

Inverse problems and sparse representations

Rémi Gribonval, DR INRIA
EPI METISS (Speech and Audio Processing)
INRIA Rennes - Bretagne Atlantique

remi.gribonval@inria.fr

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



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Structure of the course

- Session 1: Panorama
 - ◆ role of sparsity for compression and inverse problems
 - ◆ introduction to compressed (random) sensing
- Session 2: Algorithms
 - ◆ Review of main algorithms & complexities
 - ◆ Success guarantees for L1 minimization to solve under-determined inverse linear problems
- Session 3: Deterministic vs Random dictionaries
 - ◆ Comparison of guarantees for different algorithms
 - ◆ Robust guarantees & Restricted Isometry Property
 - ◆ Explicit guarantees for various inverse problems
- Session 4: Bayesian (mis)interpretations

Overview of Session I

- Sparsity and compression of large-scale data
- Sparsity for source separation and inverse problems
- Sparse decomposition algorithms
 - ◆ L1 minimisation
 - ◆ Matching Pursuits
- Provably good algorithms
- Sparsity and compressed sensing

Sparsity & data compression



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



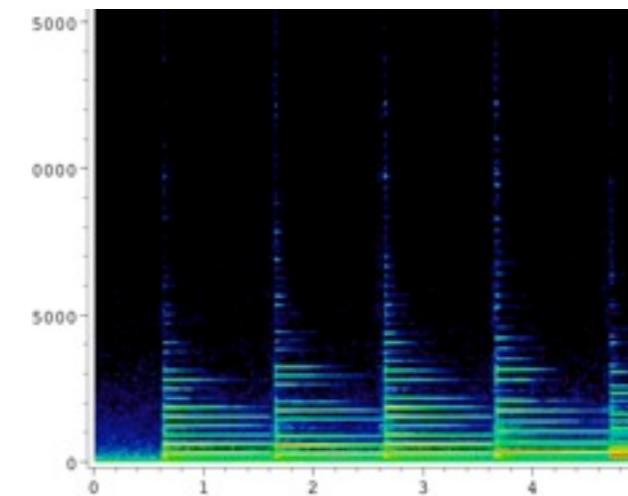
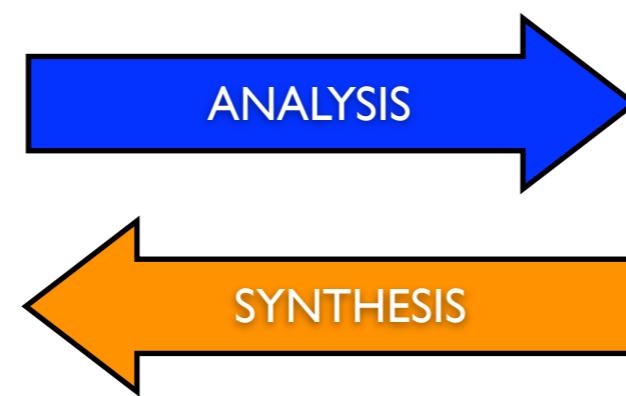
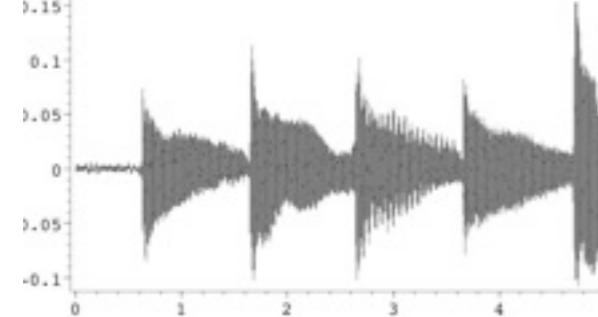
centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Large-scale data

- **Fact** : digital data = large volumes
 - ◆ 1 second stereo audio, CD quality = 1,4 Mbits
 - ◆ 1 uncompressed 10 Mpixels picture = 240 Mbits
- **Need** : «concise» data representations
 - ◆ storage & transmission (volume / bandwidth) ...
 - ◆ manipulation & processing (algorithmic complexity)

Sparse representations

- Audio : time-frequency representations (MP3)

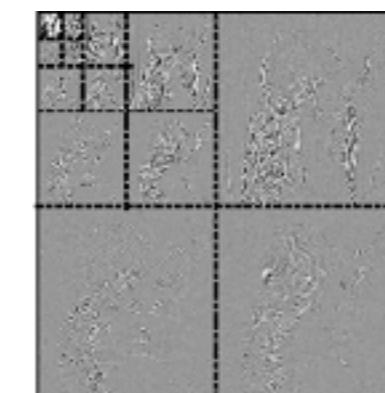
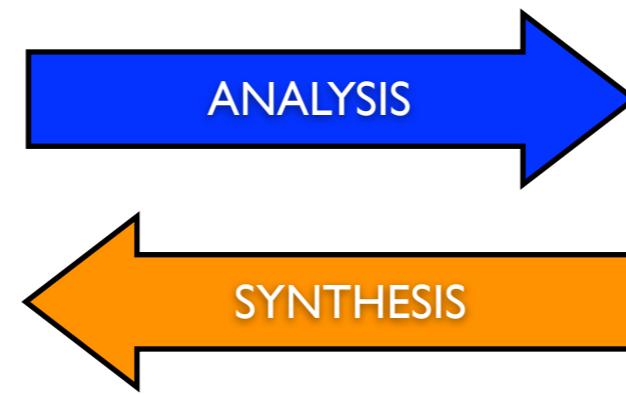


Black
= zero

- Images : wavelet transform (JPEG2000)



128, 129, 125, 64, 65,



4123, -12.4, -96.7, 45,

Gray
= zero

Mathematical expression

- Signal / image = high dimensional vector

$$y \in \mathbb{R}^N$$

- **Model** = linear combination of basis vectors
(ex: *time-frequency atoms, wavelets*)

$$y \approx \sum_k x_k \varphi_k = \Phi x$$

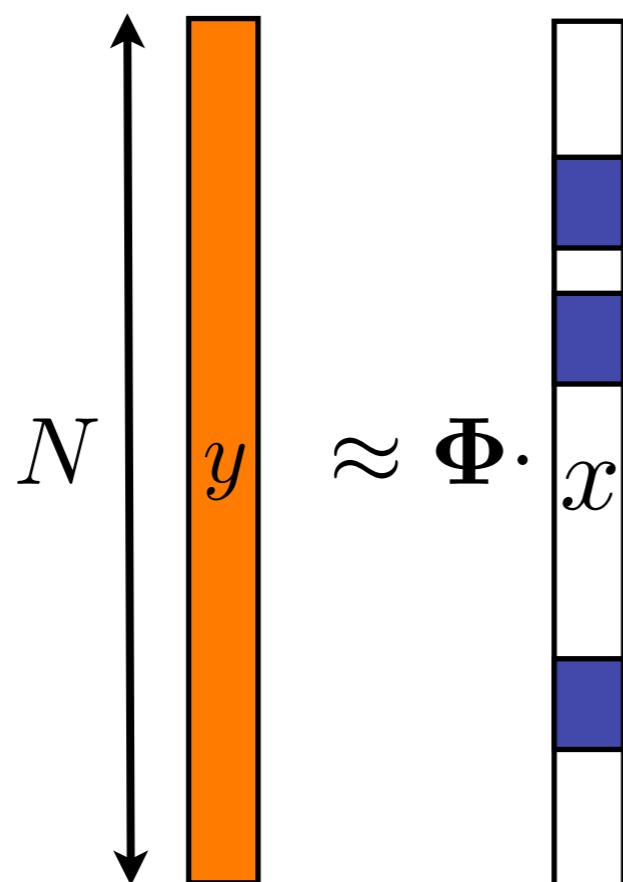
atoms
dictionary

- **Sparsity** = small L0 (quasi)-norm

$$\|x\|_0 = \sum_k |x_k|^0 = \text{card}\{k, x_k \neq 0\}$$

Sparsity & compression

- Full vector

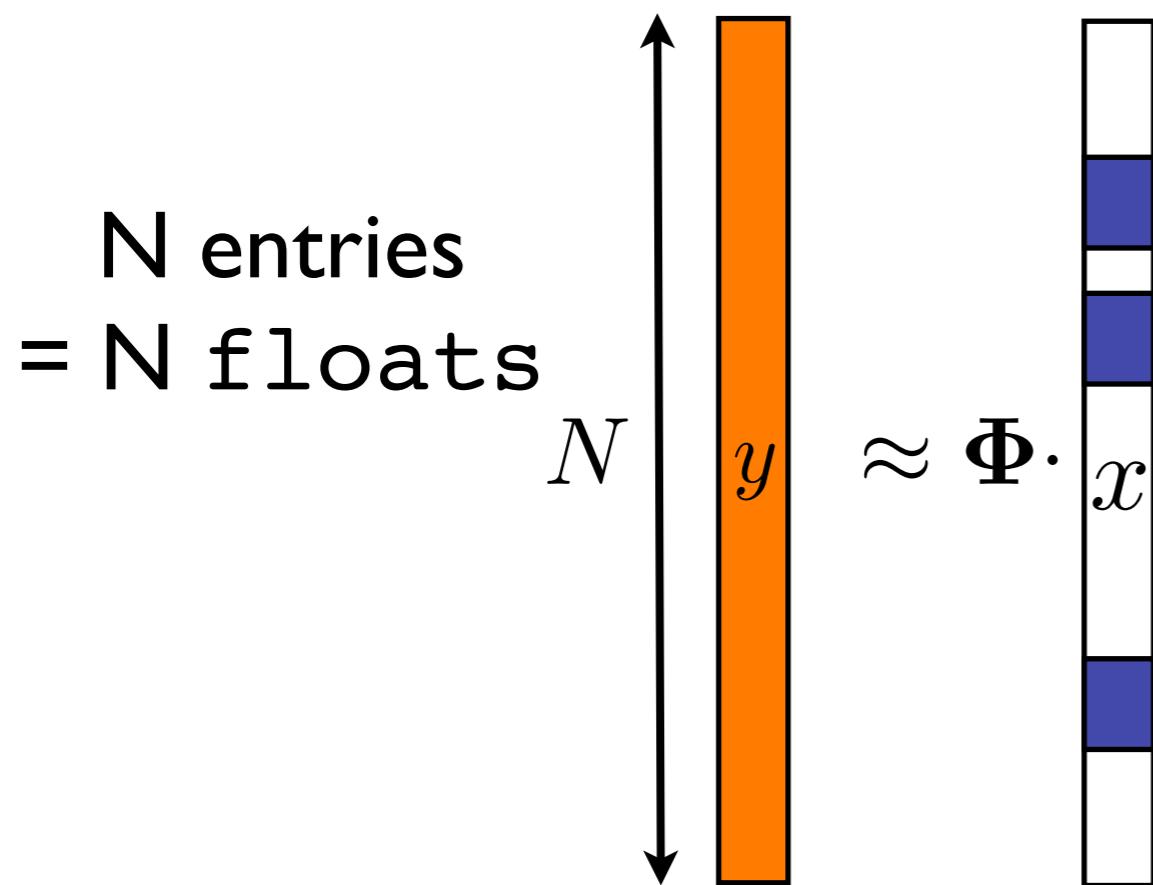


- Sparse vector

Key practical issues: choose dictionary

Sparsity & compression

- Full vector

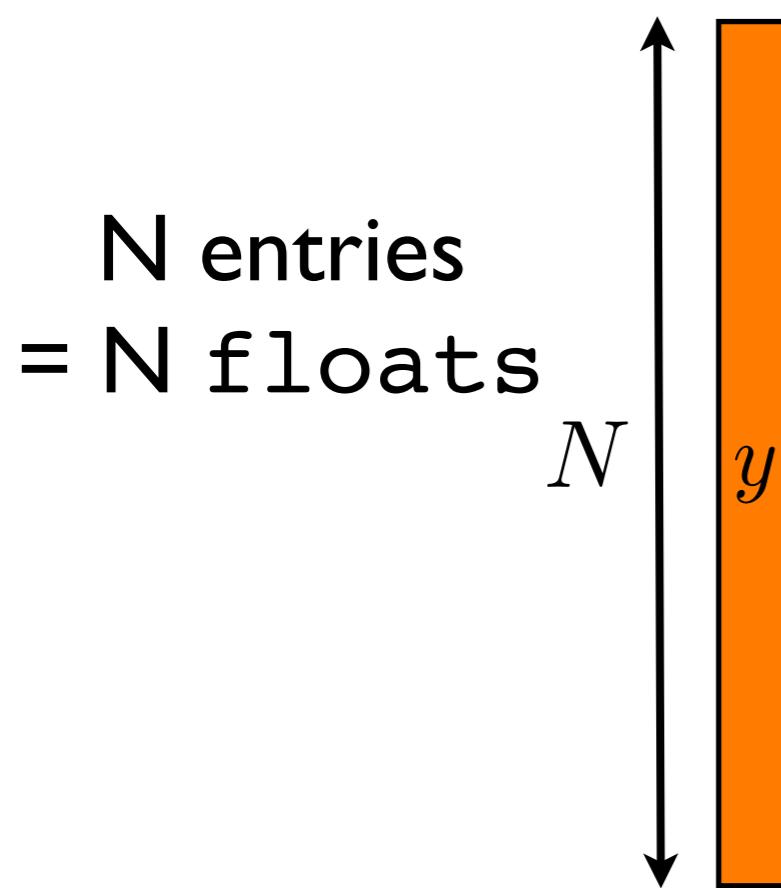


- Sparse vector

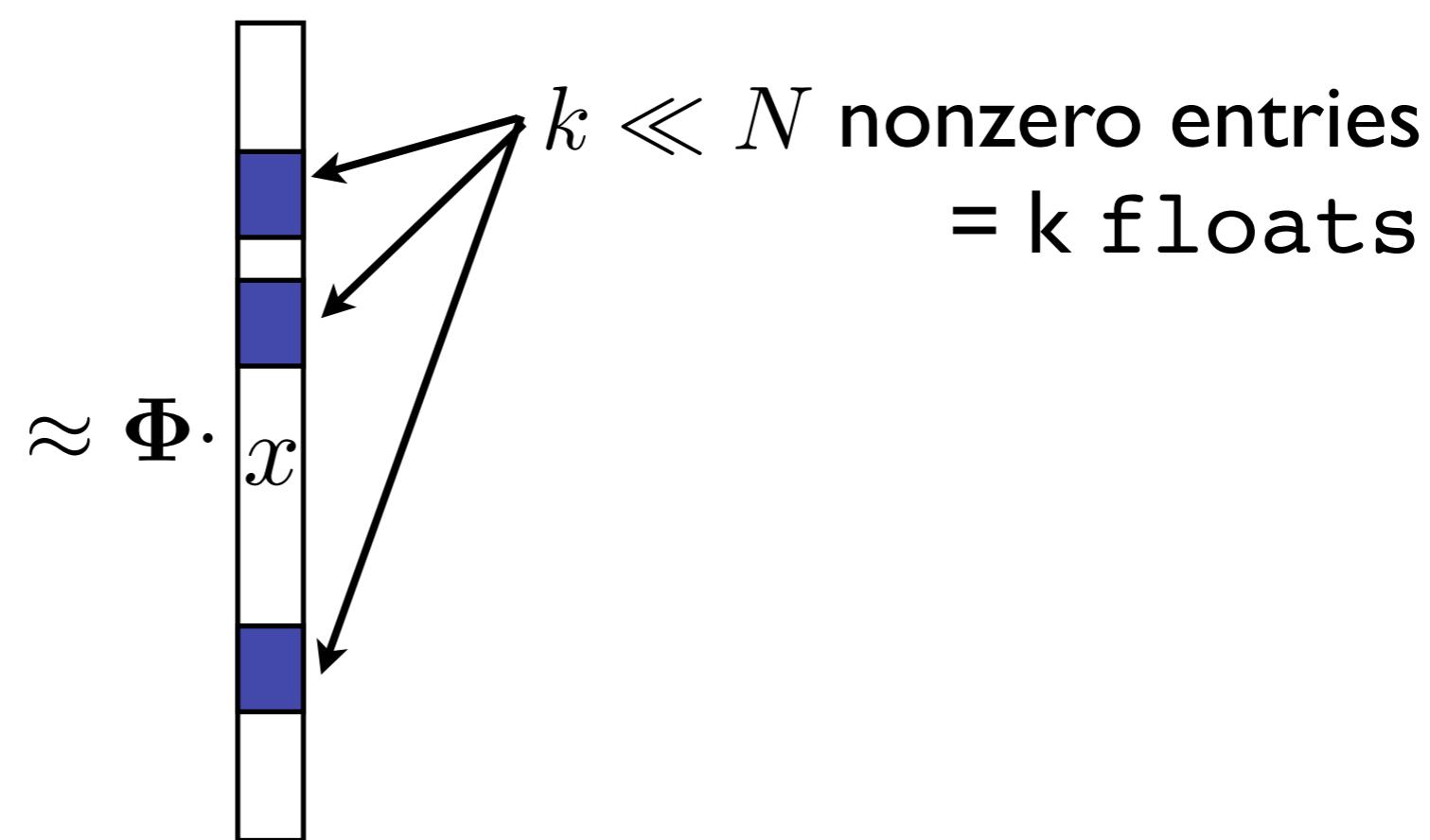
Key practical issues: choose dictionary

Sparsity & compression

- Full vector



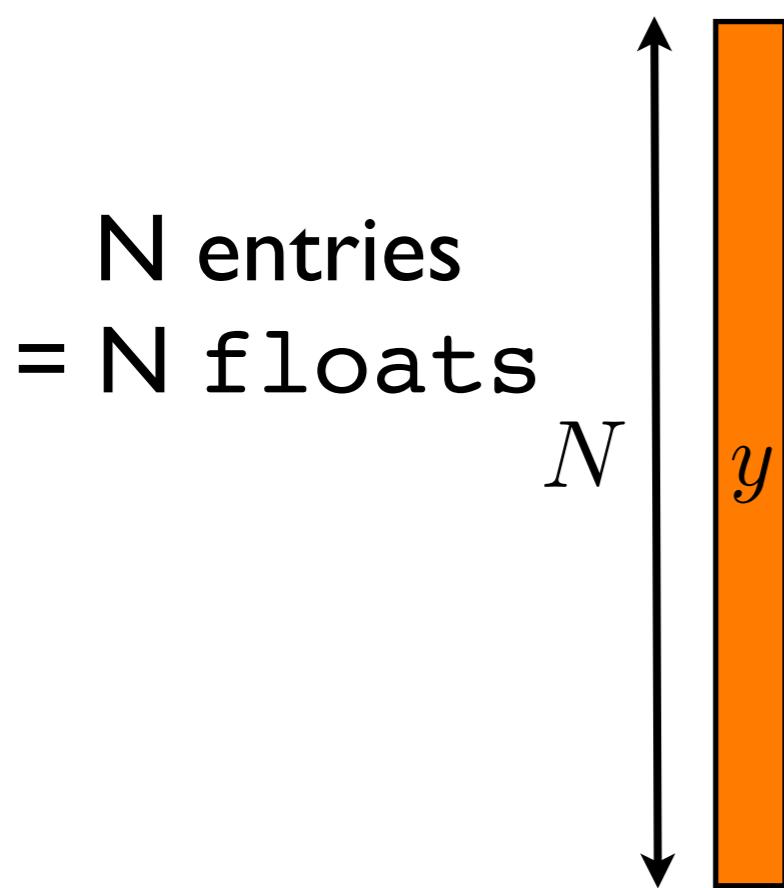
- Sparse vector



Key practical issues: choose dictionary

Sparsity & compression

- Full vector



- Sparse vector

$k \ll N$ nonzero entries
= k floats

+ k positions among N

$$= \log_2 \binom{N}{k} \approx k \log_2 \frac{N}{k} \text{ bits}$$

Key practical issues: choose dictionary

Sparsity & inverse problems



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Example: image inpainting

Courtesy of: G. Peyré, Ceremade, Université Paris 9 Dauphine



Example: image inpainting

Courtesy of: G. Peyré, Ceremade, Université Paris 9 Dauphine



Example: image inpainting

Courtesy of: G. Peyré, Ceremade, Université Paris 9 Dauphine

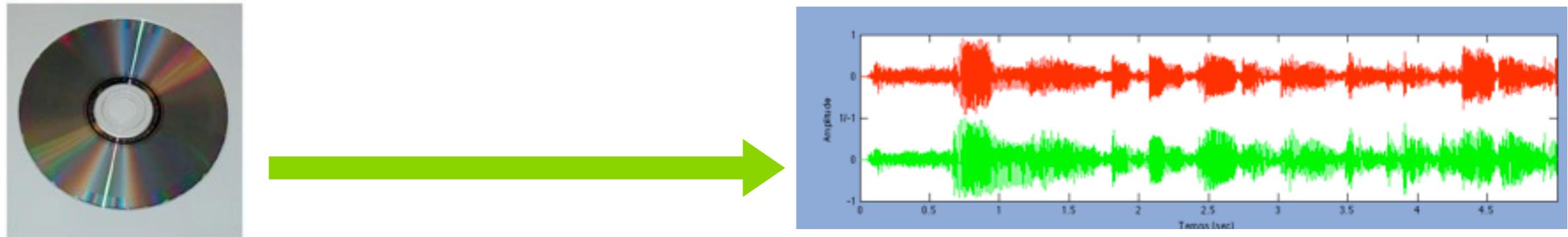


Inpainting



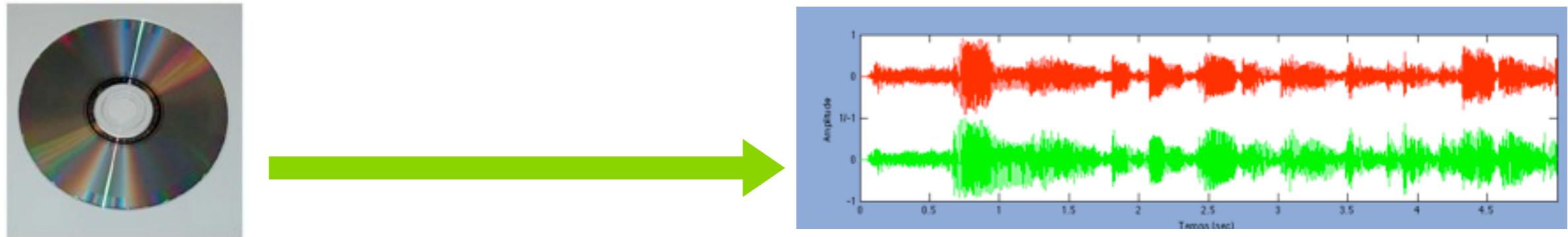
Example : audio source separation

- « Softly as in a morning sunrise »



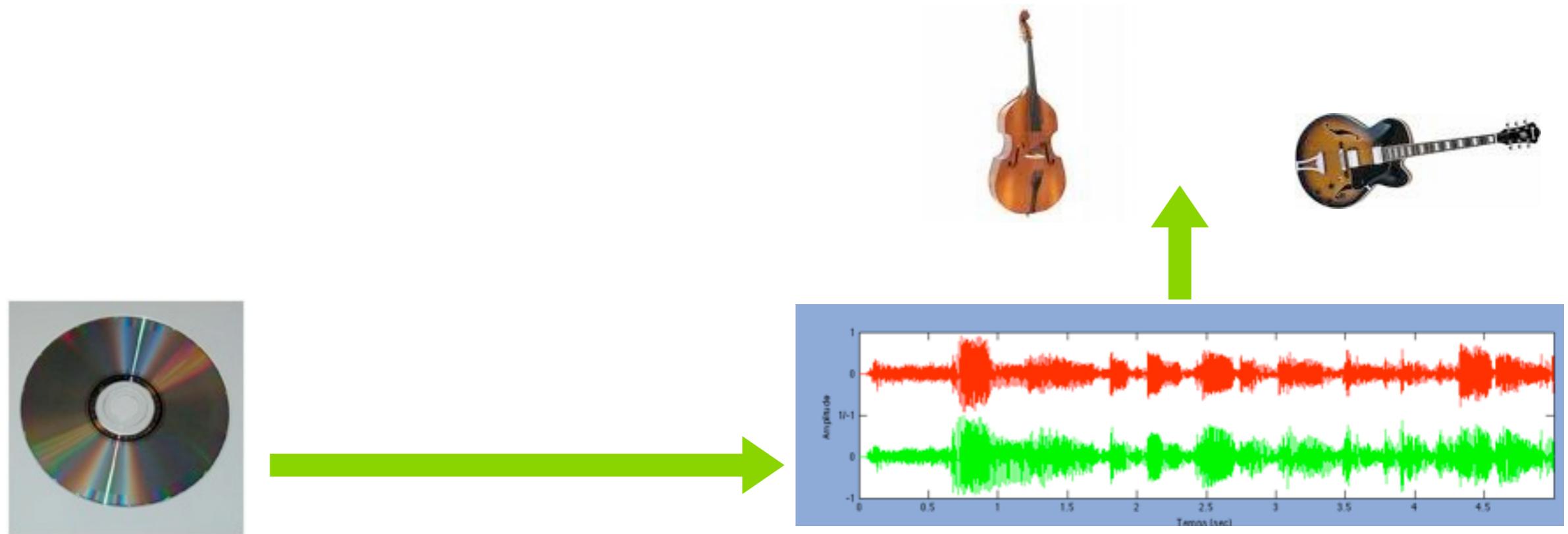
Example : audio source separation

- « Softly as in a morning sunrise »



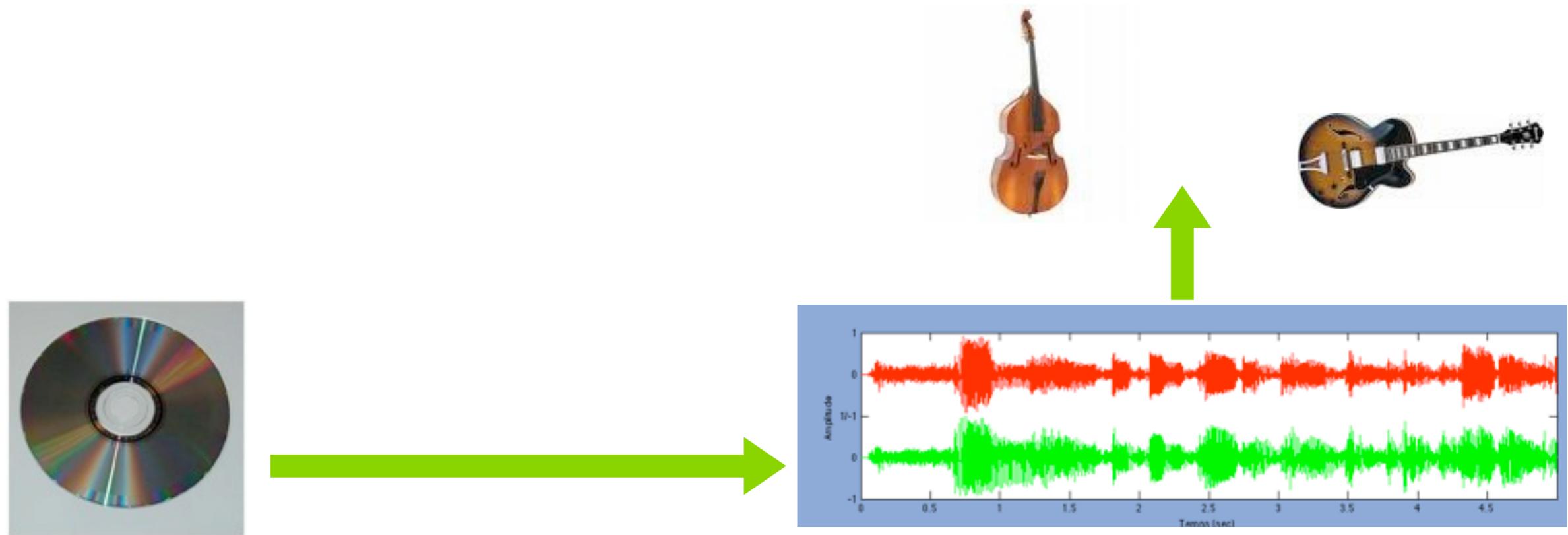
Example : audio source separation

- « Softly as in a morning sunrise »



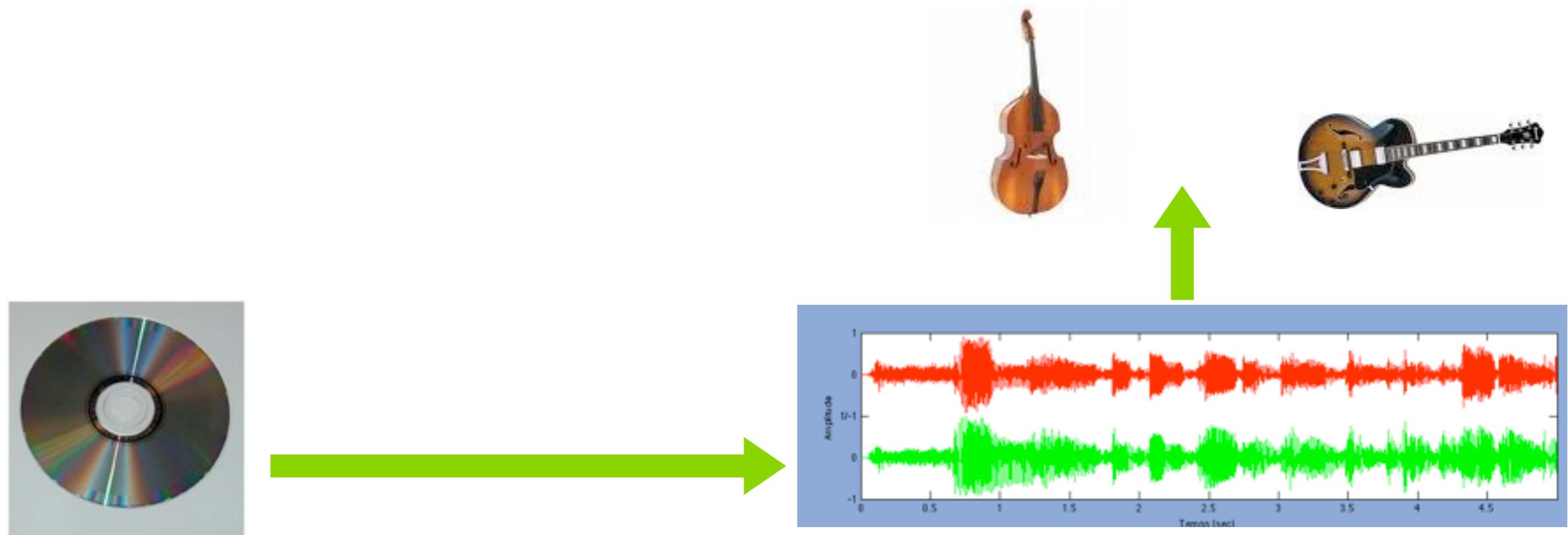
Example : audio source separation

- « Softly as in a morning sunrise »



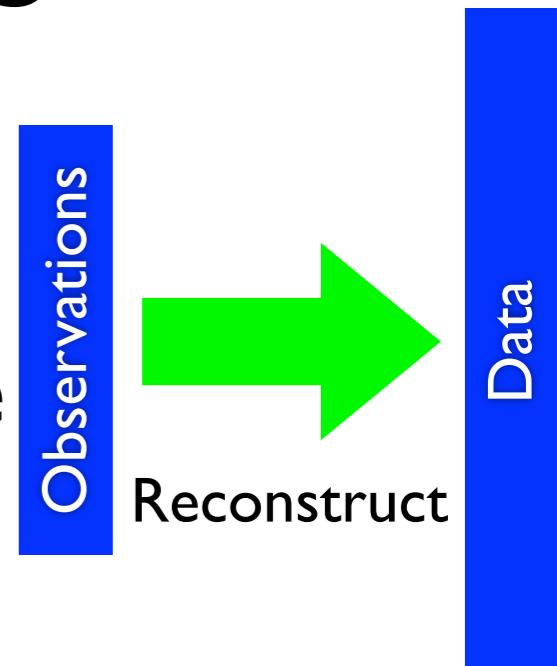
Example : audio source separation

- « Softly as in a morning sunrise »



Inverse problems

- **Inverse problem** : exploit indirect or incomplete observation to reconstruct some data



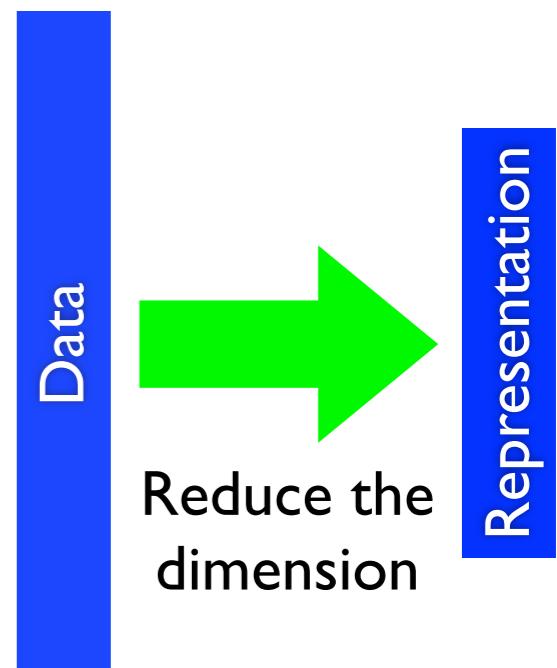
$$z = \mathbf{M}\mathbf{y}$$

↑
fewer equations than unknowns

- **Sparsity** : represent / approximate high-dimensional & complex data using few parameters

$$\mathbf{y} \approx \Phi\mathbf{x}$$

↑
few nonzero components

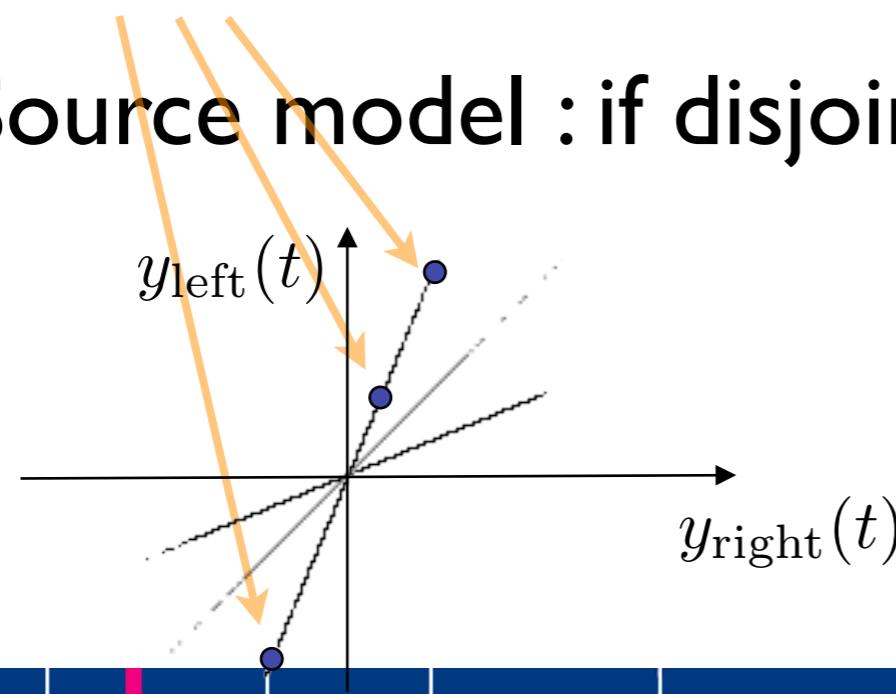


Blind Source Separation

- Mixing model : linear instantaneous mixture

$$\begin{matrix} y_{\text{right}}(t) \\ y_{\text{left}}(t) \end{matrix} = \mathbf{A} \begin{matrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{matrix}$$

- Source model : if disjoint time-supports ...



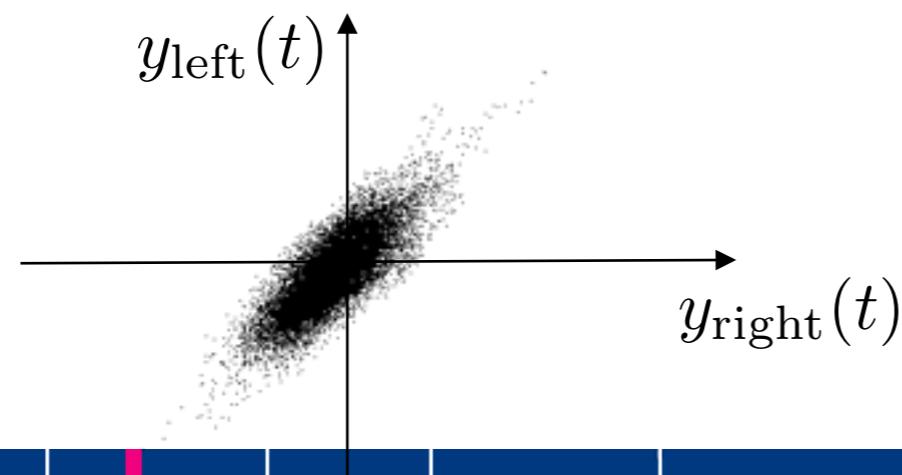
... then clustering to :
 1- identify (columns of) the mixing matrix
 2- recover sources

Blind Source Separation

- Mixing model : linear instantaneous mixture

$$\begin{matrix} y_{\text{right}}(t) \\ y_{\text{left}}(t) \end{matrix} \left(\begin{array}{c} \text{blue waveform} \\ \text{blue waveform} \end{array} \right) = \mathbf{A} \left(\begin{array}{c} \text{blue waveform} \\ \text{blue waveform} \\ \text{blue waveform} \end{array} \right) \begin{matrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{matrix}$$

- In practice ...

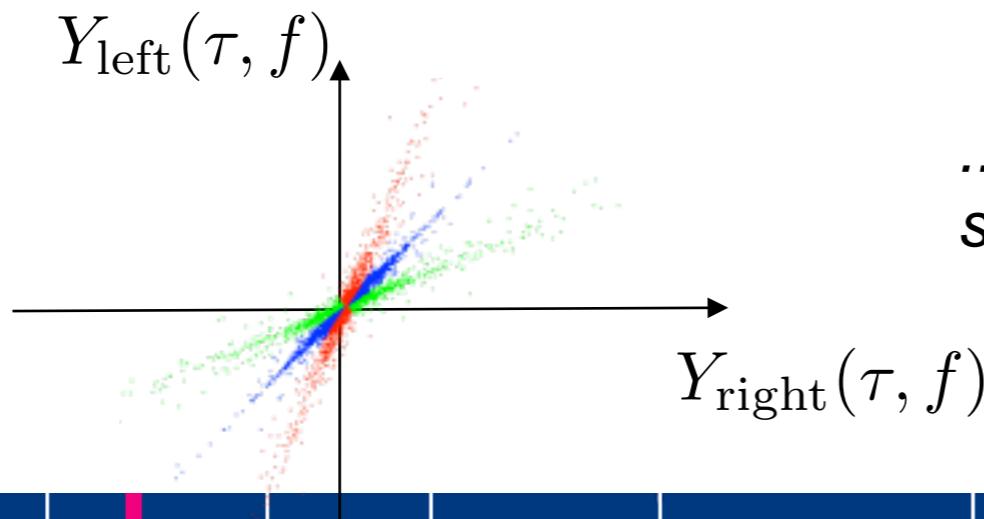


Time-Frequency Masking

- Mixing model in the time-frequency domain

$$\begin{matrix} Y_{\text{right}}(\tau, f) \\ Y_{\text{left}}(\tau, f) \end{matrix} \left(\begin{array}{c} \text{[Heatmap]} \\ \text{[Heatmap]} \end{array} \right) = \mathbf{A} \mathbf{S}(\tau, f)$$

- And “miraculously” ...



*... time-frequency representations of audio signals are (often) **almost disjoint**.*

Mathematical foundations

- Bottleneck 1990-2000 : fewer equations than unknowns

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

- Novelty 2001-2006 :

- ♦ Uniqueness of sparse solution:

- ♦ if x_0, x_1 are “sufficiently sparse”,

- ♦ then $\mathbf{A}x_0 = \mathbf{A}x_1 \Rightarrow x_0 = x_1$

- ♦ Recovery of x_0 with practical algorithms

- ♦ Thresholding, Matching Pursuits, Minimisation of L_p norms $p \leq 1, \dots$

Algorithmic principles for sparse approximation

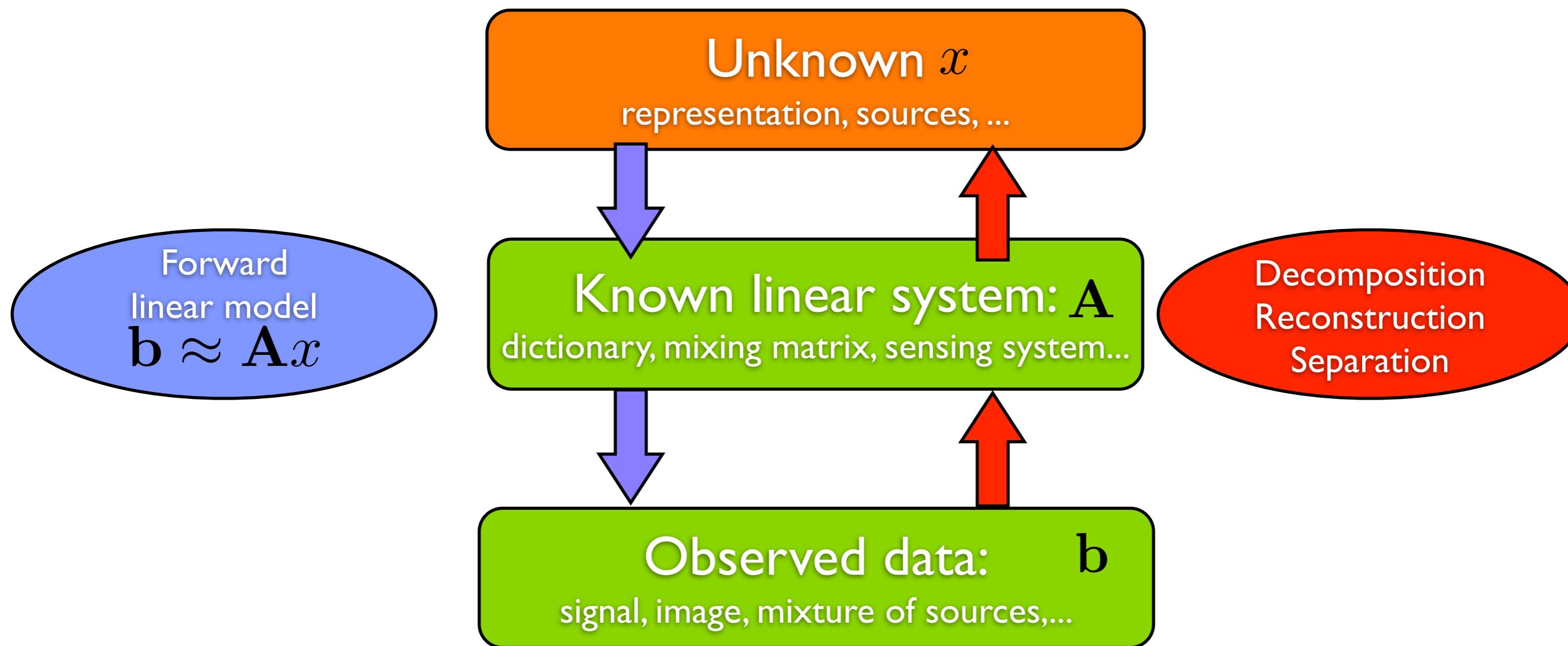


INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Vocabulary



Ideal sparse approximation

- Input:
 $m \times N$ matrix **A**, with $m < N$, m -dimensional vector **b**

- Possible objectives:
 find the sparsest approximation within tolerance

$$\arg \min_x \|x\|_0, \text{ s.t. } \|\mathbf{b} - \mathbf{A}x\| \leq \epsilon$$

find best approximation with given sparsity

$$\arg \min_x \|\mathbf{b} - \mathbf{A}x\|, \text{ s.t. } \|x\|_0 \leq k$$

find a solution x to

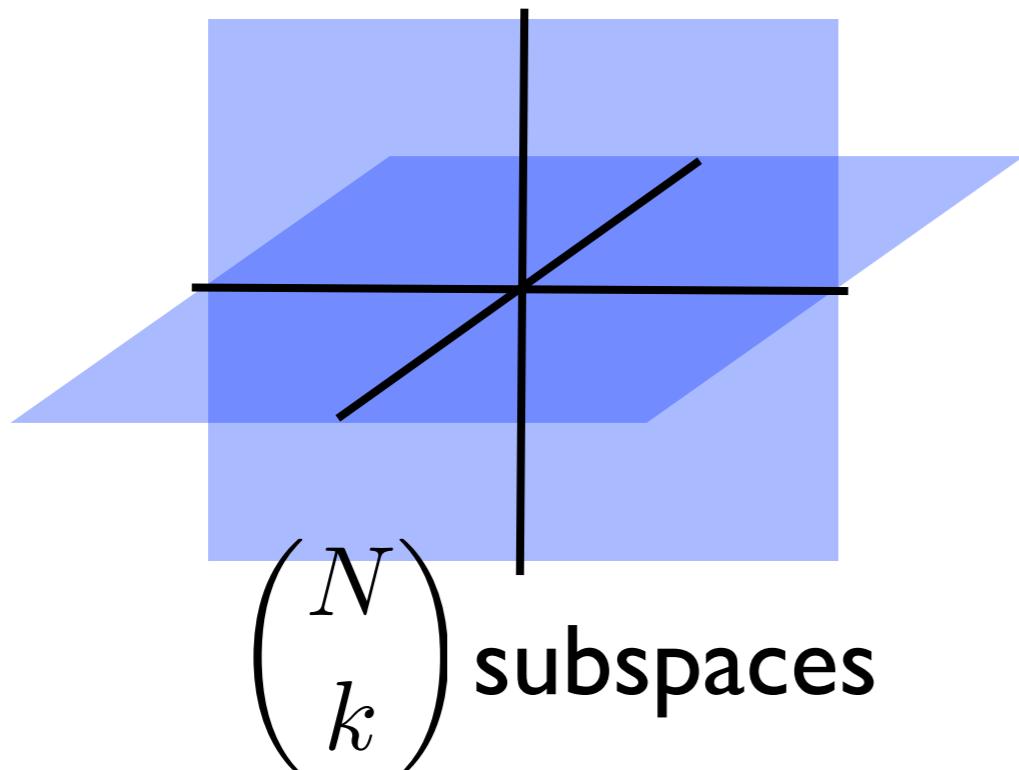
$$\|\mathbf{b} - \mathbf{A}x\| \leq \epsilon, \text{ and } \|x\|_0 \leq k$$



Geometric interpretation of sparse approximation

- Coefficient domain \mathbb{R}^N :
 - ◆ set Σ_k of sparse vectors

$$\|x\|_0 \leq k$$



- Set $\mathbf{A}\Sigma_k = \binom{N}{k}$ subspaces in signal domain
- Ideal sparse approximation
= find nearest subspace among $\binom{N}{k}$

Combinatorial search!
Actual complexity ?
NP-complete!

Practical approaches: Optimization *principles*



Overall compromise

- Approximation quality

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$$

- Ideal sparsity measure : ℓ^0 “norm”

$$\|x\|_0 := \#\{n, x_n \neq 0\} = \sum_n |x_n|^0$$

- “Relaxed” sparsity measures

$$0 < p < \infty, \|x\|_p := \left(\sum_n |x_n|^p \right)^{1/p}$$

L^p norms / quasi-norms

- **Norms** when $1 \leq p < \infty$ = convex

$$\|x\|_p = 0 \Leftrightarrow x = 0$$

$$\|\lambda x\|_p = |\lambda| \|x\|_p, \forall \lambda, x$$

Triangle inequality $\|x + y\|_p \leq \|x\|_p + \|y\|_p, \forall x, y$

- **Quasi-norms** when $0 < p < 1$ = nonconvex

$$\|x + y\|_p \leq 2^{1/p} (\|x\|_p + \|y\|_p), \forall x, y$$

Quasi-triangle inequality

$$\|x + y\|_p^p \leq \|x\|_p^p + \|y\|_p^p, \forall x, y$$

- “Pseudo”-norm for $p=0$

$$\|x + y\|_0 \leq \|x\|_0 + \|y\|_0, \forall x, y$$

Optimization problems

- Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2 \text{ s.t. } \|x\|_p \leq \tau$$

- Sparsification

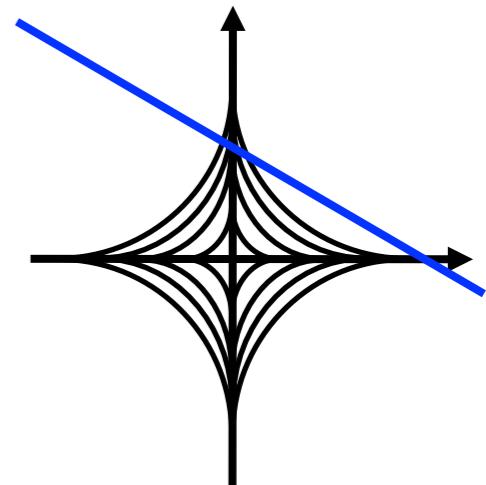
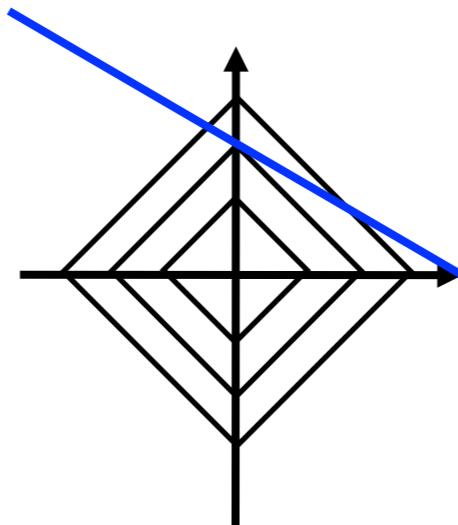
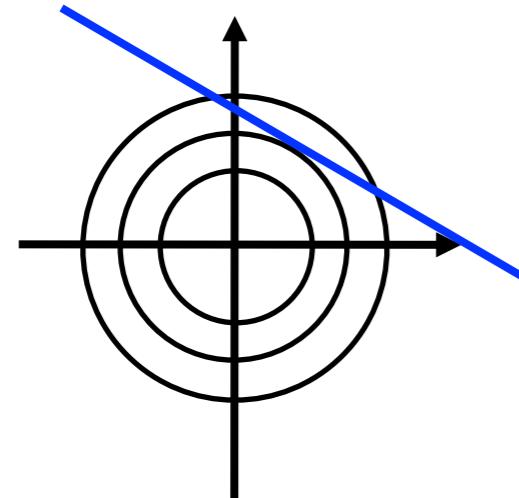
$$\min_x \|x\|_p \text{ s.t. } \|\mathbf{b} - \mathbf{A}x\|_2 \leq \epsilon$$

- Regularization

$$\min_x \frac{1}{2} \|\mathbf{b} - \mathbf{A}x\|_2 + \lambda \|x\|_p$$

L^p “norms” level sets

- Strictly convex when $p > 1$
- Convex $p=1$
- Nonconvex $p < 1$



Observation: the minimizer is sparse

— $\{x \text{ s.t. } b = Ax\}$

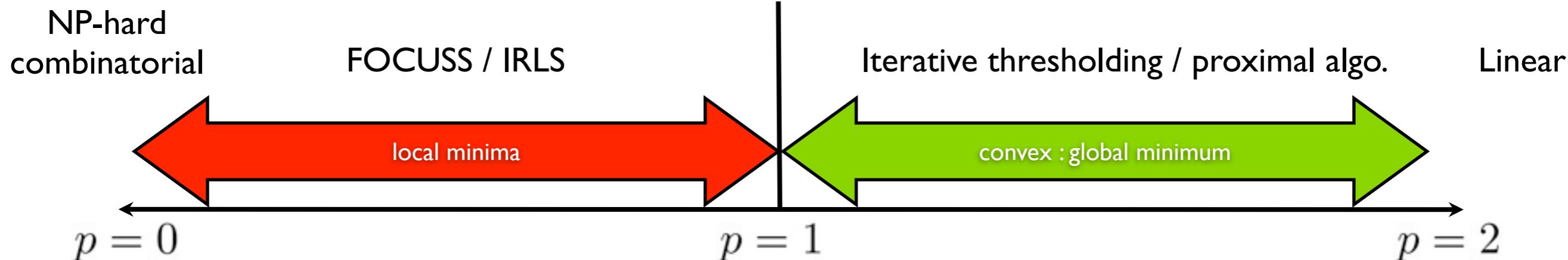
Global Optimization : from Principles to Algorithms

- Optimization principle

$$\min_x \frac{1}{2} \|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda \|x\|_p^p$$

- ♦ Sparse representation
- ♦ Sparse approximation

$$\begin{aligned} \lambda \rightarrow 0 & \quad \mathbf{A}x = \mathbf{b} \\ \lambda > 0 & \quad \mathbf{A}x \approx \mathbf{b} \end{aligned}$$



Lasso [Tibshirani 1996], Basis Pursuit (Denoising) [Chen, Donoho & Saunders, 1999]
 Linear/Quadratic programming (interior point, etc.)
 Homotopy method [Osborne 2000] / Least Angle Regression [Efron & al 2002]
 Iterative / proximal algorithms [Daubechies, de Frise, de Mol 2004, Combettes & Pesquet 2008, ...]

Greedy Algorithms



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Greedy algorithms

- Observation: when \mathbf{A} is orthonormal,
 - ◆ the problem

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 \text{ s.t. } \|x\|_0 \leq k$$

- ◆ is equivalent to

$$\min_x \sum_n (\mathbf{a}_n^T \mathbf{b} - x_n)^2 \text{ s.t. } \|x\|_0 \leq k$$

- Let Λ_k index the k largest inner products

$$\min_{n \in \Lambda_k} |\mathbf{a}_n^T \mathbf{b}| \geq \max_{n \notin \Lambda_k} |\mathbf{a}_n^T \mathbf{b}|$$

- ◆ an optimum solution is

$$x_n = \mathbf{a}_n^T \mathbf{b}, n \in \Lambda_k; \quad x_n = 0, n \notin \Lambda_k$$

Greedy algorithms

- Iterative algorithm (= *Matching Pursuit*)
 - ◆ Initialize a residual to $\mathbf{r}_0 = \mathbf{b}$ $i = 1$
 - ◆ Compute all inner products

$$\mathbf{A}^T \mathbf{r}_{i-1} = (\mathbf{a}_n^T \mathbf{r}_{i-1})_{n=1}^N$$

- ◆ Select the largest in magnitude

$$n_i = \arg \max_n |\mathbf{a}_n^T \mathbf{r}_{i-1}|$$

- ◆ Compute an updated residual

$$\mathbf{r}_i = \mathbf{r}_{i-1} - (\mathbf{a}_{n_i}^T \mathbf{r}_{i-1}) \mathbf{a}_{n_i}$$

- ◆ If $i \geq k$ then stop, otherwise increment i and iterate

Matching Pursuit (MP)

- Matching Pursuit (*aka* Projection Pursuit, CLEAN)

- ◆ Initialization $\mathbf{r}_0 = \mathbf{b}$ $i = 1$
- ◆ Atom selection:

$$n_i = \arg \max_n |\mathbf{a}_n^T \mathbf{r}_{i-1}|$$

- ◆ Residual update

$$\mathbf{r}_i = \mathbf{r}_{i-1} - (\mathbf{a}_{n_i}^T \mathbf{r}_{i-1}) \mathbf{a}_{n_i}$$

- Energy preservation (Pythagoras theorem)

$$\|\mathbf{r}_{i-1}\|_2^2 = |\mathbf{a}_{n_i}^T \mathbf{r}_{i-1}|^2 + \|\mathbf{r}_i\|_2^2$$

Summary

Global optimization

Iterative greedy algorithms

Principle	$\min_x \frac{1}{2} \ \mathbf{A}x - \mathbf{b}\ _2^2 + \lambda \ x\ _p^p$	iterative decomposition $\mathbf{r}_i = \mathbf{b} - \mathbf{A}x_i$ <ul style="list-style-type: none"> • select new components • update residual
Tuning quality/sparsity	regularization parameter λ	stopping criterion (nb of iterations, error level, ...) $\ x_i\ _0 \geq k \quad \ \mathbf{r}_i\ \leq \epsilon$
Variants	<ul style="list-style-type: none"> • choice of sparsity measure p • optimization algorithm • initialization 	<ul style="list-style-type: none"> • selection criterion (weak, stagewise ...) • update strategy (orthogonal ...)



Provably good algorithms

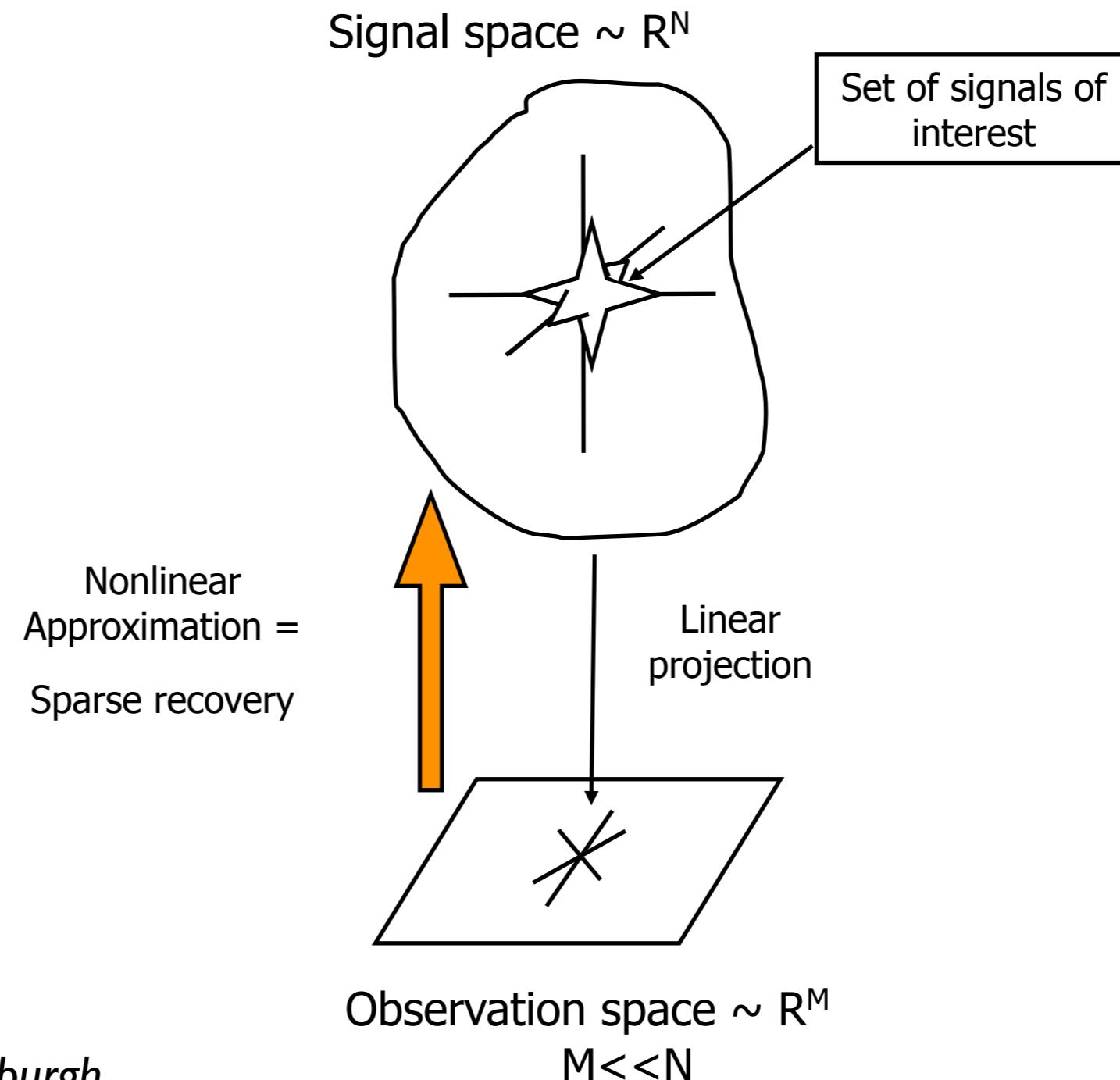


INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

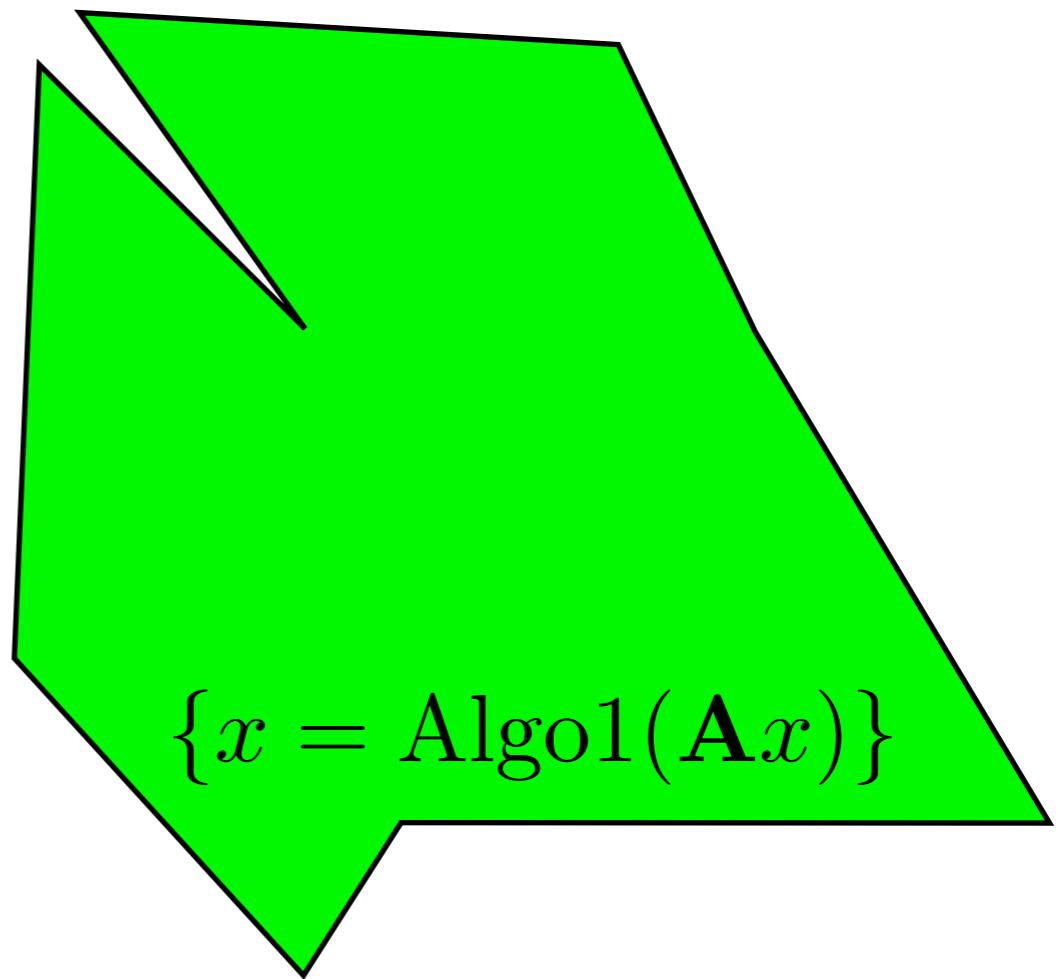
Inverse problems



Courtesy: M. Davies, U. Edinburgh

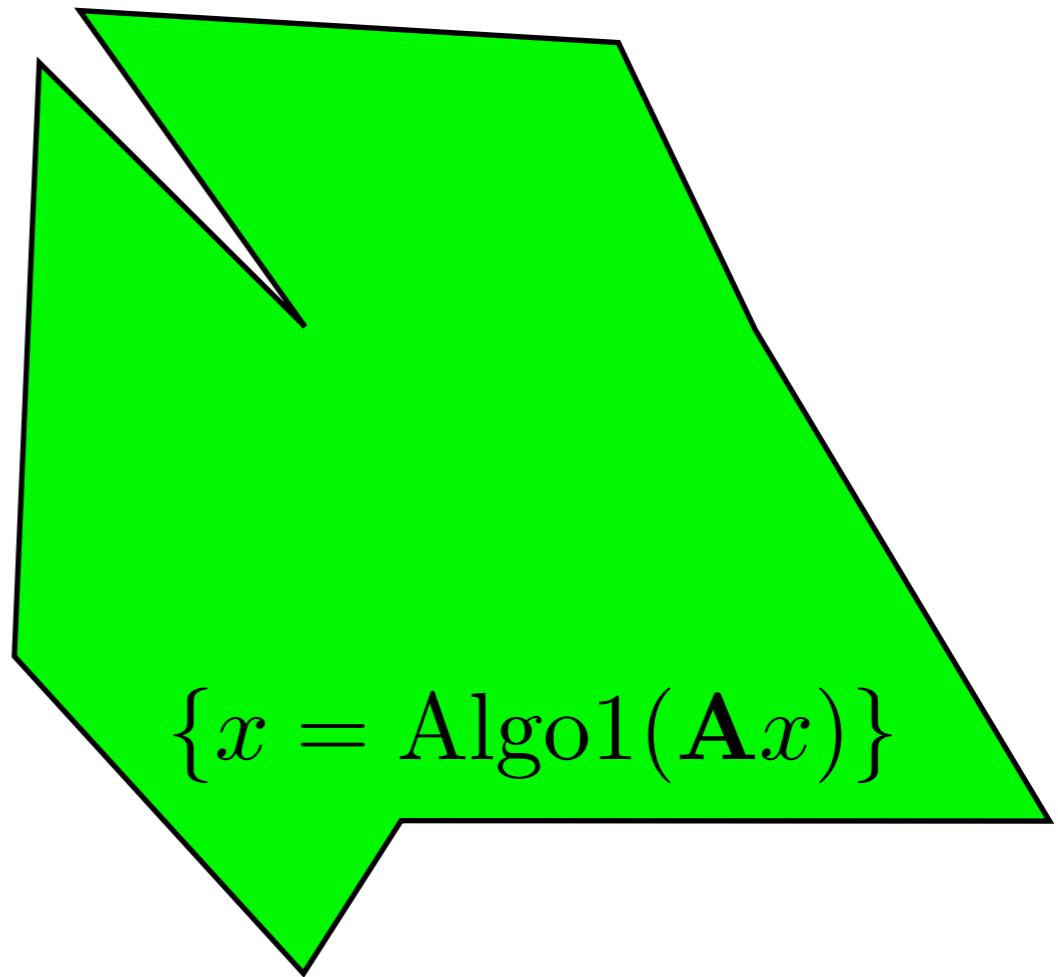
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm



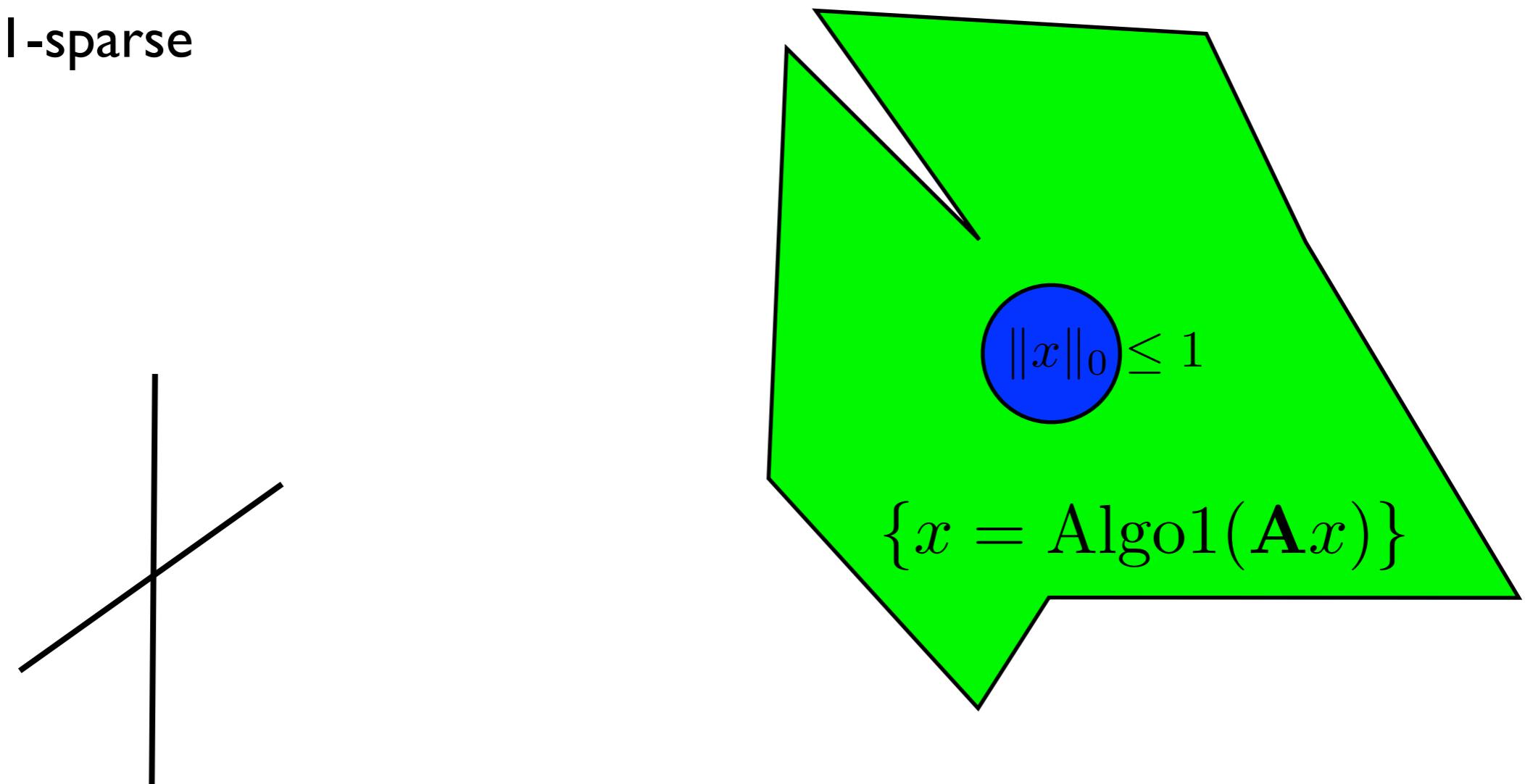
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm



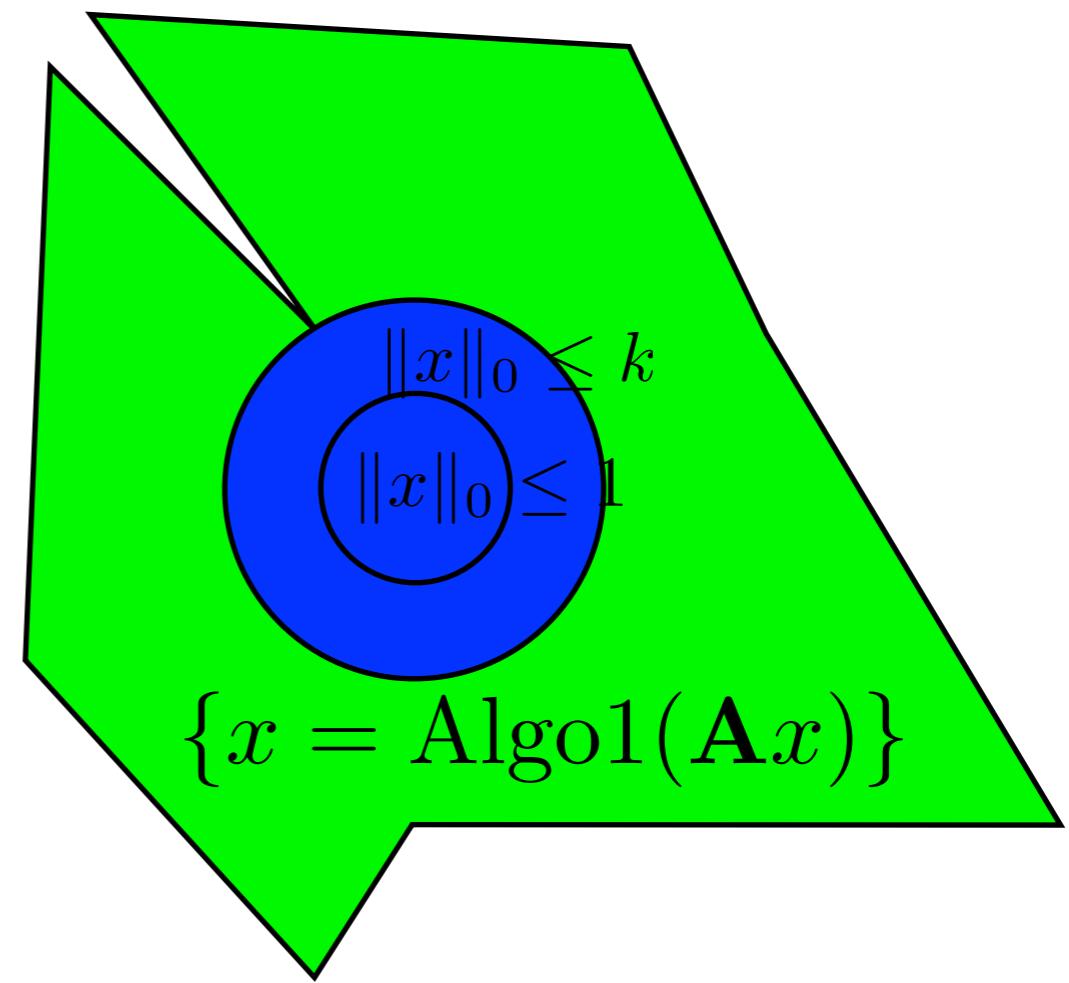
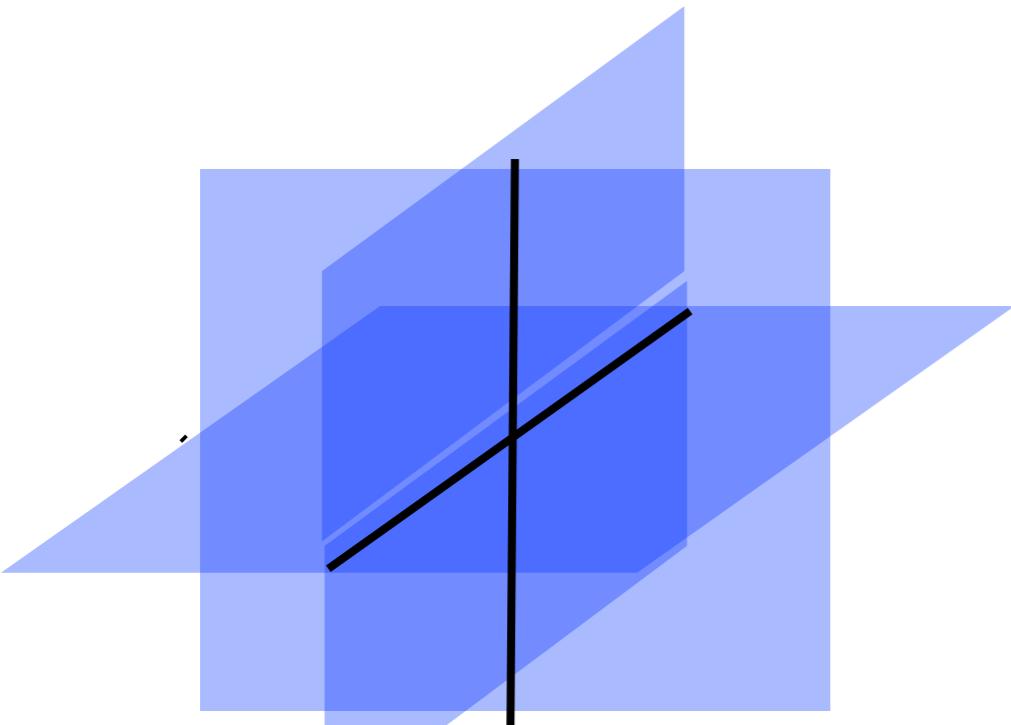
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ♦ ℓ -sparse



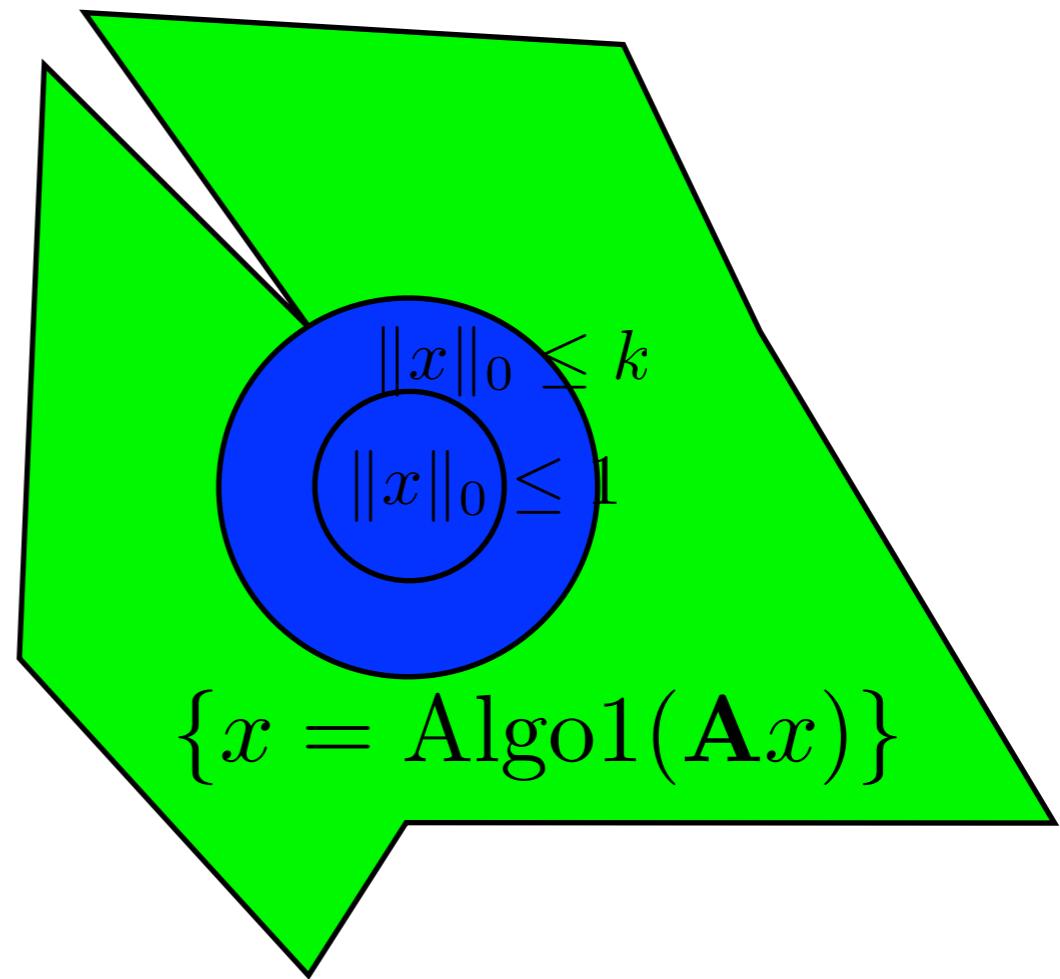
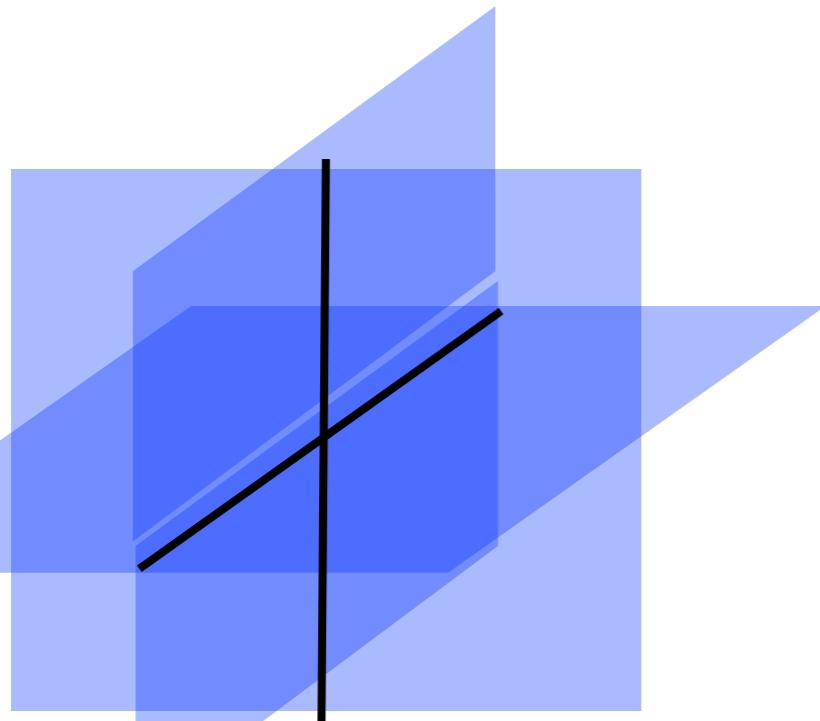
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ◆ 1-sparse
 - ◆ 2-sparse



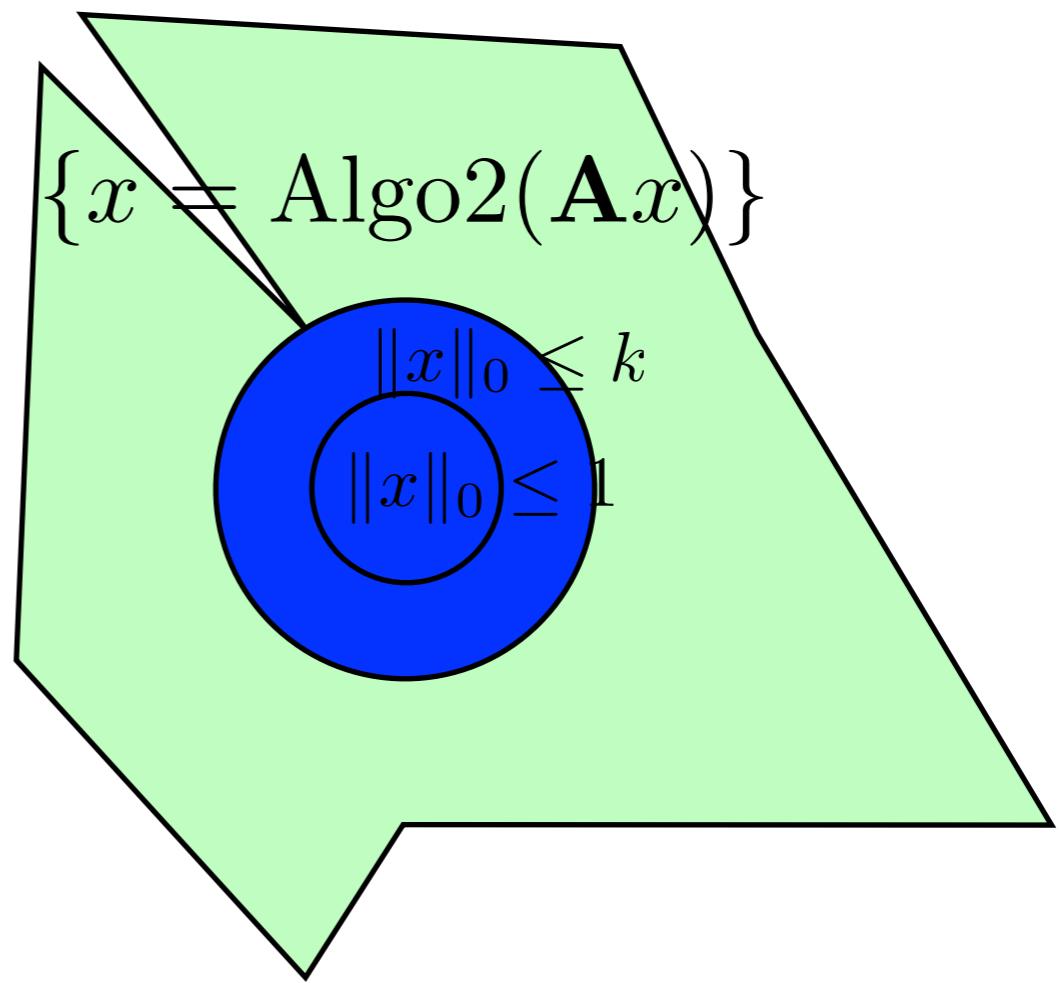
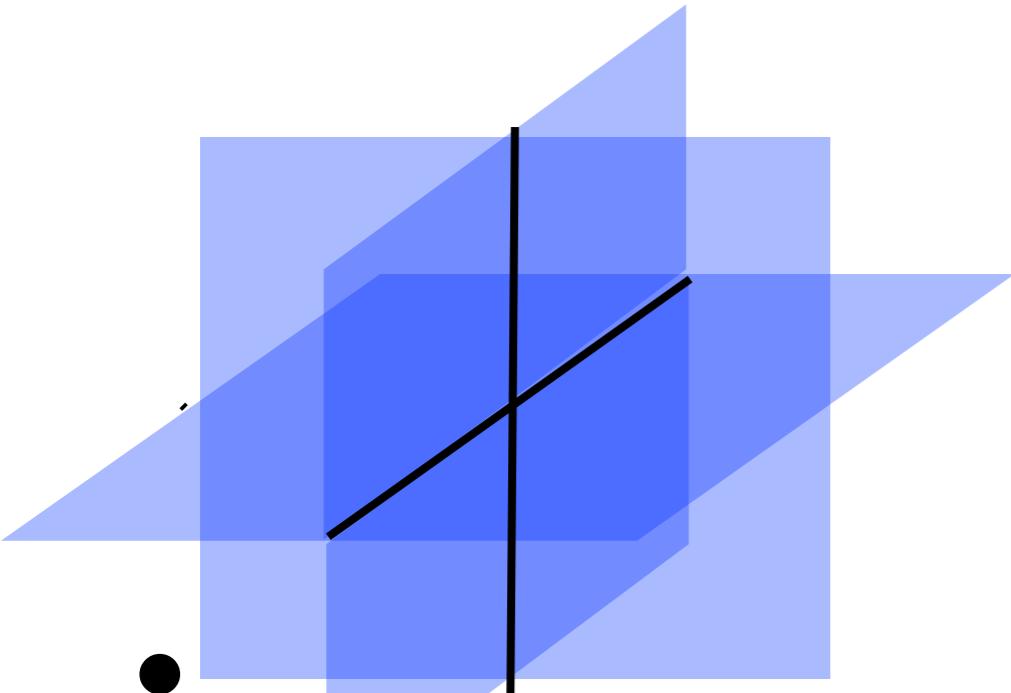
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ◆ 1-sparse
 - ◆ 2-sparse
 - ◆ 3-sparse ...



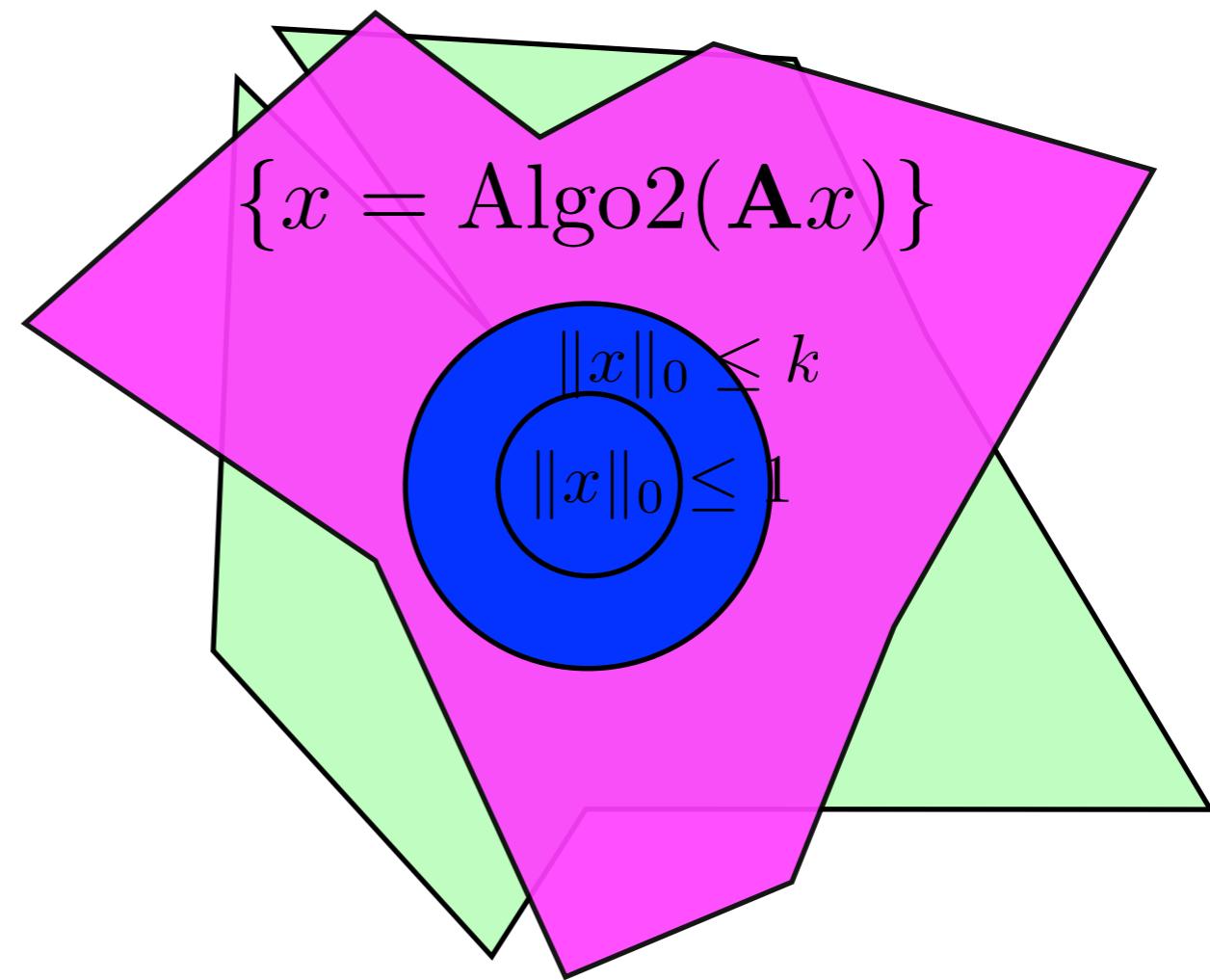
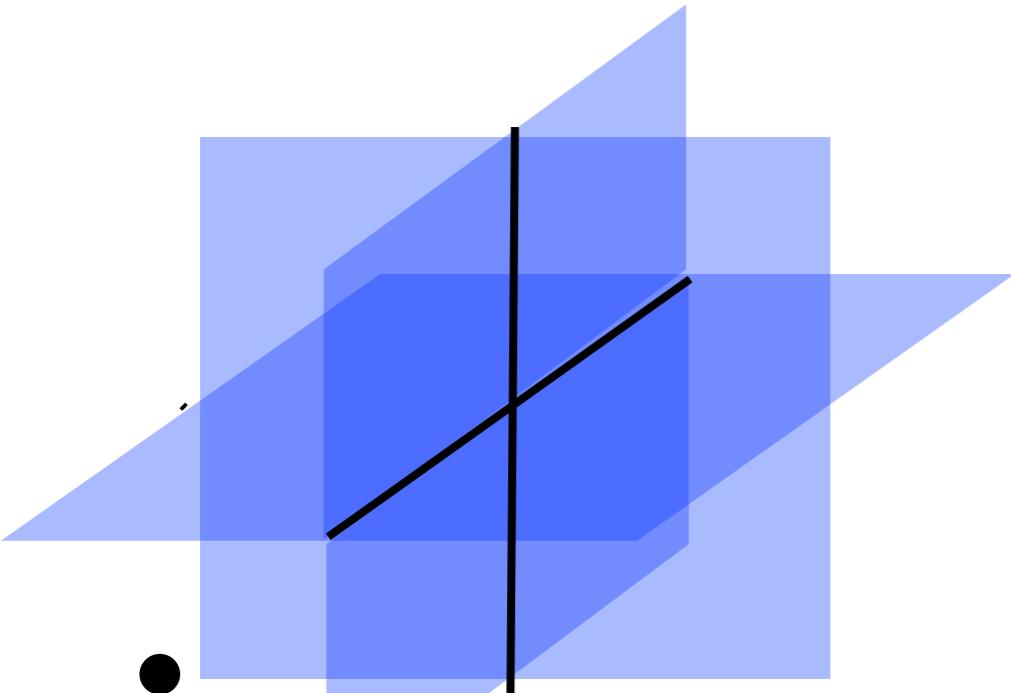
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ◆ 1-sparse
 - ◆ 2-sparse
 - ◆ 3-sparse ...



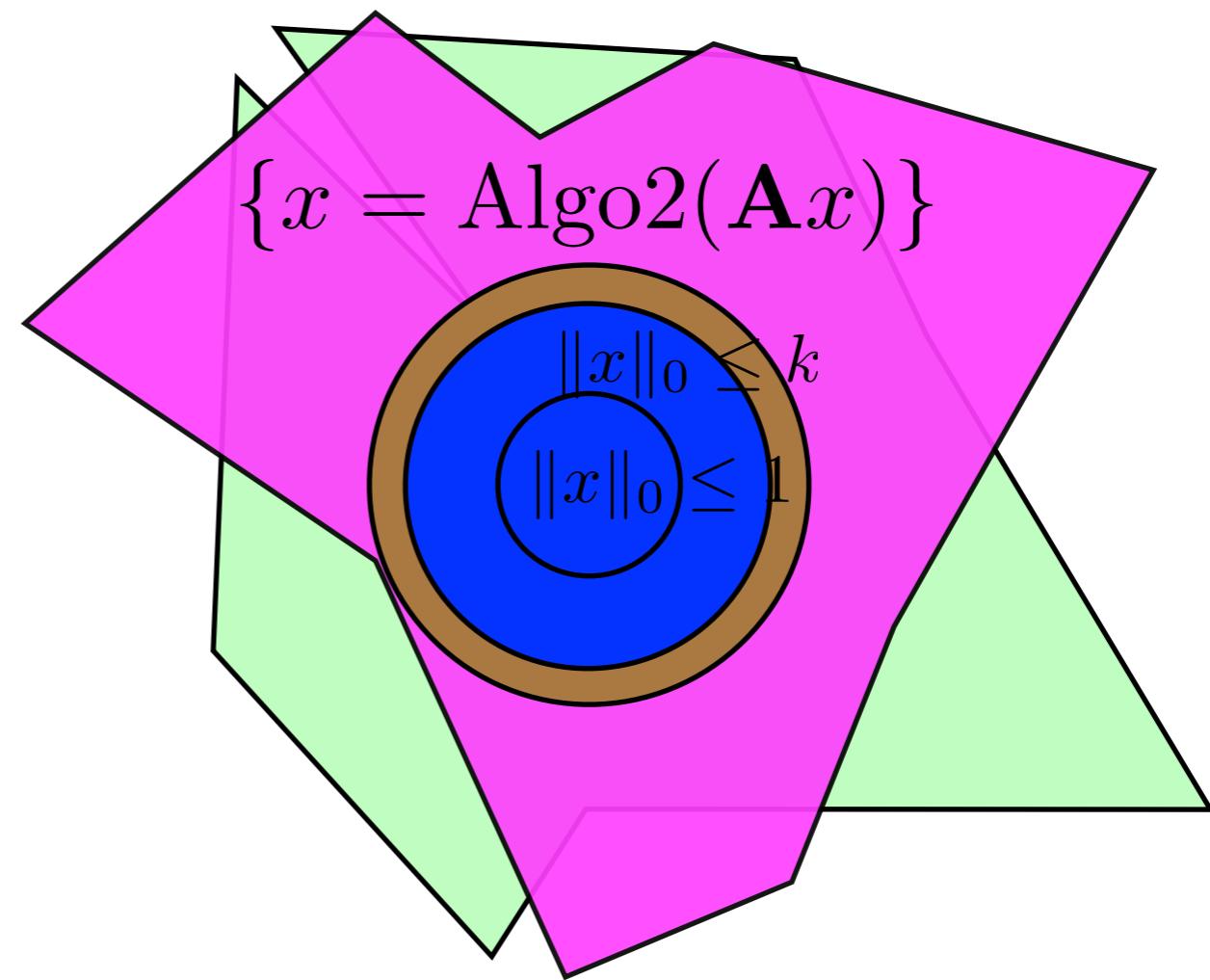
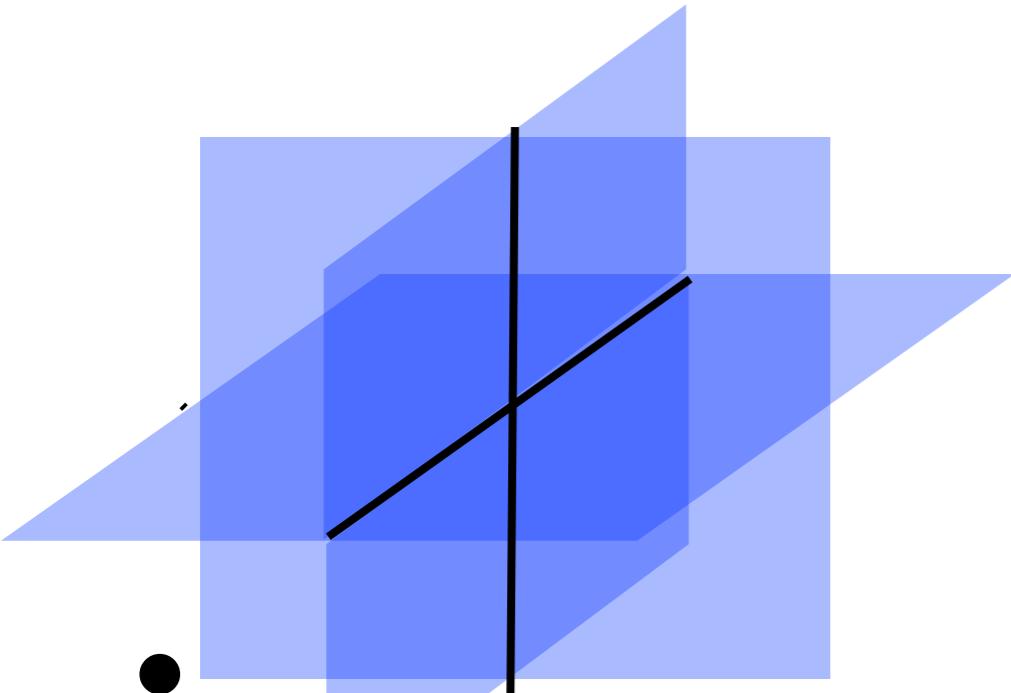
Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ◆ 1-sparse
 - ◆ 2-sparse
 - ◆ 3-sparse ...



Recovery analysis for inverse problem $b = Ax$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
 - ◆ 1-sparse
 - ◆ 2-sparse
 - ◆ 3-sparse ...



Equivalence between L0, L1, OMP

- **Theorem** : assume that $\mathbf{b} = \mathbf{A}x_0$

- ♦ if $\|x_0\|_0 \leq k_0(\mathbf{A})$ then $x_0 = x_0^*$
- ♦ if $\|x_0\|_0 \leq k_1(\mathbf{A})$ then $x_0 = x_1^*$

where $x_p^* = \arg \min_{\mathbf{A}x=\mathbf{A}x_0} \|x\|_p$

- *Donoho & Huo 01 : pair of bases, coherence*
- *Donoho & Elad, Gribonval & Nielsen 2003 : dictionary, coherence*
- *Tropp 2004 : Orthonormal Matching Pursuit, cumulative coherence*
- *Candes, Romberg, Tao 2004 : random dictionaries, restricted isometry constants*

Compressed sensing



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

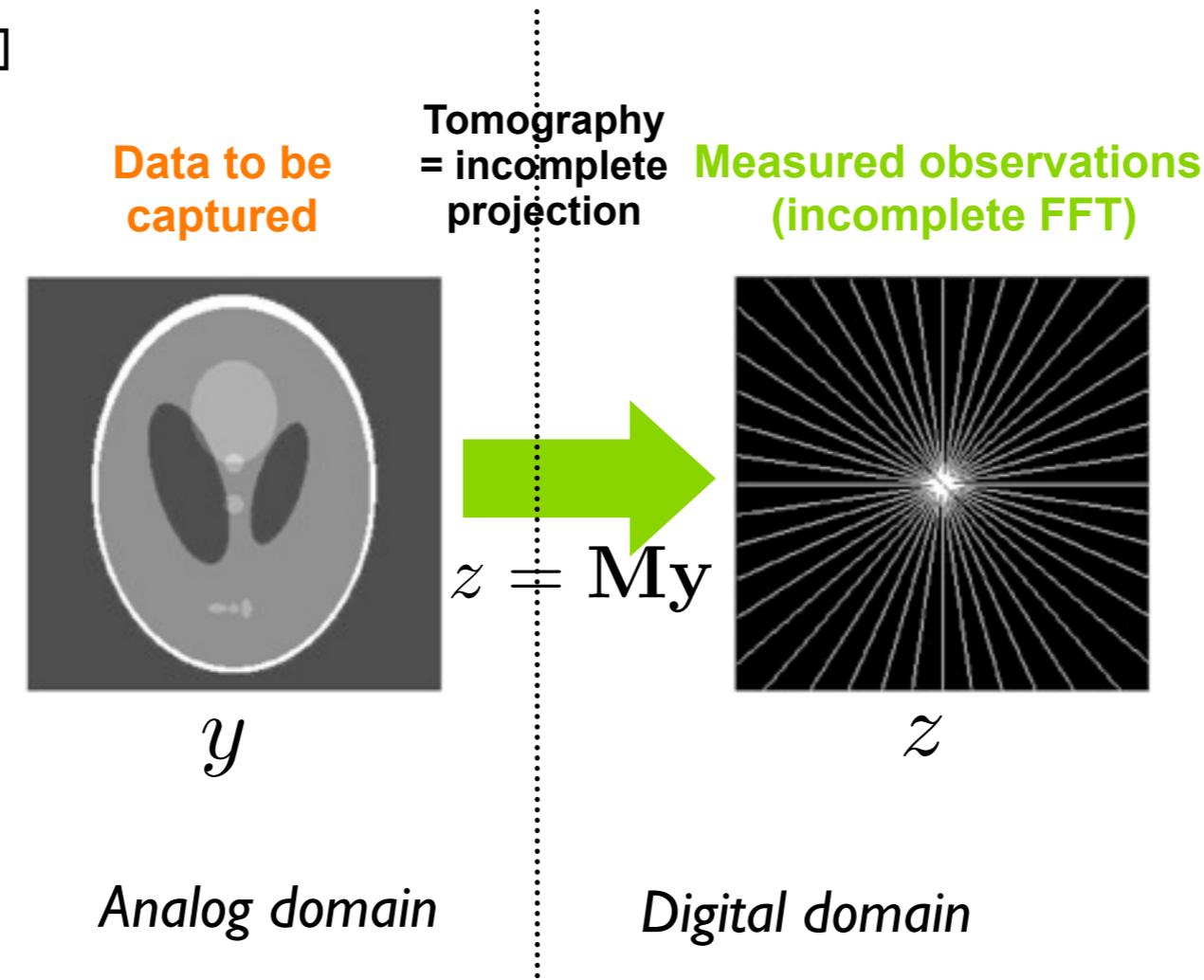


centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Example: tomography

- MRI from incomplete data

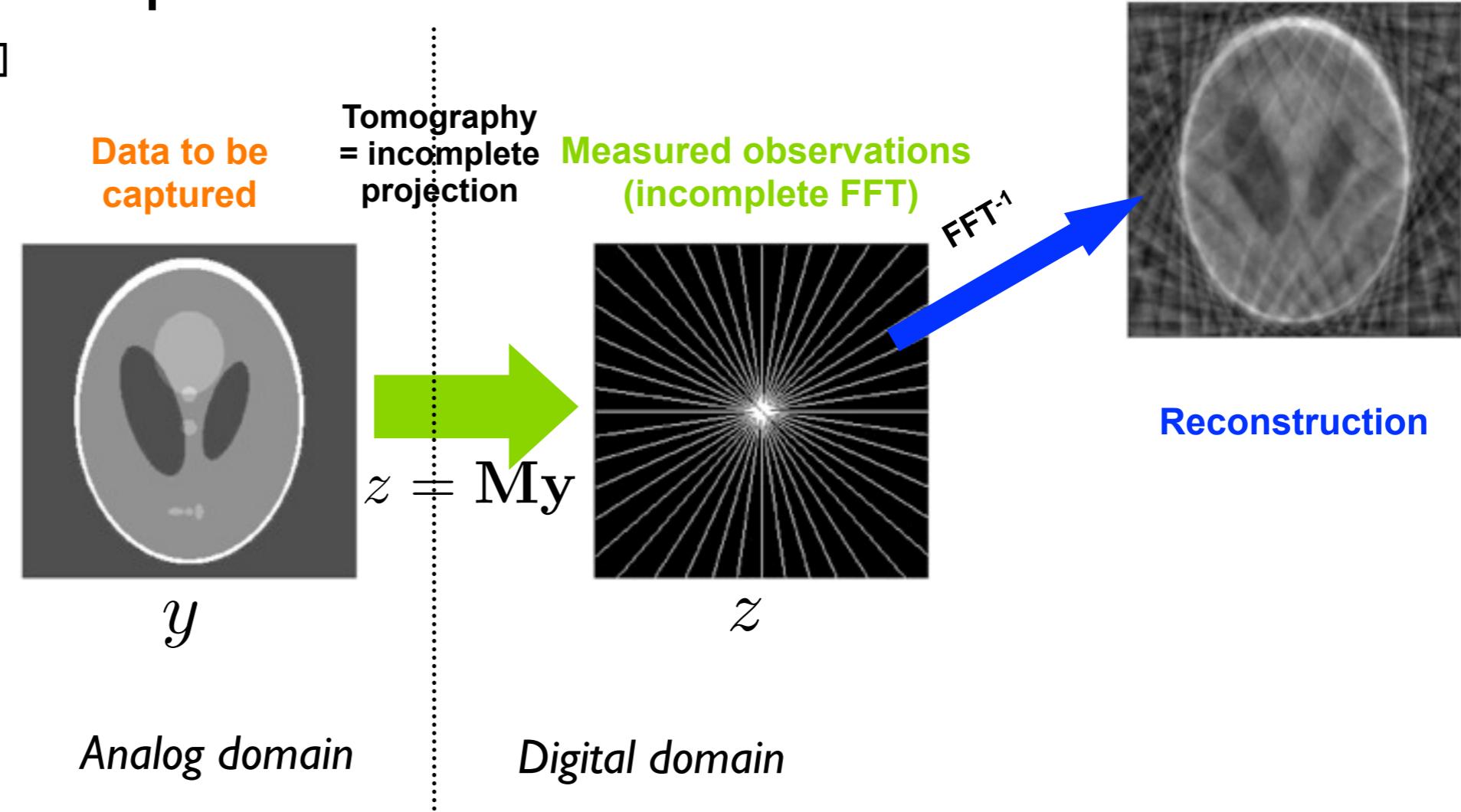
[Candès, Romberg & Tao]



Example: tomography

- MRI from incomplete data

[Candès, Romberg & Tao]



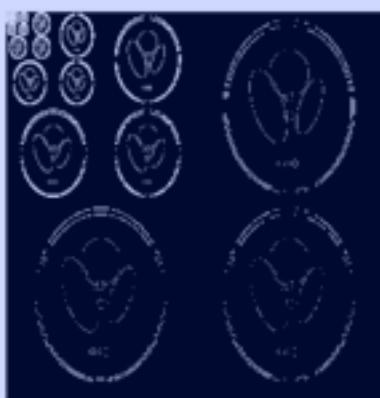
Example: tomography

- MRI from incomplete data

[Candès, Romberg & Tao]

Model / knowledge

The (unknown) wavelet transform is sparse



x

$$y = \Phi x$$

Data to be captured



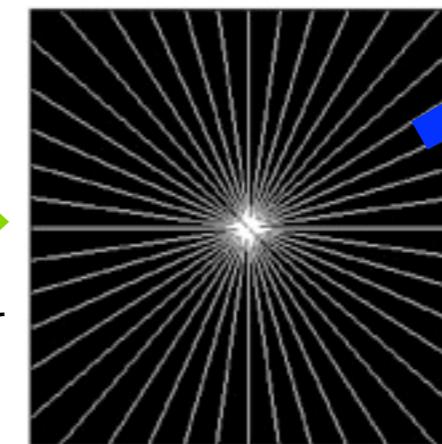
y

Tomography
= incomplete projection

$$z = M y$$

Analog domain

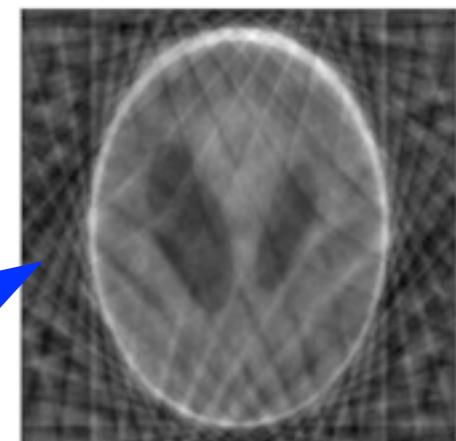
Measured observations
(incomplete FFT)



z

Digital domain

$$\text{FFT}^{-1}$$



Reconstruction

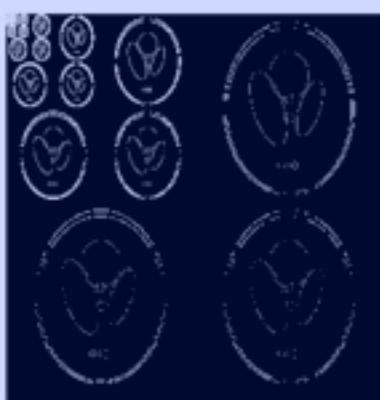
Example: tomography

- MRI from incomplete data

[Candès, Romberg & Tao]

Model / knowledge

The (unknown) wavelet transform is sparse



x

$$y = \Phi x$$

Data to be captured



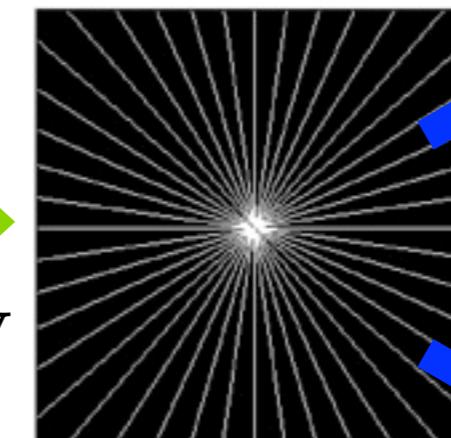
y

Analog domain

Tomography
= incomplete projection

$$z = M y$$

Measured observations (incomplete FFT)

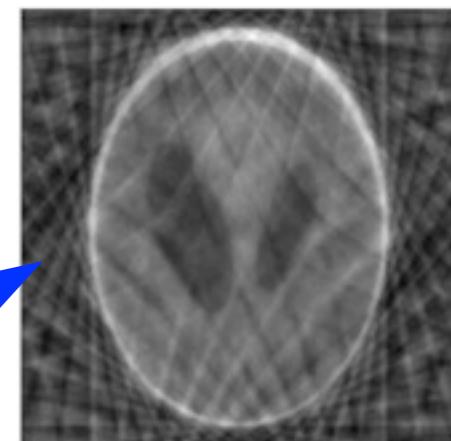


z

Digital domain

$$\text{FFT}^{-1}$$

Sparse reconstruction
(Candes et al 2004)



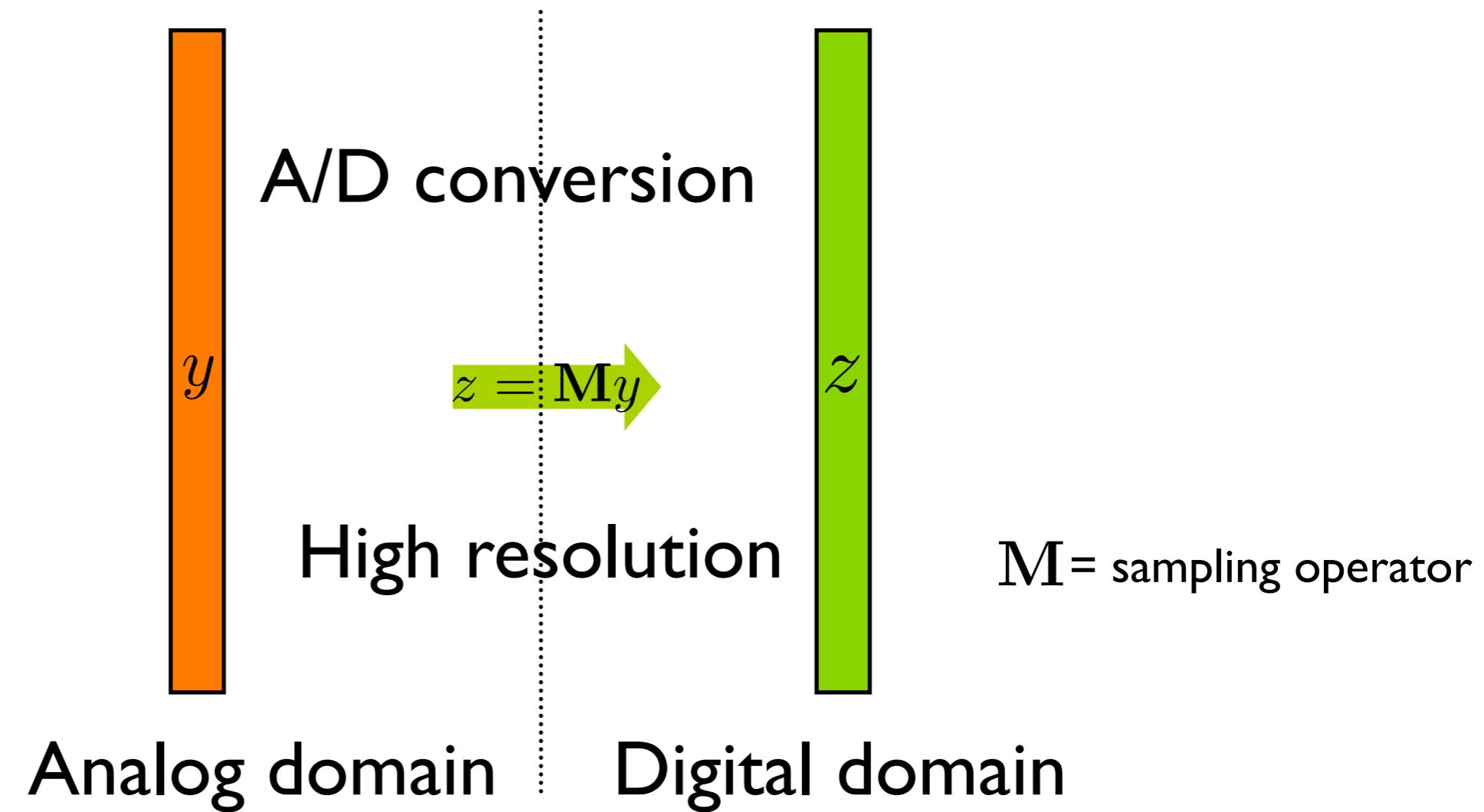
Reconstruction



$$\min \|x\|_1 \text{ s.t. } z = M \Phi x$$

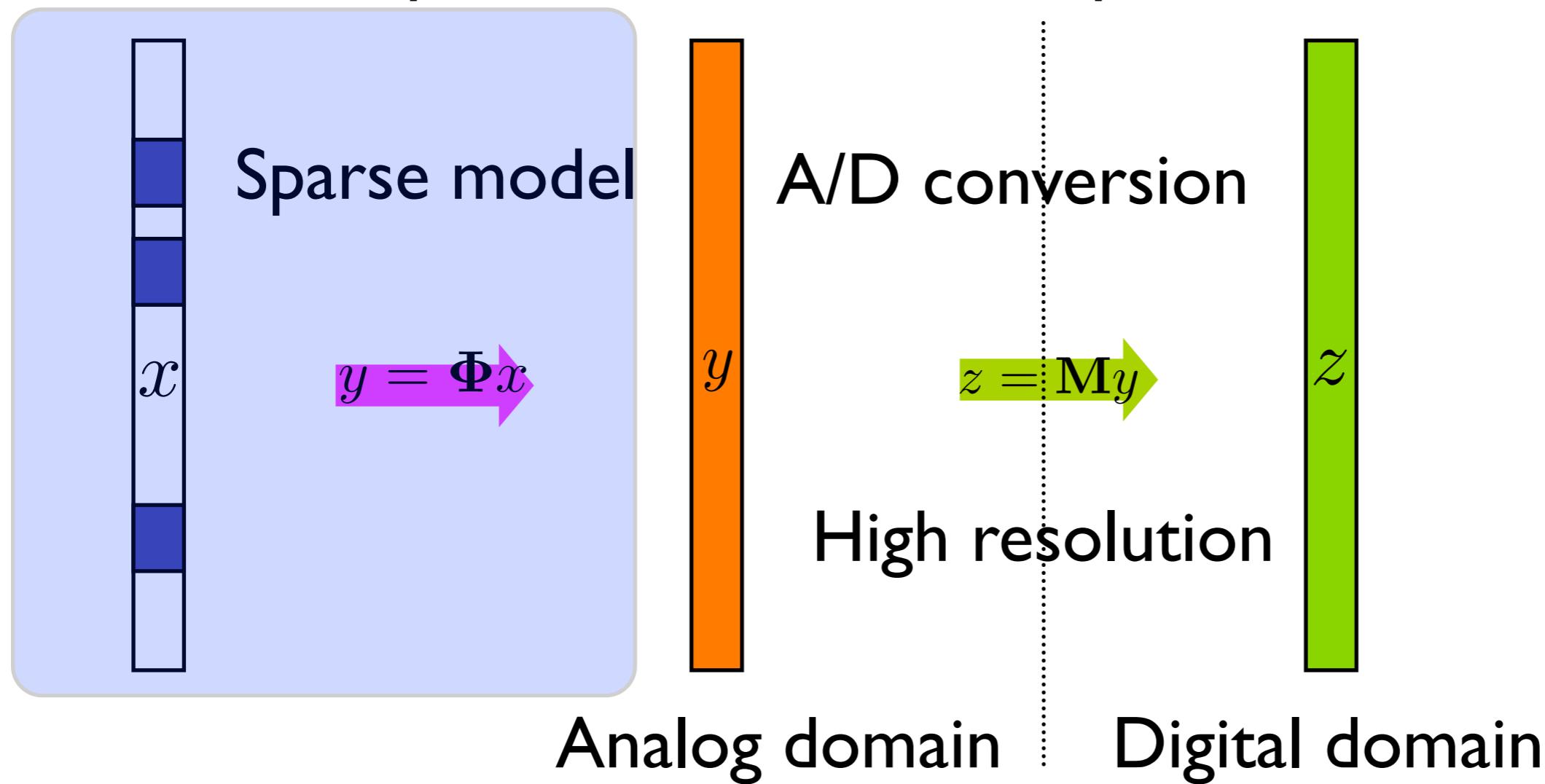
Classical Shannon Sampling

- « Sample first, think and compress afterwards »



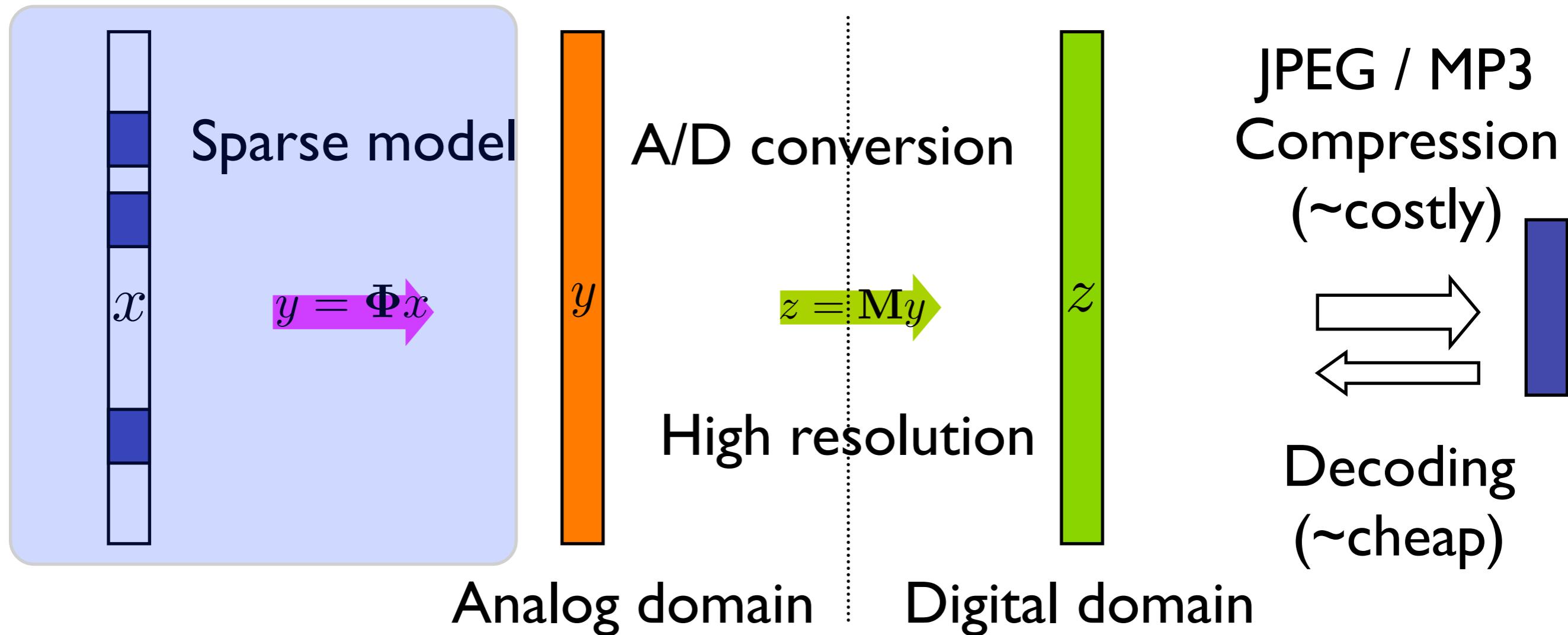
Classical Shannon Sampling

- « Sample first, think and compress afterwards »



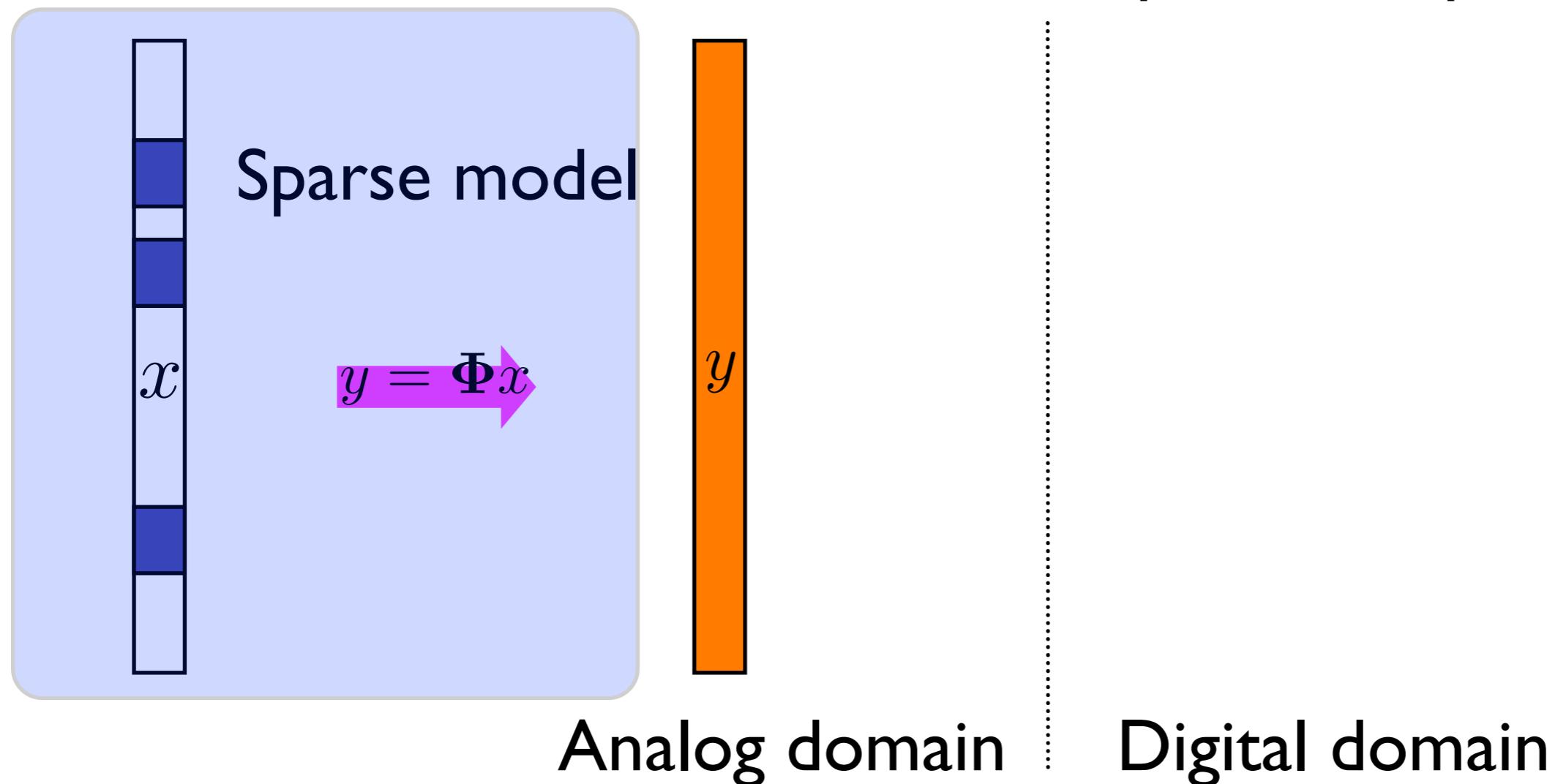
Classical Shannon Sampling

- « Sample first, think and compress afterwards »



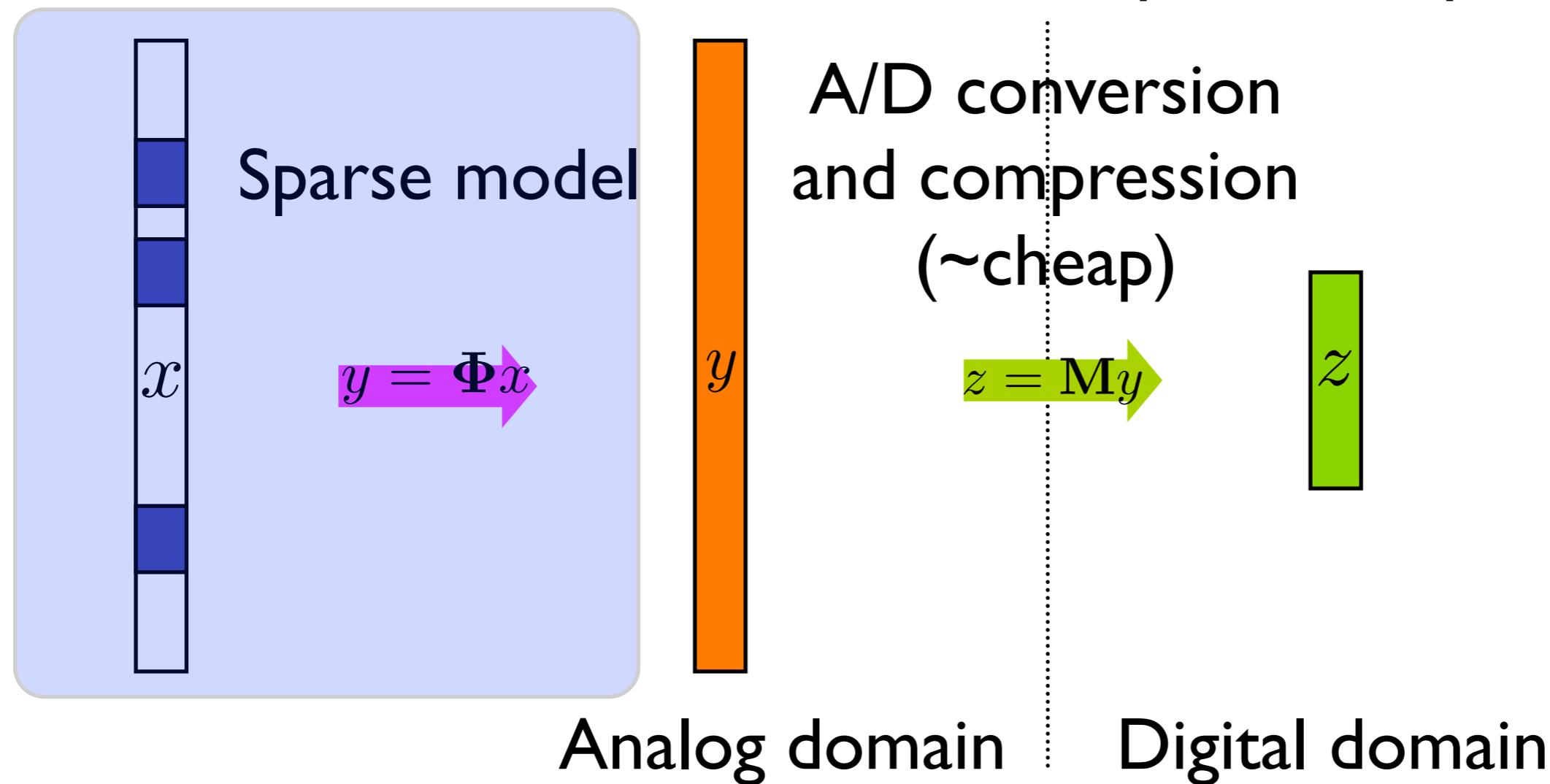
Compressed Sensing

- First model the data, then sample & compress



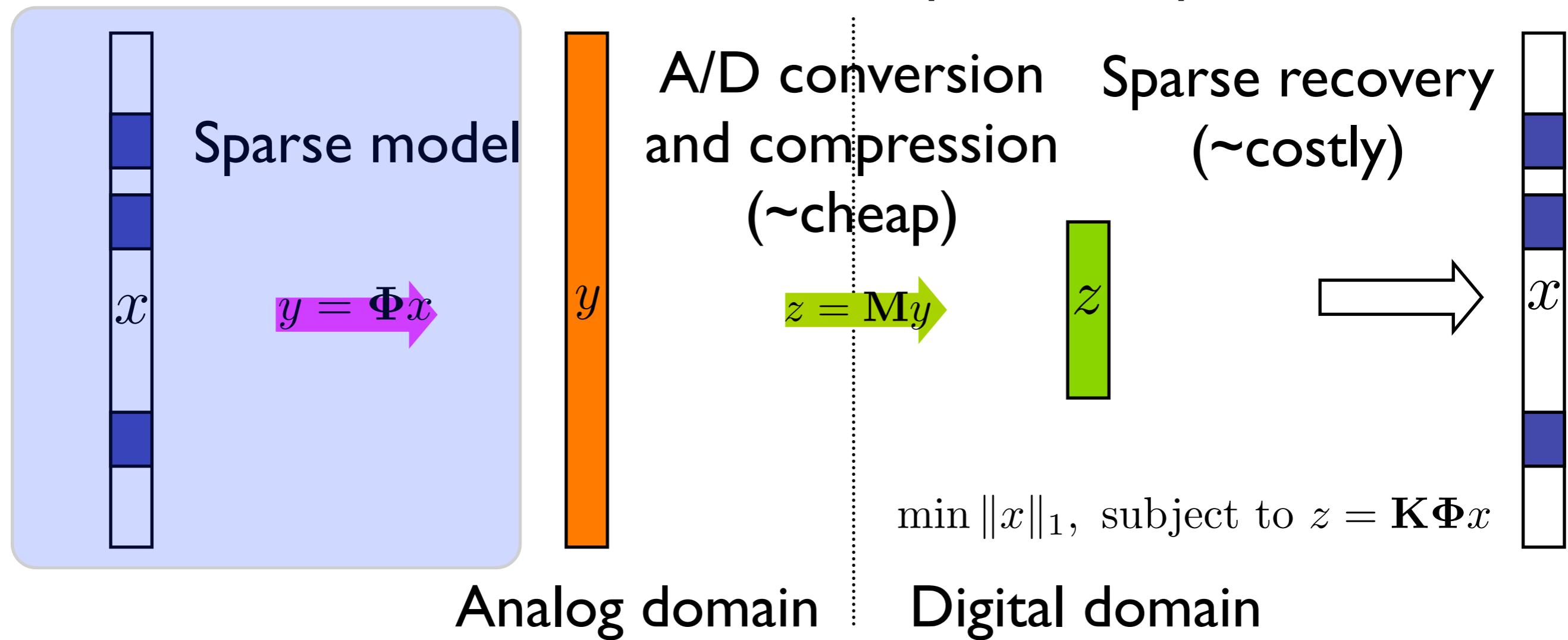
Compressed Sensing

- First model the data, then sample & compress



Compressed Sensing

- First model the data, then sample & compress



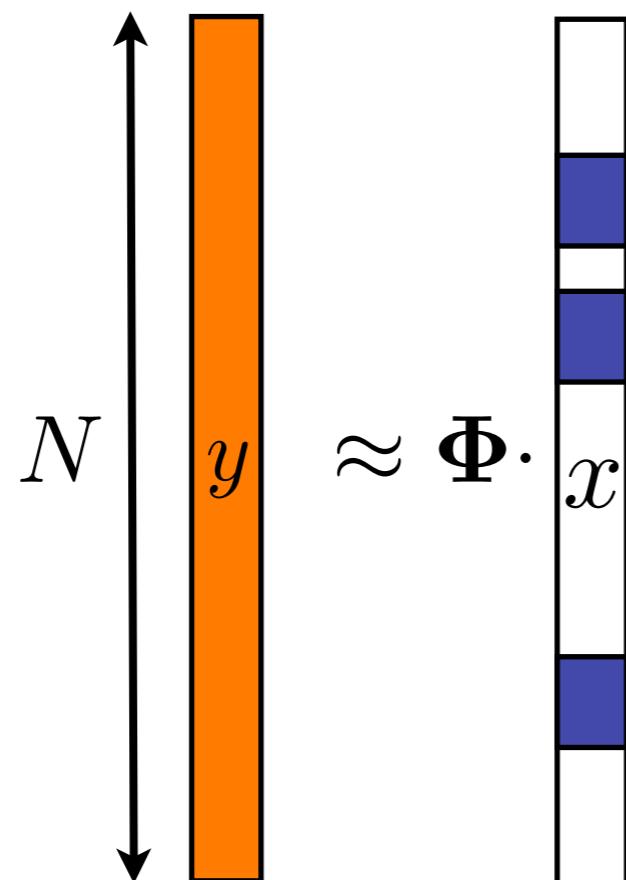
Conditions of success of Compressed Sensing

- Knowledge: transform domain where data is sparse
- «Incoherence» between measurement domain and sparse domain (uncertainty principle à la Heisenberg), for example:
 - ◆ time domain / frequency domain
 - ◆ spatial domain / frequency domain
 - ◆ **random measurements!**
- Sufficiently many measures
 - ◆ necessary
 - ◆ sufficient (with random Gaussian measures)

$$m \geq Ck \log_2 \frac{N}{k}$$

Why the log factor ?

- Full vector

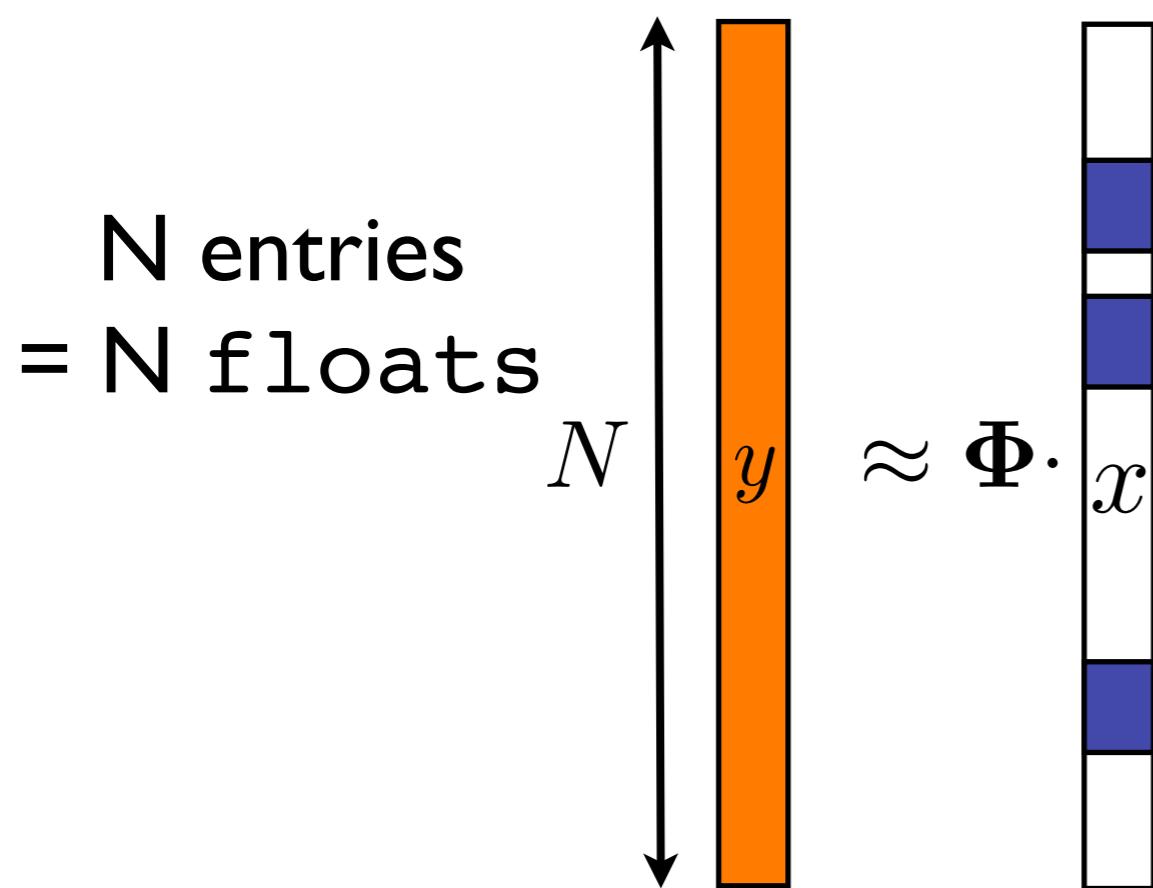


- Sparse vector

Key practical issues: choose dictionary

Why the log factor ?

- Full vector

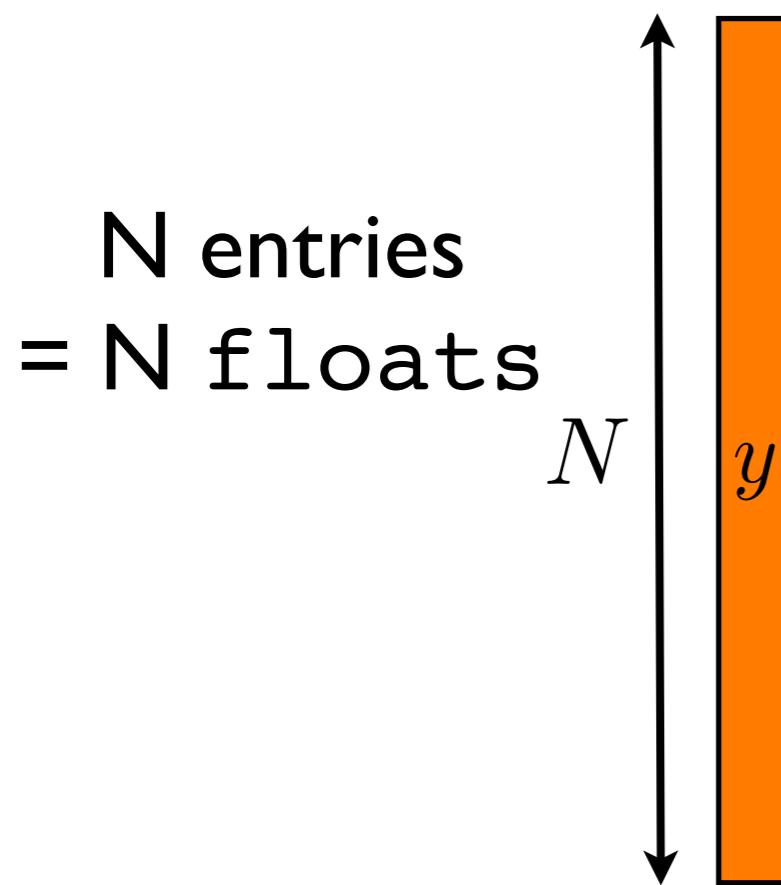


- Sparse vector

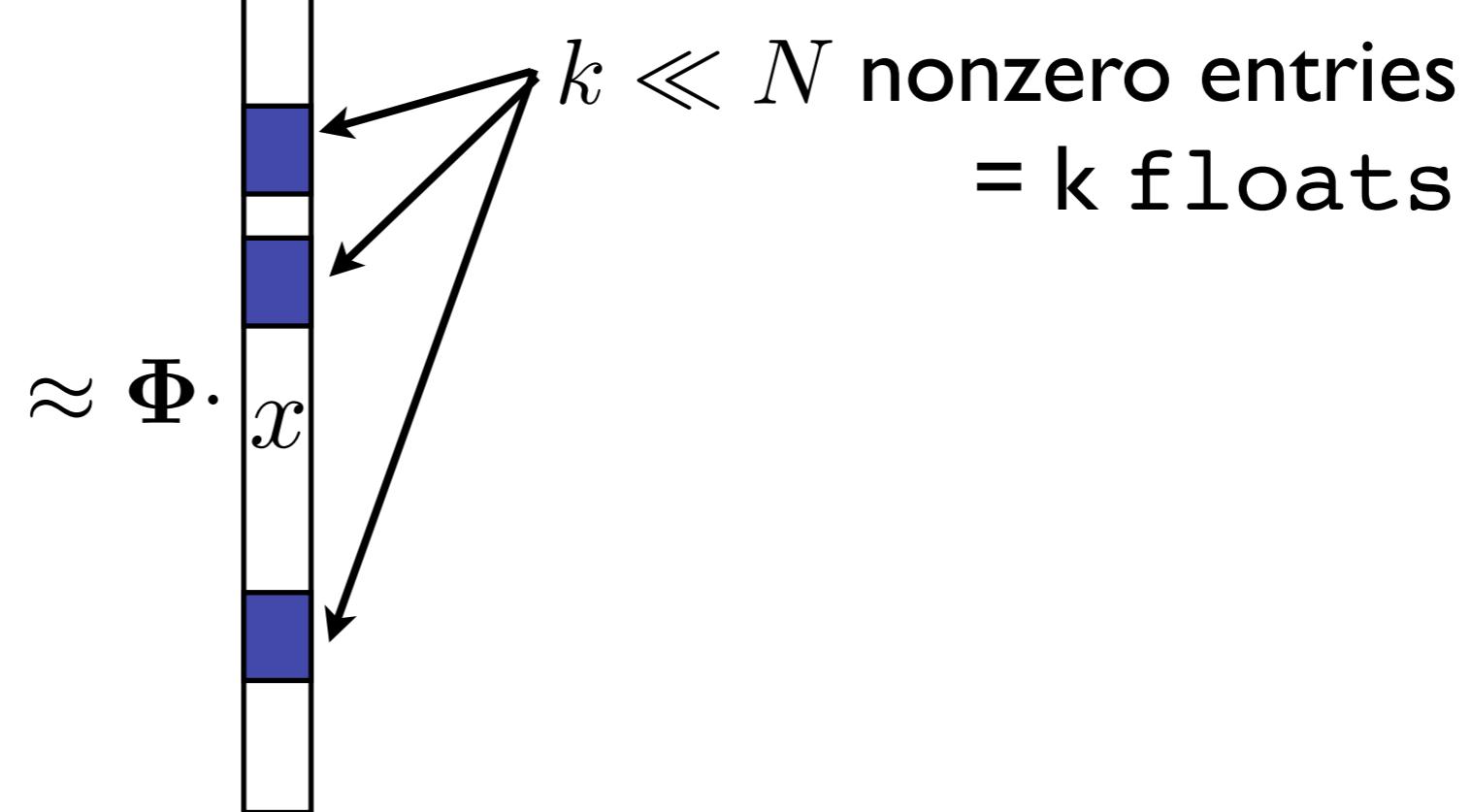
Key practical issues: choose dictionary

Why the log factor ?

- Full vector



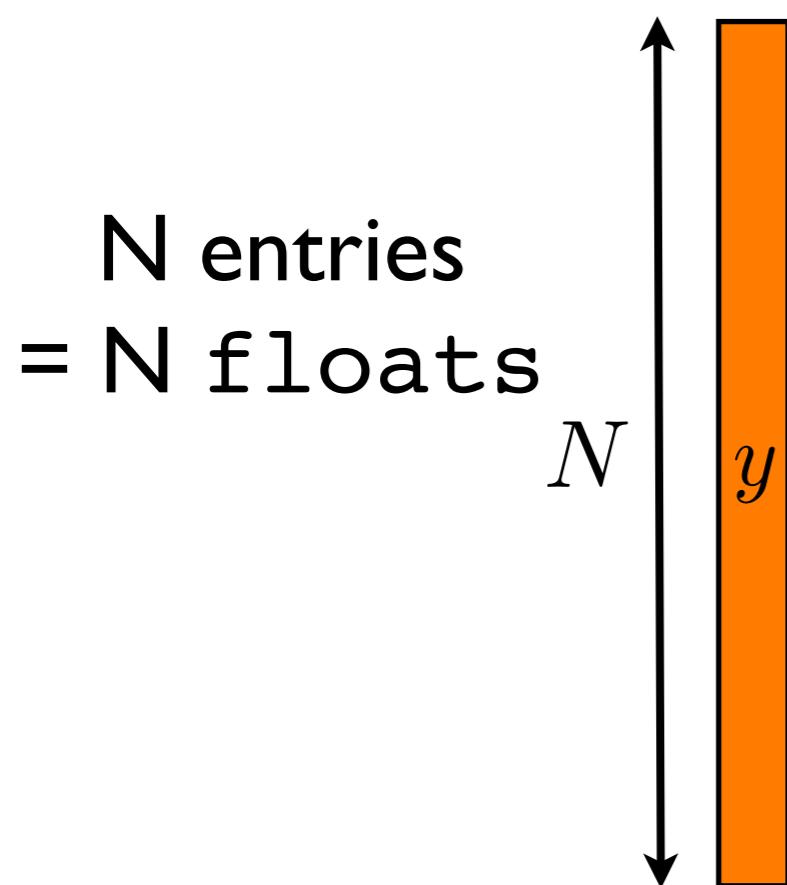
- Sparse vector



Key practical issues: choose dictionary

Why the log factor ?

- Full vector



- Sparse vector

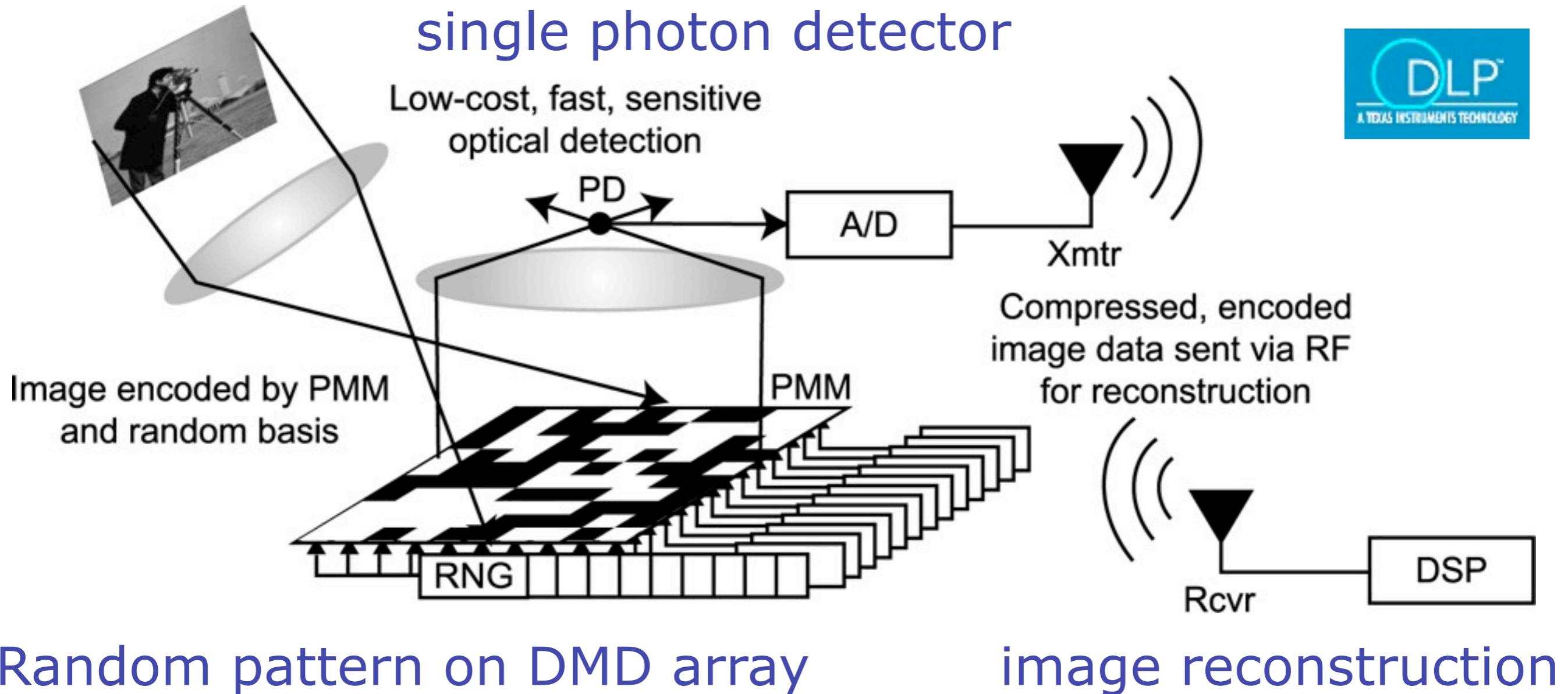
$k \ll N$ nonzero entries
= k floats

+ k positions among N

$$= \log_2 \binom{N}{k} \approx k \log_2 \frac{N}{k} \text{ bits}$$

Key practical issues: choose dictionary

Example : single-pixel camera, Rice University



Summary

Compression
Representation
Description
Classification

Notion of sparsity
(Fourier, wavelets, ...)

Sparsity

Denoising
Blind source separation
Compressed sensing
...

Natural / traditional role

Sparsity = low cost (bits, computations, ...)

Direct objective

Novel indirect role

Sparsity = prior knowledge, regularization

Tool for inverse problems