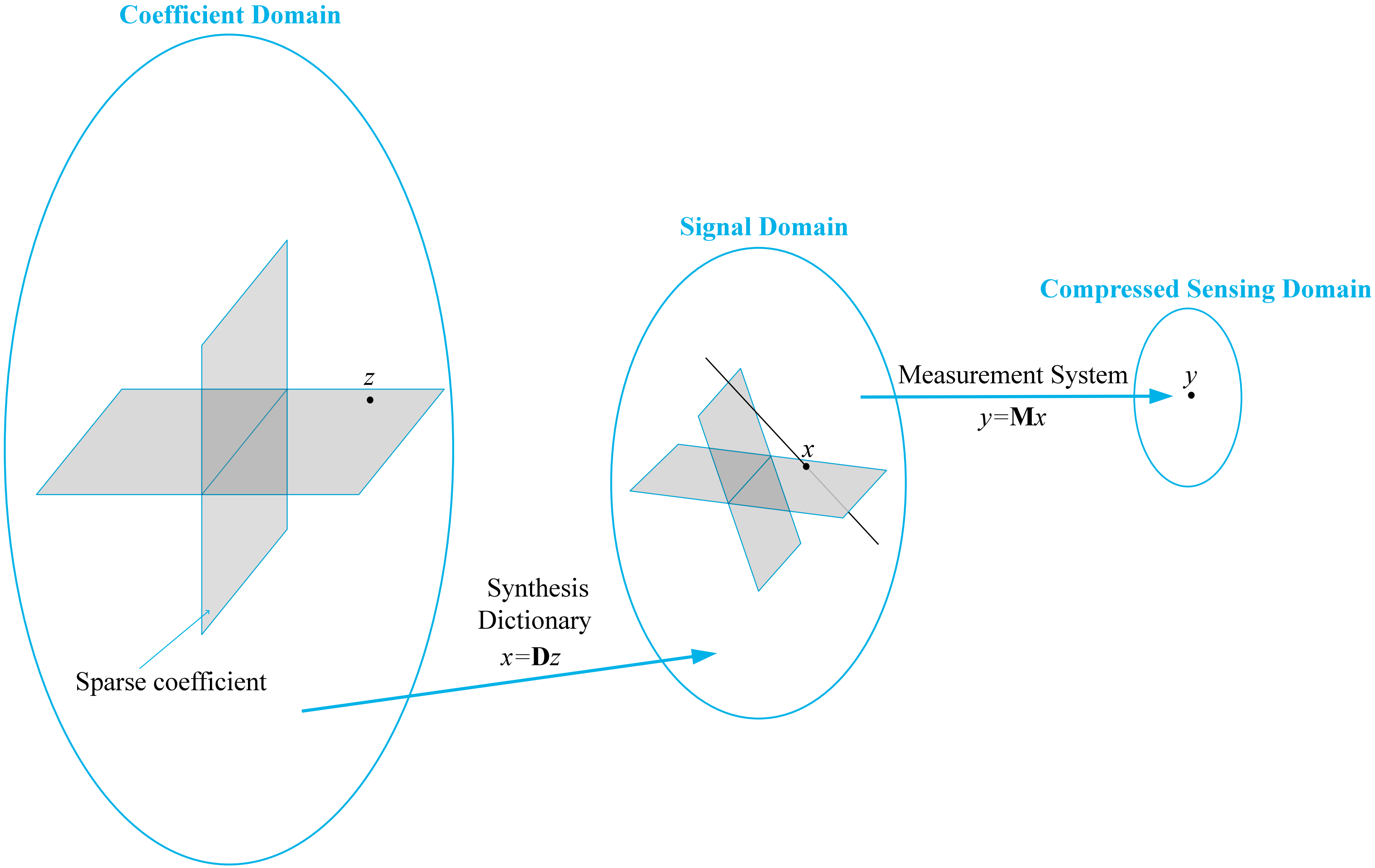
- Theoretical & algorithmic breakthrough (2001-2006) :
  - ✓ **Uniqueness** of the sparse solution = well-posed!
    - ✦ if x_0, x_1 are "sparse enough",
    - ✦ then $$\mathbf{A}x_0 = \mathbf{A}x_1 \Rightarrow x_0 = x_1$$
  - ✓ **Bounded complexity** algorithms to reconstruct x_0
    - ✦ Thresholding, Matching Pursuits, Convex Optimization,...

[Mallat & Zhang 1993] [Gorodnistky & Rao 1997] [Chen Donoho & Saunders 1999]

Sparse models and inverse problems



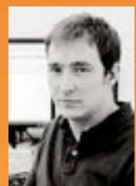
Sparse models and inverse problems



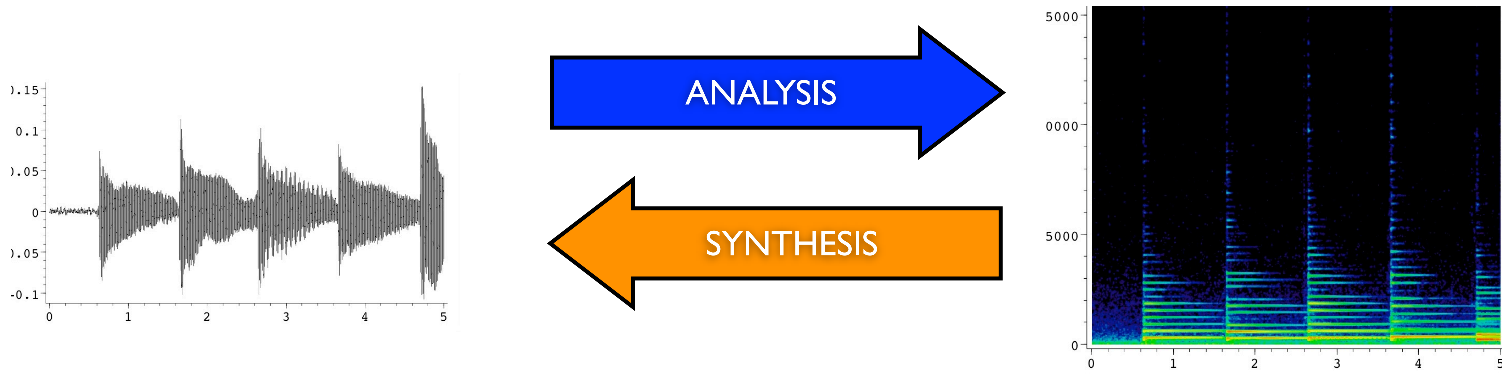
Analysis *vs* Synthesis: *Cosparsity*

with S. Nam, M. Davies, M. Elad

small-project.eu


Sparsifying Transforms
= Atomic Decompositions ?



Black
= zero

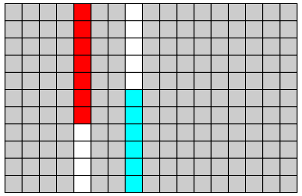
- **Analysis** = transform $\mathbf{z} = \mathbf{\Omega}\mathbf{z}$
  - ✦ small number of nonzero coefficients $\|\mathbf{z}\|_0 := \sum_k |z_k|^0$
- **Synthesis** = decomposition $\mathbf{x} = \mathbf{D}\mathbf{z} = \sum_k z_k \mathbf{d}_k$

Analysis priors
for regularization / data separation

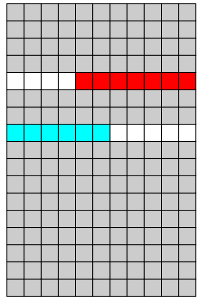
$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2^2 + \lambda\|\mathbf{\Omega}\mathbf{x}\|_p$$

- **Much empirical evidence of success**
  - with $\mathbf{\Omega}$ associated to TV norm, curvelet coefficients, undecimated Haar ...
  - Starck & al 2003, Elad & al 2007, Portilla 2009, Selesnick & Figueiredo 2009, Afonso, Bioucas-Dias & Figueiredo 2010, Pustelnik & al 2011, ...
- **Few theoretical guarantees**
  - Donoho & Kutyniok 2010, Candès & al 2010, Nam & al 2011, Vaiter & al 2011.
- **Which signals can be well recovered ?**
  - ✓ Better understand analysis model
  - ✓ (New) analysis algorithms with performance

Equivalence ? With *tight frames*

Dictionary $\mathbf{D}$ 

: Tight frame

Operator Ω  $= \mathbf{D}^T$

$$\mathbf{x} = \mathbf{D}\Omega\mathbf{x}.$$

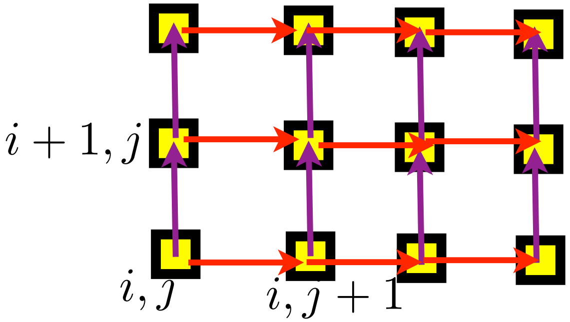
$\mathbf{x}$: *sparse in* $\mathbf{D}$ $\Longleftrightarrow$ $\Omega\mathbf{x}$: *sparse*

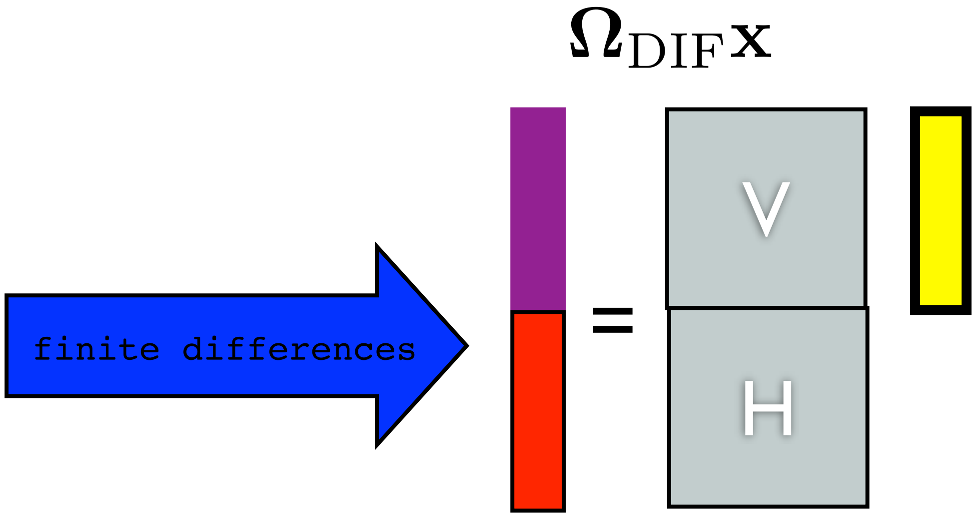
- Yes ... but some troubling facts:
  - ✓ **only one** analysis representation
  - ✓ **infinitely many** synthesis representations

Analysis *without* synthesis
the 2D finite difference operator

- Cousin of TV norm
  - Rudin, Osher, Fatemi 1992

$\mathbf{x} = (x_{ij})$





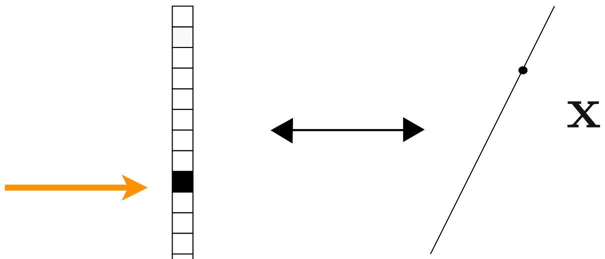
- TV like reconstruction

$$\min_{\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2 \leq \epsilon$$

- **There is NO dictionary such that** $\mathbf{x} = \mathbf{D}\mathbf{\Omega}\mathbf{x},\ \forall \mathbf{x}$

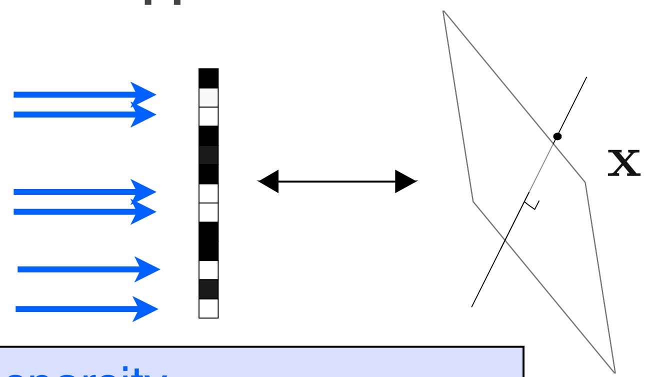
Introducing the cosparse model

- **Sparse synthesis model**
  - ✓ Synthesis dictionary $\mathbf{D}$
  - ✓ Representation $\mathbf{z}$ s.t. $\mathbf{x} = \mathbf{D}\mathbf{z}$
  - ✓ **Support** = location of nonzeroes

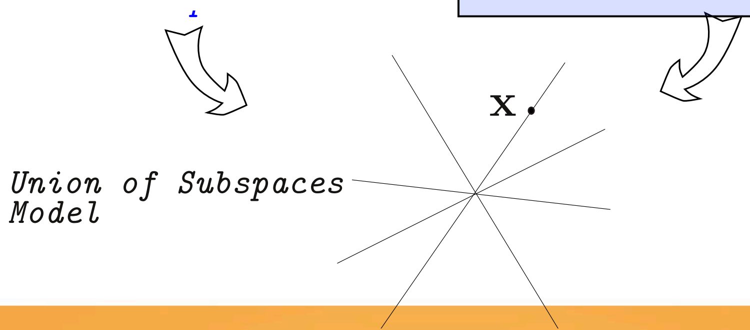


k := sparsity
= dimension of subspace

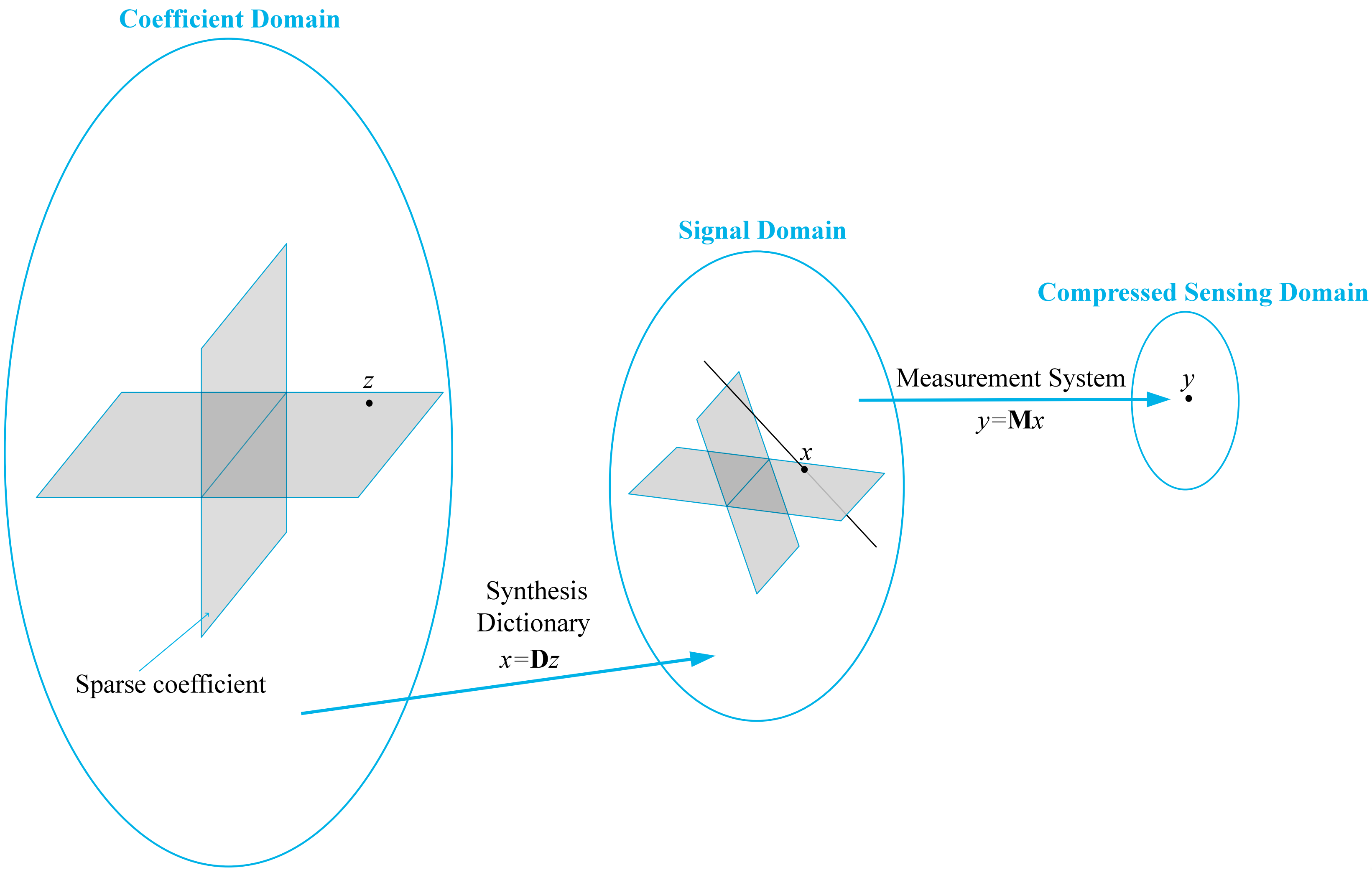
- **Cosparse analysis model**
  - ✓ Analysis operator $\mathbf{\Omega}$
  - ✓ Representation $\mathbf{\Omega x}$
  - ✓ **Cosupport** = location of zeroes



ℓ := cosparsity
= codimension of subspace

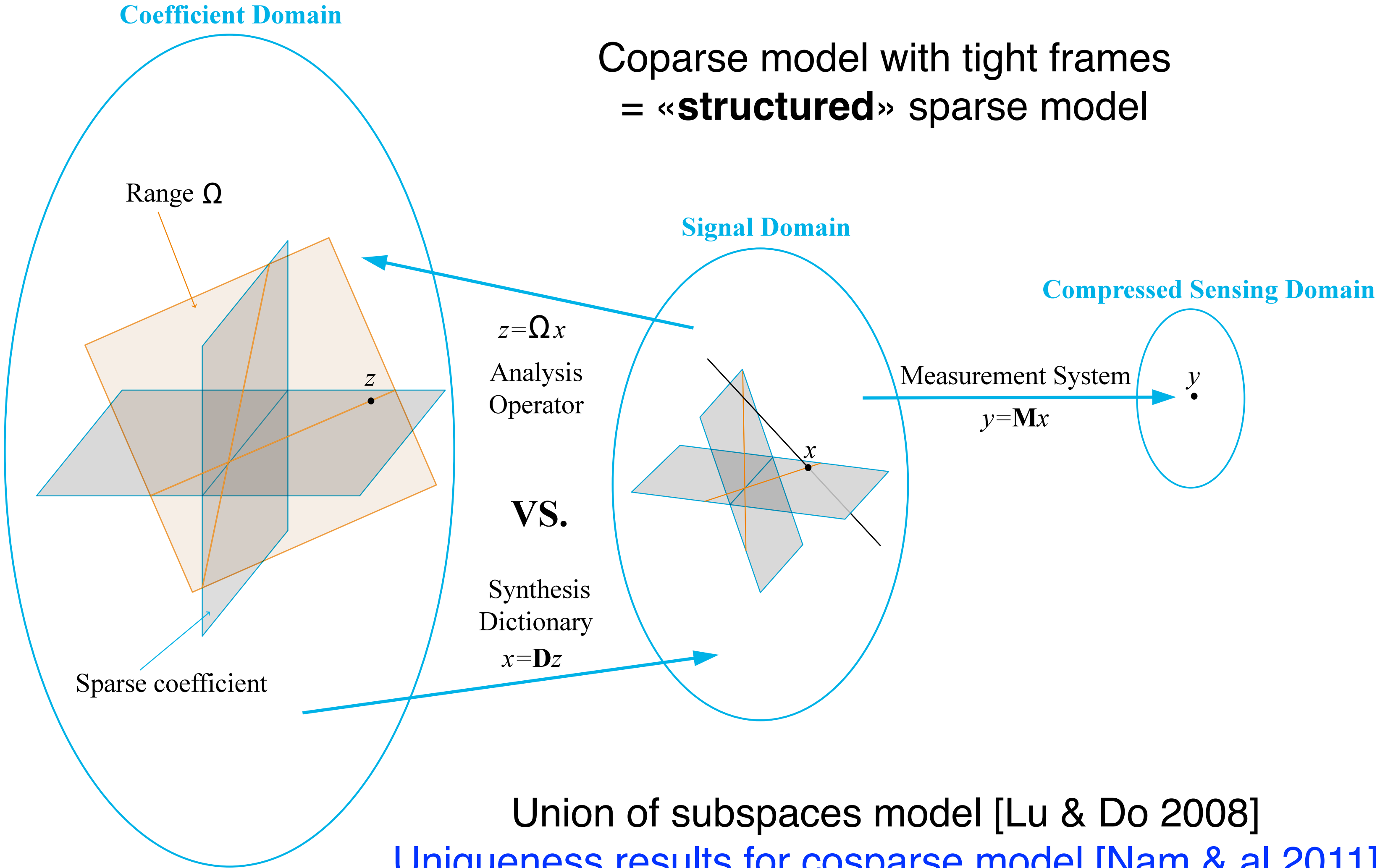


Sparse models and inverse problems



CoSparse models and inverse problems

Coparse model with tight frames
= «**structured**» sparse model



Union of subspaces model [Lu & Do 2008]
Uniqueness results for cosparse model [Nam & al 2011]

(Co)sparse recovery algorithms

Optimization principles / algorithms

- Idealized problem
  - Portilla 2009, Selesnick & Figueiredo 2009, Afonso & al 2010
  - Nam & al 2011 : **uniqueness guarantees**
- Convex relaxation
  - Starck & al 2003, Elad & al 2007, Kutyniok & Donoho 2010, Candès & al 2010
  - Nam & al 2011, Vaiter & al 2011 : **recovery guarantees**
- Greedy analysis pursuit (GAP)
  ~ analysis-OMP
  ~ agressive IRLS
  - Nam & al 2011 : **definition + recovery guarantees**
- Iterative cosparse projections
  ~ analysis-IHT
  - Gyries & al 2011: definition + recovery guarantees

$$\hat{\mathbf{x}}_{A-L0} := \arg\min_{\mathbf{x}:\mathbf{y}=\mathbf{M}\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_0$$

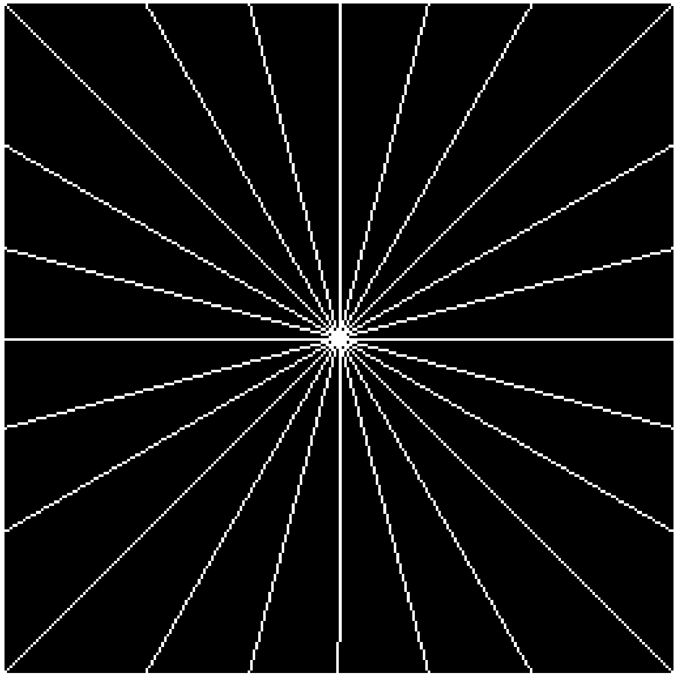
$$\hat{\mathbf{x}}_{A-L1} := \arg\min_{\mathbf{x}:\mathbf{y}=\mathbf{M}\mathbf{x}} \|\mathbf{\Omega}\mathbf{x}\|_1$$

- + Noise aware variants with

$$\|\mathbf{y} - \mathbf{M}\mathbf{x}\|_2 \leq \epsilon$$

- + Structured co-sparsity

Results: finite difference operator

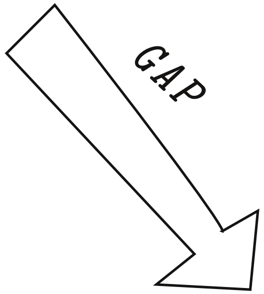


Fourier subsampling

$d = 256 \times 256 = 65536$

12 slices = radial lines

$m = 3032$ = 4.7%

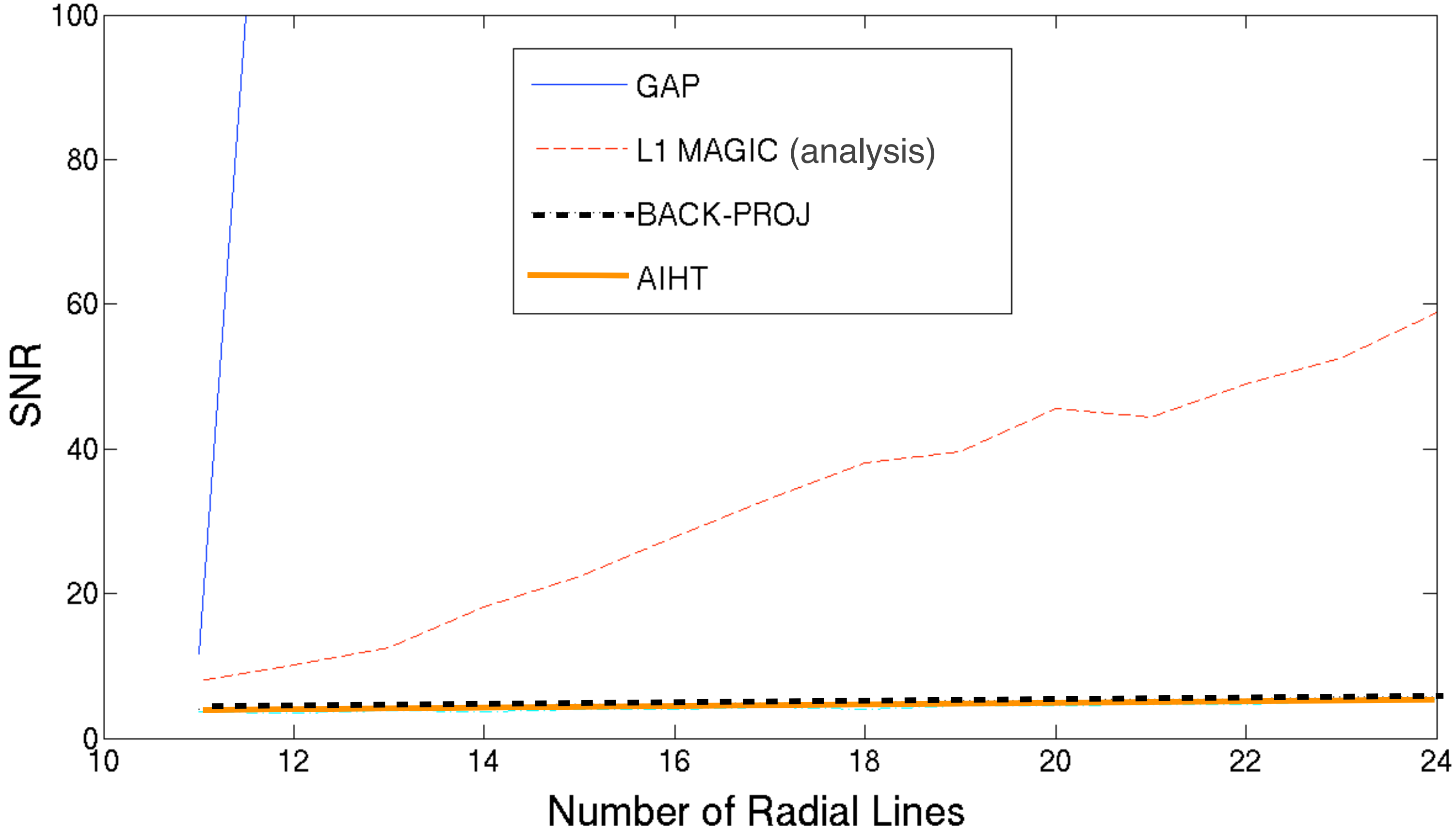


Uniqueness guarantee

$m \geq 2554$

Recovered Image

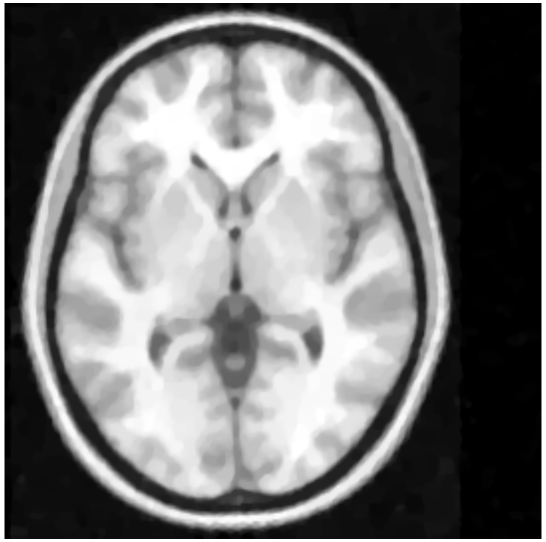
Results: finite difference operator



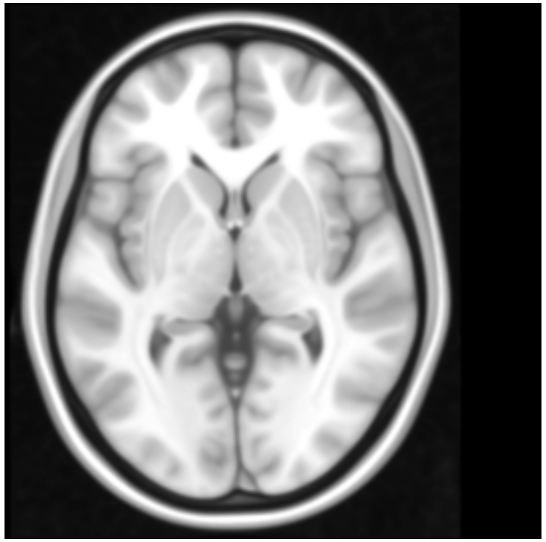
Linear dependencies = fewer small-dimensional subspaces (links with **matroids**?)

Robustness to approximate cosparsity

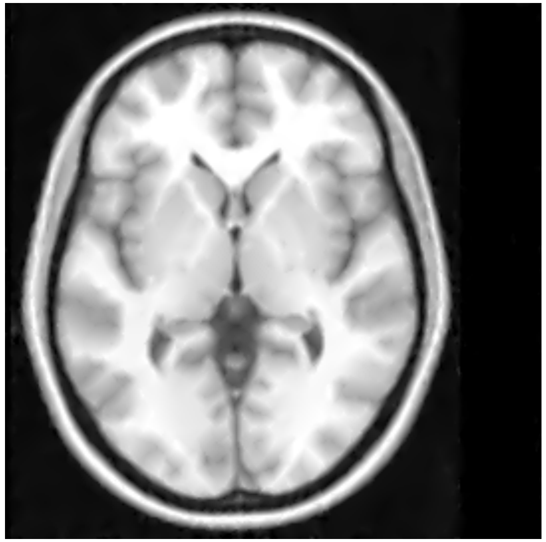
- Second order 2D finite difference operator
- 50 Fourier lines



✓ Original



TV



GAP

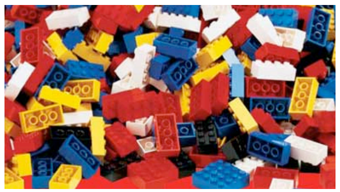
What's next ?

Summary

- Traditional **Synthesis Model**
  - ✓ **Synthesis dictionary** of *atoms*

$$\mathbf{x} = \mathbf{D}\mathbf{z} = \sum_i z_i \mathbf{d}_i \qquad \|\mathbf{z}\|_0 \ll \text{dimension}$$

  - ✓ «Lego» model: building blocks



  - ✓ Low-dimension = few atoms
    - ✦ Ex: man-made codes in communications

- Alternate **Analysis Model**
  - ✓ **Analysis operator**

$\langle \omega_i, \mathbf{x} \rangle = 0$ for many *rows* of $\mathbf{\Omega}$

$$\|\mathbf{\Omega}\mathbf{x}\|_0 \ll \text{dimension}$$

  - ✓ «Carving out» model: constraints

  - ✓ Low-dimension = many constraints
    - ✦ Ex: coupling with laws of physics

$$\left(\mathbf{\Delta x} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{x}\right)_{|\dot{\mathcal{D}}} = 0$$

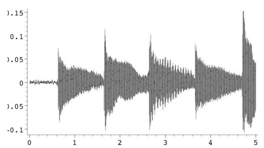
What's next ?

- New cosparse recovery **algorithms** with guarantees
  - GAP & Analysis-L1 [Nam & al 2011 a, Vaiter & al 2011 ...]
  - Analysis Iterative Hard Thresholding [Gyries & al 2011]
  - ...
- **Hybrid** sparse / cosparse models [Figueiredo & al 2011]
- **Group cosparsity** [Nam & al 2011 b]
- **Learning / designing** dictionary / analysis operator

  [Ophir & al 2011, Yaghoobi & al 2011, Peyré & Fadili 2011] [Nam & al 2011 b]

Data Deluge

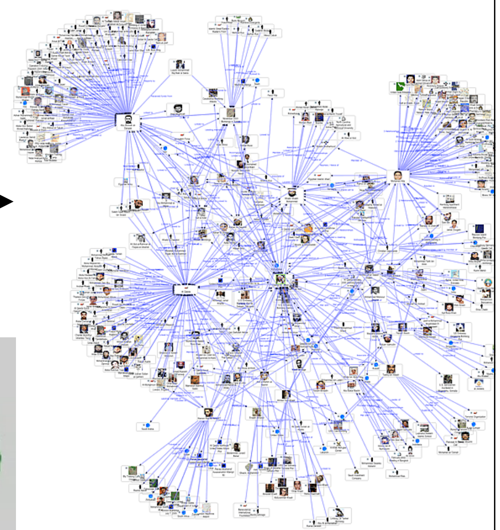
- Sparsity: historically for signals & images
  - ✓ bottleneck = **large-scale** algorithms

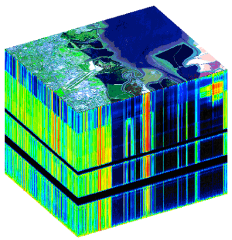


Signals

Images

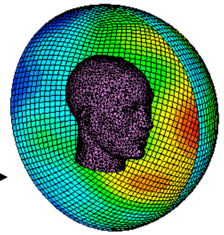
Graph data
Social networks
Brain connectivity

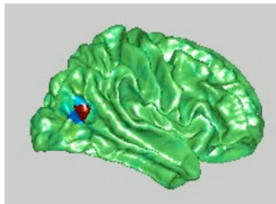




Hyperspectral
Satellite imaging

Spherical geometry
Cosmology, HRTF (3D audio)





Vector valued
Diffusion tensor

- New “exotic” or composite data
  - ✓ bottleneck = **dictionary/operator design/learning**

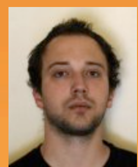
Scientific challenges ...

- *Fundamental objectives*
  1. Understand the **coupling** between sparsity and underlying **physical phenomena**
  2. Exploit underlying **structures** & **invariances** (graphs, geometric manifolds,...)
  3. Build **generic formalisms** to model **composite data** (multichannel, multimodal, ...)
- *A key challenge :*
  - ✓ **Data-adaptive models** through **learning** from a corpus

Cosparsity and physical models ?

with S. Nam, N. Bertin, G. Chardon, L. Daudet

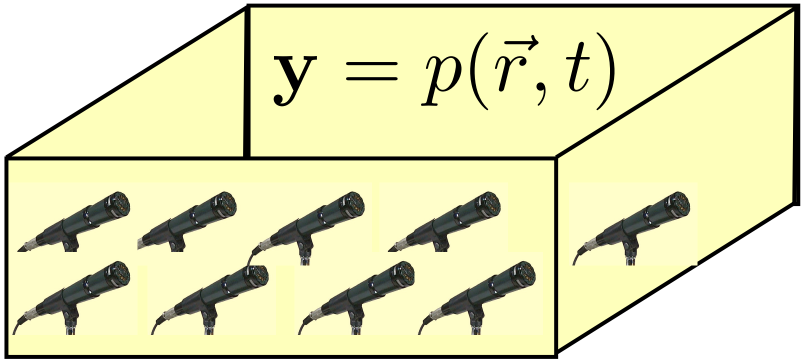





echange.inria.fr

3D scene acquisition (ANR project ECHANGE)

- **Acoustic pressure field**



Nyquist sampling →

✓ ***High dimension***

10^6 microphones / m^3

10^{12} samples / second

cosparse model from the wave equation

$$(\boldsymbol{\Omega}\mathbf{y})_{|\mathcal{D}} = 0$$

$\mathcal{D} =$ domain without acoustic source

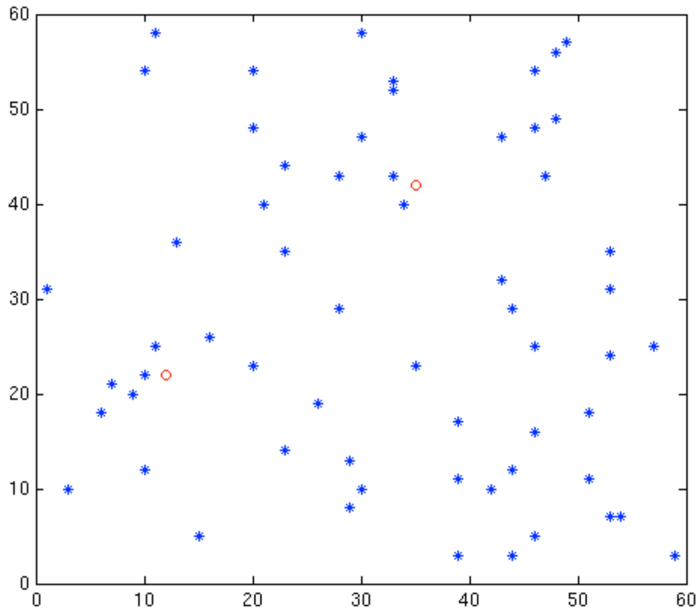
- operator = from Laplacian

$$\boldsymbol{\Omega}\mathbf{y} = \Delta_{\vec{r}}\mathbf{y} - \frac{1}{c^2}\frac{\partial^2 \mathbf{y}}{\partial t^2}$$

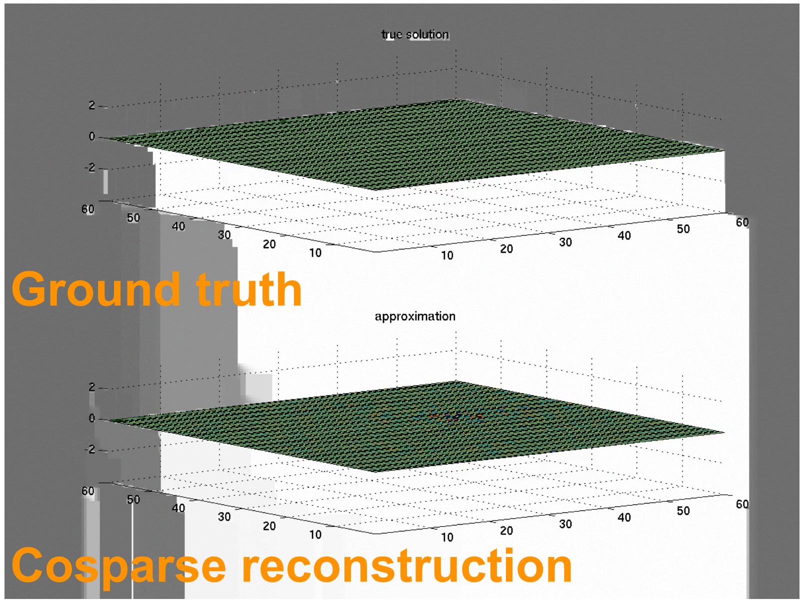
- + initial & boundary conditions

2D field reconstruction (simulation)

- 2D+t vibrating plate 60x60x120
  - ✓ 2 sources
  - ✓ 60 microphones, random location
  - ✓ known boundaries

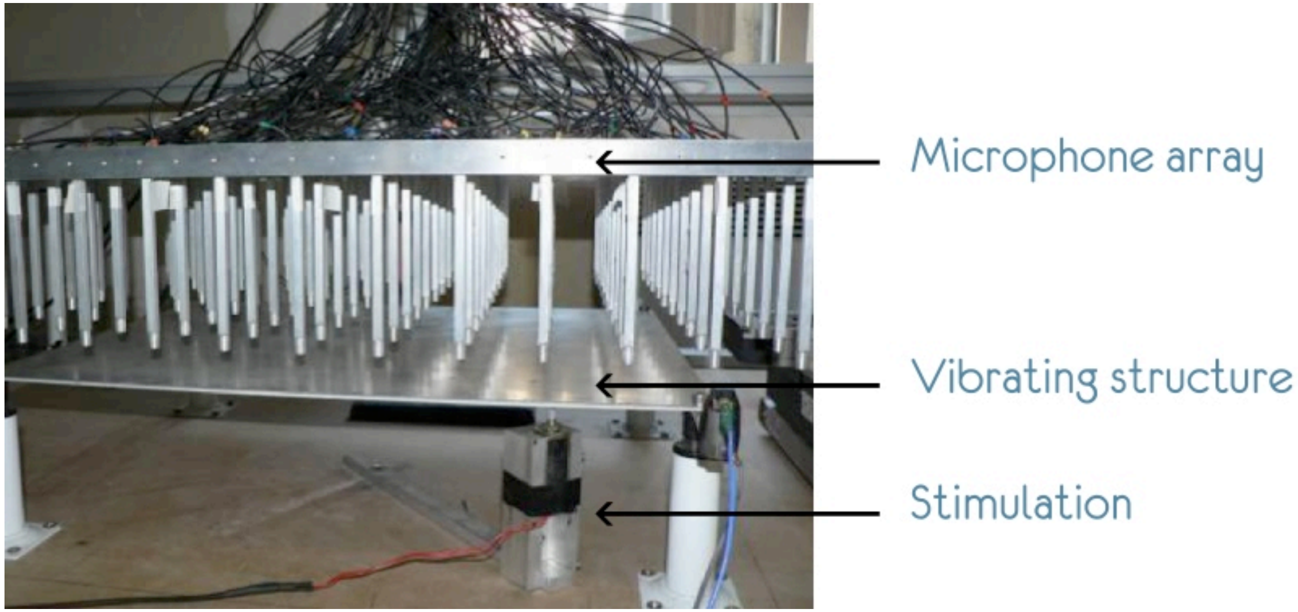


- Reconstructed 2D+t field
  - ✓ Structured-GAP (GRASP)



Nearfield Acoustic Holography

- Random antenna & sparsity: from 1920 to 80 microphones [Peillot & al 2011]
- Cosparsity?



Dictionary learning

with K. Schnass, F. Bach, R. Jenatton

Dictionary learning for sparse representations

- Sparse modeling = choose a dictionary

Training image database

patch extraction

Training patches

$$\mathbf{y}_n = \mathbf{A} x_n, \quad 1 \leq n \leq N$$

Unknown dictionary	Unknown sparse coefficients

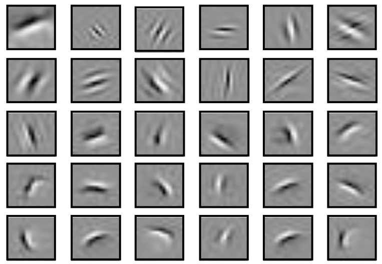
sparse learning

$\hat{\mathbf{A}}$ = edge-like atoms

[Olshausen & Field 96, Aharon et al 06, Mairal et al 09, ...]

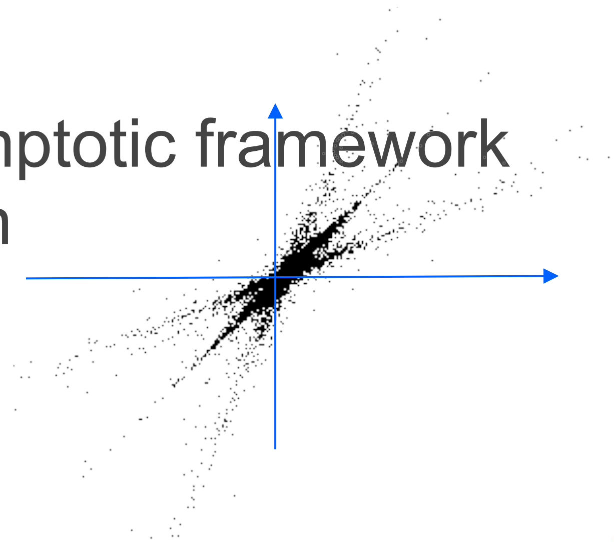
= shifts of edge-like motifs

[Blumensath 05, Jost et al 05, ...]



Theoretical Dictionary Learning

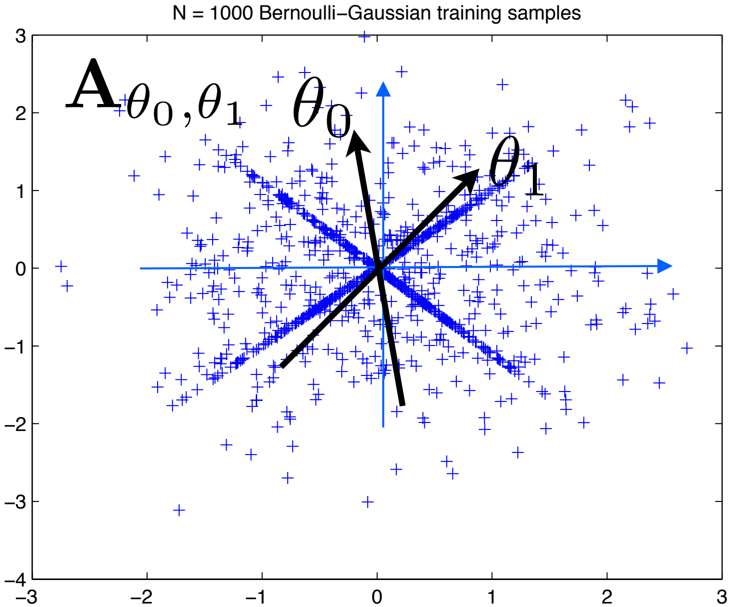
- Formalisation of the problem :
  - ✓ N observed data $\mathbf{y}_n = \mathbf{A}x_n \in \mathbb{R}^d \longrightarrow \mathbf{Y} = \mathbf{A}X$
  - ✓ Comon unknown dictionary **A**,
  - ✓ Unknown coefficients X + *sparsity hypothesis*
  - ✓ Goal: to identify **A**
- Appraoches
  - ✓ Independent Component Anal. = asymptotic framework
  - ✓ Proposed framework : L1 minimisation

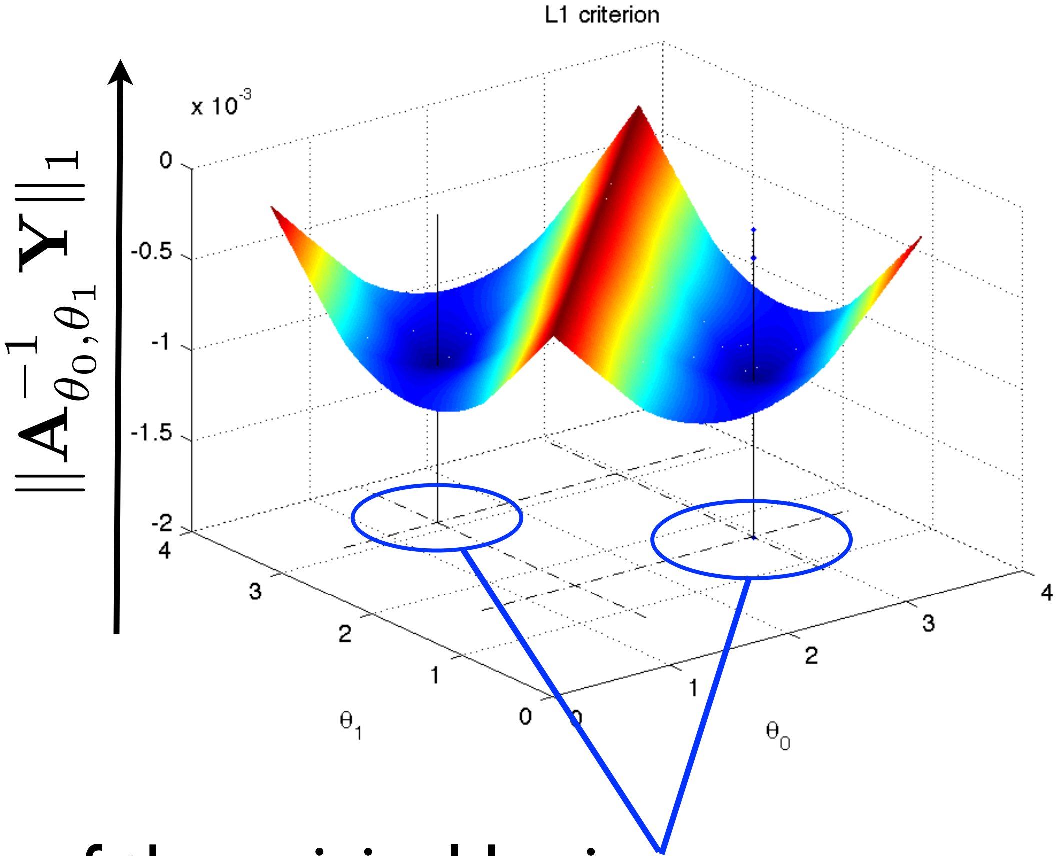


$$\min_{\mathbf{A},X|\mathbf{Y}=\mathbf{A}X} \|X\|_1$$

2D Numerical example

$$\mathbf{Y} = \mathbf{A}_0 X_0$$



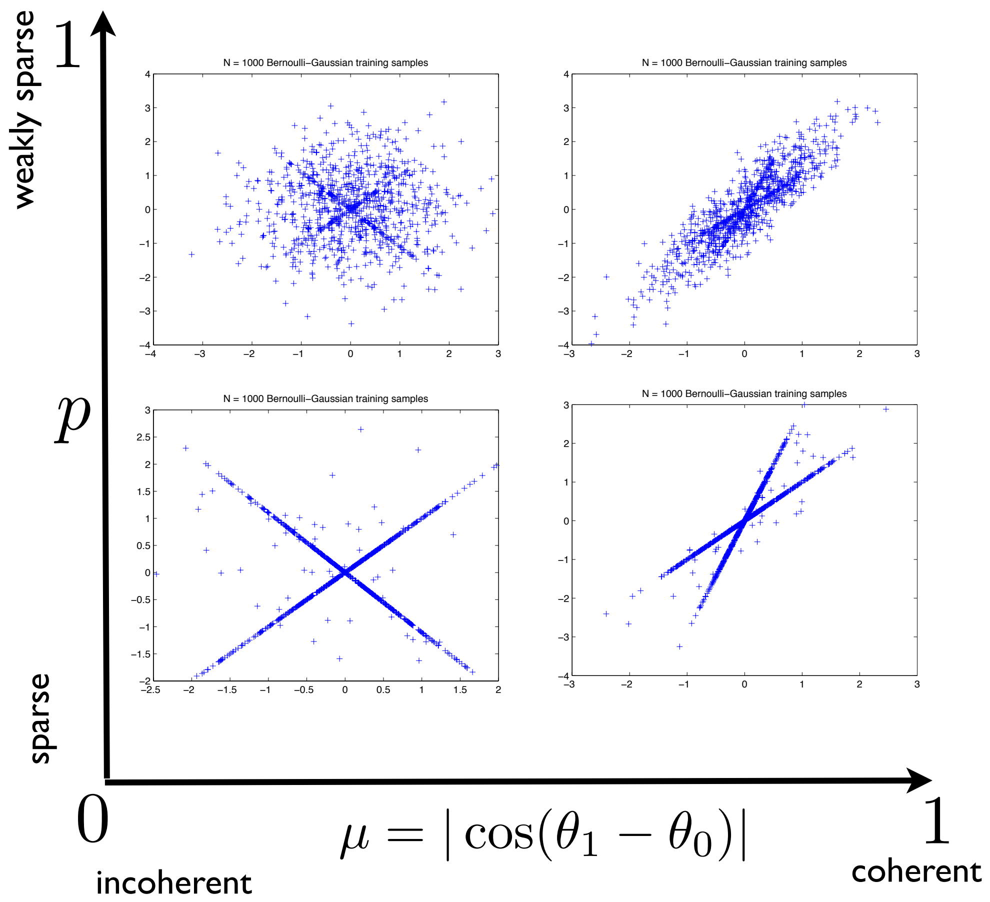


Empirical observations

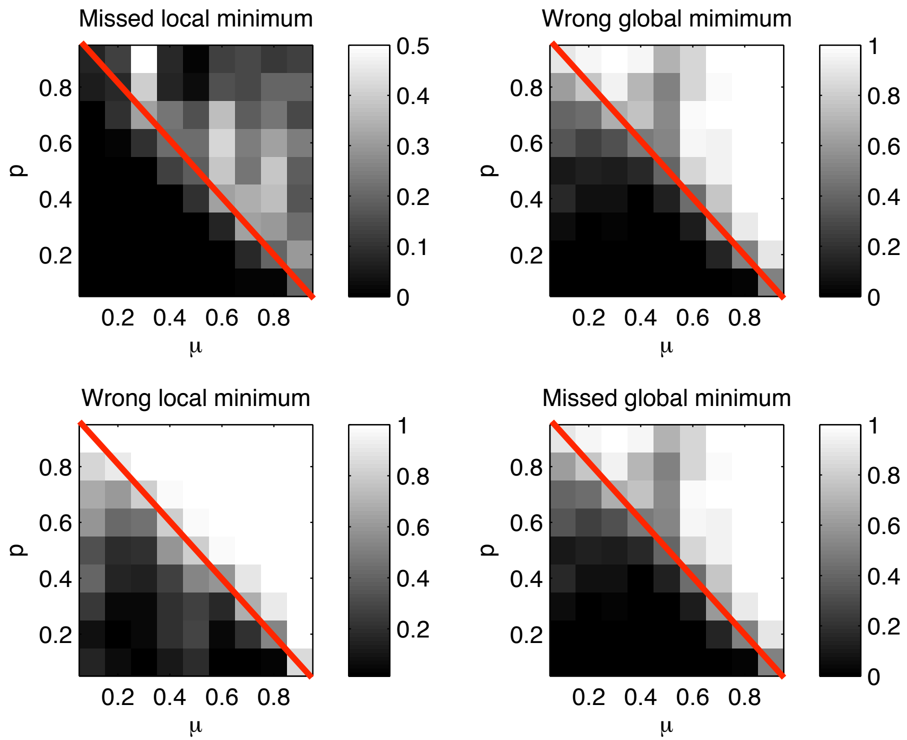
a) Global minima match angles of the original basis

b) There is no other local minimum.

Sparsity *vs* coherence



Empirical probability of erroneous minima



Rule of thumb: perfect recovery if:

a) $\mu < 1 - p$

b) N is large enough

(many training samples)

Recent Theoretical Guarantees

- [G. & Schnass 2010]:
  - ✓ Identifiability conditions for invertible **A**
  - ✓ Control on sufficient number of training samples *N* for identifiability with high probability (under Bernoulli-Gaussian model on X) in dimension *d*

$$N \geq C d \log d$$

- *[Bach,Jenatton, Gribonval] (work in progress)*
  - ✓ *Robustness to noise*
  - ✓ *Robustness to outliers*

$$\min_{\mathbf{A},X} \frac{1}{2}\|\mathbf{Y} - \mathbf{A}X\|_F^2 + \lambda\|X\|_1$$

INRIA & Machine Learning?

INRIA, Machine Learning & Signal Processing!

Signal Processing & Machine Learning

- Exploit **Sparsity**

✓ *Analog level*: object = data vector	...	✓ *«Semantic» level*: object = function

$$\mathbf{y} \approx \sum_{\mathbf{few}\ k} x_k \mathbf{a}_k = \mathbf{A}x$$

✓ Data model = atoms of a *dictionary*

✓ *Compression, representation, ...*

$$f(\mathbf{y}) \approx \sum_{\mathbf{few}\ n} \alpha_n K(\mathbf{y}, \mathbf{y}_n)$$

✓ Function model = *kernel*

✓ *Classification, regression, ...*

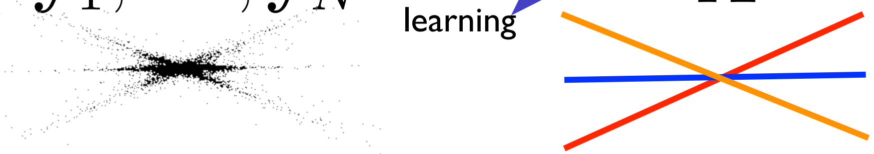
Signal Processing & Machine Learning

- **Learn** from a collection

✓ *Analog level*: object = data vector …

✓ *«Semantic» level*: object = function

$\mathbf{y}_1, \ldots, \mathbf{y}_N$ → learning → $\hat{\mathbf{A}}$



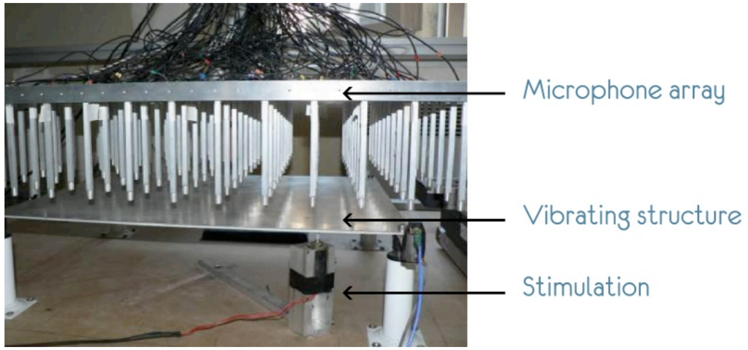
✓ *Dictionary Learning*:
infer model from examples

$\mathbf{y}_1, \ldots, \mathbf{y}_N$ → learning → $\hat{f}(\cdot)$

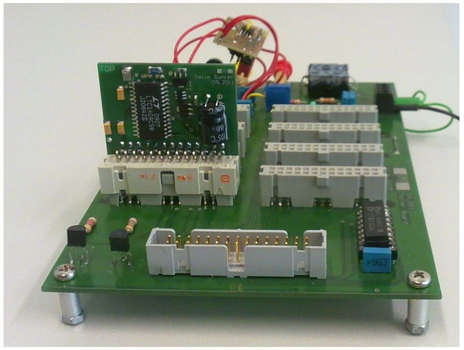
✓ *Machine Learning*:
infer function from examples

Signal Processing *vs* Machine Learning ?

- Signal Processing meets Machine Learning
  - ✓ model / learn from data
  - ✓ algorithms (scalability)
  - ✓ high-dimensional statistics, large-scale optimization
  - ✓ exploit sparsity
- +SP specific: hardware constraints/design



random acoustic antenna



compressed ECG A/D converter

Machine Learning *vs* Signal Processing?

- Machine Learning meets Signal Processing
  - ✓ model / learn from data
  - ✓ algorithms (scalability)
  - ✓ high-dimensional statistics, large-scale optimization
  - ✓ exploit sparsity
- +Machine Learning :
  - ✓ fundamental complexity / performance tradeoffs
    - ✦ Statistical Learning != Machine Learning

THANKS

Learning @ INRIA?

- Strong existing task force but
- Spectrum of required skills ever increasing...
  - ✓ data specific expertise (biomedical / audiovisual / ...)
  - ✓ + applied math
  - ✓ + large-scale optim.
  - ✓ + probability & stats
  - ✓ + complexity theory...
- Need to secure / consolidate INRIA position
  - ✓ Hiring: Junior Researchers & R&D Engineers
  - ✓ Revisit organisation of Themes ?

Reconstruction parcimonieuse: garanties théoriques

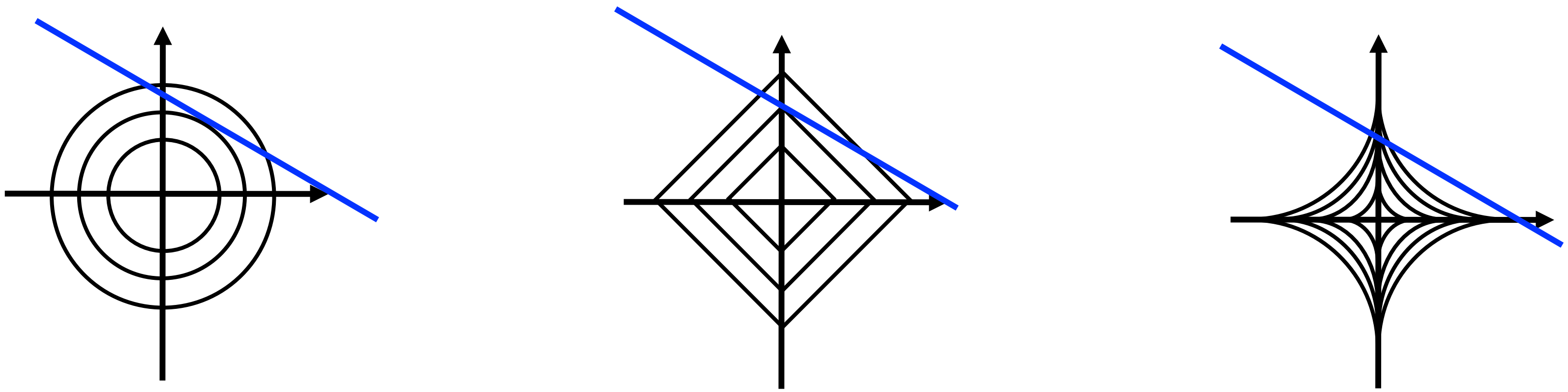
- **Théorème** : soit $z = \mathbf{A}x_0$
  - ✓ si $\|x_0\|_0 \leq k_0(\mathbf{A})$ alors $x_0 = x_0^\star$
  - ✓ si $\|x_0\|_0 \leq k_p(\mathbf{A})$ alors $x_0 = x_p^\star$

avec $x_p^\star = \arg \min_{\mathbf{A}x = \mathbf{A}x_0} \|x\|_p$

- [Donoho & Huo 01] : *paires de bases, notion de cohérence, p=1*
- [Gribonval & Nielsen 2003] [Donoho & Elad 2003] : *dictionnaires, cohérence*
- [Tropp 2004] : *Orthonormal Matching Pursuit, cohérence cumulative*
- [Candès, Romberg & Tao 2004] : *dictionnaires aléatoires, isométrie restreintes*
- [Gribonval & Nielsen 2007] : *résultats d'interpolation 0<p<1*

Géométrie des “normes” Lp

- Strictement convexe $p>1$
- Convexe pour $p=1$
- Non convexe $p<1$



Fait: *la solution est parcimonieuse*

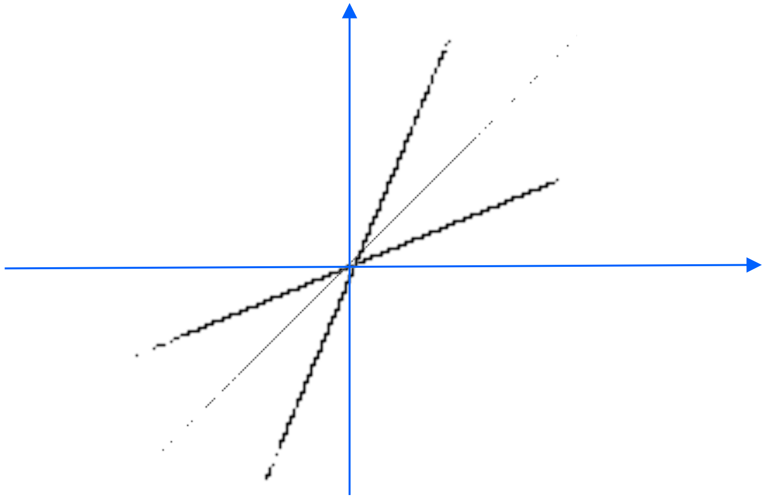
$\{x \text{ tel que } \mathbf{b} = \mathbf{A}x\}$

Stabilité

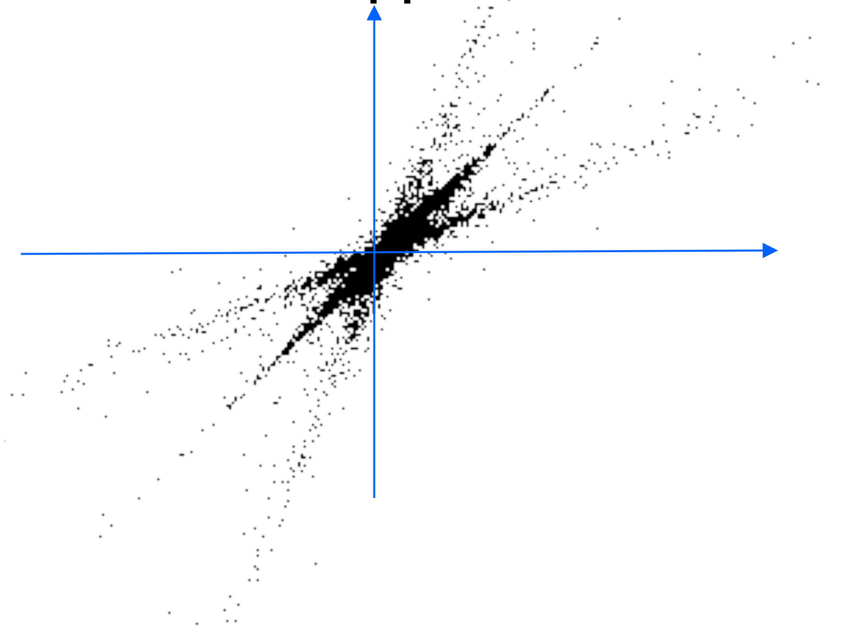
$$y = \mathbf{A}x, \|x\|_0 = 1$$

$$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$$

Parcimonie idéale

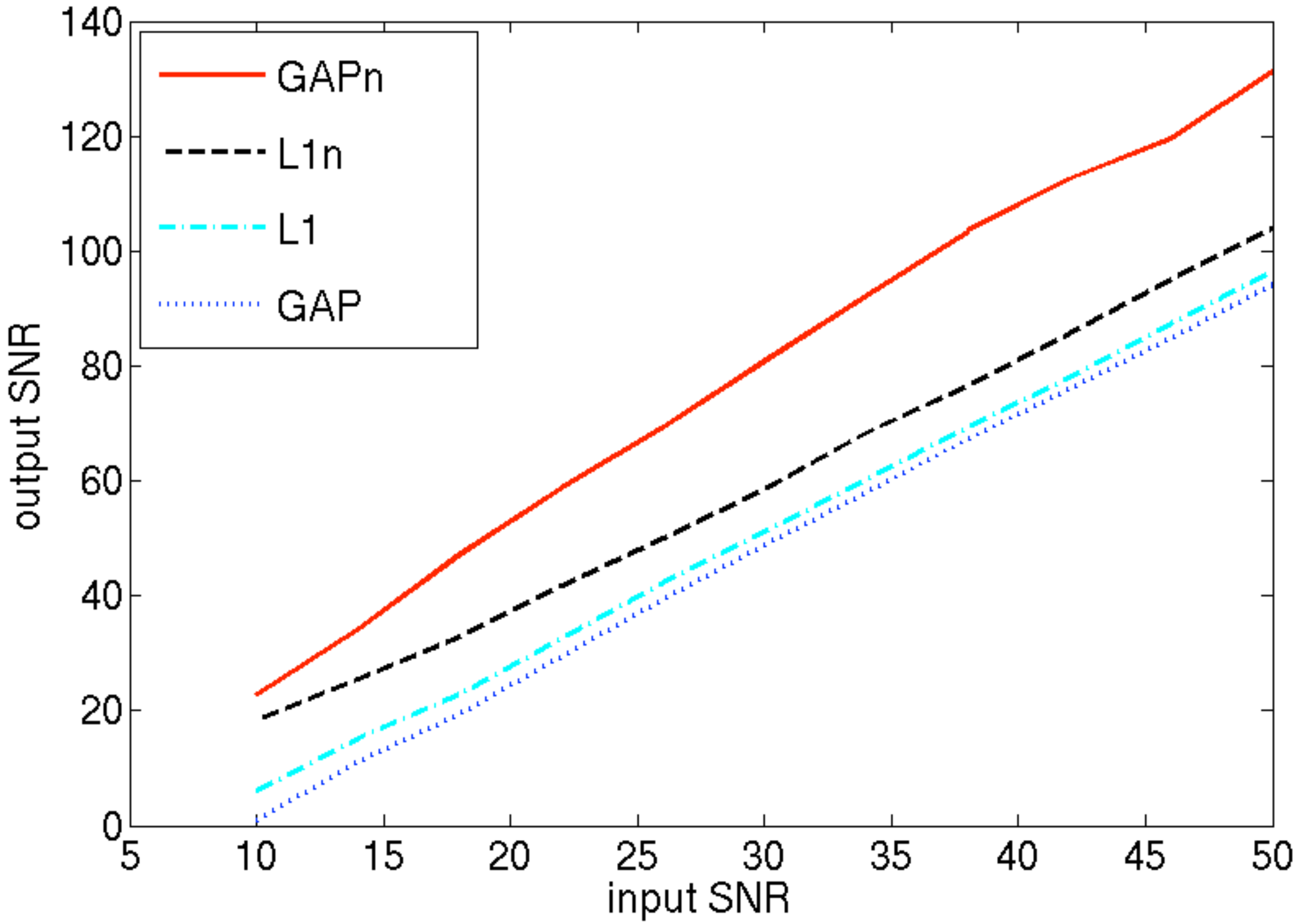


Parcimonie approchée



Robustness

Robustness to noise



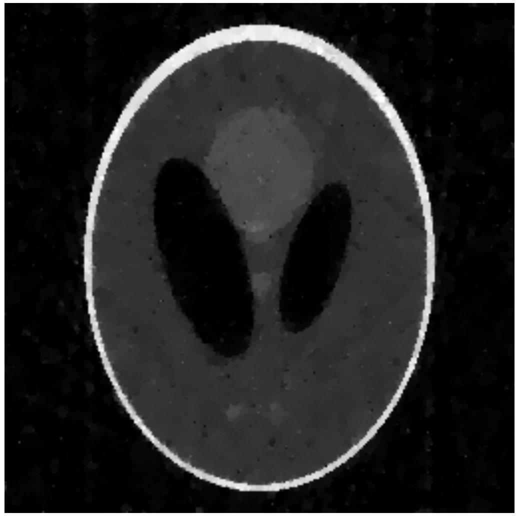
d=200
n=240
m=160
l=180

Robustness to noise

- Original

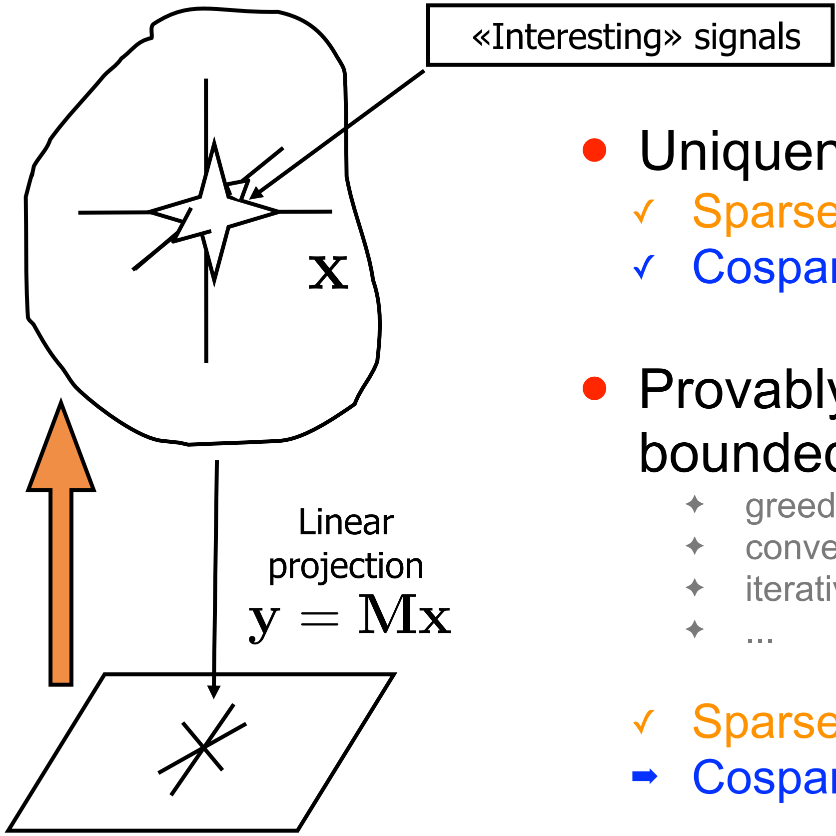
- TVDantzig

- TVQC



- GAP TV

Geometry of Inverse Problems



- Uniqueness given the model
  - ✓ Sparse model (*spark* of **D**)
  - ✓ Cosparse model
- Provably good algorithms, with bounded complexity
  - greedy methods
  - convex optimization
  - iterative hard thresholding
  - ...

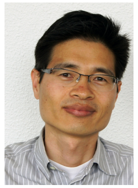
  ✓ Sparse model

  ➡ Cosparse model ????

Thanks to M. Davies, U. Edinburgh

THANKS

- Cosparse model:
  - Sangnam Nam (INRIA, France)
  - Mike Davies (University of Edinburgh, UK)
  - Miki Elad (The Technion, Israel)

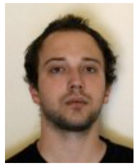


small-project.eu

SMALL

SEVENTH FRAMEWORK PROGRAMME

- Acoustic imaging
  - Laurent Daudet, Gilles Chardon
  - François Ollivier, Antoine Peillot
  - Nancy Bertin


...

echange.inria.fr/

echange

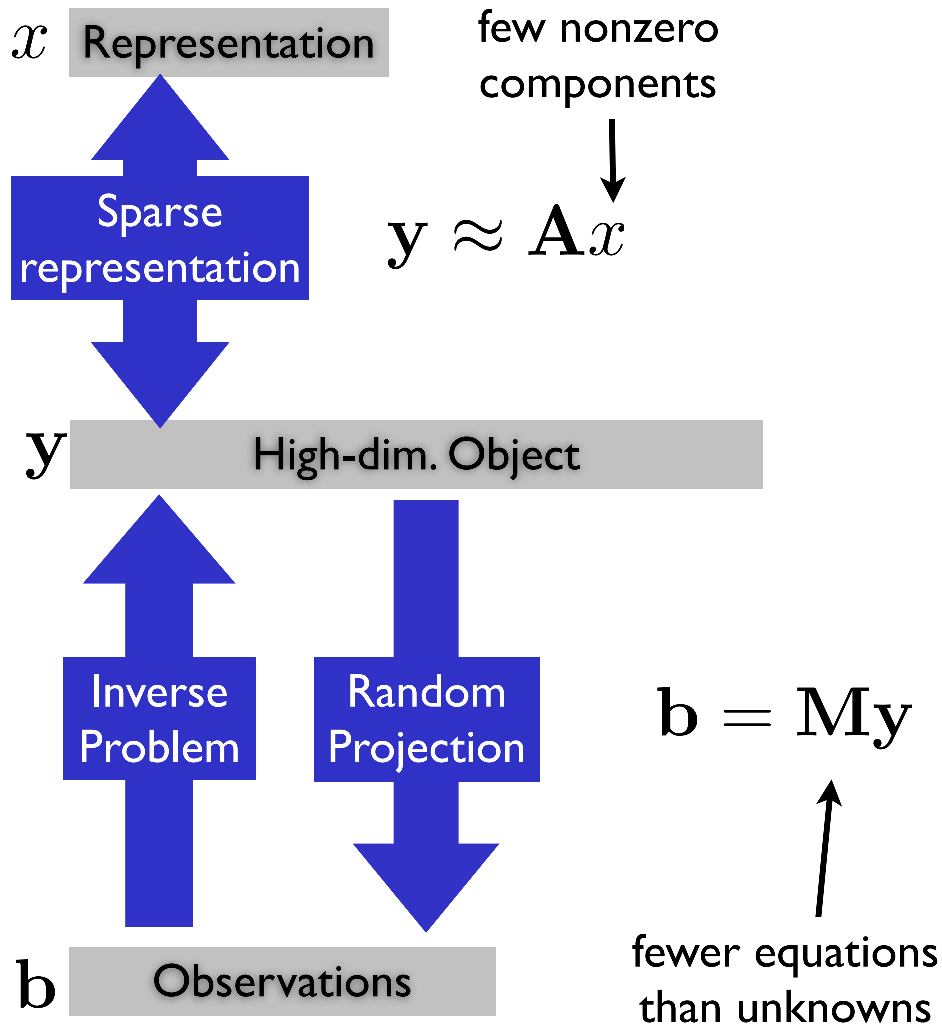
SUPPORTED BY ANR

- Graphical design:
  - Jules Espiau (INRIA, France)

erc European Research Council

Sparsity & Projections
In Signal Processing **and Machine Learning**

- **Sparsity**:
  approximate *high-dimensional object* with few parameters
  - ex: signal, image
- **Inverse problem**:
  accurately reconstruct object from *incomplete observation*
  - ex: tomography
- **Random projections**
  intentionally reduce dimension, with stable reconstruction
  - ✓ theoretical guarantees
  - ✓ algorithms of bounded complexity
    - ex: compressed sensing,
      compressed machine learning



What's next ?

- New cosparse recovery **algorithms** with guarantees
  - GAP & Analysis-L1 [Nam & al 2011 a, Vaiter & al 2011 ...]
  - Analysis Iterative Hard Thresholding [Gyries & al 2011]
  - ...
- **Hybrid** sparse/cosparse models [Figueiredo & al 2011]
- **Group cosparsity** [Nam & al 2011 b]
- **Learning/designing** analysis operators [Ophir & al 2011, Yaghoobi & al 2011, Peyré & Fadili 2011]
- Connections with PDEs ... [Nam & al 2011 b]
- More applications ...