Sparse Representations: from Source Separation to Compressed Sensing

Rémi Gribonval
METISS project-team (audio signal processing, indexing, source separation)
INRIA, Rennes, France

IRISA
UNE UNITÉ DE RECHERCHE À LA POINTE DES SCIENCES ET DES TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

Ecole d'été en Traitement du Signal
Peyresq, Juillet 2009

UNIVERSITE DE RENNES 1 INSA CNRS INRIA RENNES

Structure of the course

- Part I: Overview
- Part II: Algorithms, complexity & convergence
  - Lp minimization
  - Greedy Algorithms
- Part III: Recovery, stability, robustness
  - Null Space Properties and Lp minimization
  - Exact Recovery Condition and greedy algorithms
  - Restricted Isometry Constants, stability and robustness
- Part IV: Compressed Sensing and Random Matrices

Introduction

Sparsity = old concept!
(wavelets, ...)

Compression
Adaptive representation
Feature extraction
Kernel methods (SVM ...)

Sparsity

Denoising
Blind source separation
Compressed sensing
...

Natural / traditional role :

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

Novel indirect role

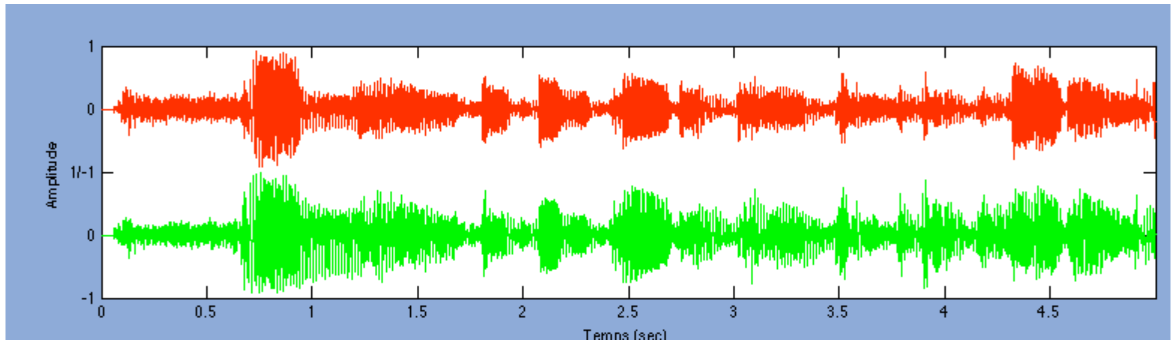
Sparsity = prior knowledge
Tool for inverse problems

Overview

- Introduction : source separation and inverse problems
- Sparse decomposition algorithms
  - L1 minimisation
  - Matching Pursuits
- Provably good algorithms to recover sparse representations
- Compressed sensing & random sampling

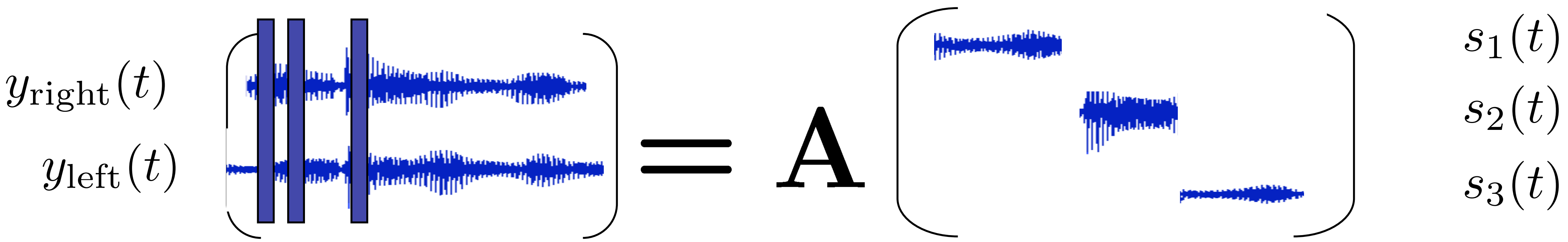
« Blind » Audio Source Separation

- « Softly as in a morning sunrise »

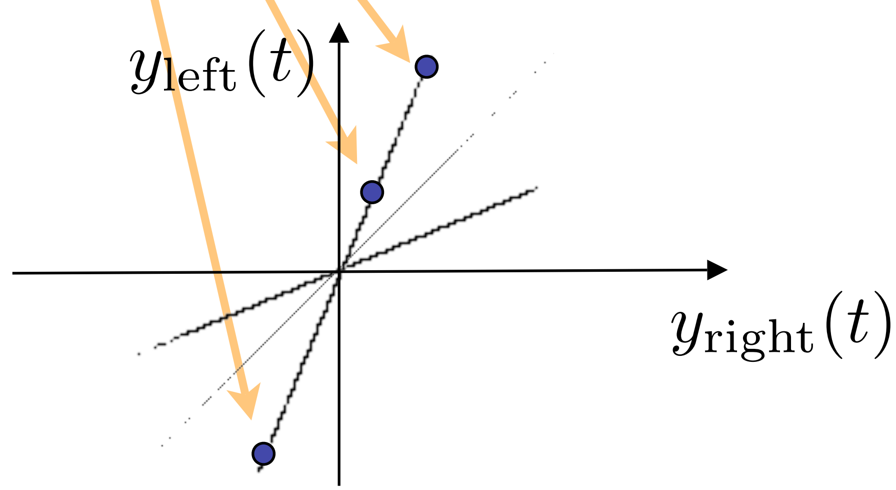


Blind Source Separation

- Mixing model : linear instantaneous mixture



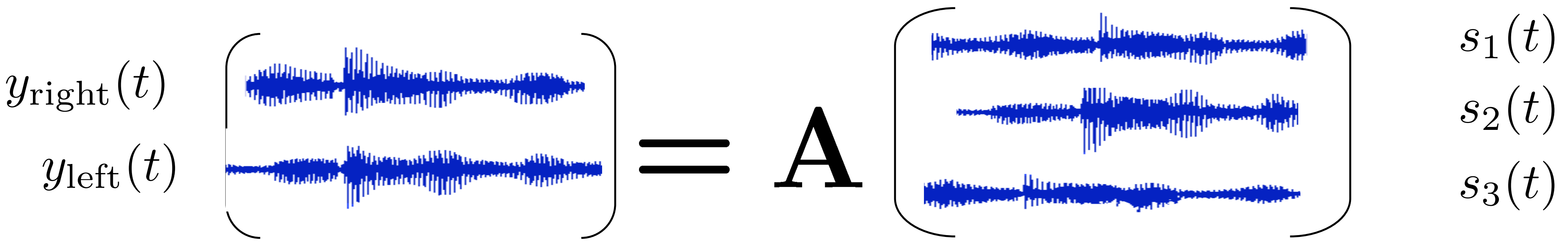
- Source model : if disjoint time-supports …



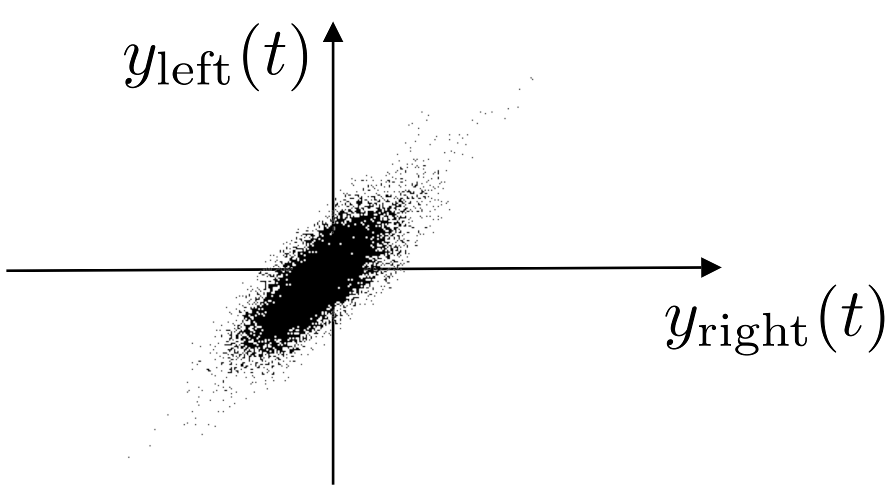
... then clustering to :
1- identify (columns of) the mixing matrix
2- recover sources

Blind Source Separation

- Mixing model : linear instantaneous mixture



- In practice ...

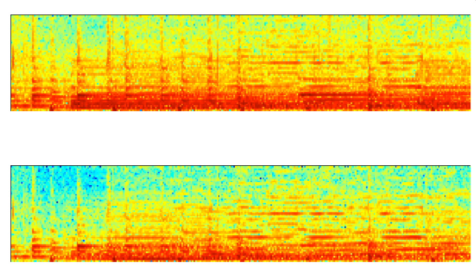


Time-Frequency Masking

- Mixing model in the time-frequency domain

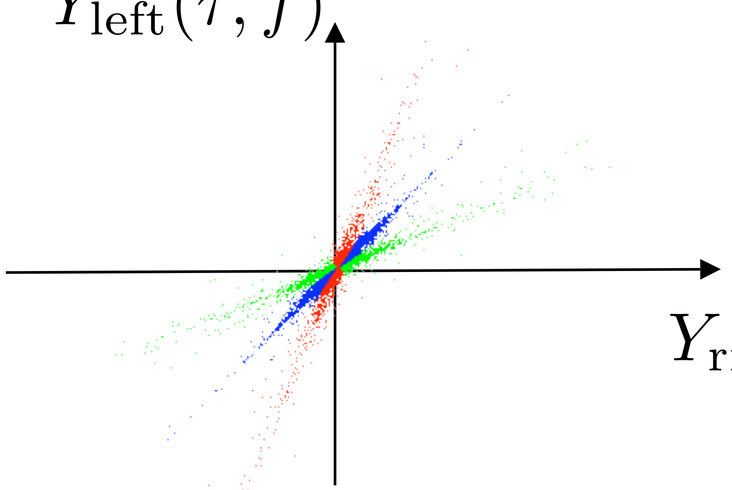
$Y_{\text{right}}(\tau, f)$

$Y_{\text{left}}(\tau, f)$



$$= \mathbf{A}\,\mathbf{S}(\tau, f)$$

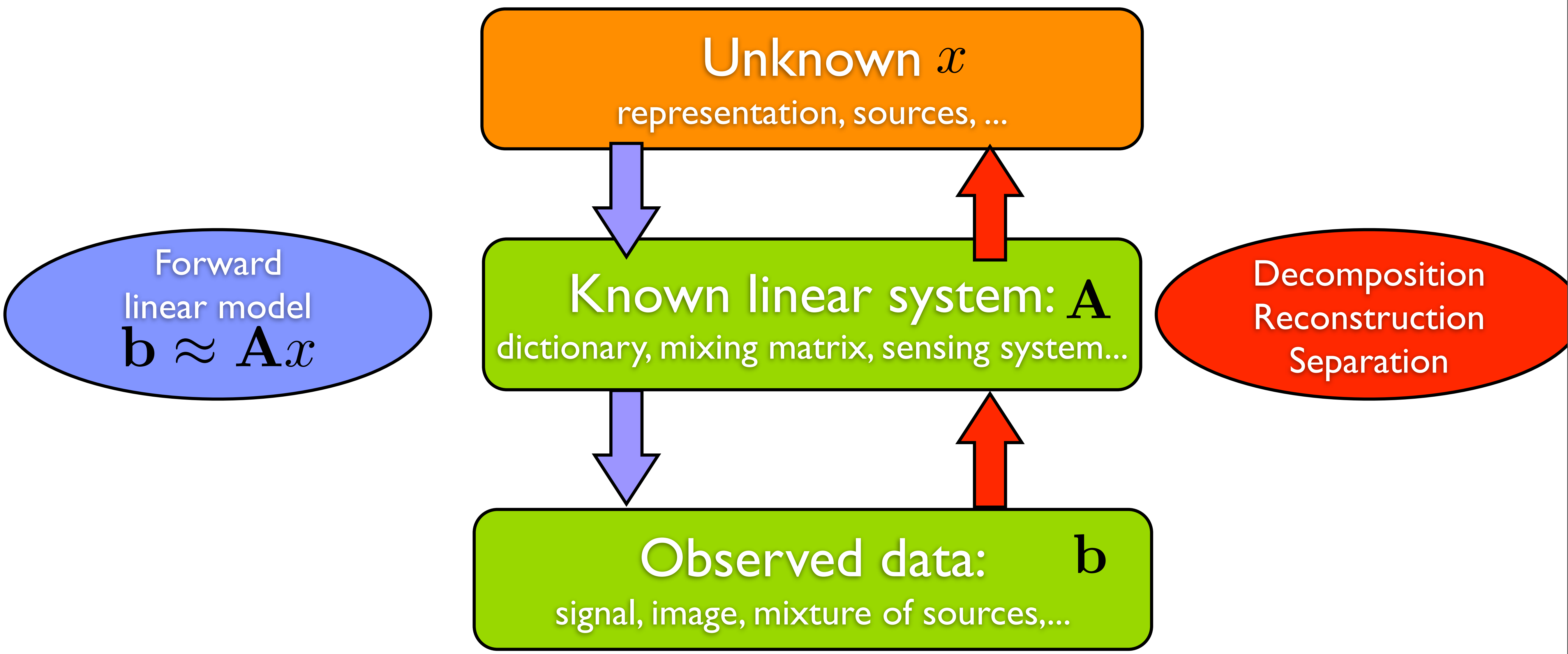
- And "miraculously" ...



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

Sparse decomposition algorithms

Vocabulary



Unknown x
representation, sources, ...
Forward
linear model
$\mathbf{b} \approx \mathbf{A}x$
Known linear system: $\mathbf{A}$
dictionary, mixing matrix, sensing system...
Decomposition
Reconstruction
Separation
Observed data: $\mathbf{b}$
signal, image, mixture of sources,...

Sparsity and Ill-Posed Inverse Problems

- Ill-posedness if more unknowns than equations

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

- Uniqueness of sparse solutions:
  - if x_0, x_1 are "sufficiently sparse",
  - then $\mathbf{A}x_0 = \mathbf{A}x_1 \Rightarrow x_0 = x_1$

- Recovery with practical algorithms
  - Thresholding, Matching Pursuits, Lp minimization p<=1,...

Overall compromise

- Approximation quality

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

- Ideal sparsity measure : ℓ^0 "norm"

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

  - Relaxed sparsity measure

$$\|x\|_p := (\sum_n |x_n|^p)^{1/p}$$

IRISA

Algorithms

	Global optimization	Iterative greedy algorithms
Principle	$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$	
Tuning quality/sparsity	regularization parameter λ	
Variants	• choice of sparsity measure p • optimization algorithm • initialization	

Global Optimization

- Principle $$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$$
  - Sparse representation $\lambda \to 0$
  - Sparse approximation $\lambda > 0$

Global Optimization

- Principle $$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$$
  - ✦ Sparse representation $\lambda \to 0$
  - ✦ Sparse approximation $\lambda > 0$

Linear

$p = 0$ $\longleftrightarrow$ $p = 1$ $\longleftrightarrow$ $p = 2$

Global Optimization

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

  - ✦ Sparse representation $\lambda \to 0$
  - ✦ Sparse approximation $\lambda > 0$

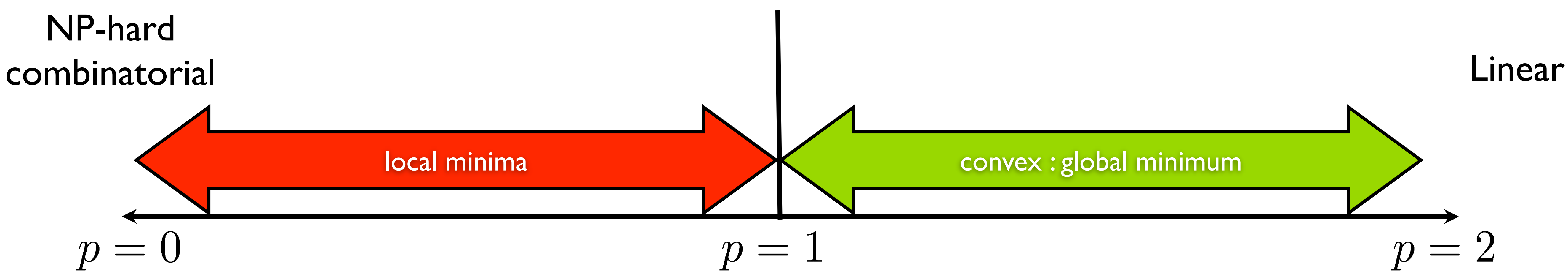
NP-hard
combinatorial

Linear

$p = 0$ $p = 1$ $p = 2$

Global Optimization

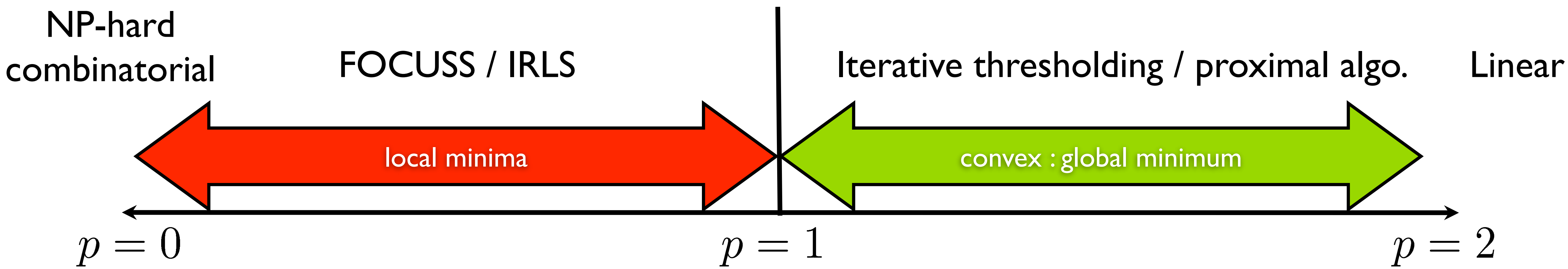
- Principle $$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$$
  - ✦ Sparse representation $\lambda \to 0$
  - ✦ Sparse approximation $\lambda > 0$



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, Defrise, de Mol 2004, Combettes, & Pesquet 2008 ...]

Global Optimization

- Principle $$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$$
  - ✦ Sparse representation $\lambda \to 0$
  - ✦ Sparse approximation $\lambda > 0$



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, Defrise, de Mol 2004, Combettes, & Pesquet 2008 ...]

Algorithms

	Global optimization	Iterative greedy algorithms
Principle	$\min_x \frac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2 + \lambda\|x\|_p^p$	iterative decomposition $\mathbf{r}_i = \mathbf{b} - \mathbf{A}x_i$ • select new components • update residual
Tuning quality/sparsity	regularization parameter λ	stopping criterion (nb of iterations, error level, ...) $\|x_i\|_0 \geq k$ $\|\mathbf{r}_i\| \leq \epsilon$
Variants	• choice of sparsity measure p • optimization algorithm • initialization	•selection criterion (stagewise ...) •update strategy (orthogonal ...)

Main greedy algorithms

$$\mathbf{b} = \mathbf{A}x_i + \mathbf{r}_i \qquad\qquad \mathbf{A} = [\mathbf{A}_1, \ldots \mathbf{A}_N]$$

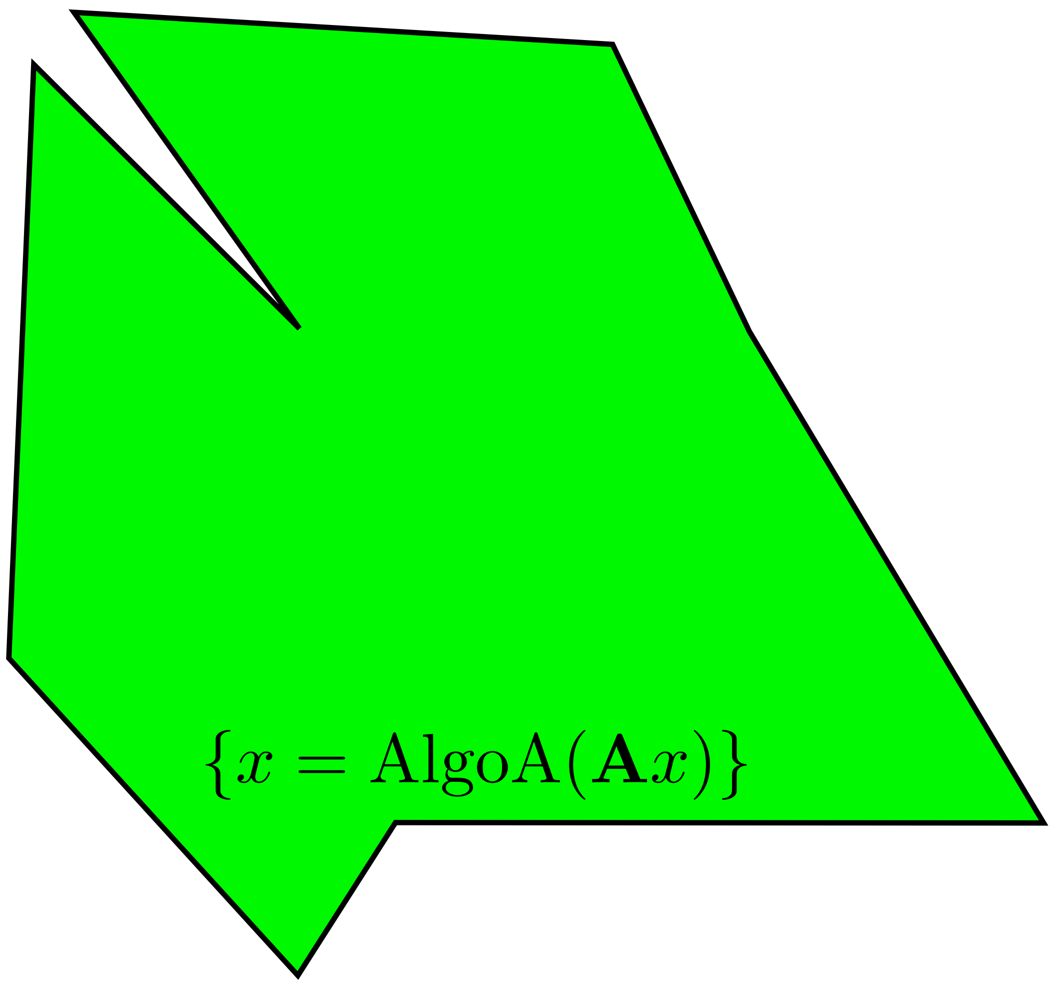
	Matching Pursuit	OMP	Stagewise
Selection	$\Gamma_i := \arg\max_n \lvert \mathbf{A}_n^T \mathbf{r}_{i-1} \rvert$		$\Gamma_i := \{n \mid \lvert \mathbf{A}_n^T \mathbf{r}_{i-1} \rvert > \theta_i\}$
Update	$\Lambda_i = \Lambda_{i-1} \cup \Gamma_i$ $x_i = x_{i-1} + \mathbf{A}_{\Gamma_i}^+ \mathbf{r}_{i-1}$ $\mathbf{r}_i = \mathbf{r}_{i-1} - \mathbf{A}_{\Gamma_i} \mathbf{A}_{\Gamma_i}^+ \mathbf{r}_{i-1}$	$\Lambda_i = \Lambda_{i-1} \cup \Gamma_i$ $x_i = \mathbf{A}_{\Lambda_i}^+ \mathbf{b}$ $\mathbf{r}_i = \mathbf{b} - \mathbf{A}_{\Lambda_i} x_i$	

MP & OMP: *Mallat & Zhang 1993*
StOMP: *Donoho & al 2006* (similar to MCA, *Bobin & al 2006*)

Provably good algorithms

Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

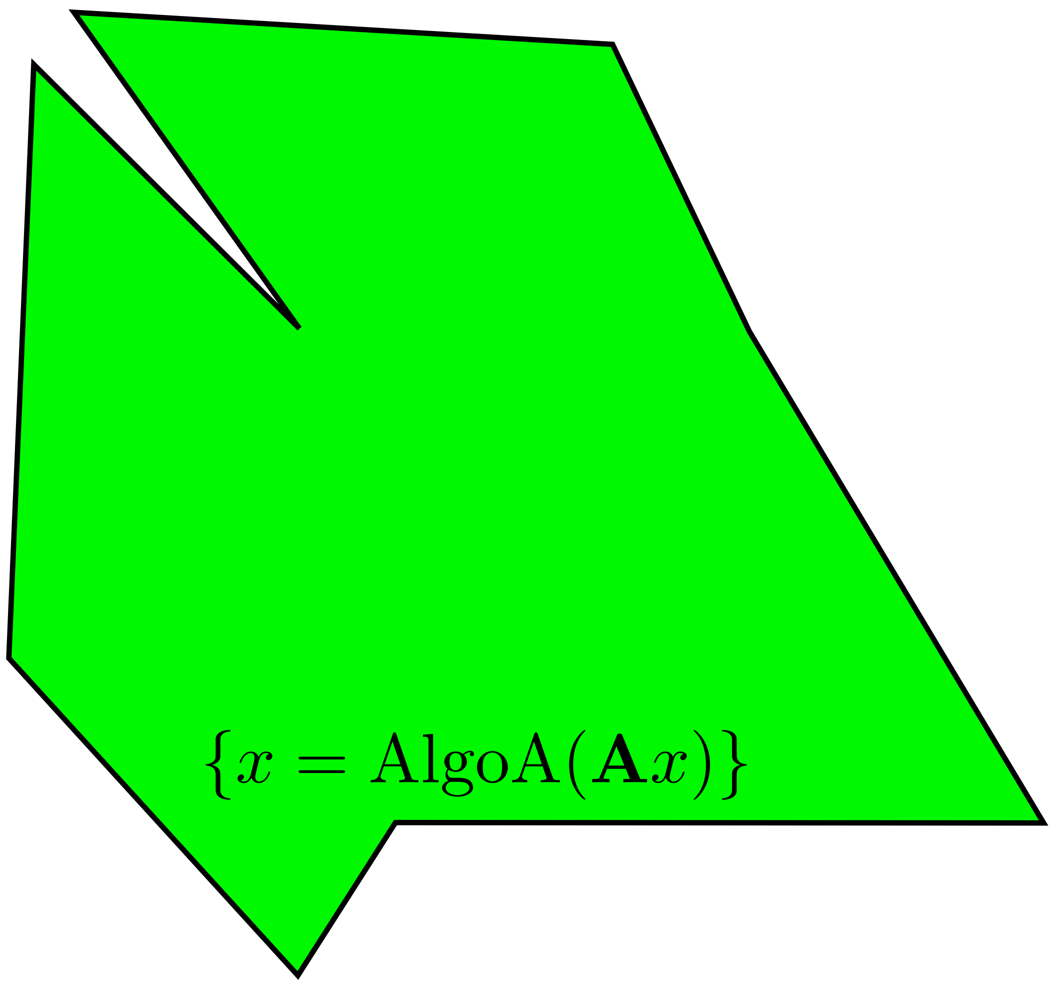
- Recoverable set for a given "inversion" algorithm



IRISA

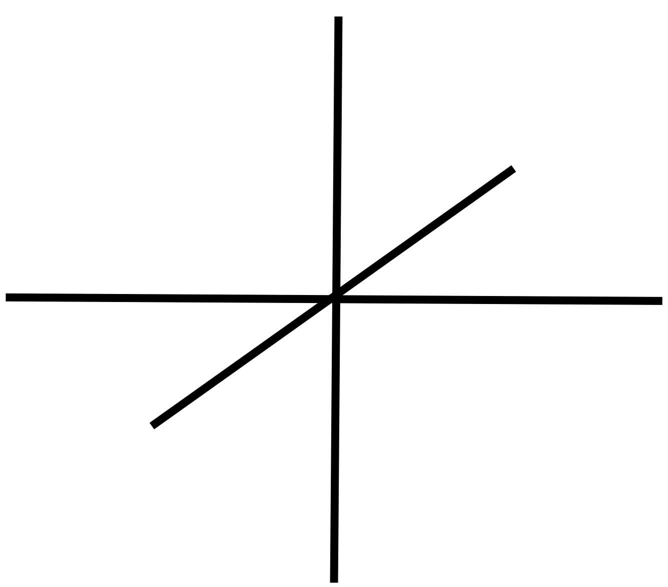
Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

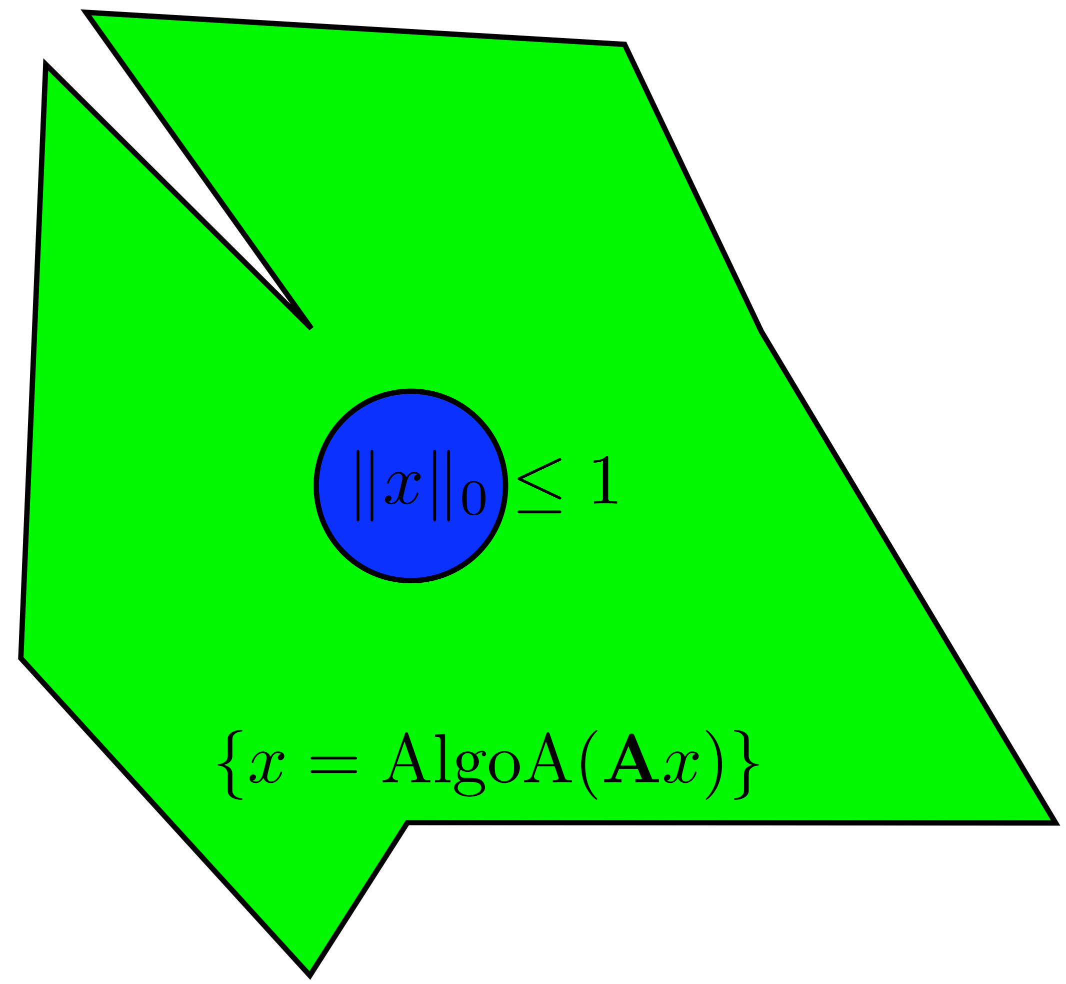
- Recoverable set for a given "inversion" algorithm
- Level sets of L0-norm



Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

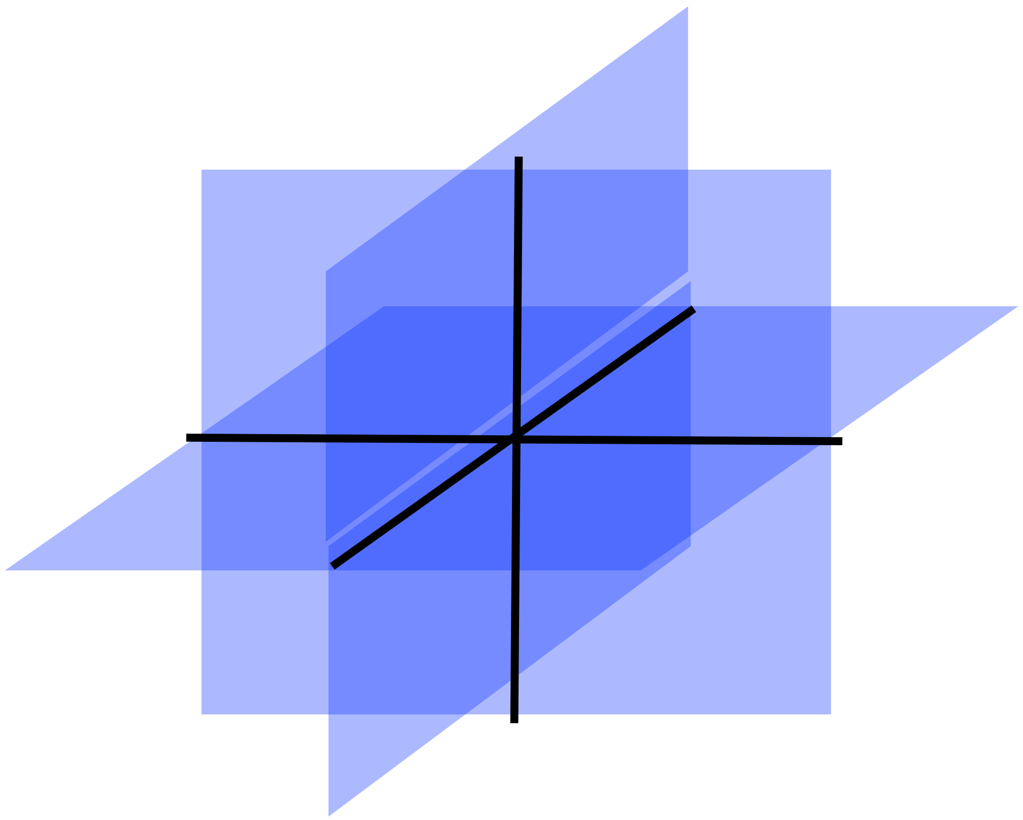
- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
  - 1-sparse

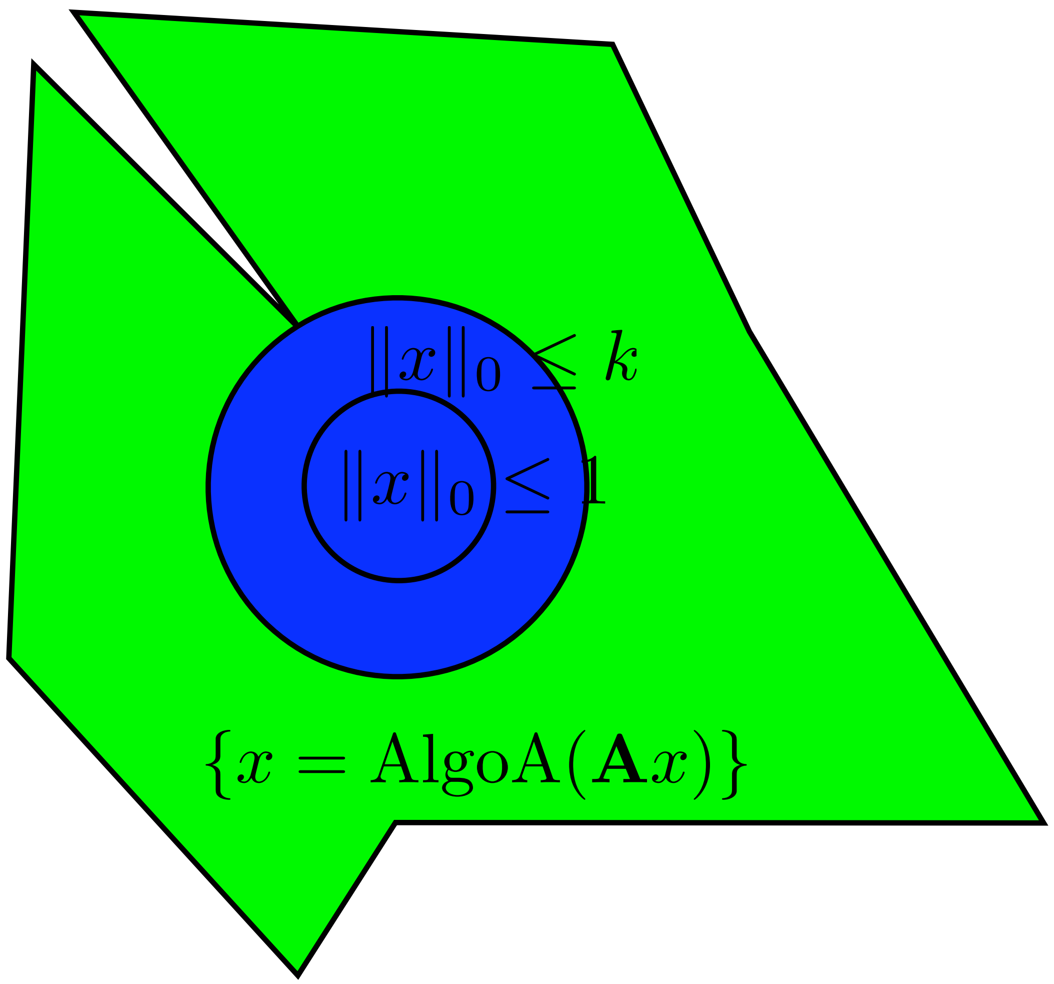




Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

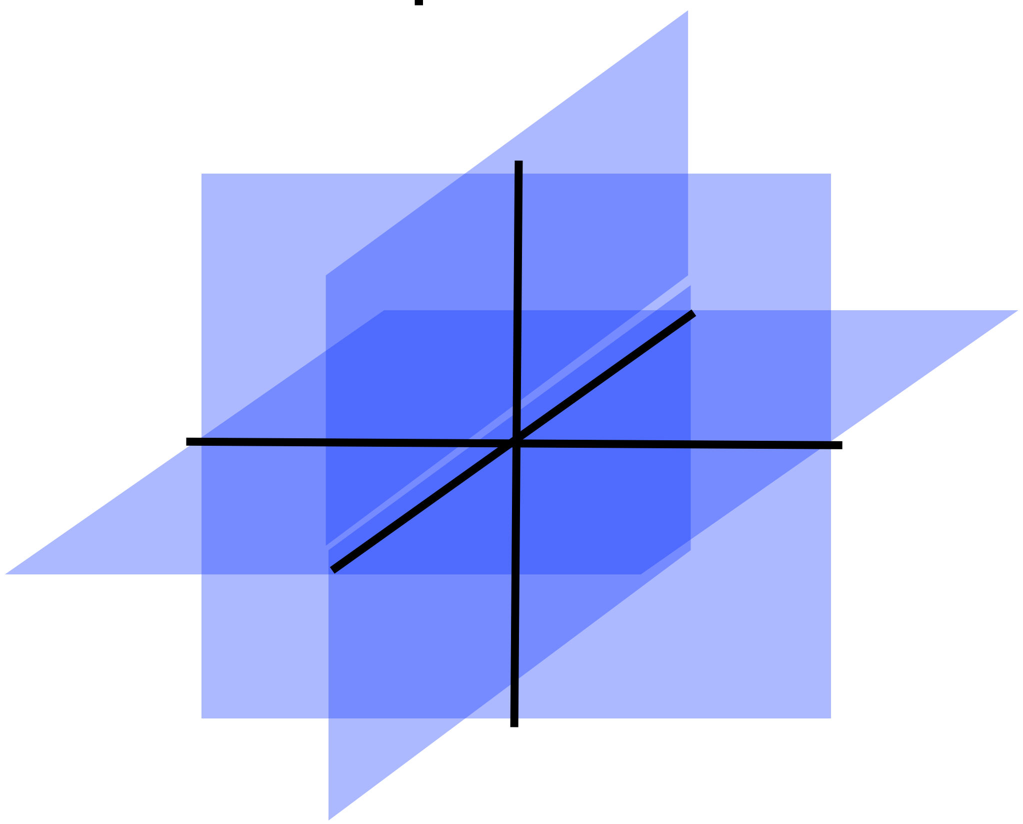
- Recoverable set for a given "inversion" algorithm
- Level sets of L0-norm
  - 1-sparse
  - 2-sparse

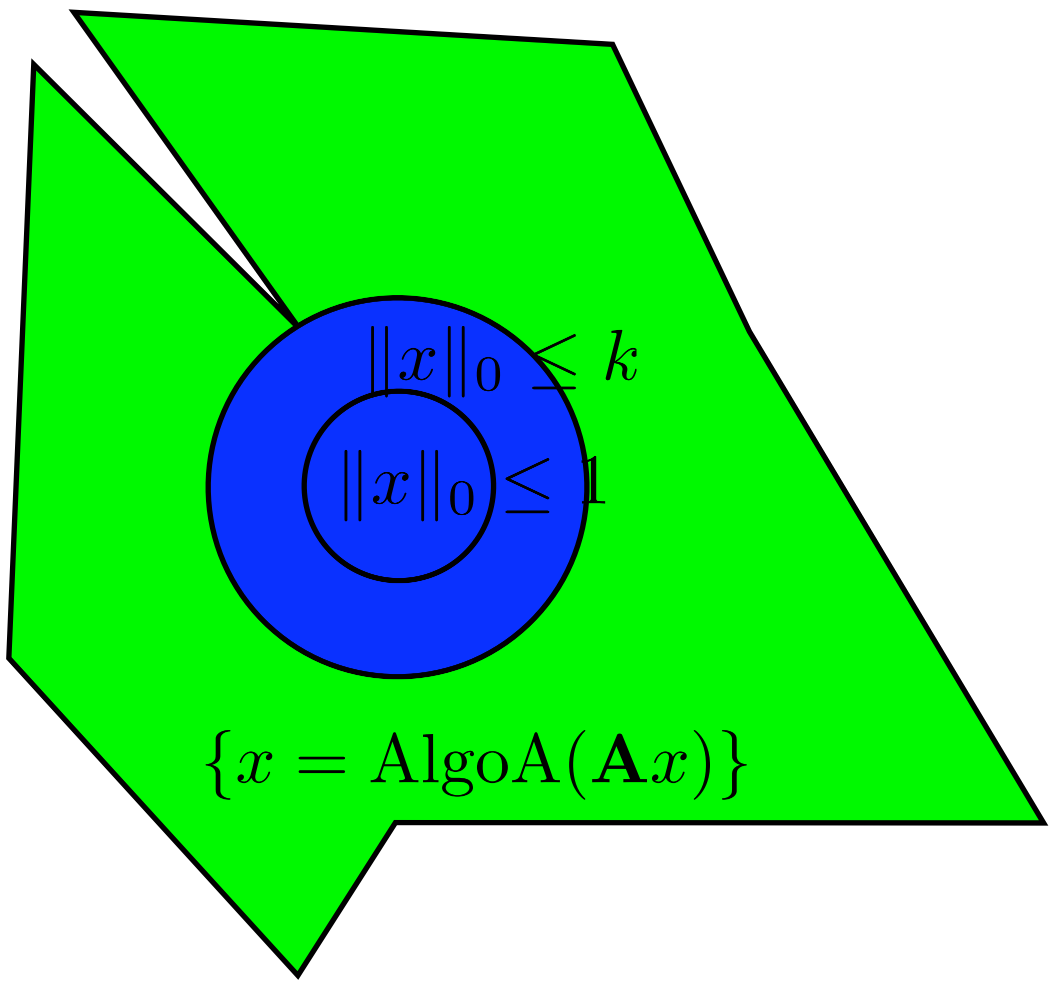




Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

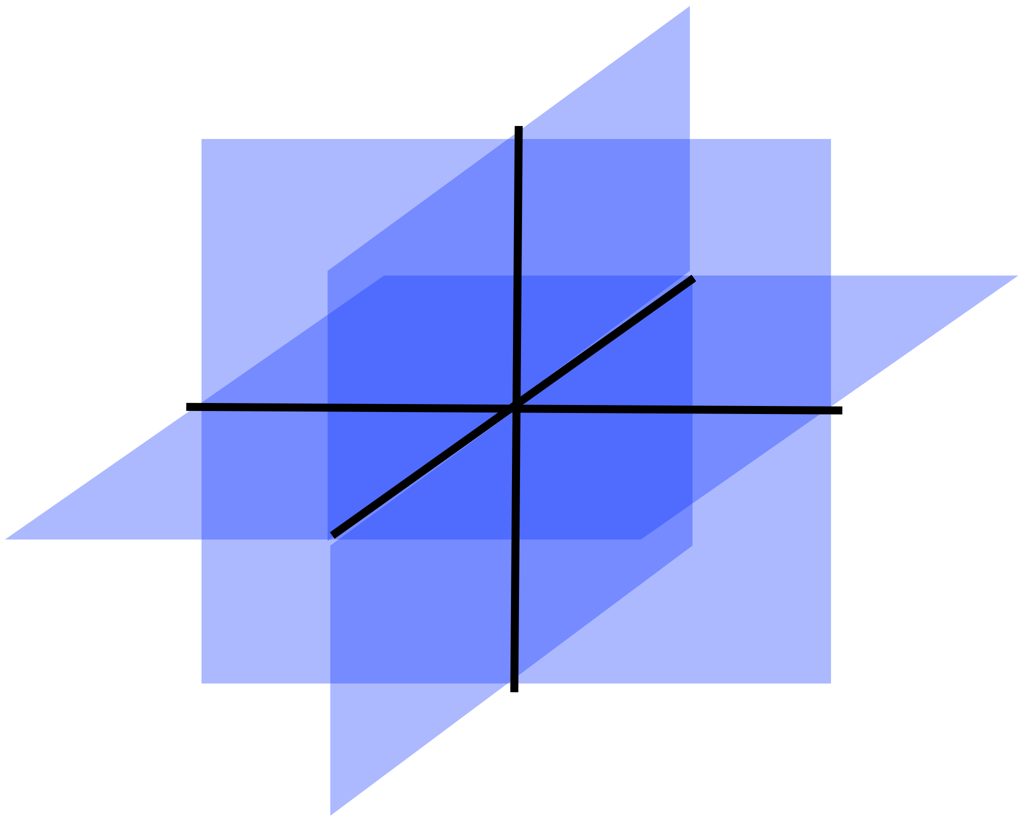
- Recoverable set for a given "inversion" algorithm
- Level sets of L0-norm
  - 1-sparse
  - 2-sparse
  - 3-sparse ...

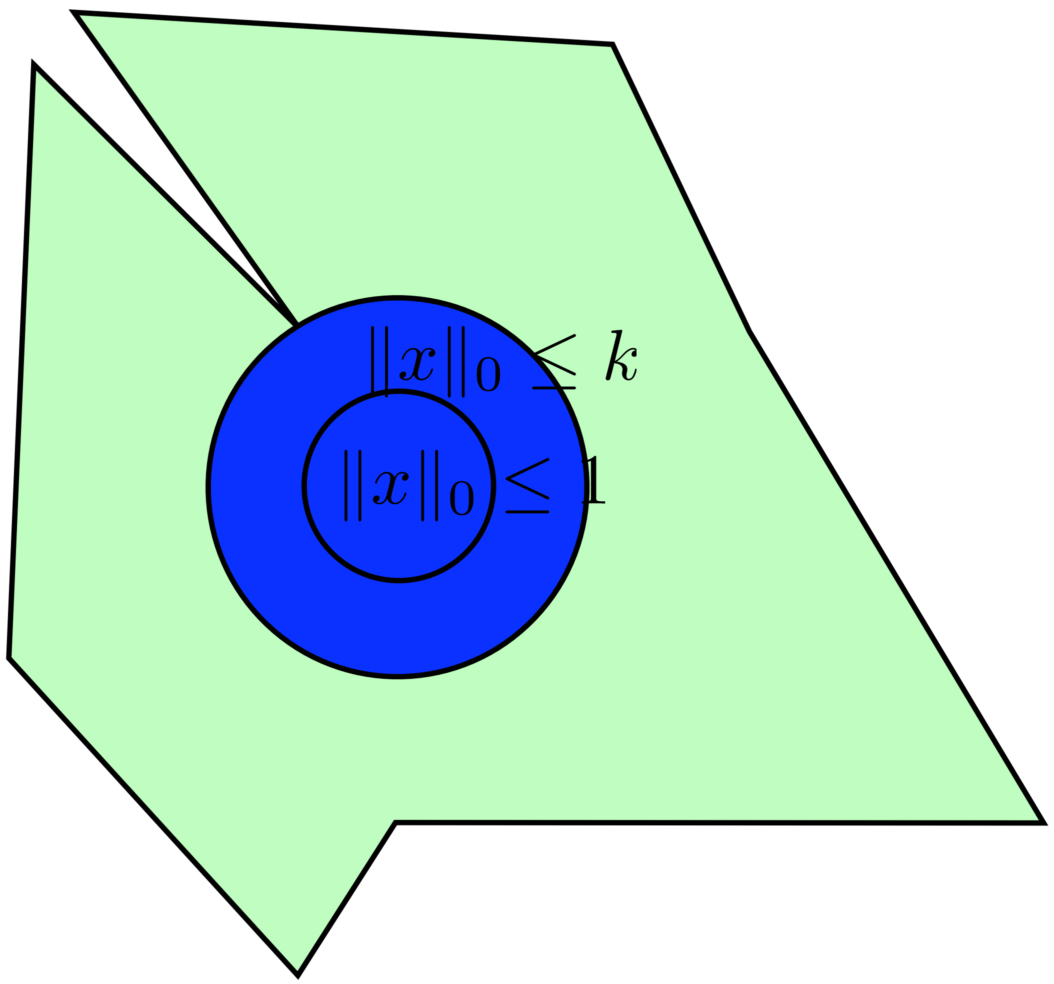




Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

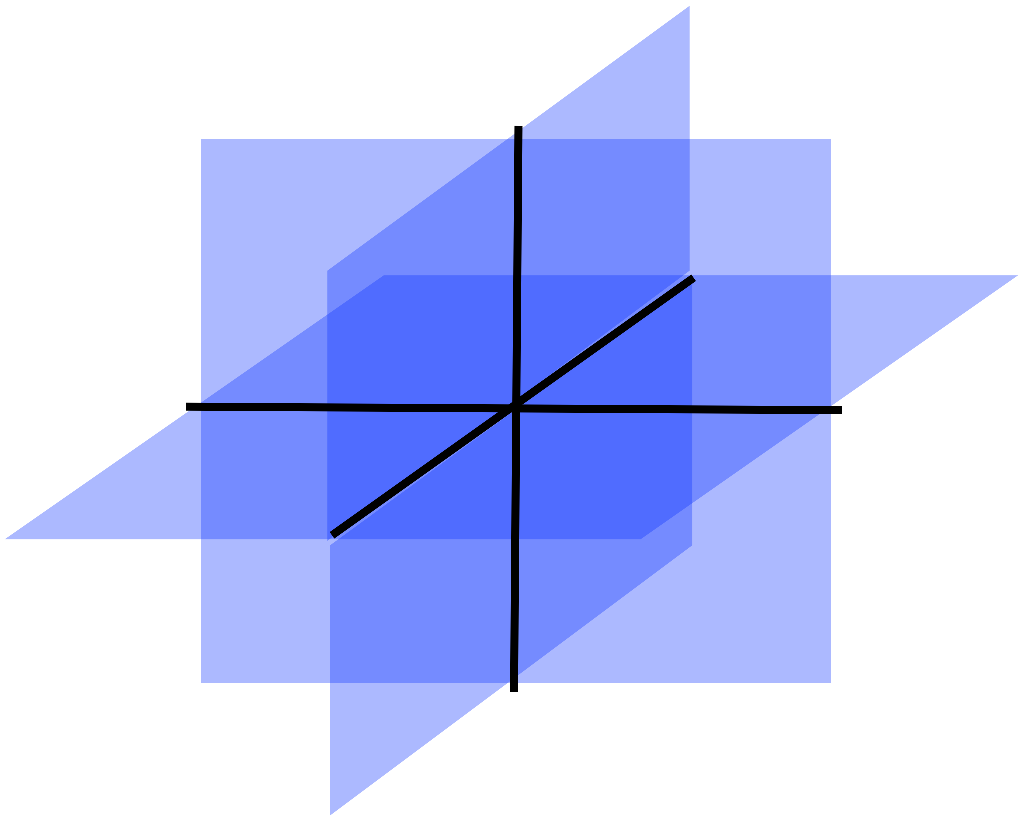
- Recoverable set for a given "inversion" algorithm
- Level sets of L0-norm
  - 1-sparse
  - 2-sparse
  - 3-sparse ...

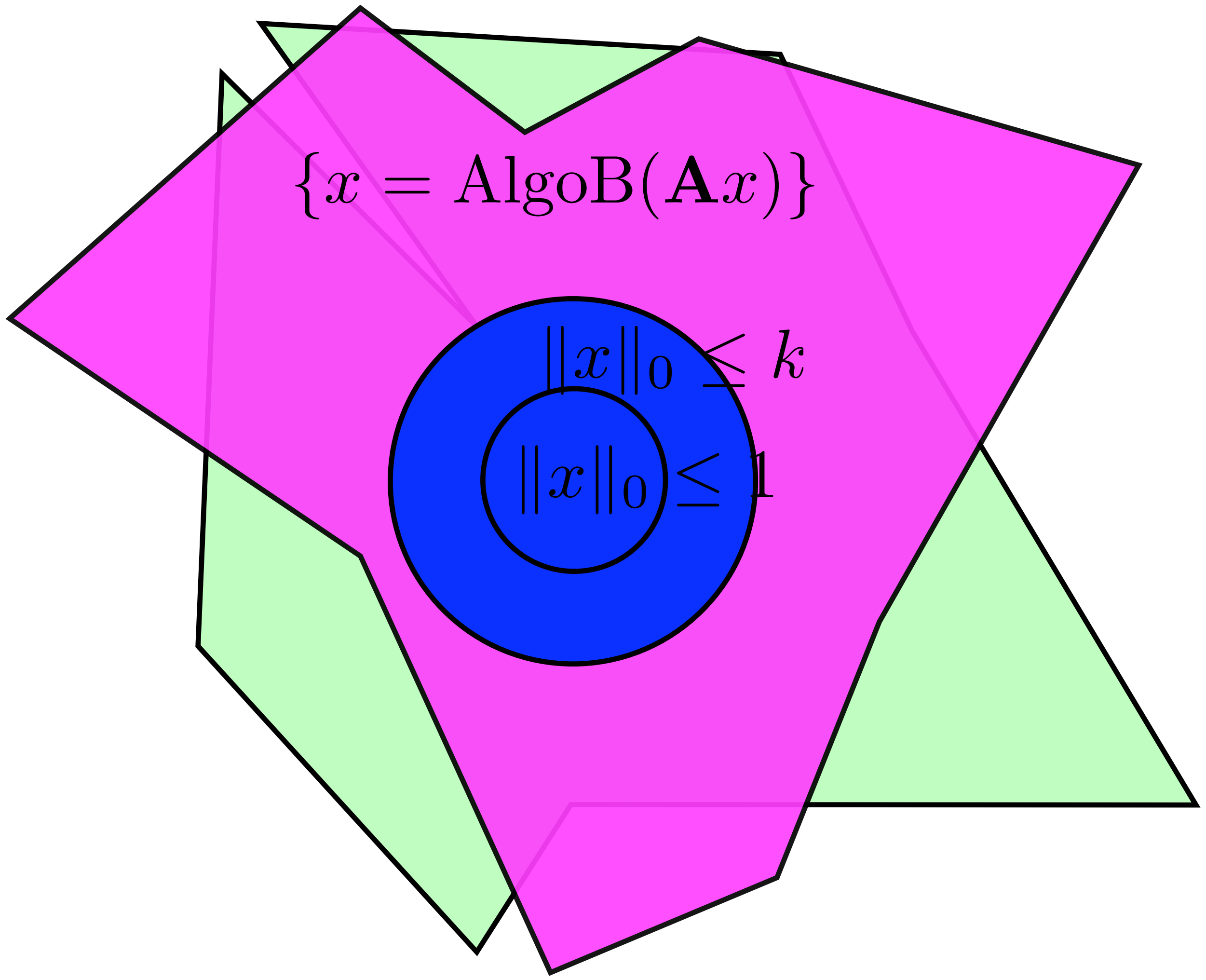




Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

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

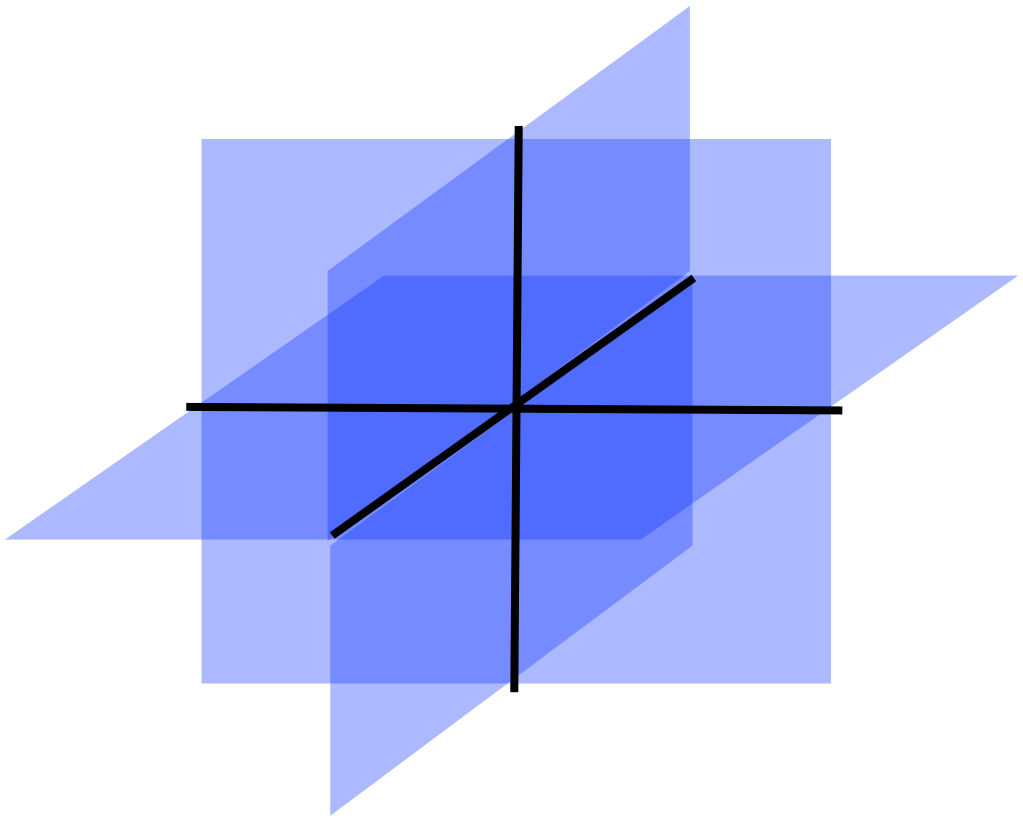


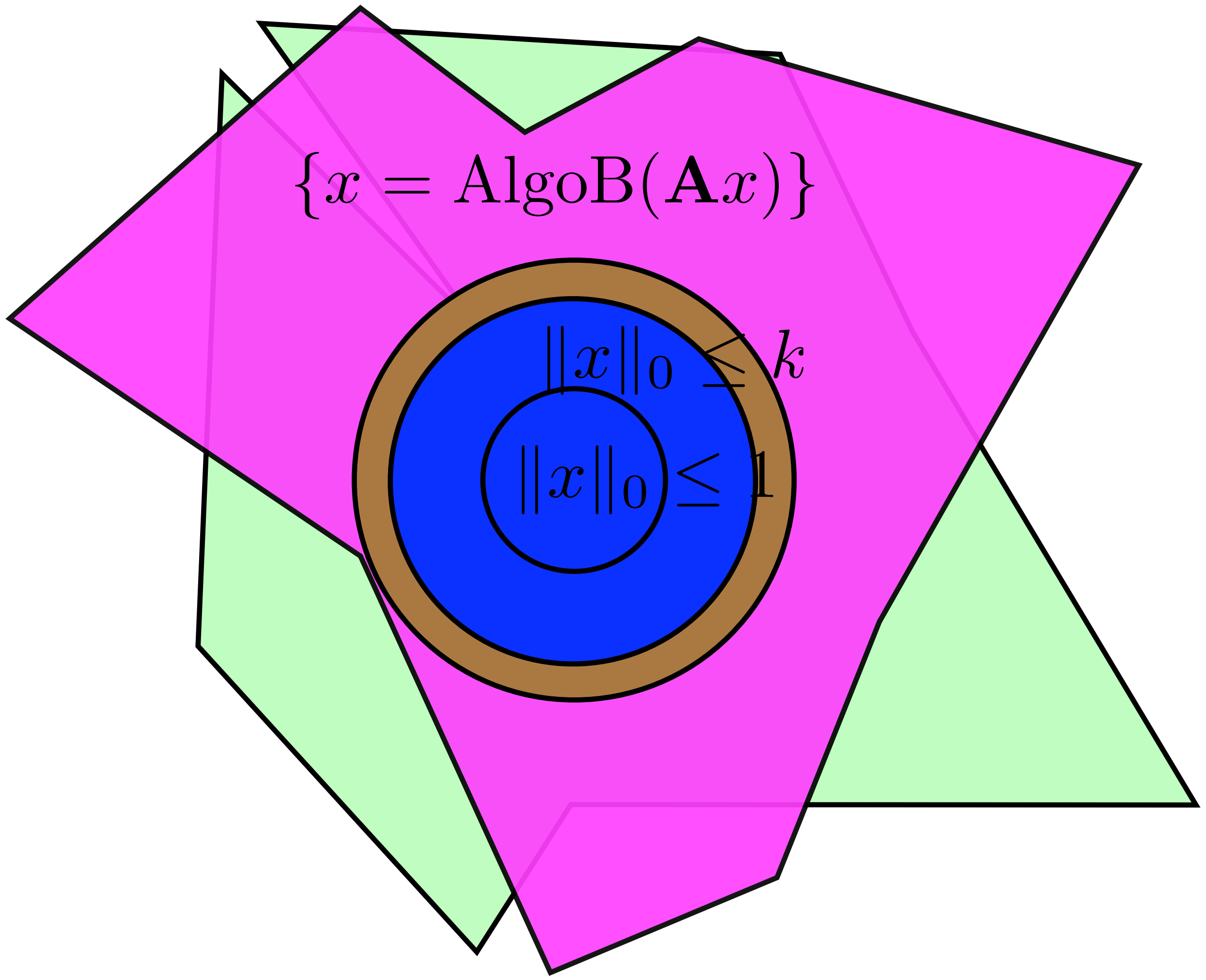


IRISA

Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

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

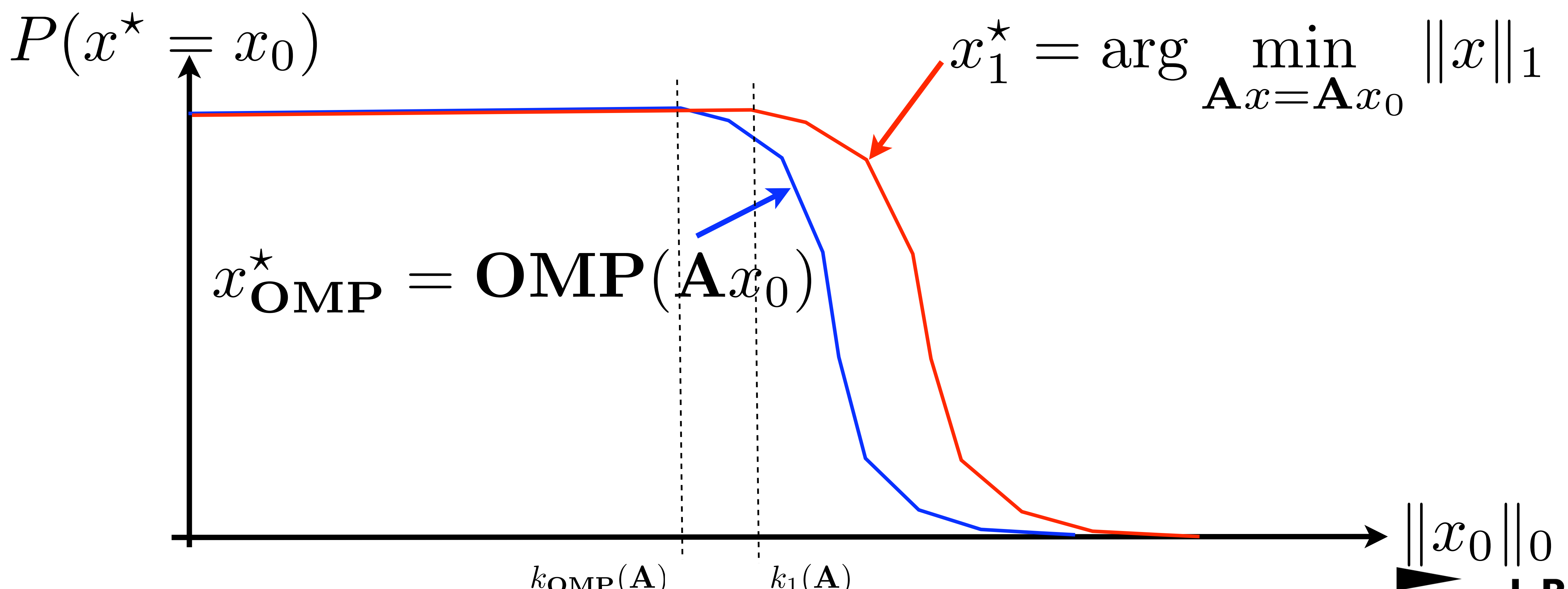




IRISA

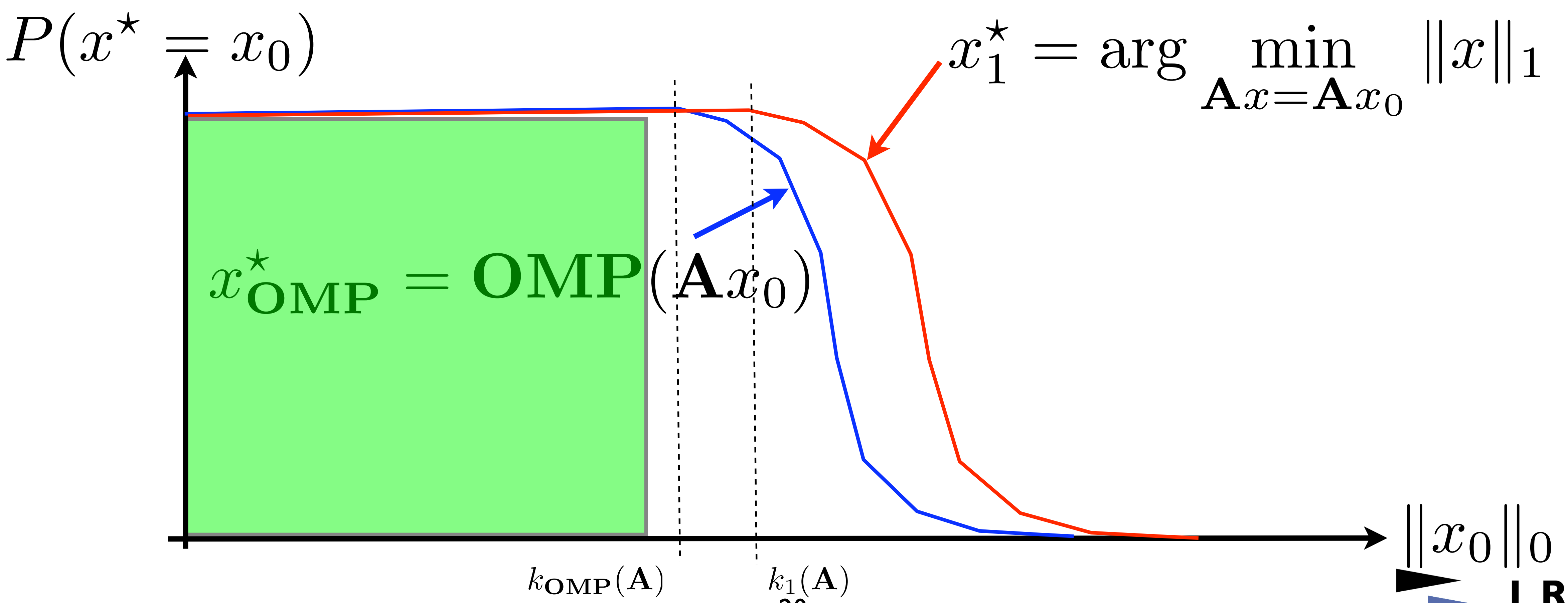
Empirical facts

- Highly sparse vectors : always recovered by
  - Orthonormal Matching Pursuit
  - Basis Pursuit (L1 minimization)
- Relatively sparse vectors : likely to be recovered



Empirical facts

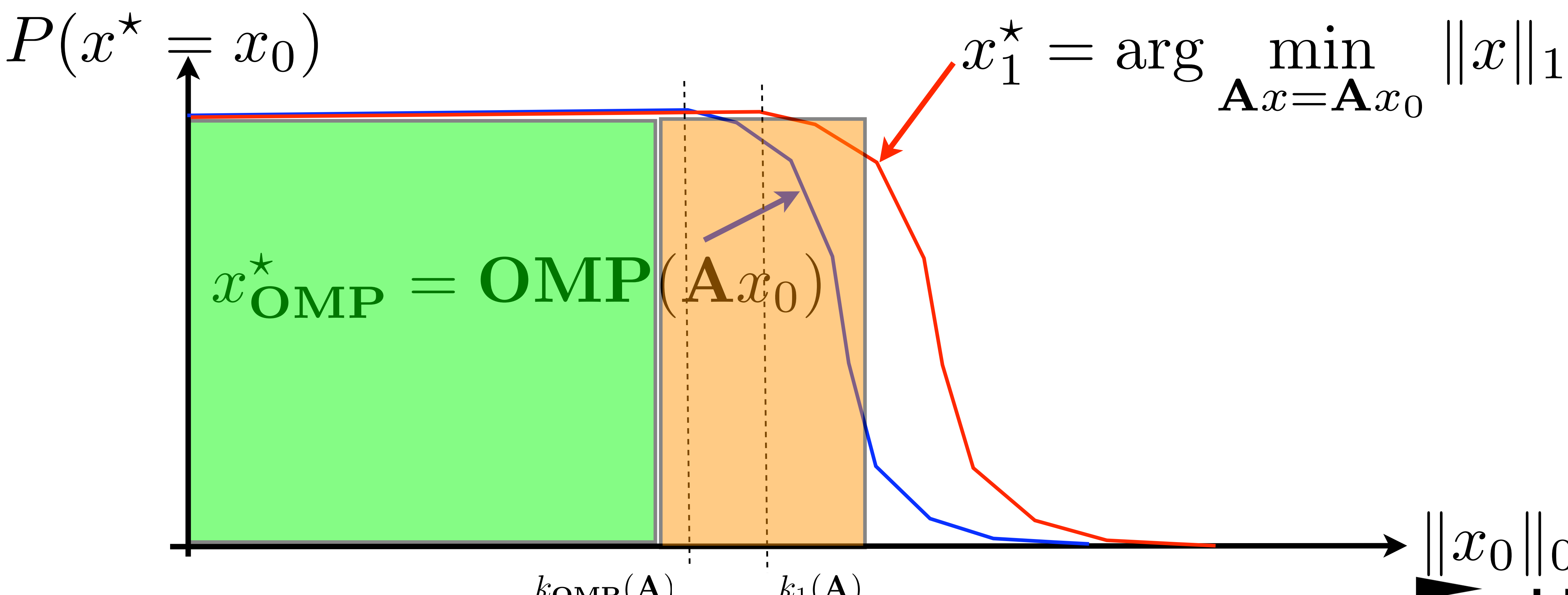
- Highly sparse vectors : always recovered by
  - Orthonormal Matching Pursuit
  - Basis Pursuit (L1 minimization)
- Relatively sparse vectors : likely to be recovered



IRISA

Empirical facts

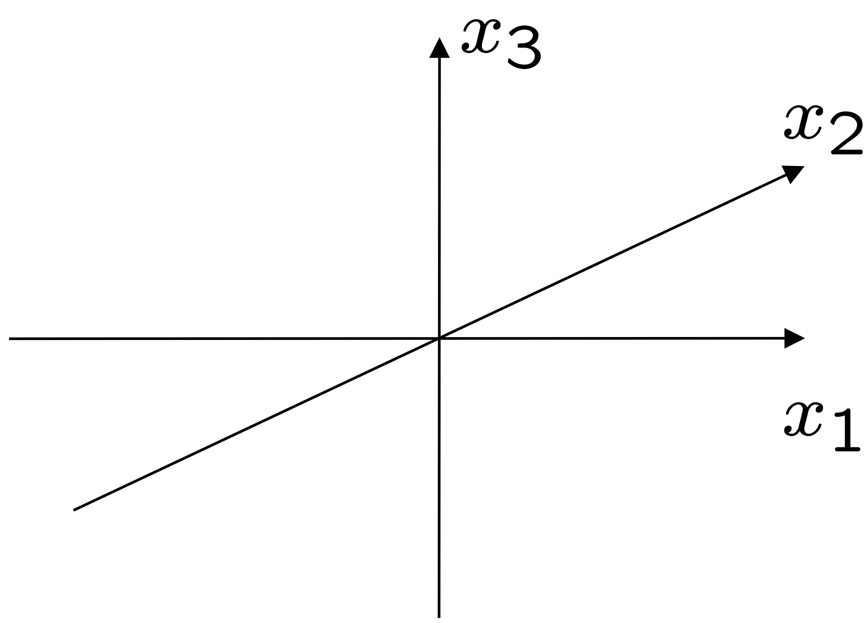
- Highly sparse vectors : always recovered by
  - ✦ Orthonormal Matching Pursuit
  - ✦ Basis Pursuit (L1 minimization)
- Relatively sparse vectors : likely to be recovered



Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

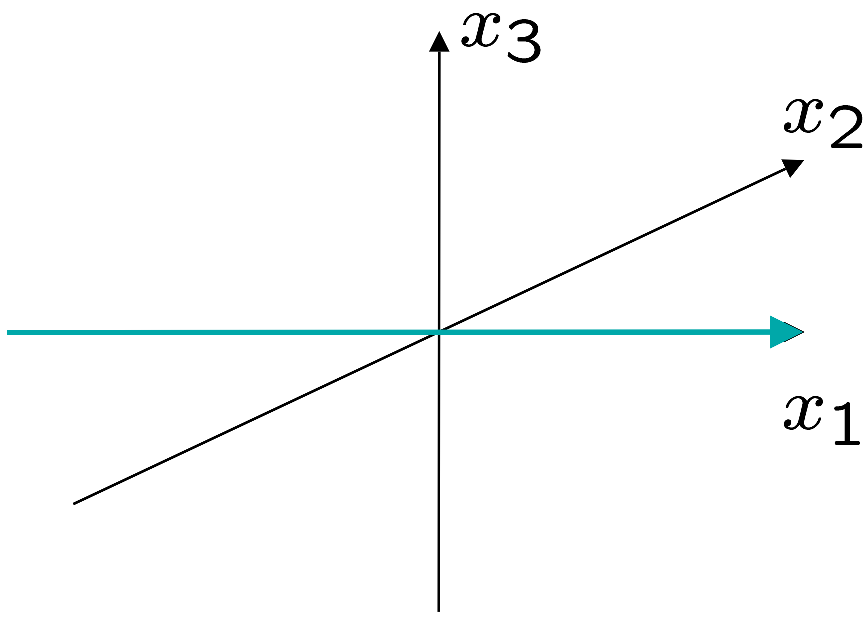
  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

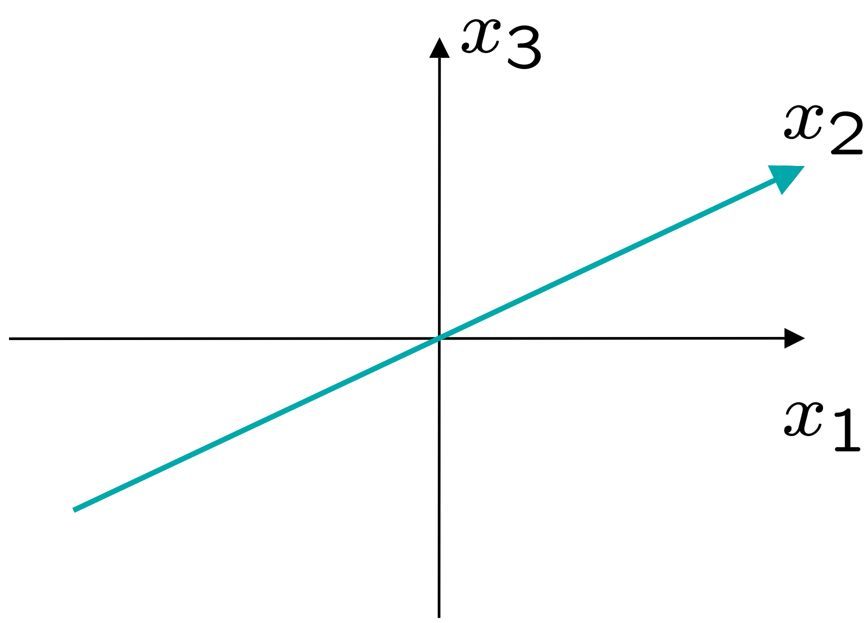
  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

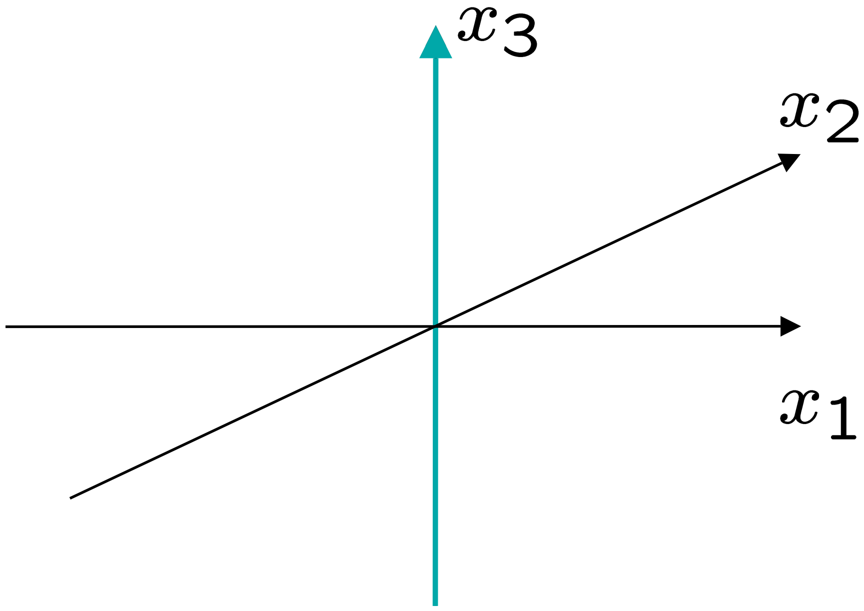
  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

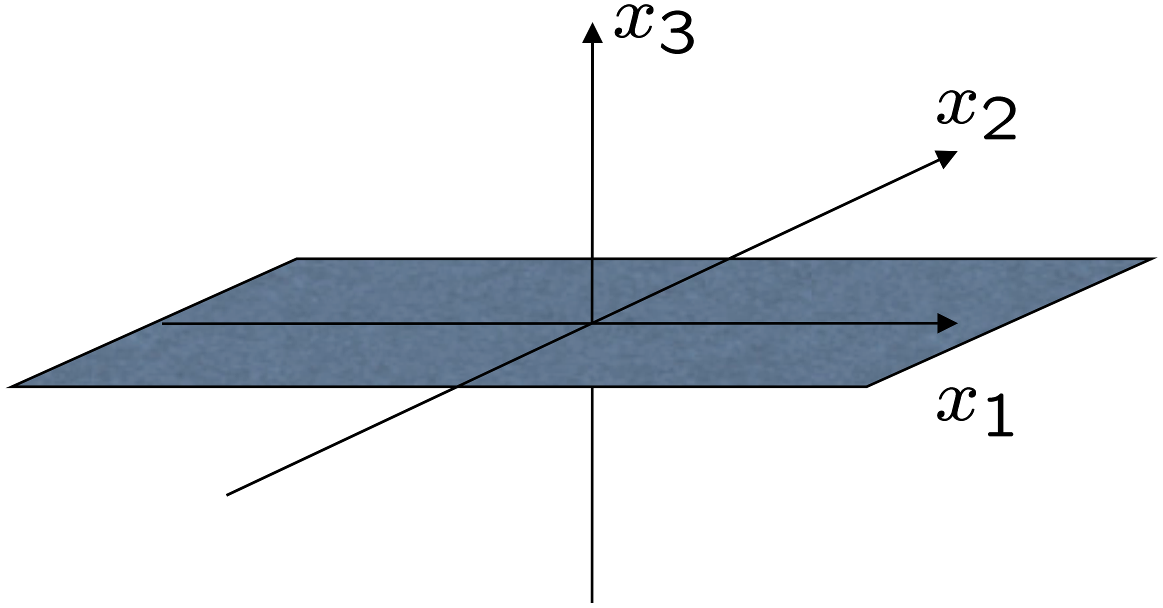
- Are there other more efficient alternatives ?

IRISA

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

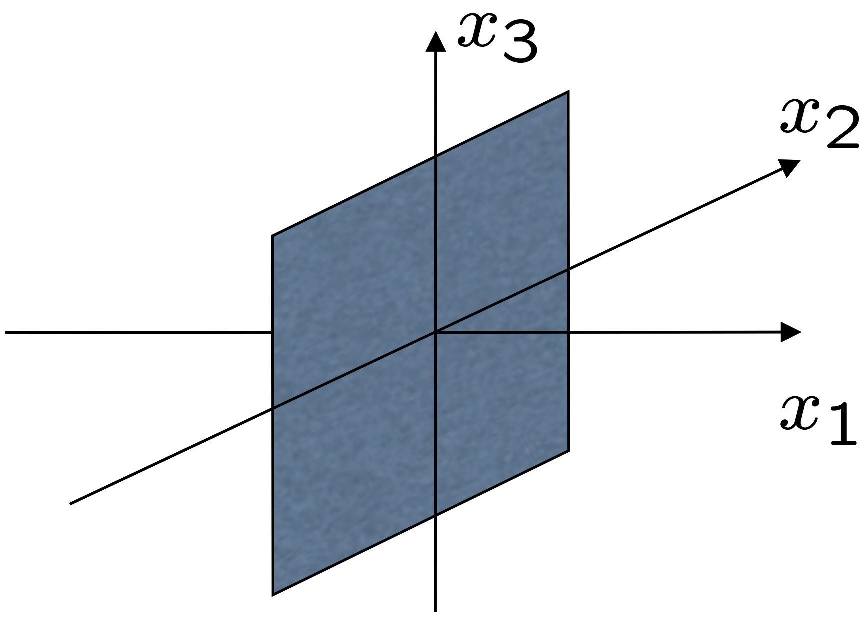
  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda\|x\|_0$$

- Brute force search



- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

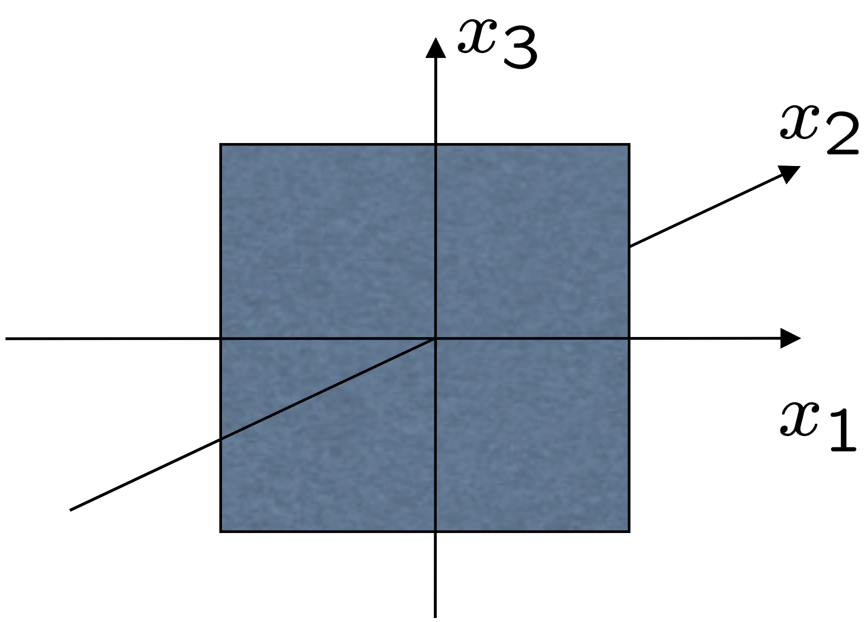
- Are there other more efficient alternatives ?

IRISA

Ideal Sparse Approximation

$$\min_x \|\mathbf{b} - \mathbf{A}x\|_2^2 + \lambda \|x\|_0$$

- Brute force search



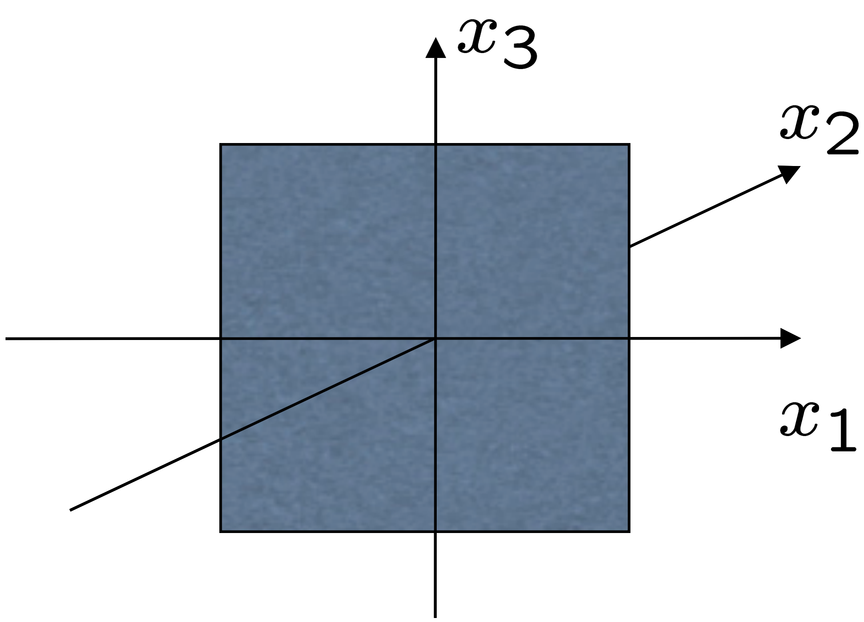
- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

- Brute force search



$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{b}$$

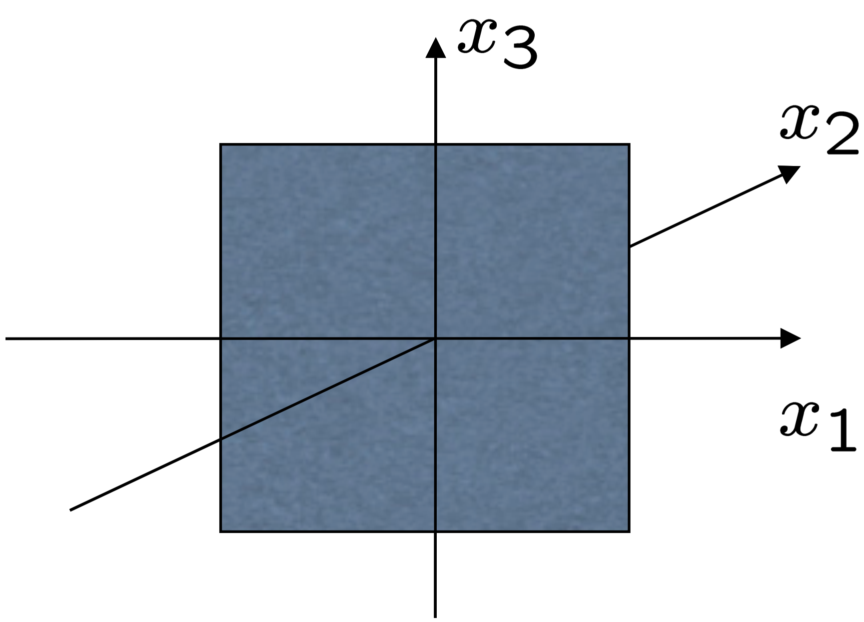
- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

- Brute force search



$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ 0 \\ x_3 \end{pmatrix} = \mathbf{b}$$

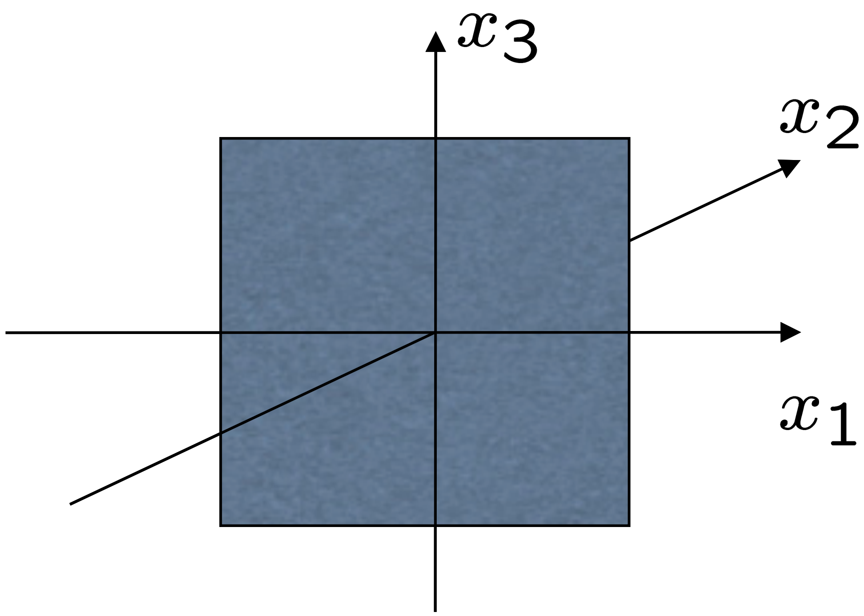
- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

- Brute force search



$$\begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \mathbf{b}$$

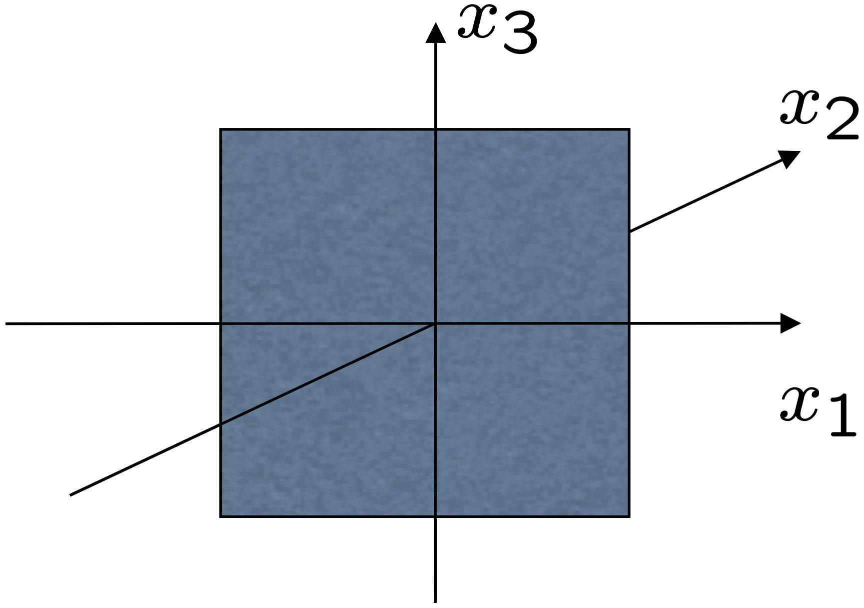
- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

- Brute force search



$$\begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \mathbf{b}$$

$$\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix}^{-1} \cdot \mathbf{b}$$

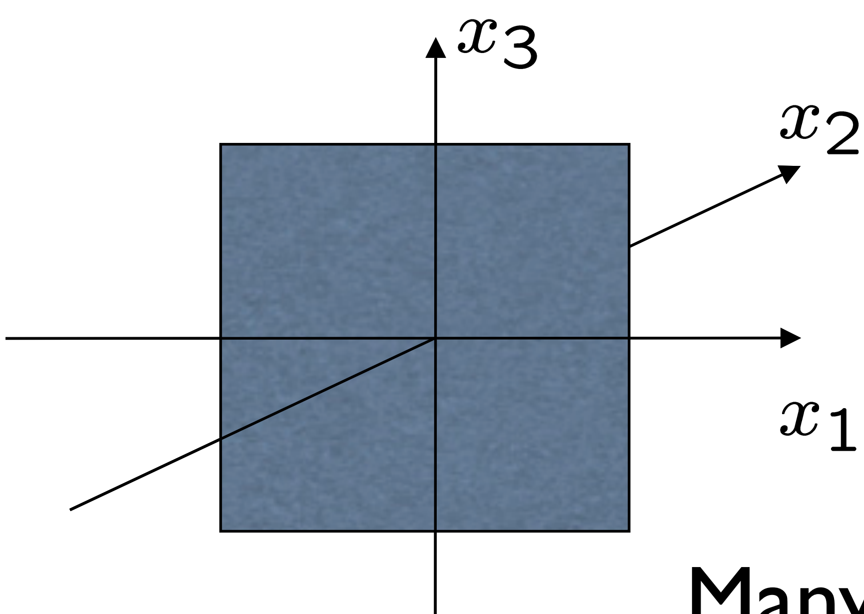
- **Theorem** *(Davies et al, Natarajan)*

  It is NP-hard!

- Are there other more efficient alternatives ?

Ideal Sparse Approximation

- Brute force search



$$\begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \mathbf{b}$$

$$\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix}^{-1} \cdot \mathbf{b}$$

Many n-tuples to try!

- **Theorem** *(Davies et al, Natarajan)*

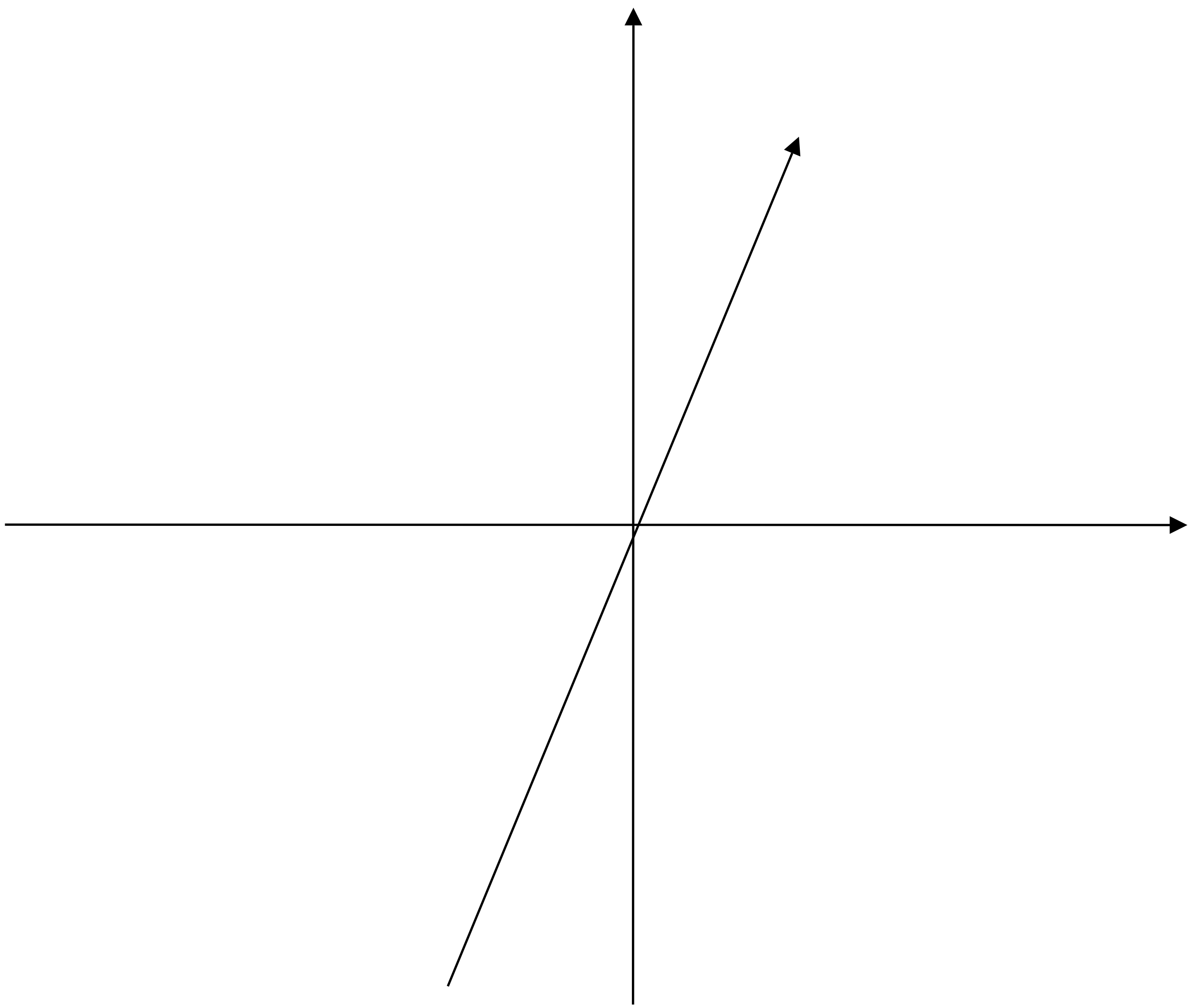
  It is NP-hard!

- Are there other more efficient alternatives ?

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry



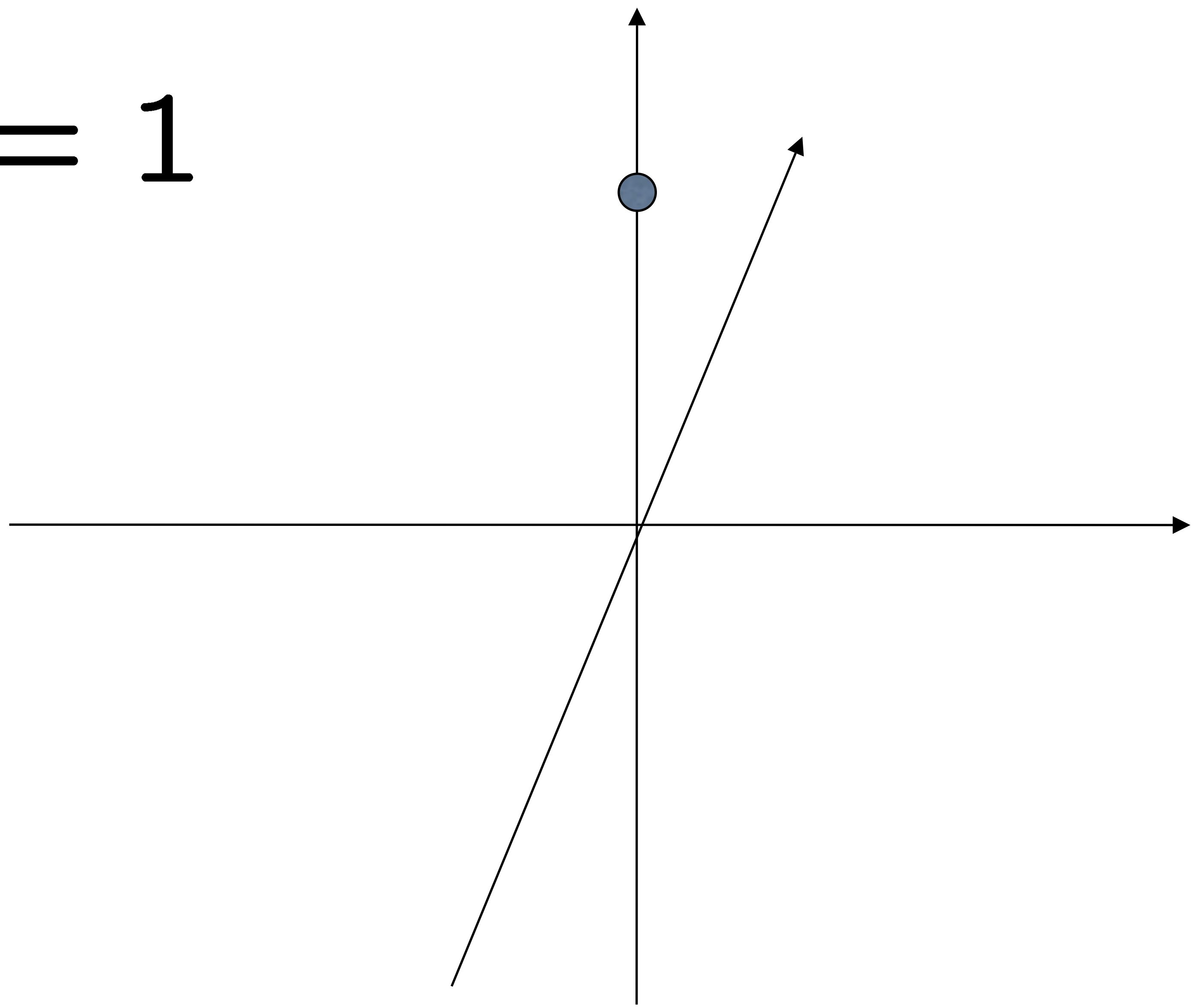
IRISA

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$\|x_0\|_0 = 1$$



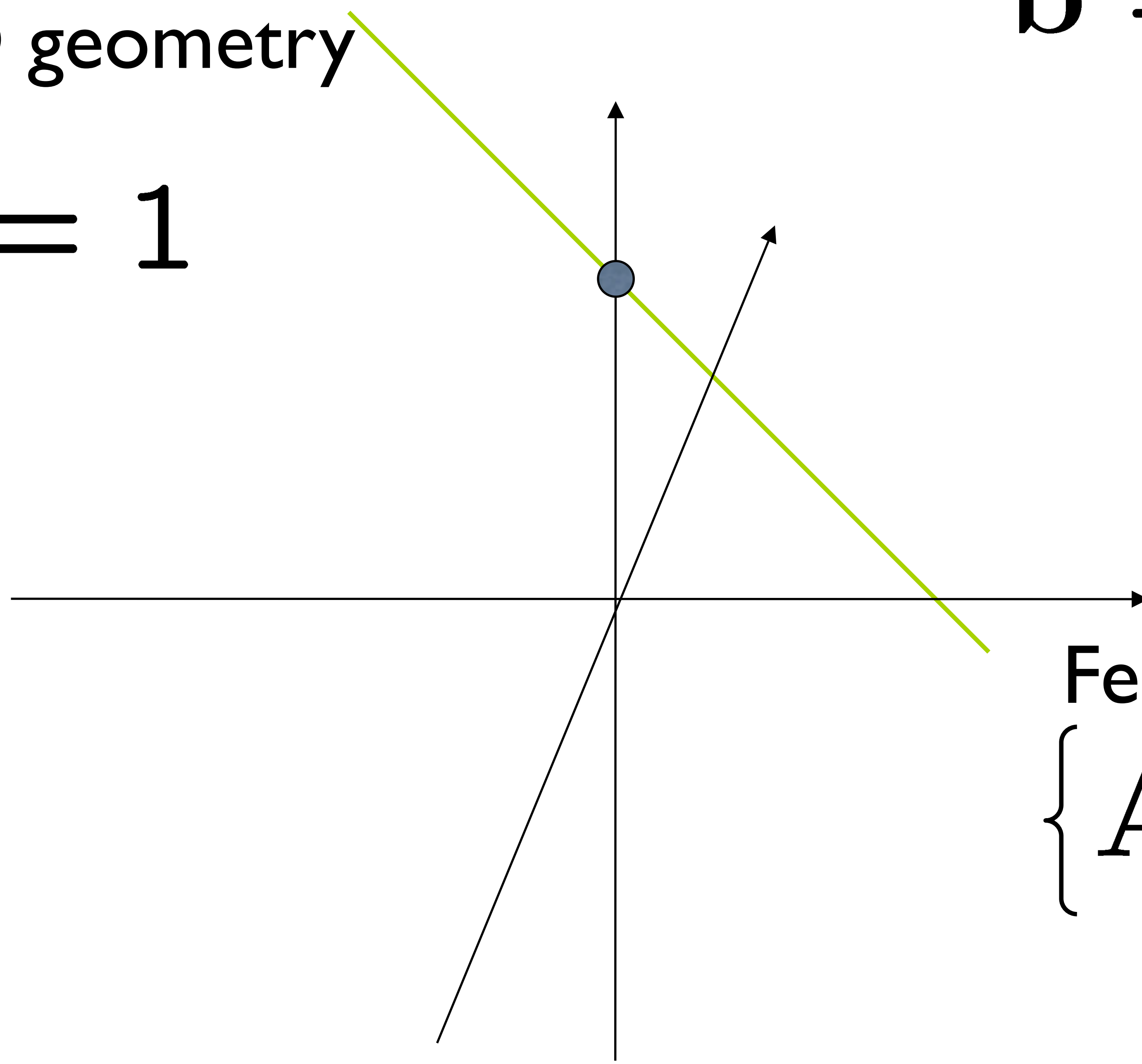
IRISA

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$\|x_0\|_0 = 1$$



Feasible solutions

$$\{\mathbf{A}x = \mathbf{A}x_0\}$$

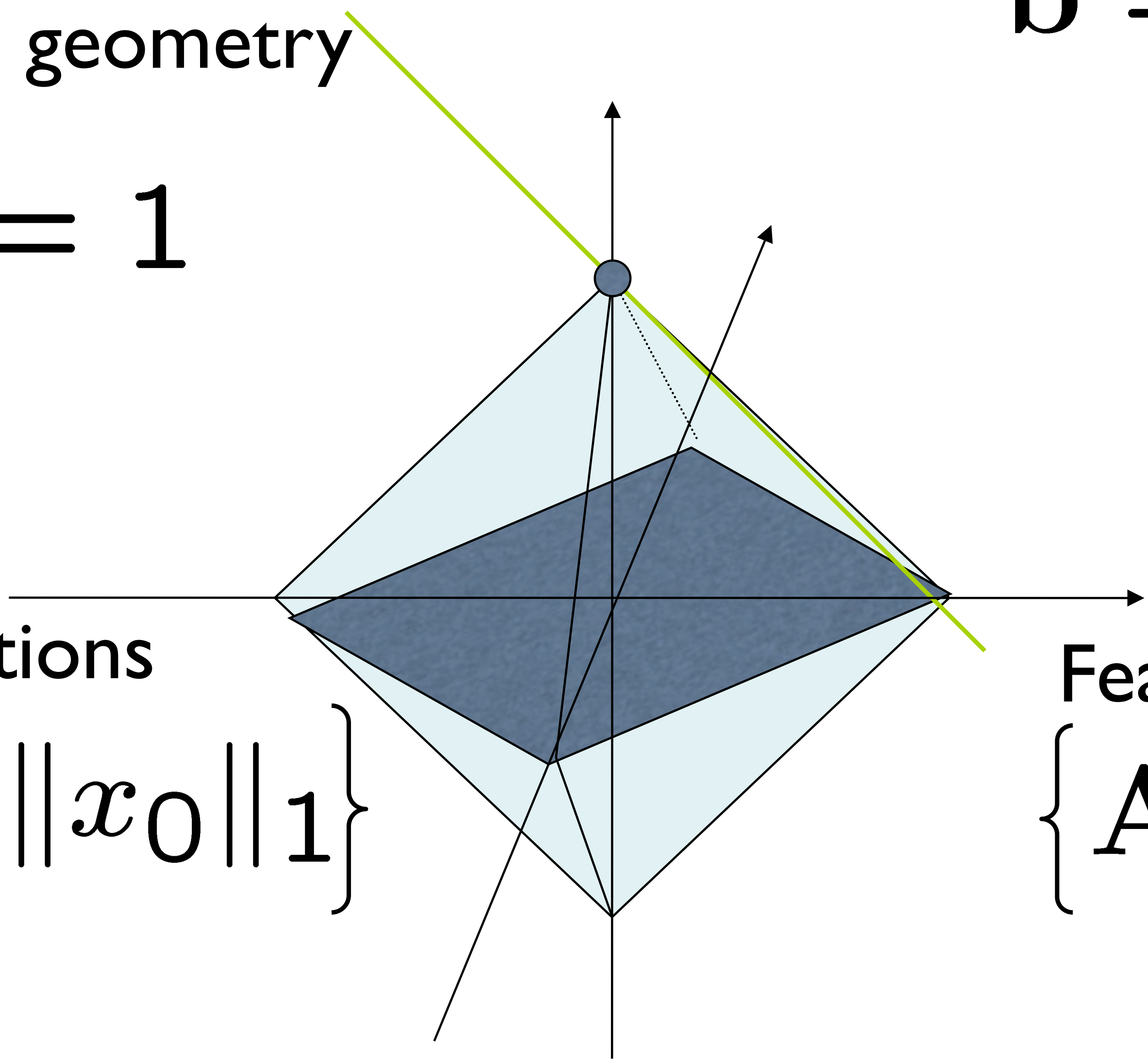
IRISA

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$\|x_0\|_0 = 1$$



Sparse solutions

$$\left\{\|x\|_1 \leq \|x_0\|_1\right\}$$

Feasible solutions

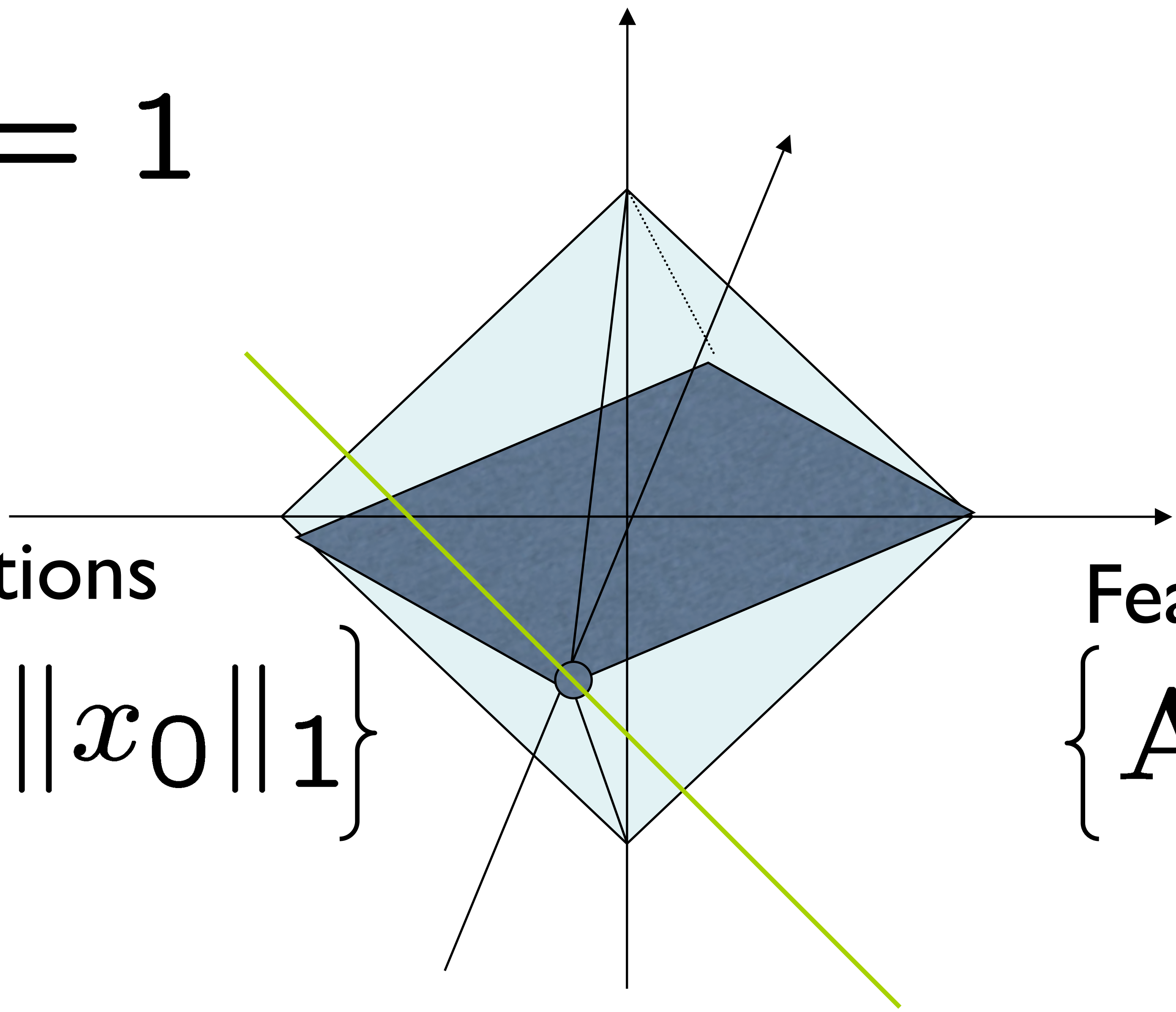
$$\left\{\mathbf{A}x = \mathbf{A}x_0\right\}$$

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$\|x_0\|_0 = 1$$



Sparse solutions

$$\left\{\|x\|_1 \leq \|x_0\|_1\right\}$$

Feasible solutions

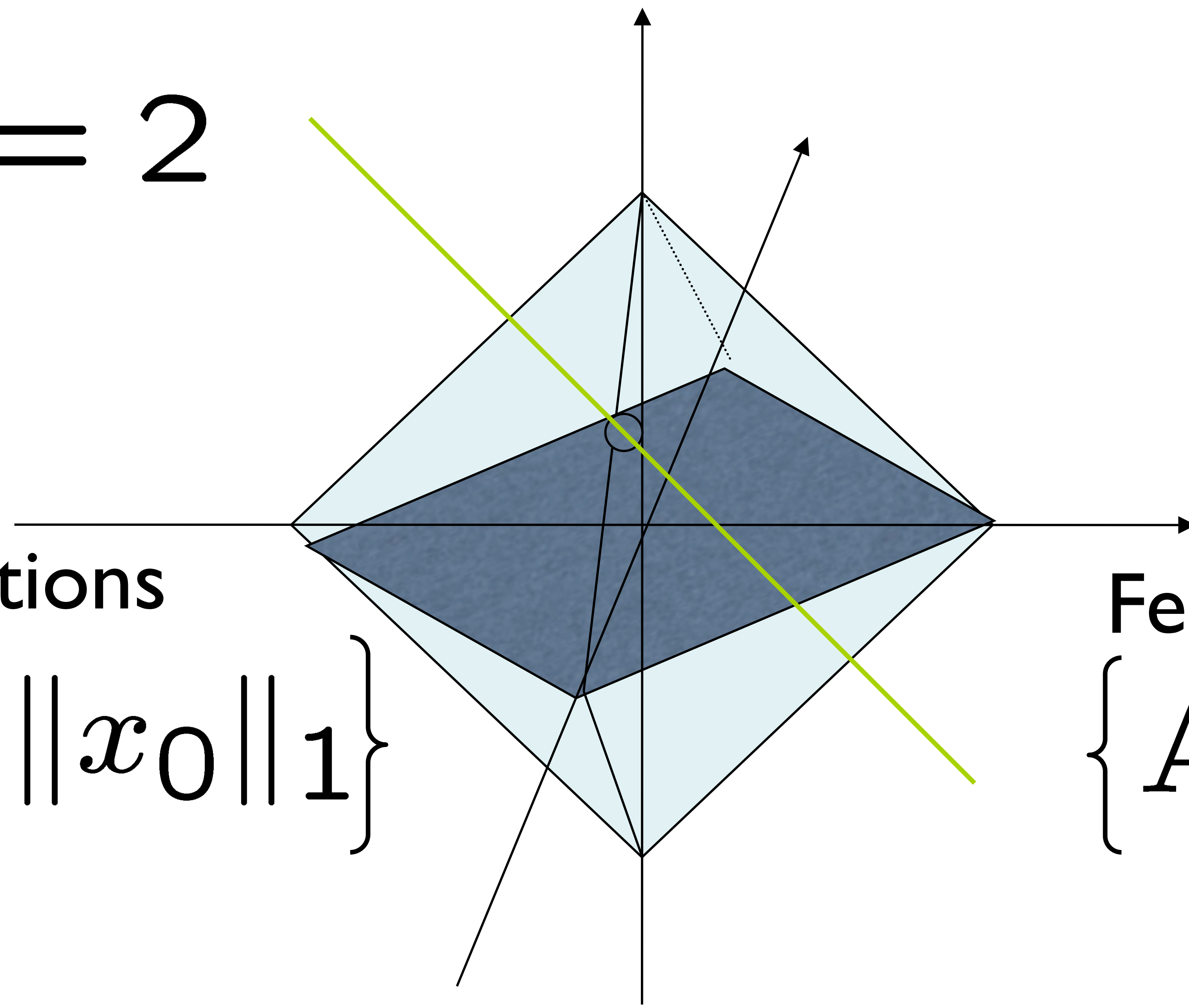
$$\left\{\mathbf{A}x = \mathbf{A}x_0\right\}$$

Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$\|x_0\|_0 = 2$$



Sparse solutions

$$\left\{\|x\|_1 \leq \|x_0\|_1\right\}$$

Feasible solutions

$$\left\{\mathbf{A}x = \mathbf{A}x_0\right\}$$

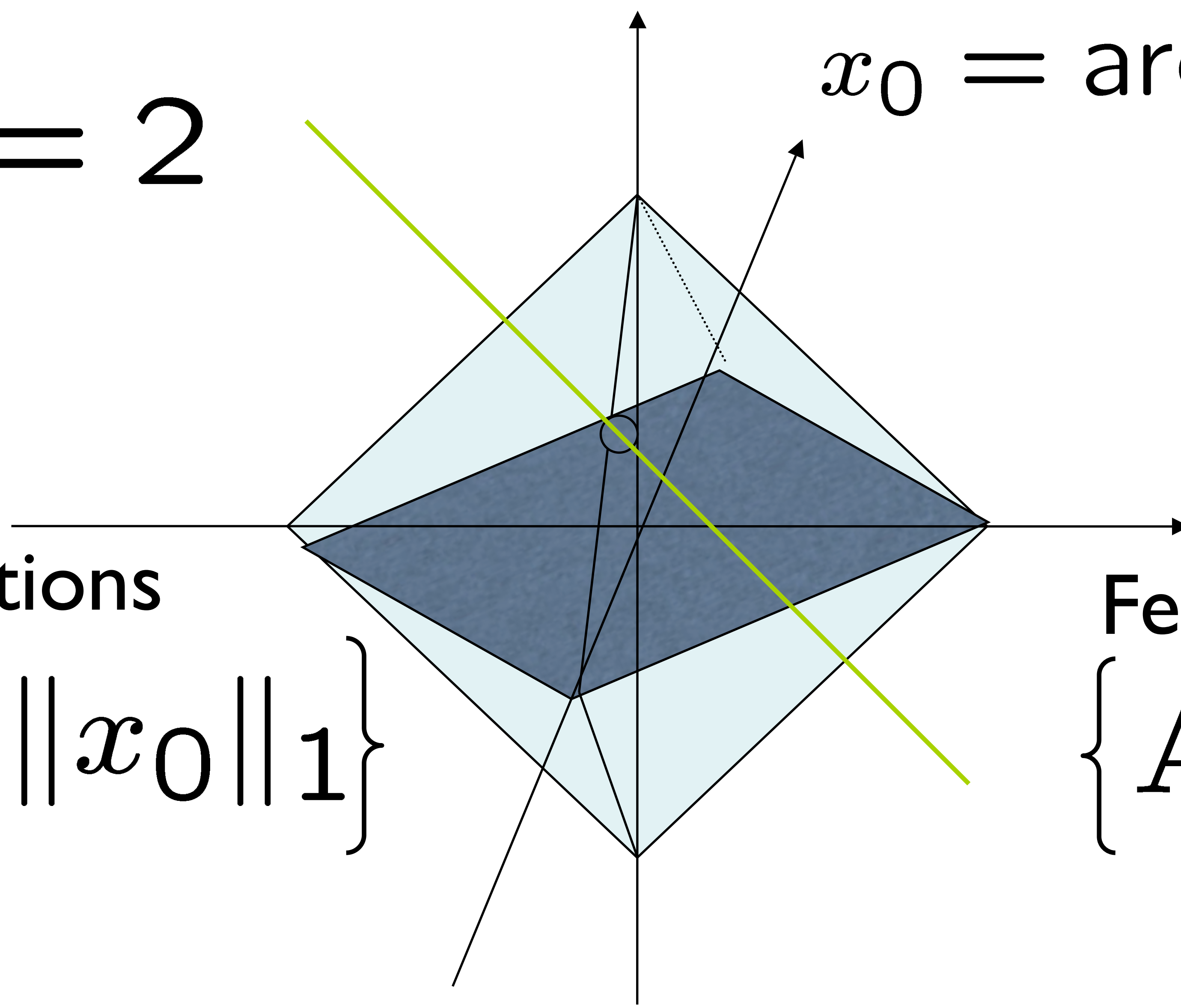
Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$x_0 = \arg \min_{x:\mathbf{A}x=\mathbf{b}} \|x\|_1$$

$$\|x_0\|_0 = 2$$



Sparse solutions

$$\left\{\|x\|_1 \leq \|x_0\|_1\right\}$$

Feasible solutions

$$\left\{\mathbf{A}x = \mathbf{A}x_0\right\}$$

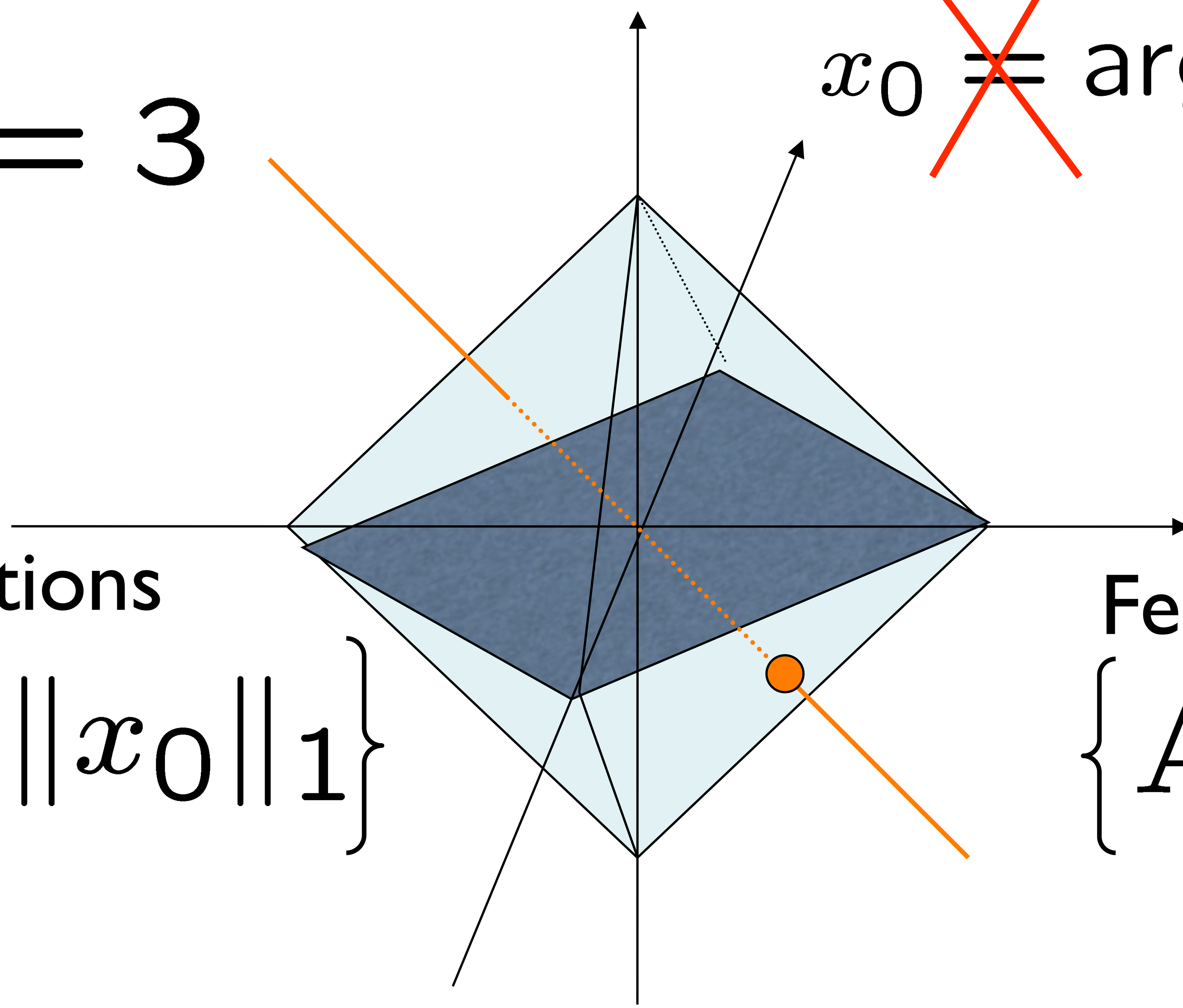
Recovery with Basis Pursuit

$$\mathbf{b} = \mathbf{A}x_0$$

- Some 3D geometry

$$x_0 \cancel{=} \arg\min_{x:\mathbf{A}x=\mathbf{b}} \|x\|_1$$

$$\|x_0\|_0 = 3$$



Sparse solutions

$$\left\{\|x\|_1 \leq \|x_0\|_1\right\}$$

Feasible solutions

$$\left\{\mathbf{A}x = \mathbf{A}x_0\right\}$$

Equivalence between L0, L1, OMP

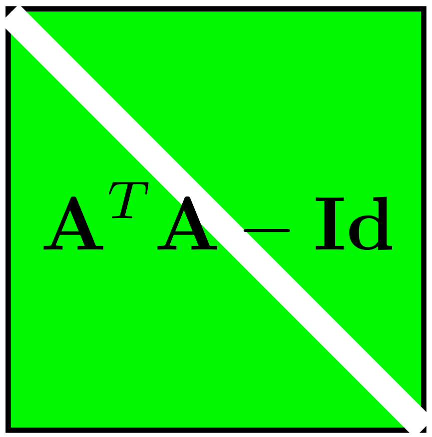
- **Theorem** : assume that $\mathbf{b} = \mathbf{A}x_0$
  - ✦ if $\|x_0\|_0 \leq k_0(\mathbf{A})$ then $x_0 = x_0^\star$
  - ✦ if $\|x_0\|_0 \leq k_1(\mathbf{A})$ then $x_0 = x_1^\star$

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

- *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*

IRISA

State of the art tools to estimate $k_p(\mathbf{A})$

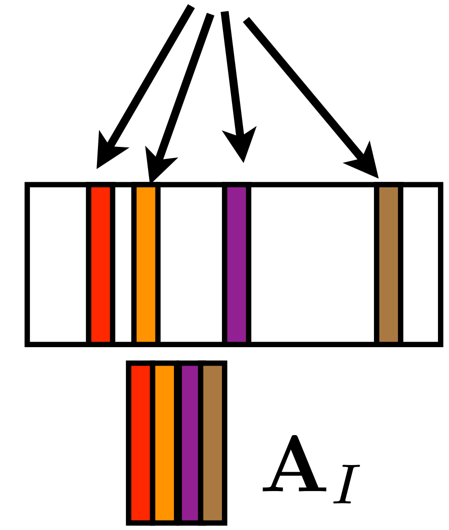


$\mathbf{A}$

N unit columns

$\|\mathbf{A}_n\|_2 = 1$

$n \in I, \sharp I \leq k$



max over N(N-1) entries

$$\mu = \mu(\mathbf{A}) := \max_{k \neq l} |\langle \mathbf{A}_k, \mathbf{A}_l \rangle|$$

max over $\frac{N!}{k!(N-k)!}$ subsets *I*

$$\delta_k := \sup_{\sharp I \leq k,\ c \in \mathbb{R}^k} \left| \frac{\|\mathbf{A}_I c\|_2^2}{\|c\|_2^2} - 1 \right|$$

$$\widehat{k}(\mathbf{A}) = (1 + 1/\mu)/2$$

L0-recovery (identifiability) $\delta_{2k_0} < 1$

L1-recovery (identification) $\delta_{2k_1} < \sqrt{2} - 1$

(Cumulative) coherence
Low cost
"Coarse / pessimistic"

Common beliefs

Isometry constants
Hard to compute
"Almost sharp" ?

Compressed sensing

Compressed Sensing

- MRI from incomplete measures

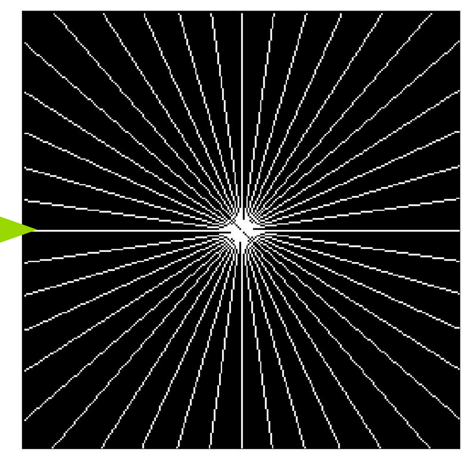
[from Candès, Romberg & Tao]

Not observed

Data

Lossy measurement
= tomography

Measured data
(FFT minus lost data)



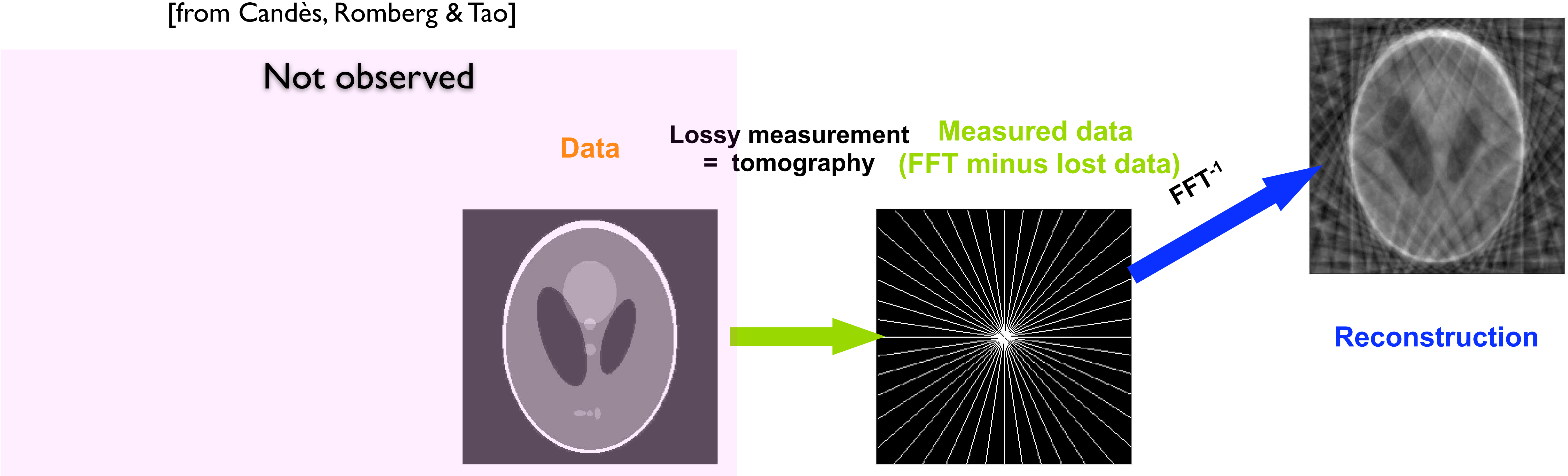
$$z = \mathbf{K}y$$

IRISA

Compressed Sensing

- MRI from incomplete measures

[from Candès, Romberg & Tao]

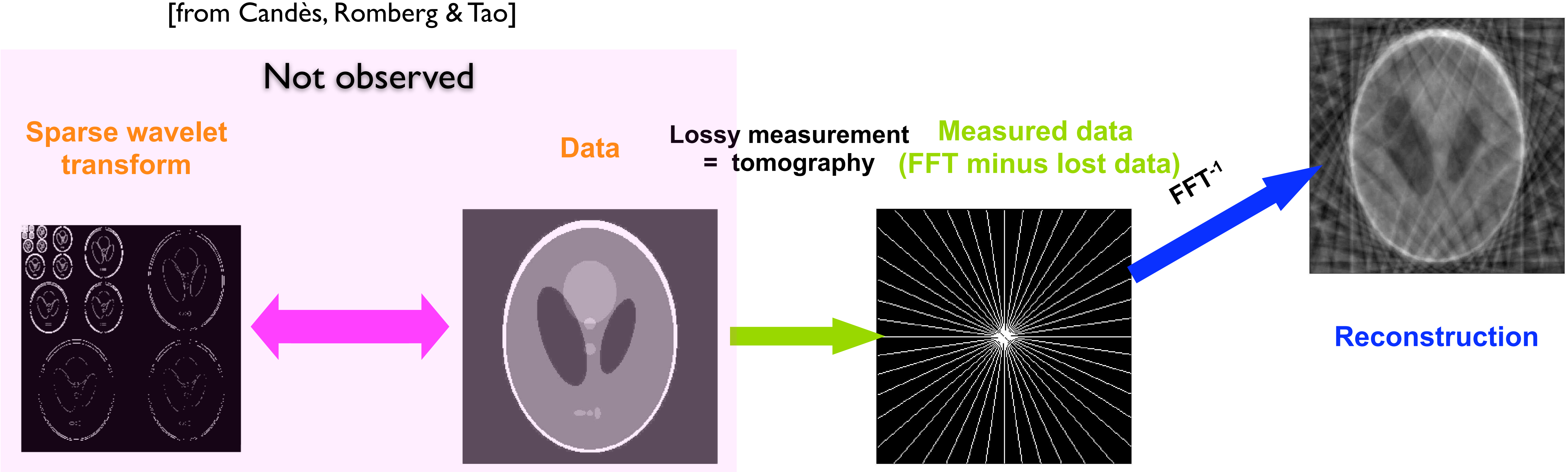


$$z = \mathbf{K} y$$

Compressed Sensing

- MRI from incomplete measures

[from Candès, Romberg & Tao]

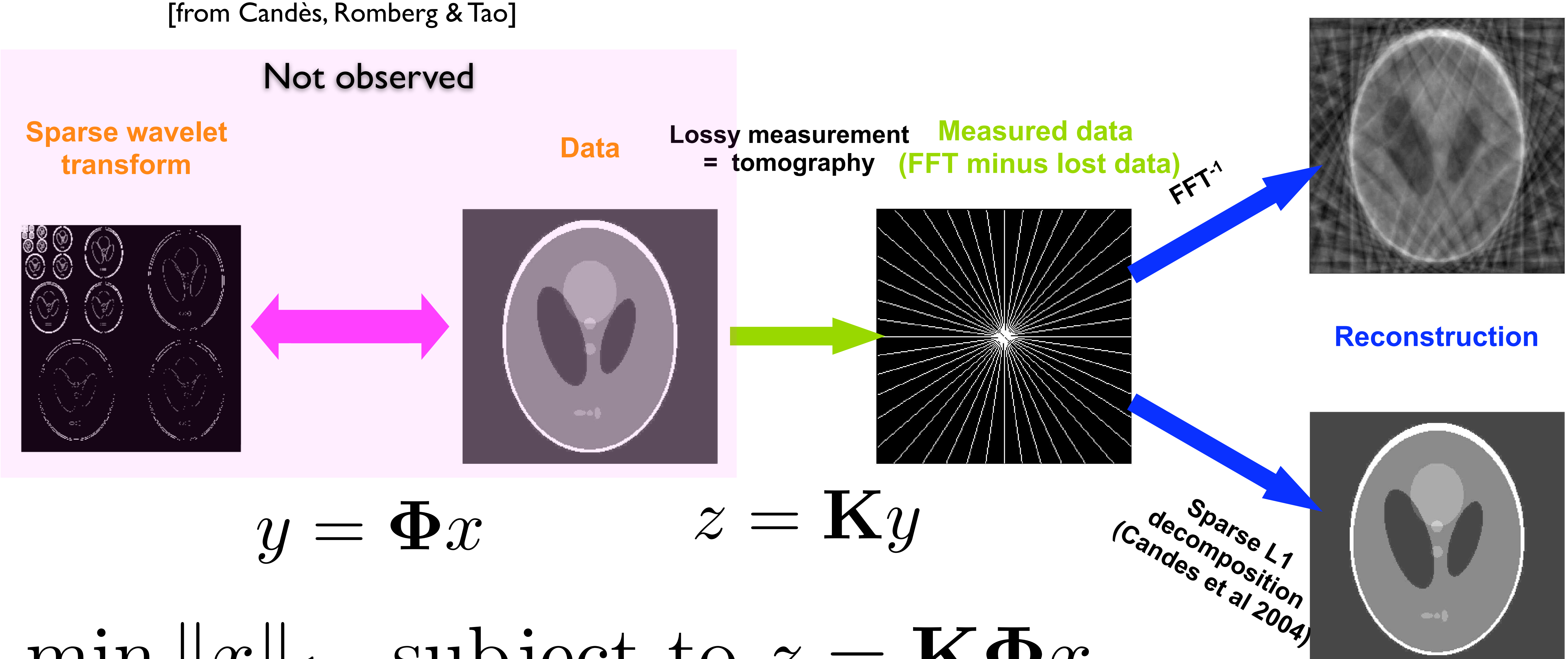


$$y = \mathbf{\Phi} x \qquad z = \mathbf{K} y$$

Compressed Sensing

- MRI from incomplete measures

[from Candès, Romberg & Tao]

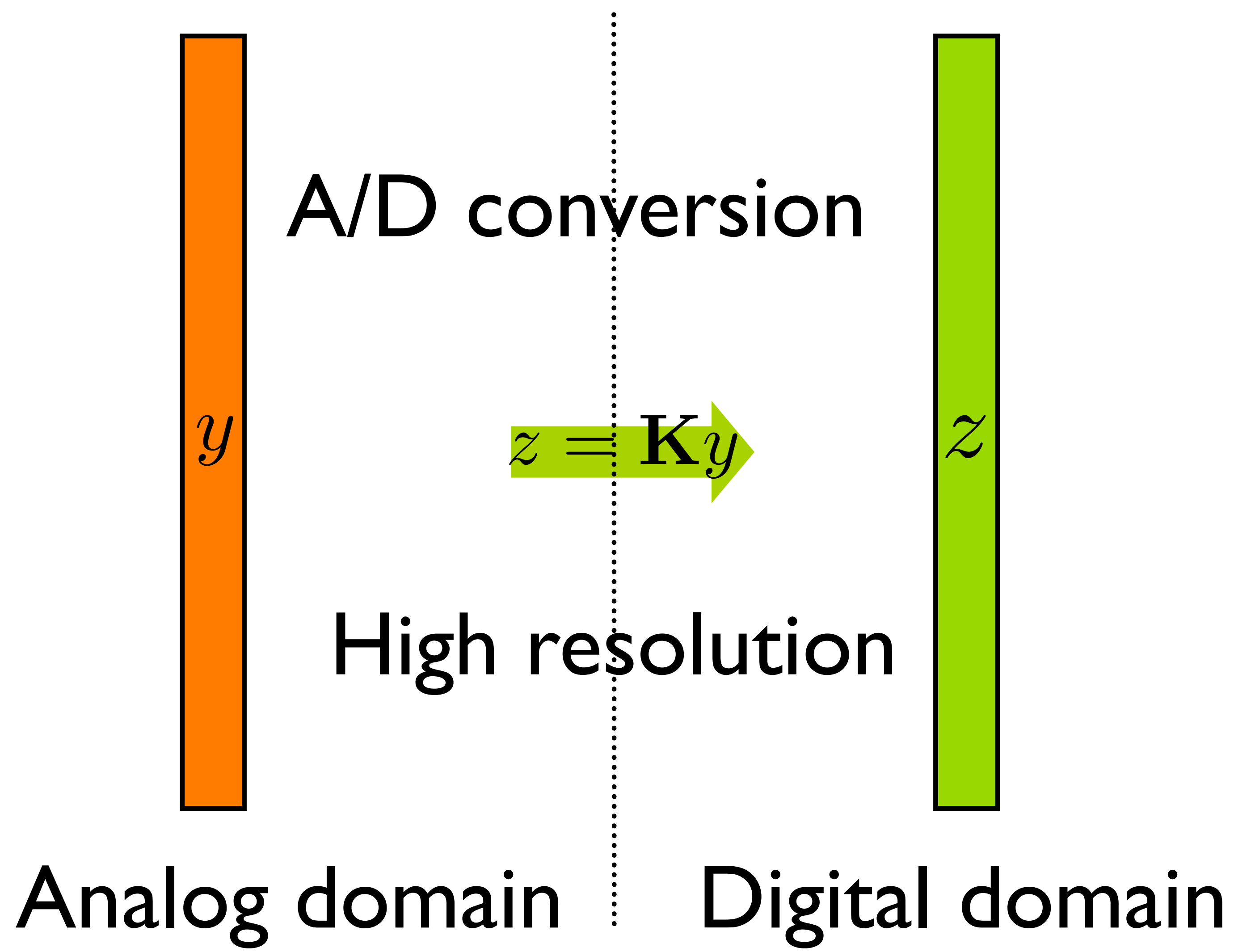


$$y = \mathbf{\Phi} x \qquad z = \mathbf{K} y$$

$$\min \|x\|_1, \text{ subject to } z = \mathbf{K\Phi} x$$

Classical Shannon Sampling

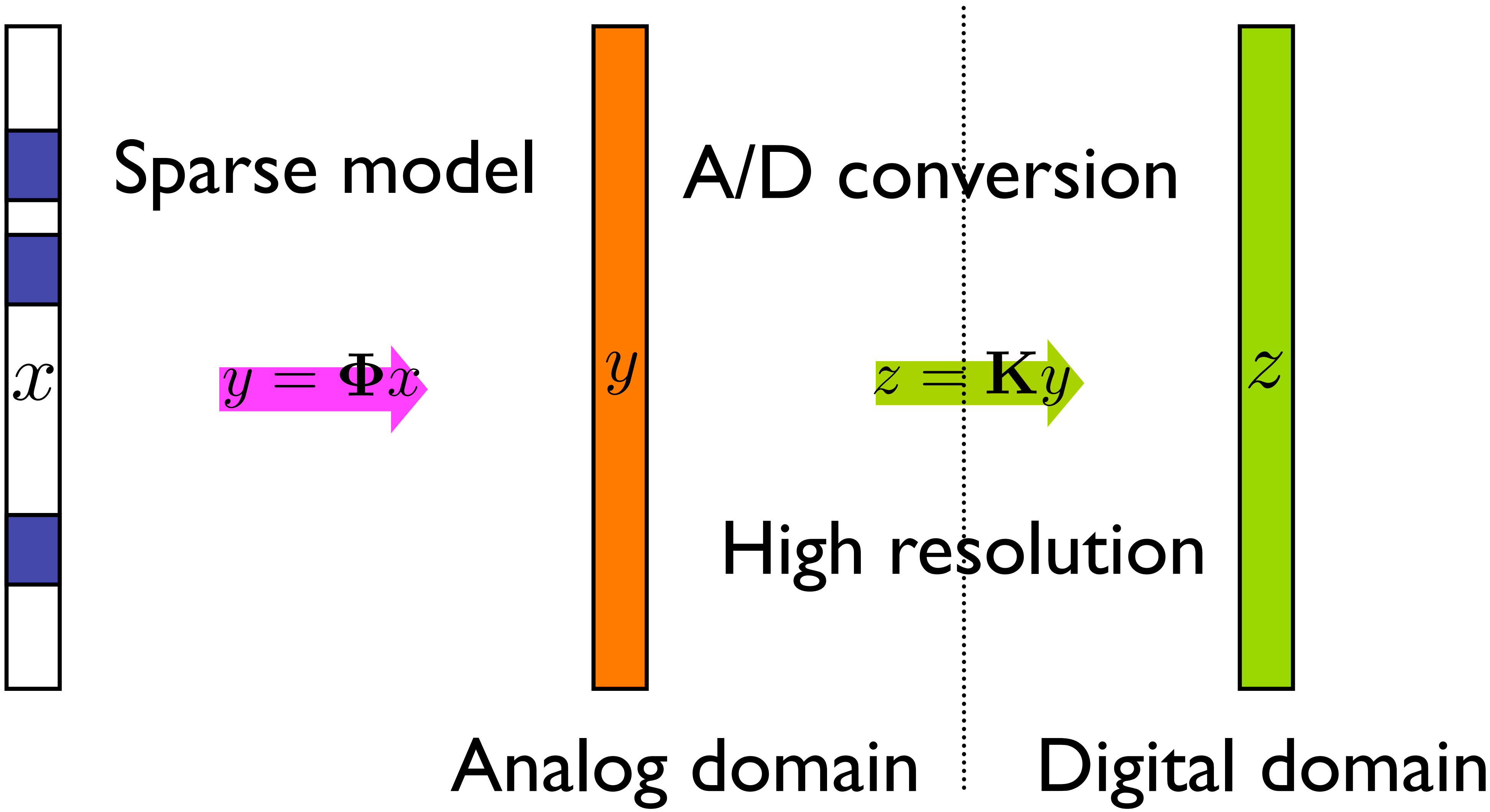
- « Sample first, think and compress afterwards »



$\mathbf{K}$ linear measures (samples)

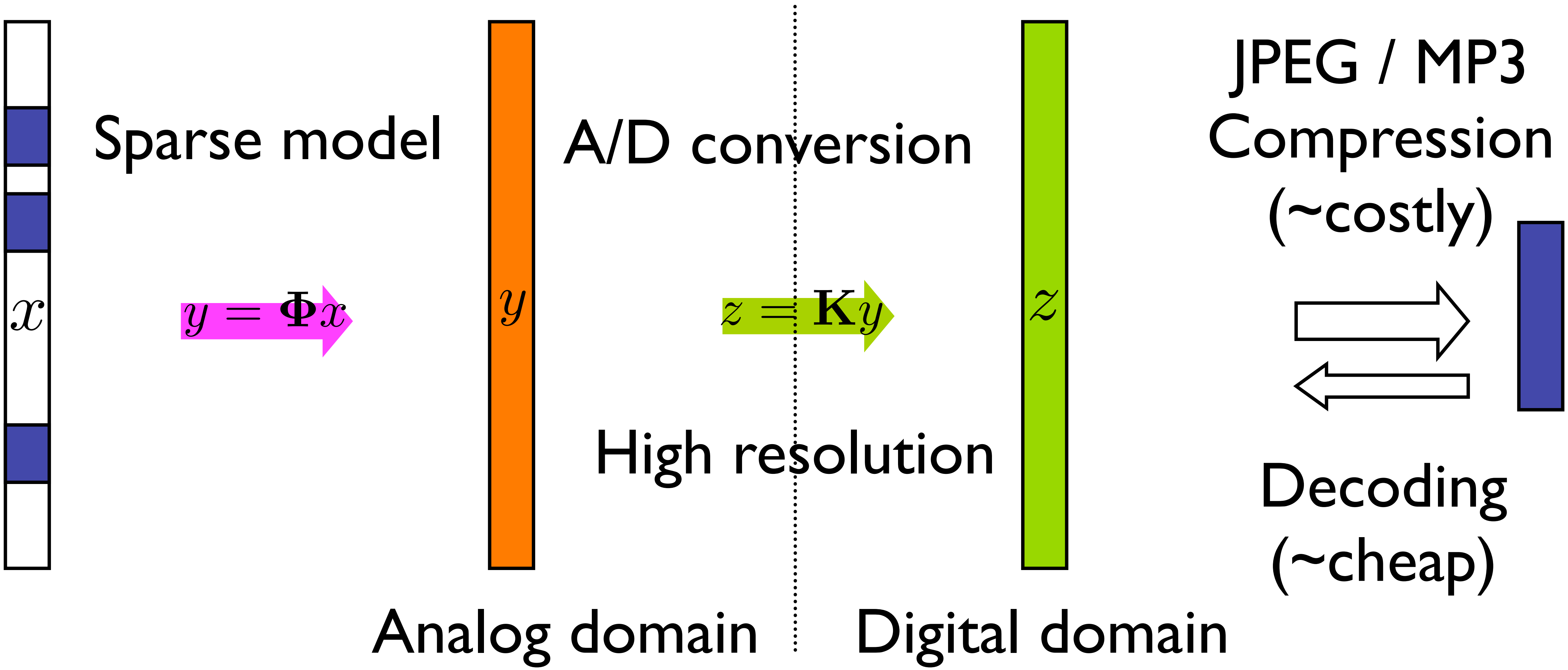
Classical Shannon Sampling

- « Sample first, think and compress afterwards »



Classical Shannon Sampling

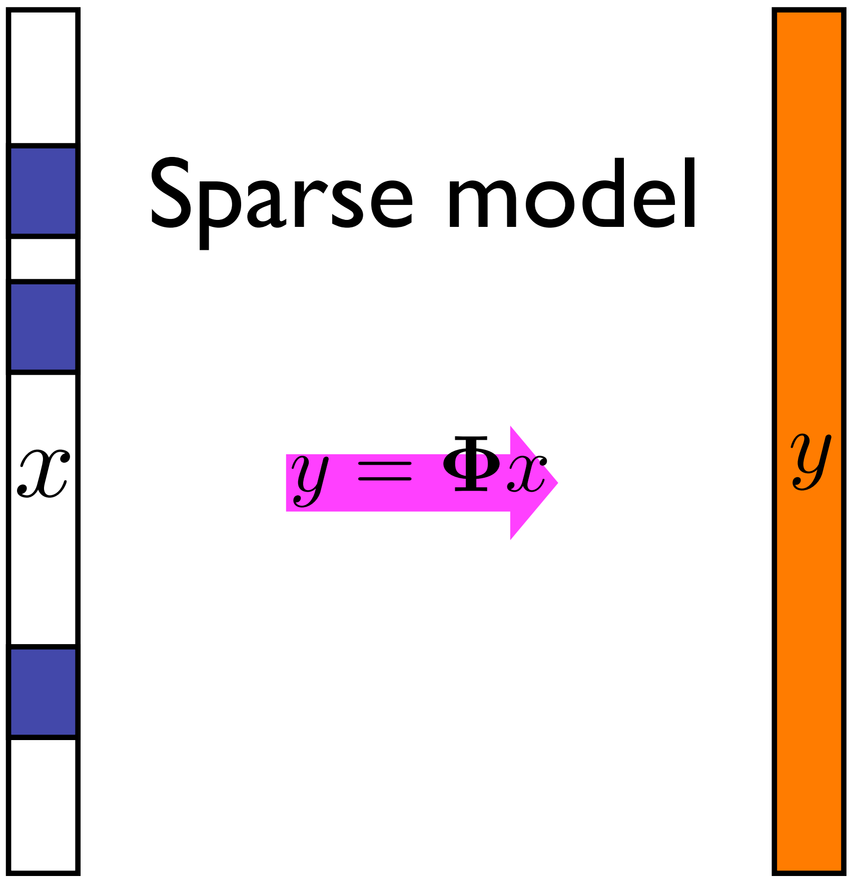
- « Sample first, think and compress afterwards »



IRISA

Compressed Sensing

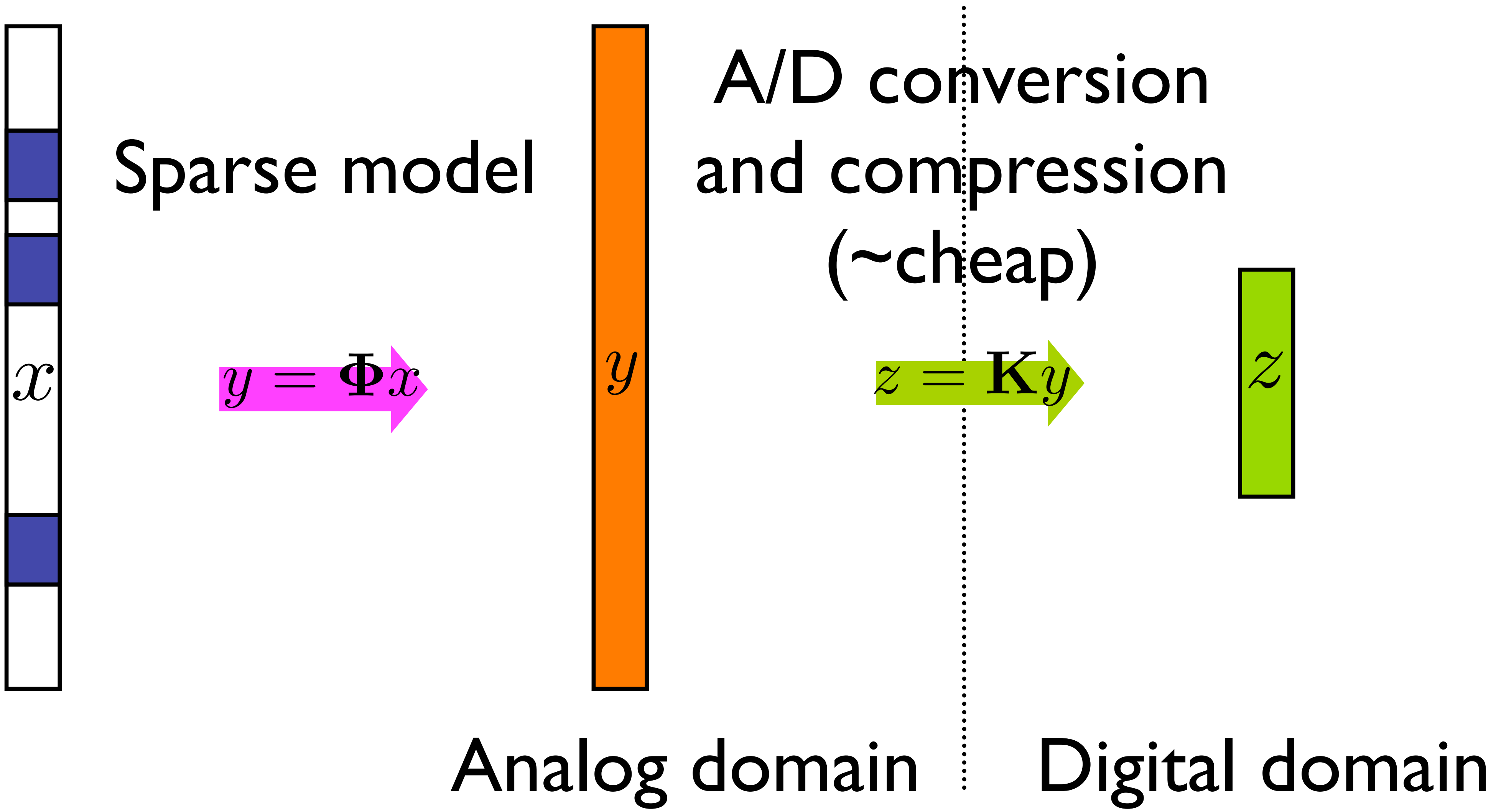
- First model the data, then sample & compress



Analog domain | Digital domain

Compressed Sensing

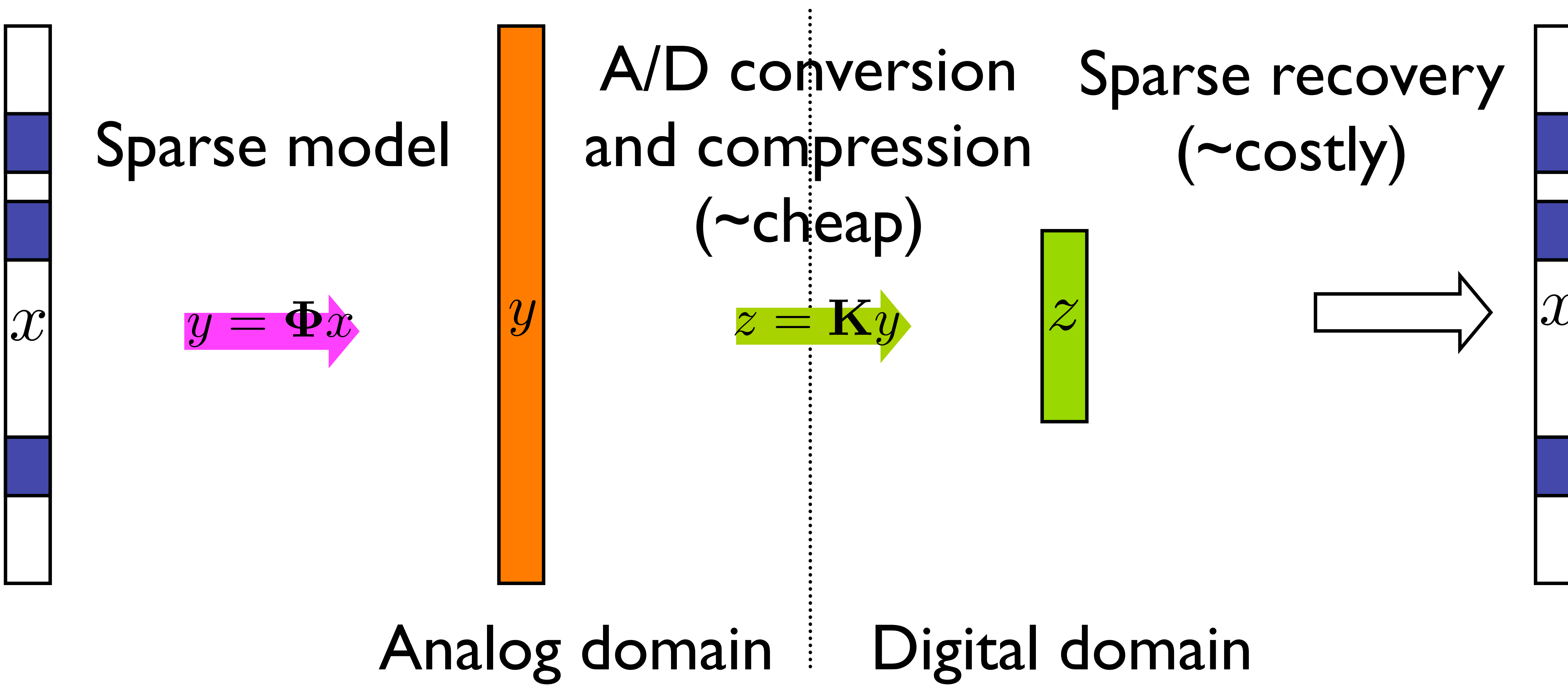
- First model the data, then sample & compress



IRISA

Compressed Sensing

- First model the data, then sample & compress



$$\min \|x\|_1, \ \text{subject to } z = \mathbf{K}\mathbf{\Phi} x$$

Partial conclusions

- Sparsity helps solve ill-posed inverse problems (more unknowns than equations)
- Sparse approximation is NP-hard but efficient sub-optimal algorithms (*pursuits*) exist
- If there is a sufficiently sparse solution, it is recovered by many of the heuristic algorithms
- This is the fundamental basis underlying the development of compressed sensing

Structure of the course

- Part I: Overview
- Part II: Algorithms, complexity & convergence
  - Lp minimization
  - Greedy Algorithms
- Part III: Recovery, stability, robustness
  - Null Space Properties and Lp minimization
  - Exact Recovery Condition and greedy algorithms
  - Restricted Isometry Constants, stability and robustness
- Part IV: Dictionaries, Random Matrices and Compressed Sensing