



## Inverse problems and sparse models (2/2)

Rémi Gribonval

INRIA Rennes - Bretagne Atlantique, France

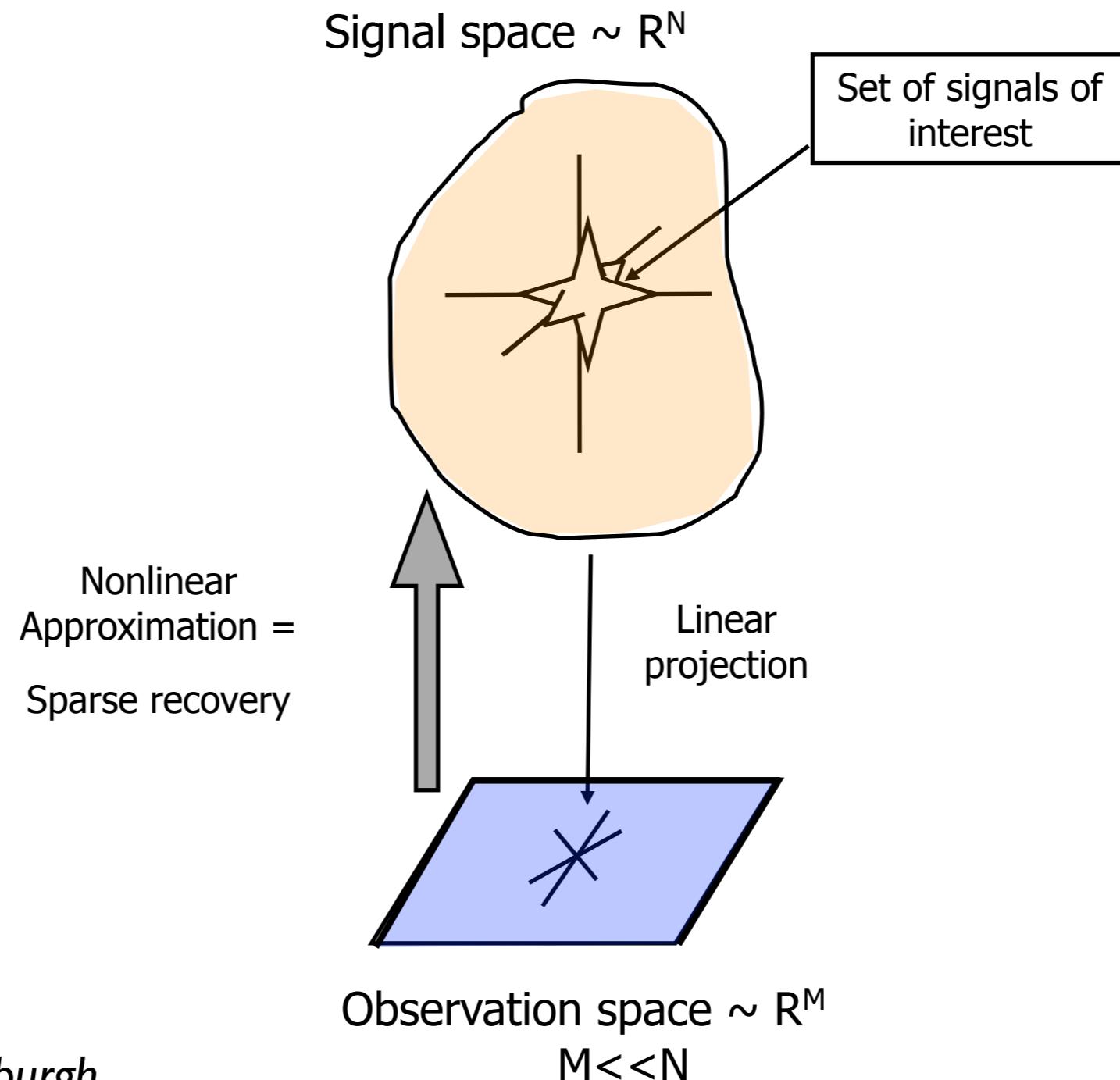
[remi.gribonval@inria.fr](mailto:remi.gribonval@inria.fr)

with L. Borup, M. Davies, R. Figueras, M. Nielsen, P. Vandergheynst

# Structure of the tutorial

- Session 1:
  - ✓ Introduction to inverse problems & sparse models
  - ✓ Sparse recovery:
  - ✓ Pursuits
- Session 2:
  - ✓ L1 recovery guarantees
    - ◆ Coherence
    - ◆ Restricted Isometry Property
    - ◆ Null Space Properties

# Inverse problems



Courtesy: M. Davies, U. Edinburgh

# Exact recovery conditions for L1

# Proved Equivalence between L0 and L1

- “Empty” theorem : assume that  $\mathbf{b} = \mathbf{A}x_0$ 
  - ✓ if  $\|x_0\|_0 \leq k_0(\mathbf{A})$  then  $x_0 = x_0^*$
  - ✓ if  $\|x_0\|_0 \leq k_1(\mathbf{A})$   $x_0 = x_1^*$
  - ✓ where  $x_p^* = \arg \min_{\mathbf{A}x=\mathbf{A}x_0} \|x\|_p$
- Content = estimation of  $k_0(\mathbf{A})$  and  $k_1(\mathbf{A})$ 
  - ✓ *Donoho & Huo 2001 :* pair of bases, coherence
  - ✓ *Donoho & Elad 2003, Gribonval & Nielsen 2003 :* dictionary, coherence
  - ✓ *Candes, Romberg, Tao 2004 : random dictionaries,* restricted isometry constants
  - ✓ *Tropp 2004 : idem for Orthonormal Matching Pursuit,* cumulative coherence

# Recovery conditions based on coherence

- **Convention:** normalized columns  $\|\mathbf{a}_i\|_2 = 1$
- **Definition:** coherence of dictionary

$$\mu(\mathbf{A}) := \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|$$

- **Theorem:**
  - ✓ Assume that
  - ✓ Then
    - ◆  $x$  minimizes the L0 and L1 norm among all solutions  $x'$  to the linear inverse problem  $\mathbf{A}x' = \mathbf{A}x$
    - ◆ k steps of **OMP** performed on  $\mathbf{b} = \mathbf{A}x$  recover  $x$

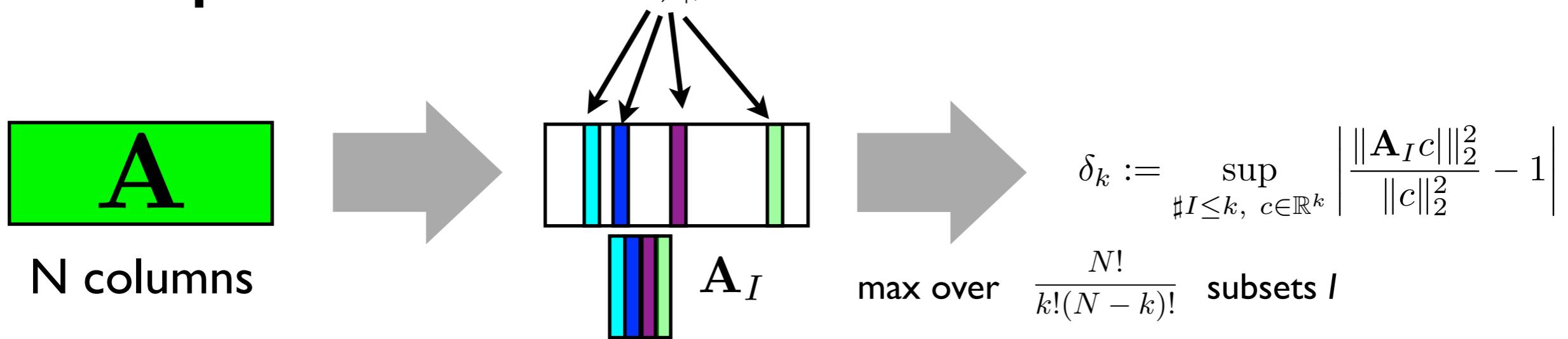
$$k = \|x\|_0 < \frac{1}{2}(1 + 1/\mu)$$

# Restricted Isometry Constants (RIC)

- **Definition:** smallest  $\delta_k$  such that for any k-sparse  $x$

$$1 - \delta_k \leq \frac{\|\mathbf{A}x\|_2^2}{\|x\|_2^2} \leq 1 + \delta_k$$

- **Computation ?**  $n \in I, \#I \leq k$



- **NP-complete** [Kloiran & Zouzias 2011, Tillmann & Pfetsch 2012, Bandeira & al 2012]
- Can be estimated for certain random matrices

# Recovery conditions based on RIC

- **Definition:** RIC of dictionary of order  $2k$

- ✓ for any  $2k$ -sparse vector  $z$

$$(1 - \delta_{2k})\|z\|_2^2 \leq \|\mathbf{A}z\|_2^2 \leq (1 + \delta_{2k})\|z\|_2^2$$

- **Theorem:**

- ✓ Assume that

$$\|x\|_0 \leq k \text{ and } \delta_{2k} < \sqrt{2} - 1 \approx 0.414$$

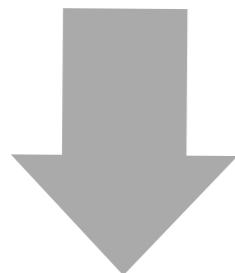
**Restricted Isometry Property (RIP)**

- ✓ Then

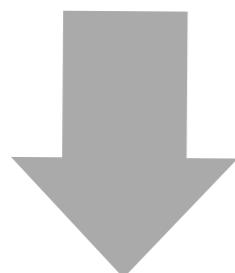
- ◆  $x$  minimizes the L0 and L1 norm among all solutions  $x'$  to the linear inverse problem  $\mathbf{A}x' = \mathbf{A}x$

In fact ... we will see that

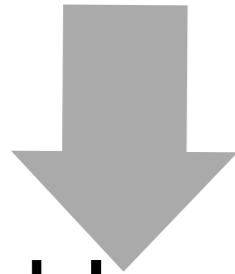
Incoherence



RIP( $k$ )



NSP( $k$ )



$$k < \frac{1}{2}(1 + 1/\mu)$$

$$\delta_{2k} < \delta$$

$$\|z_{I_k}\|_1 < \|z\|_1/2 \quad \forall z \in \text{Ker } \mathbf{A} \setminus \{0\}$$

**Null Space Property (NSP)**

L1 recovery for all  $k$ -sparse vectors

# Null space

- Null space = kernel

$$z \in \text{Ker } \mathbf{A} \Leftrightarrow \mathbf{A}z = 0$$

- Particular solution vs general solution
  - ✓ particular solution

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

- ✓ general solution

$$\mathbf{A}x' = \mathbf{b} \Leftrightarrow x' - x \in \text{Ker } \mathbf{A}$$

# L1 recovery for a given support /

- **Notation**

- ✓ restriction of vector  $z$  to index set  $I$        $z_I = (z_i)_{i \in I}$

- **Theorem 1** [Donoho & Huo 2001 for L1, G. & Nielsen 2003 for  $L_p$  & more]

- ✓ Assume the «Null Space Property» for a given support  $I$

NSP( $I$ )

$$\|z_I\|_1 < \|z\|_1/2 \quad \forall z \in \text{Ker } \mathbf{A} \setminus \{0\}$$

- ✓ Then: for every vector  $x$  **supported in  $I$**

$$x = \arg \min_{\tilde{x}} \|\tilde{x}\|_1 \text{ s.t. } \mathbf{A}\tilde{x} = \mathbf{A}x$$

- ✓ **Sharpness:** if NSP( $I$ ) fails there is **at least one failing vector  $x$**  supported in  $I$

# NSP( $I$ ) is necessary

- **Assume** there exists  $z \in \text{Ker}\mathbf{A} \setminus \{0\}$  with  $\|z_I\|_1 > \|z\|_1/2$
- Define  $\mathbf{b} := \mathbf{A}z_I = \mathbf{A}(-z_{I^c})$
- Observe that  $\|z_{I^c}\|_1 = \|z\|_1 - \|z_I\|_1 < \|z_I\|_1$
- The vector  $z_I$  is supported in  $I$  but is ***not*** the minimum L1 norm representation of  $\mathbf{b}$

# NSP( $I$ ) is sufficient

- Consider  $x$  with support set  $I$  and  $x' \neq x$  with  $\mathbf{A}x' = \mathbf{A}x$
- Denote  $z := x' - x \in \text{Ker}\mathbf{A} \setminus \{0\}$ 
  - ✓ By NSP( $I$ ) we have  $\|z_{I^c}\|_1 > \|z_I\|_1$
  - ✓ Moreover we have

$$\|x'\|_1 = \|x + z\|_1 = \|(x + z)_I\|_1 + \|(x + z)_{I^c}\|_1$$

# NSP( $I$ ) is sufficient

- Consider  $x$  with support set  $I$  and  $x' \neq x$  with  $\mathbf{A}x' = \mathbf{A}x$
- Denote  $z := x' - x \in \text{Ker}\mathbf{A} \setminus \{0\}$ 
  - ✓ By NSP( $I$ ) we have  $\|z_{I^c}\|_1 > \|z_I\|_1$
  - ✓ Moreover we have

$$\begin{aligned}\|x'\|_1 &= \|x + z\|_1 = \|(x + z)_I\|_1 + \|(x + z)_{I^c}\|_1 \\ &= \|x + z_I\|_1 + \|z_{I^c}\|_1\end{aligned}$$

# NSP( $I$ ) is sufficient

- Consider  $x$  with support set  $I$  and  $x' \neq x$  with  $\mathbf{A}x' = \mathbf{A}x$
- Denote  $z := x' - x \in \text{Ker}\mathbf{A} \setminus \{0\}$ 
  - ✓ By NSP( $I$ ) we have  $\|z_{I^c}\|_1 > \|z_I\|_1$
  - ✓ Moreover we have

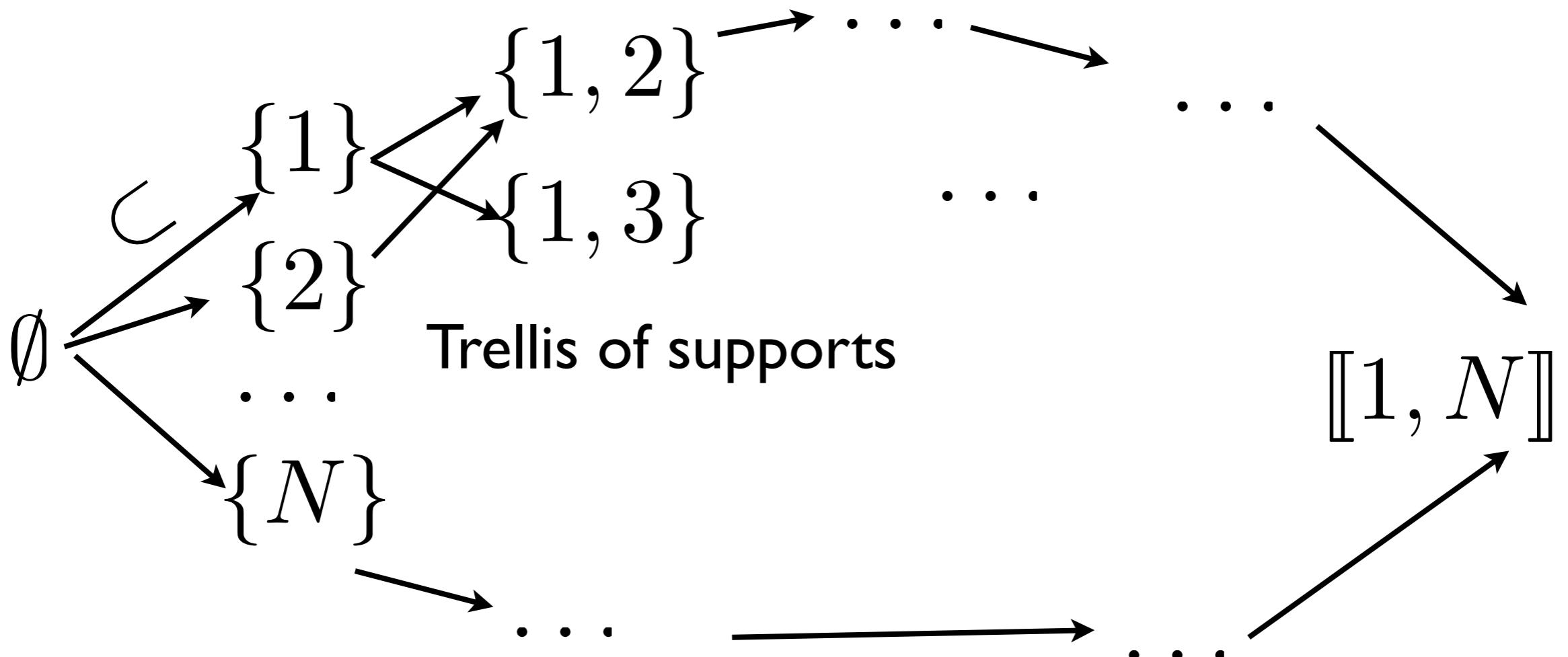
$$\begin{aligned}\|x'\|_1 &= \|x + z\|_1 = \|(x + z)_I\|_1 + \|(x + z)_{I^c}\|_1 \\ &= \|x + z_I\|_1 + \|z_{I^c}\|_1 \\ &\geq \|x\|_1 - \|z_I\|_1 + \|z_{I^c}\|_1\end{aligned}$$

# NSP( $I$ ) is sufficient

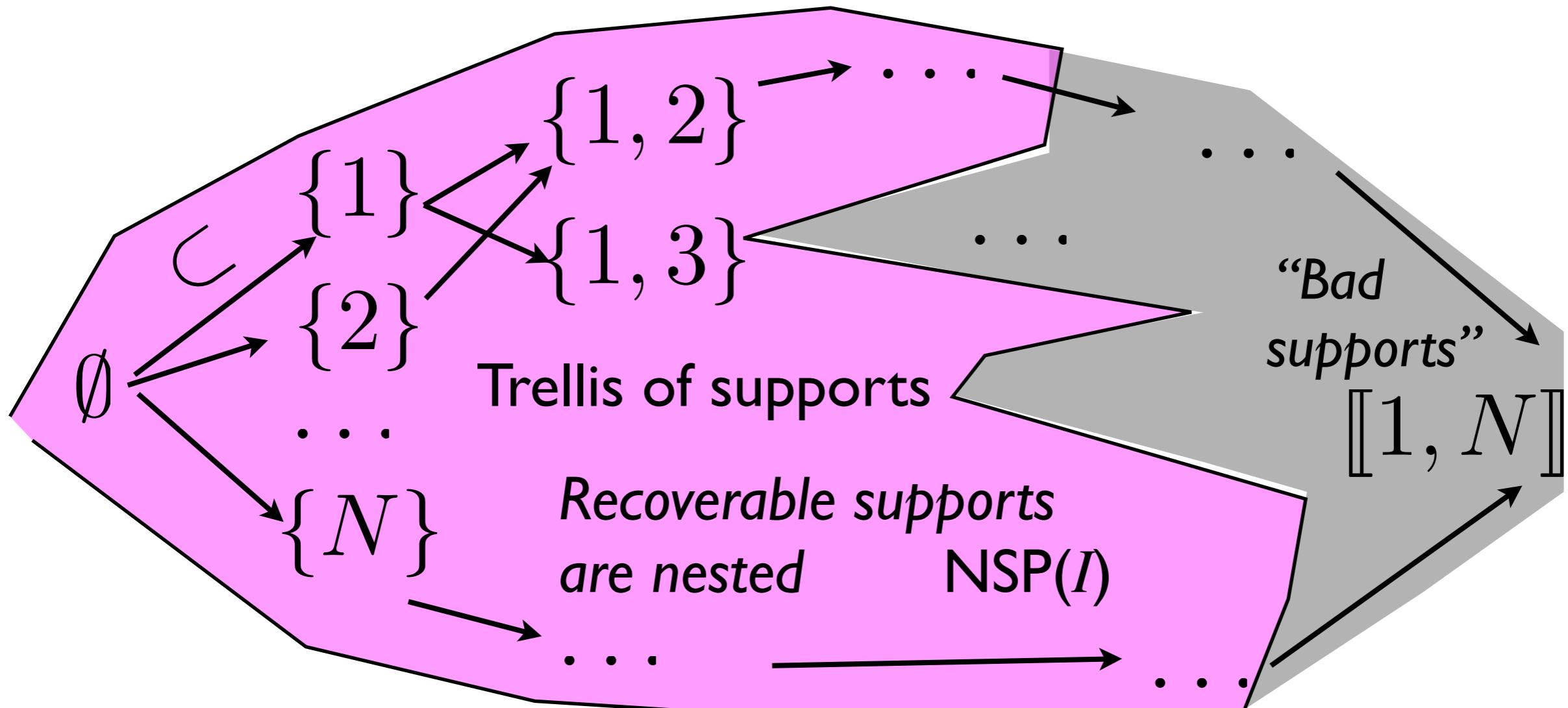
- Consider  $x$  with support set  $I$  and  $x' \neq x$  with  $\mathbf{A}x' = \mathbf{A}x$
- Denote  $z := x' - x \in \text{Ker}\mathbf{A} \setminus \{0\}$ 
  - ✓ By NSP( $I$ ) we have  $\|z_{I^c}\|_1 > \|z_I\|_1$
  - ✓ Moreover we have

$$\begin{aligned}\|x'\|_1 &= \|x + z\|_1 = \|(x + z)_I\|_1 + \|(x + z)_{I^c}\|_1 \\ &= \|x + z_I\|_1 + \|z_{I^c}\|_1 \\ &\geq \|x\|_1 - \|z_I\|_1 + \|z_{I^c}\|_1 \\ &> \|x\|_1\end{aligned}$$

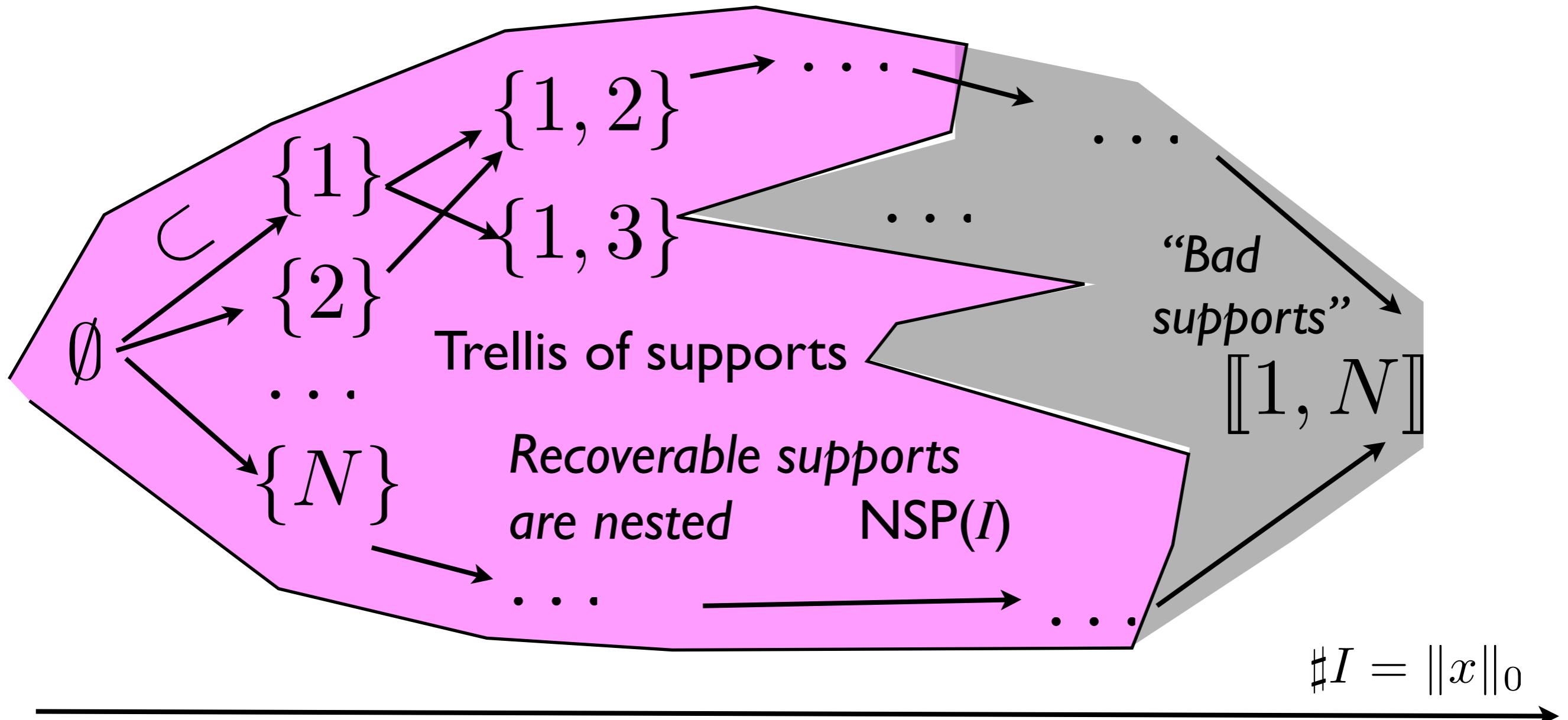
# From “recoverable” supports to “sparse” vectors



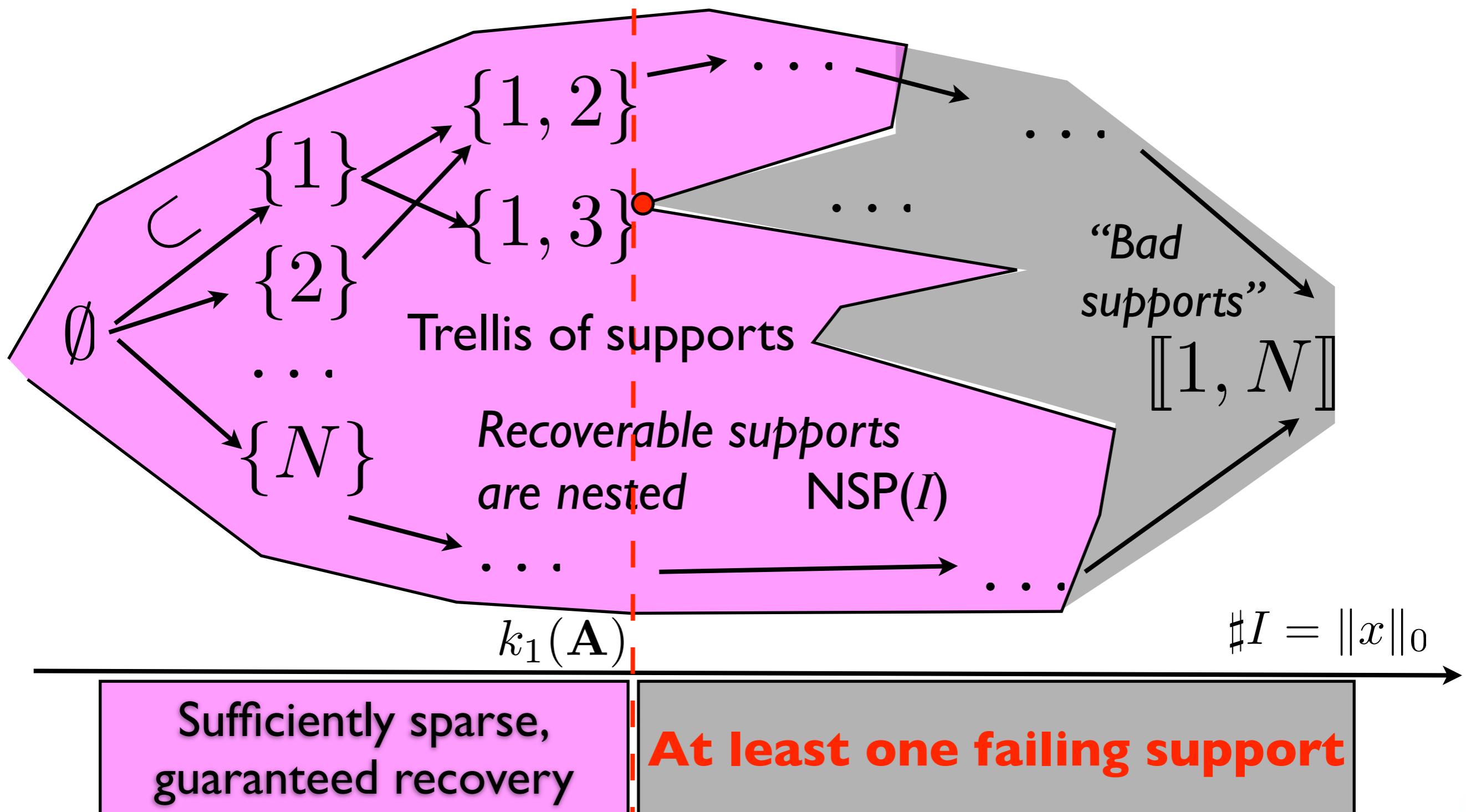
# From “recoverable” supports to “sparse” vectors



# From “recoverable” supports to “sparse” vectors



# From “recoverable” supports to “sparse” vectors



# L1 recovery for a given sparsity $k$

- **Notation**

- ✓  $I_k$  = shorthand for index of  $k$  largest entries of vector  $z$

- **Corollary 1** [*Donoho & Huo 2001 for L1, G. & Nielsen 2003 for  $L_p$  & more*]

- ✓ Assume the «Null Space Property» for a given sparsity  $k$

NSP( $k$ )

$$\|z_{I_k}\|_1 < \|z\|_1/2 \quad \forall z \in \text{Ker } \mathbf{A} \setminus \{0\}$$

- ✓ Then: for every  $k$ -sparse vector  $x$

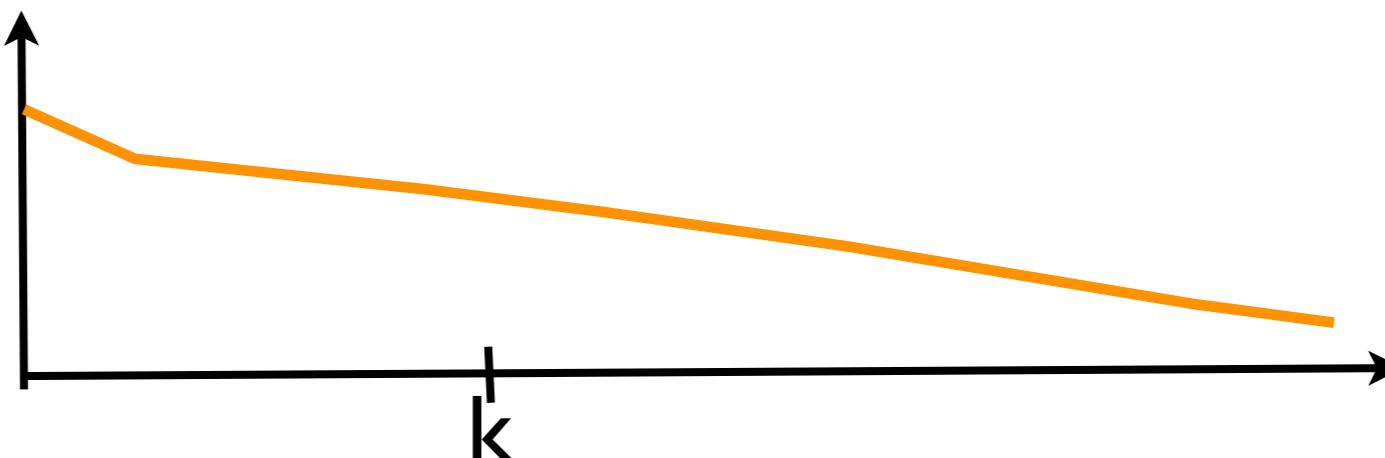
$$x = \arg \min_{\tilde{x}} \|\tilde{x}\|_1 \text{ s.t. } \mathbf{A}\tilde{x} \neq \mathbf{A}x$$

- ✓ Sharpness: if NSP( $k$ ) fails there is **at least one failing  $k$ -sparse vector  $x$**

# Interpretation of NSP

- **Geometry in coefficient space:**

- ✓ consider an element  $z$  of the Null Space of  $A$
- ✓ order its entries in decreasing order



- ✓ «mass» of the  $k$  largest terms should not exceed that of the tail

$$\|z_{I_k}\|_1 < \|z\|_1/2 \quad \Leftrightarrow \quad \|z_{I_k}\|_1 < \|z_{I_k^c}\|_1$$

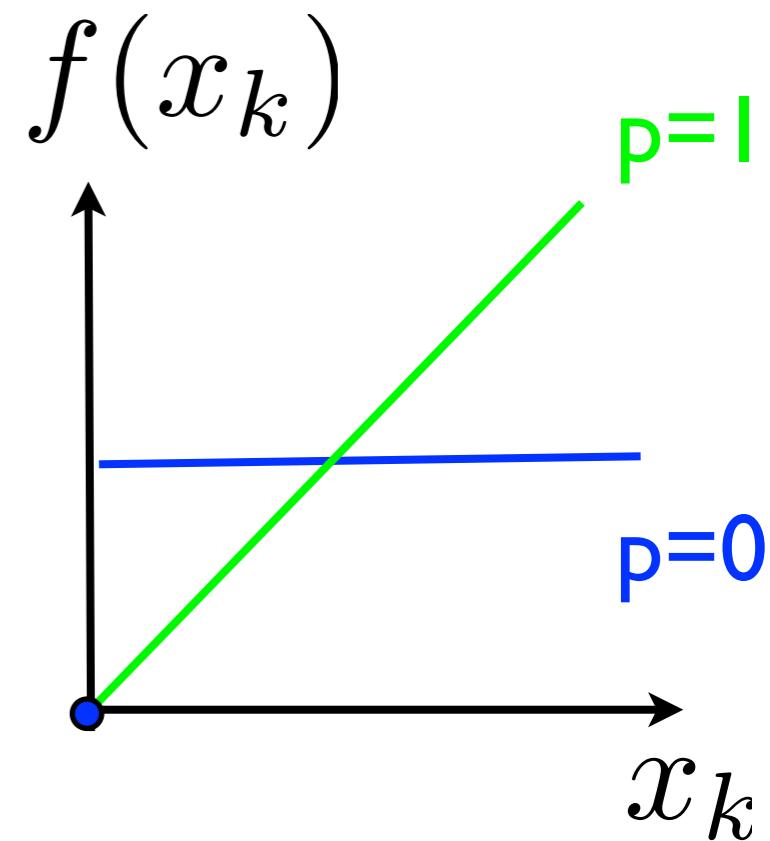
**Null space vectors must be “flat”, not sparse**

# Sparsity measures «between» L0 and L1 ?

- L<sub>p</sub>-norms       $\|x\|_p^p := \sum_k |x_k|^p, 0 \leq p \leq 1$

- f-norms!       $\|x\|_f := \sum_k f(|x_k|)$

- Constrained minimization



$$x_f^\star = x_f^\star(\mathbf{b}, \mathbf{A}) \in \arg \min_x \|x\|_f \quad \text{subject to} \quad \mathbf{b} = \mathbf{A}x$$

- When do we have  $x_f^\star(\mathbf{A}x_0, \mathbf{A}) = x_0$  ?

# NSP for sub-additive sparsity measures

- **Theorem 2** [G. & Nielsen 2003]

- ✓ **Assumption 1:** sub-additivity (for quasi-triangle inequality)

$$f(a + b) \leq f(a) + f(b), \forall a, b$$

- ✓ **Assumption 2:**

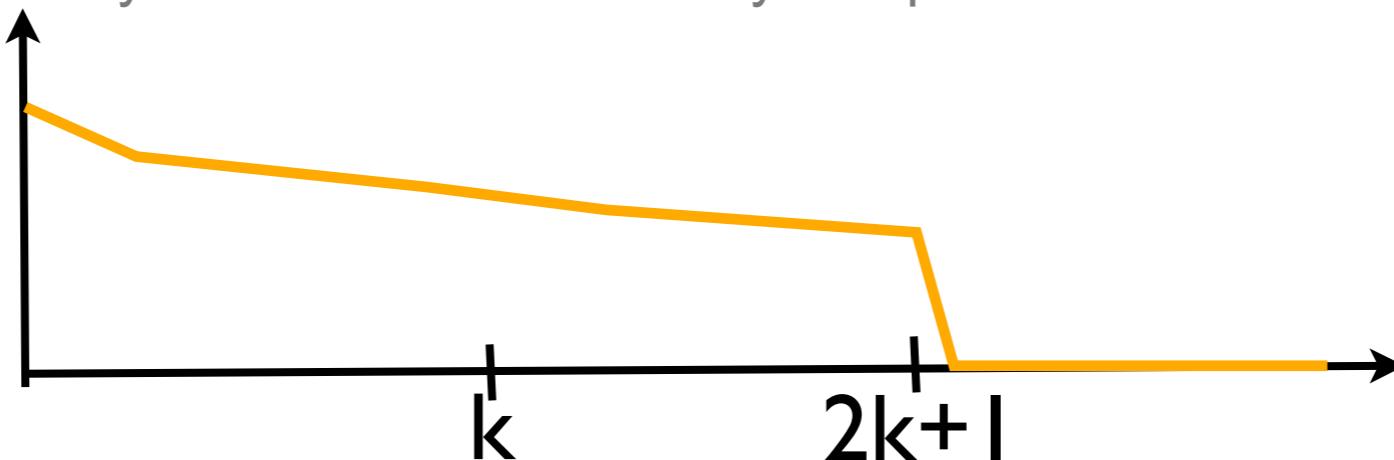
NSP( $I, f$ )

$$\|z_I\|_f < \|z\|_f / 2 \quad \forall z \in \text{Ker } \mathbf{A} \setminus \{0\}$$

- ✓ **Conclusion:**  $x_f^*$  recovers every  $x$  supported in  $I$
  - ✓ **Sharpness:** if NSP fails on support  $I$  there is **at least one failing vector**  $x$  supported in  $I$

# NSP for L0: Identifiability of sparse representations

- Case of L0: **identifiability**
  - ✓ L0 min = **guaranteed unique sparsest solution** if
    - ◆ elements in null space have at least  $2k+1$  nonzeros
    - ◆ equivalently: every  $2k$  columns are linearly independent

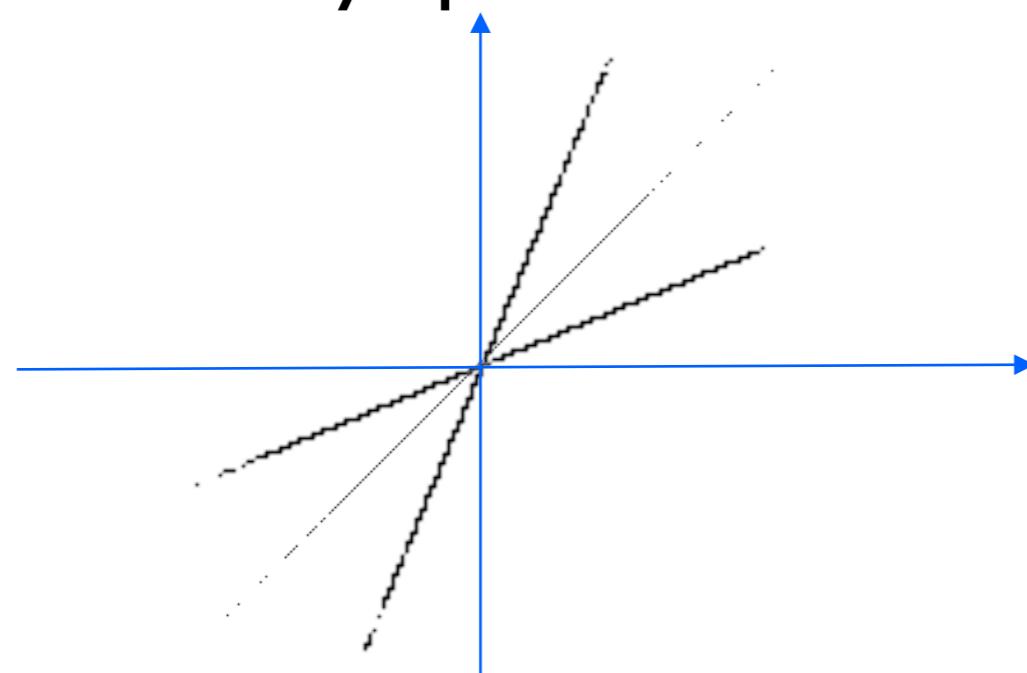


$$k = \|z_{I_k}\|_0 < \|z_{I_k^c}\|_0$$

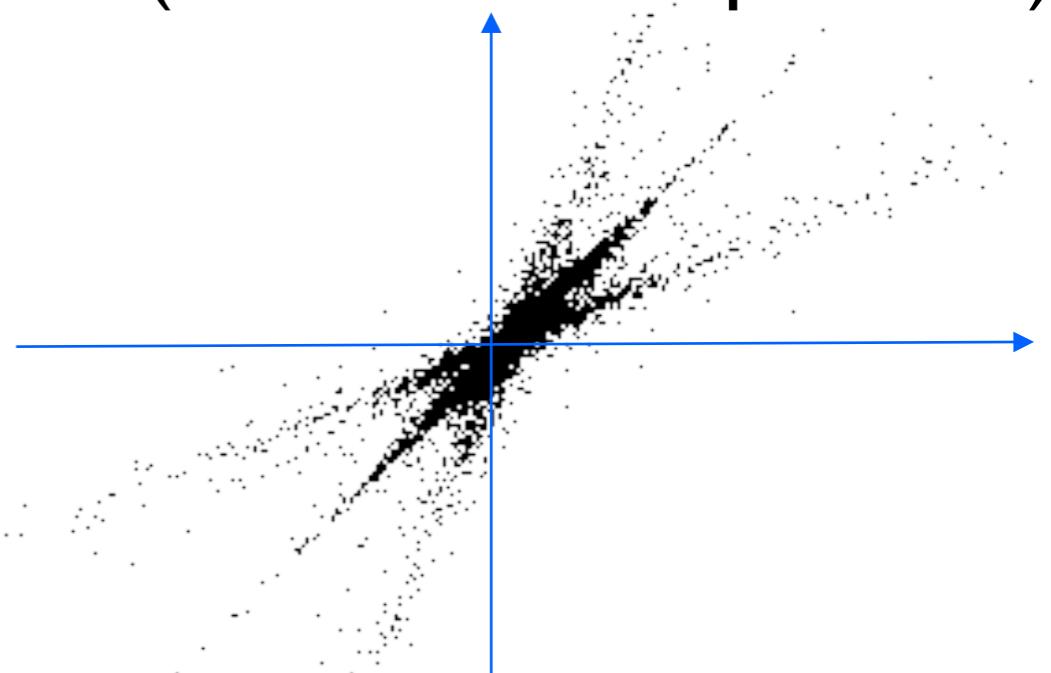
# Stability and robustness

# Need for stable recovery

Exactly sparse data



Real data (from source separation)



# Formalization of stability

- **Toy problem:** exact recovery from  $\mathbf{b} = \mathbf{A}x$ 
  - ✓ Assume sufficient sparsity  $\|x\|_0 \leq k_p(\mathbf{A}) < m$
  - ✓ Wish to obtain  $x_p^*(\mathbf{b}) = x$
- **Need to relax sparsity assumption**
  - ✓ New benchmark = best k-term approximation

$$\sigma_k(x) = \inf_{\|y\|_0 \leq k} \|x - y\|$$

- ✓ Goal = stable recovery = *instance optimality of order k*

$$\|x_p^*(\mathbf{b}) - x\| \leq C \cdot \sigma_k(x)$$

[Cohen, Dahmen & DeVore 2006]

# Stability for L<sub>p</sub> minimization

- **Assumption:** «stable Null Space Property»

$$\text{NSP}(k,p,t) \quad \|z_{I_k}\|_p^p \leq t \cdot \|z_{I_k^c}\|_p^p \quad \forall z \in \text{Ker A} \setminus \{0\}$$

- **Conclusion:** *instance optimality* for all  $x$

$$\|x_p^*(\mathbf{b}) - x\|_p^p \leq C(t) \cdot \sigma_k(x)_p^p \quad C(t) := 2 \frac{1+t}{1-t}$$

- **Sharpness:** stable NSP is *necessary*

[Davies & Gribonval, SAMPTA 2009]

# Robustness

- Toy model = noiseless
- Need to account for noise
  - ✓ measurement noise
  - ✓ modeling error
  - ✓ numerical inaccuracies ...
- Goal: predict robust estimation

$$\begin{aligned} \mathbf{b} &= \mathbf{A}\mathbf{x} \\ \mathbf{b} &= \mathbf{A}\mathbf{x} + \mathbf{e} \end{aligned}$$

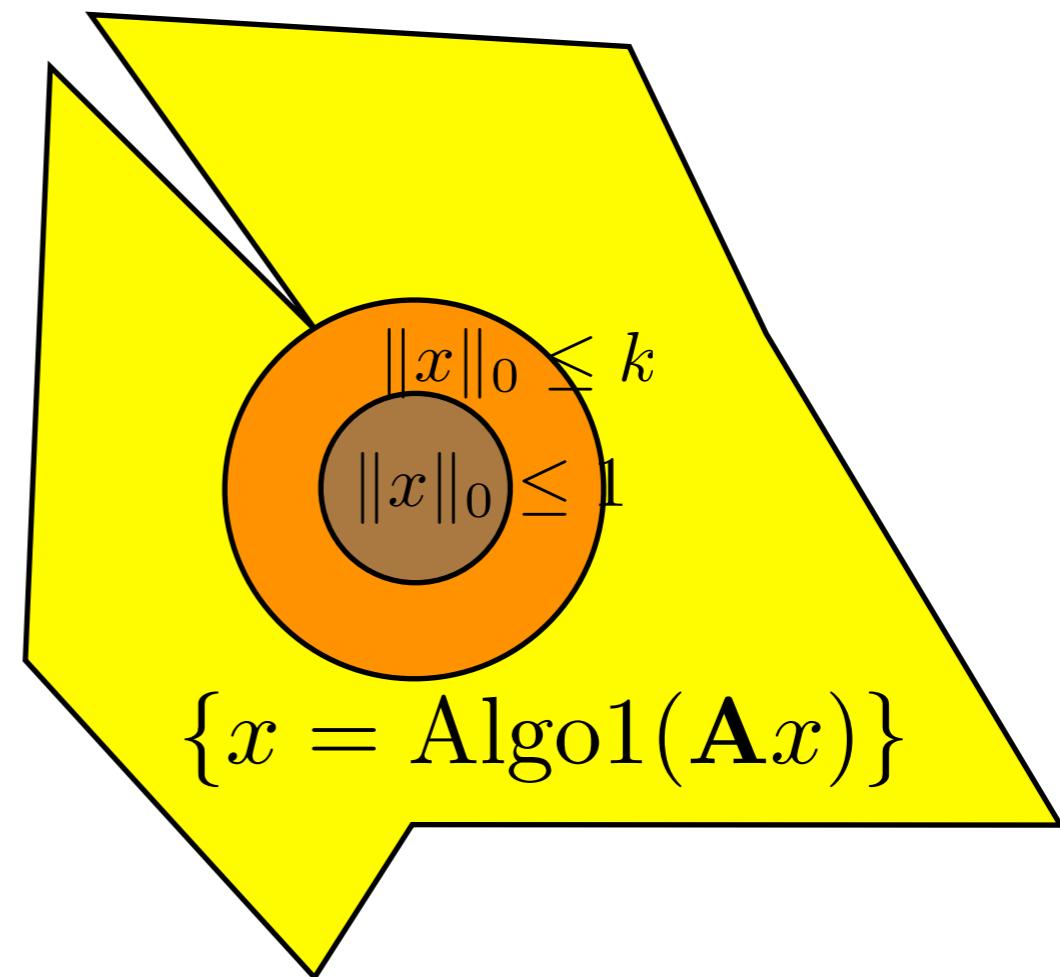
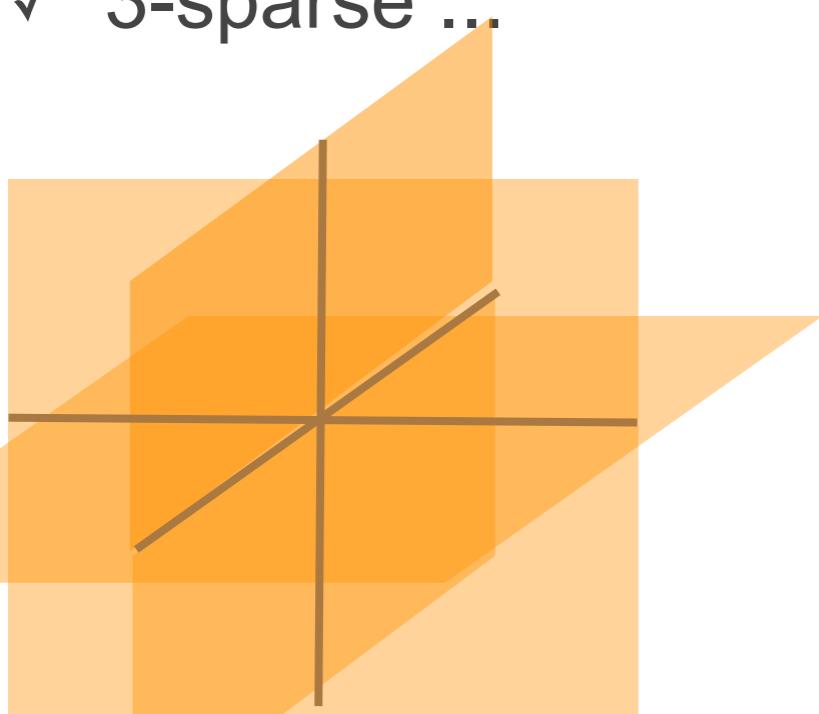
$$\|x_p^*(\mathbf{b}) - \mathbf{x}\| \leq C\|\mathbf{e}\| + C'\sigma_k(\mathbf{x})$$

- Tool: stable + robust NSP (cf. Foucart & Rauhut book)

# Comparison between algorithms ?

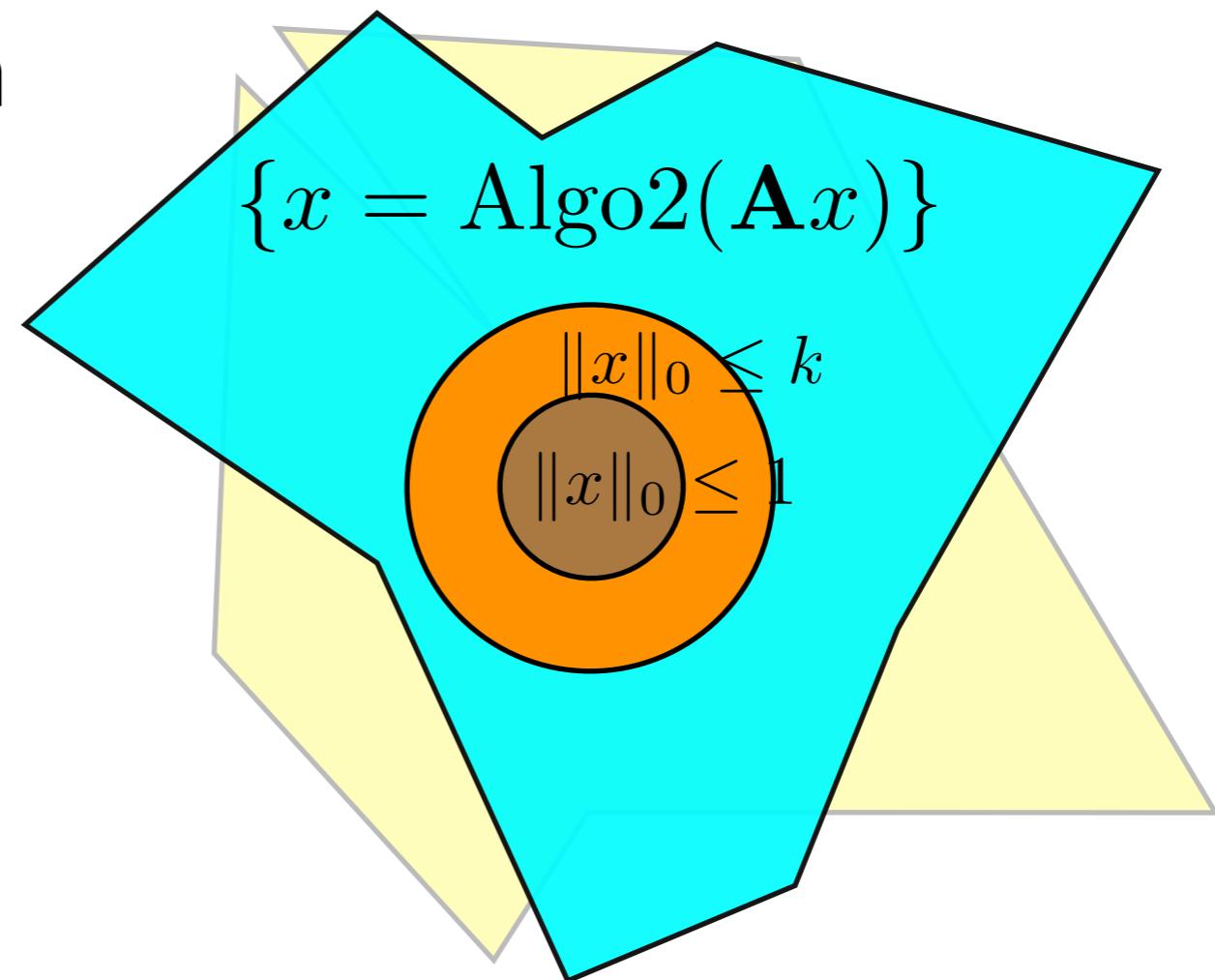
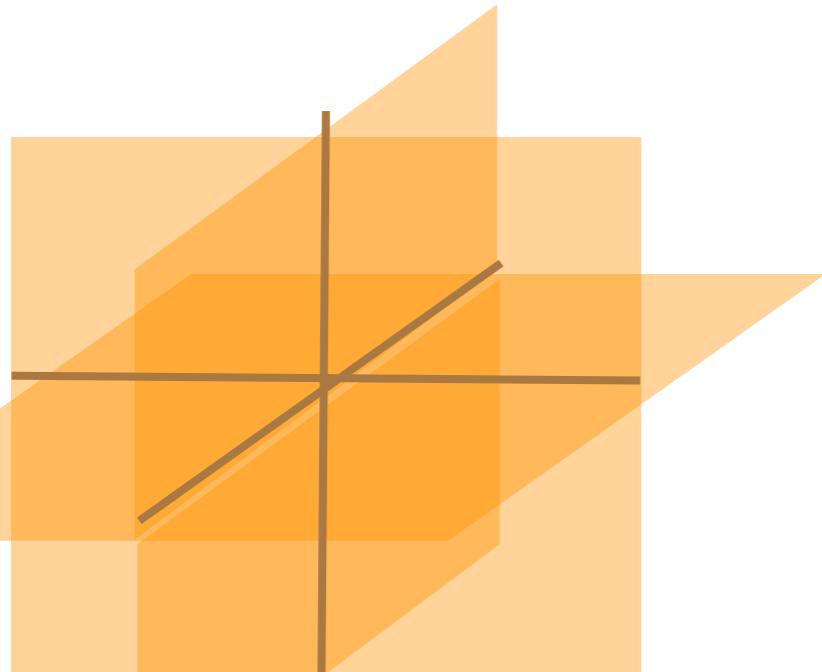
# Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

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



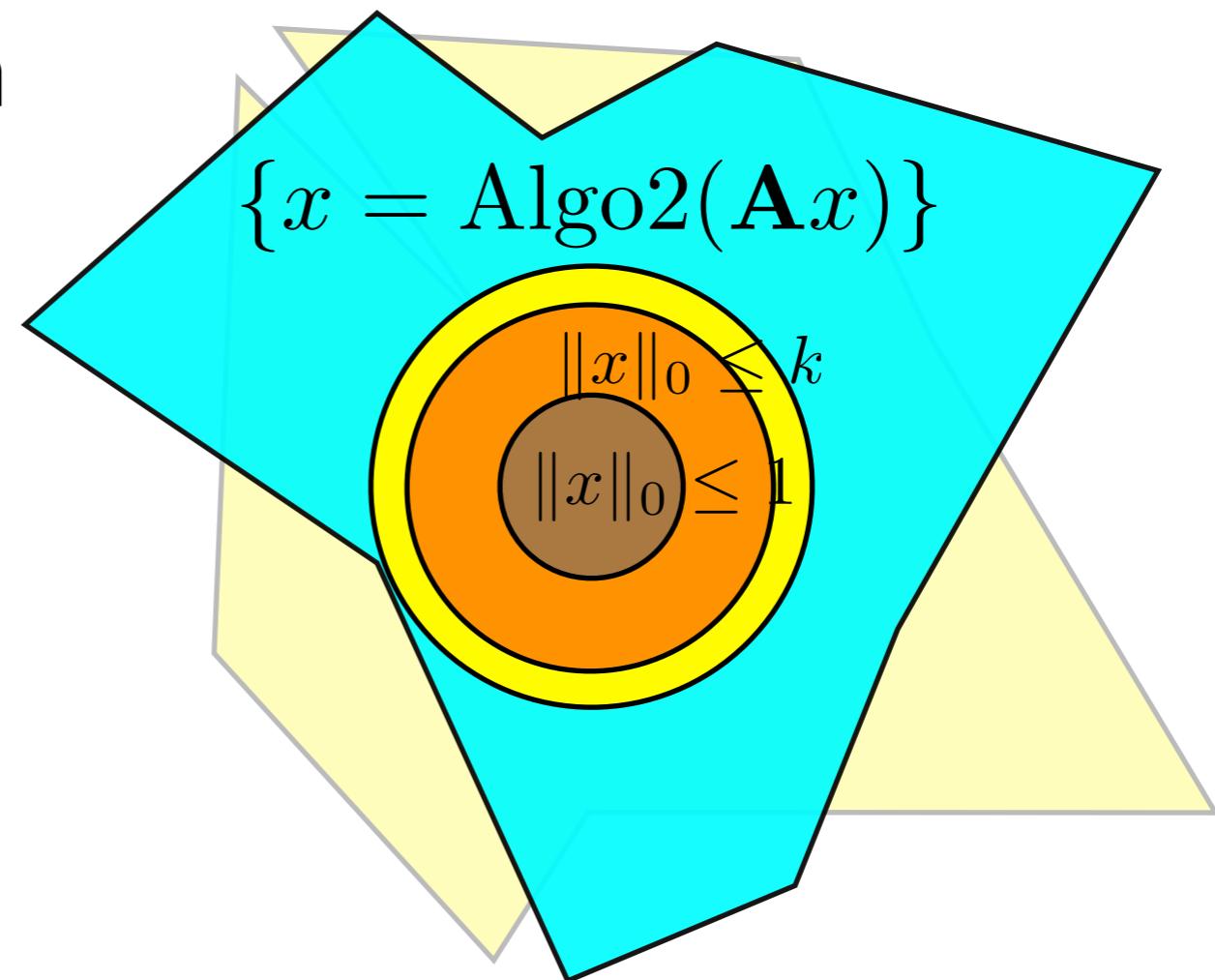
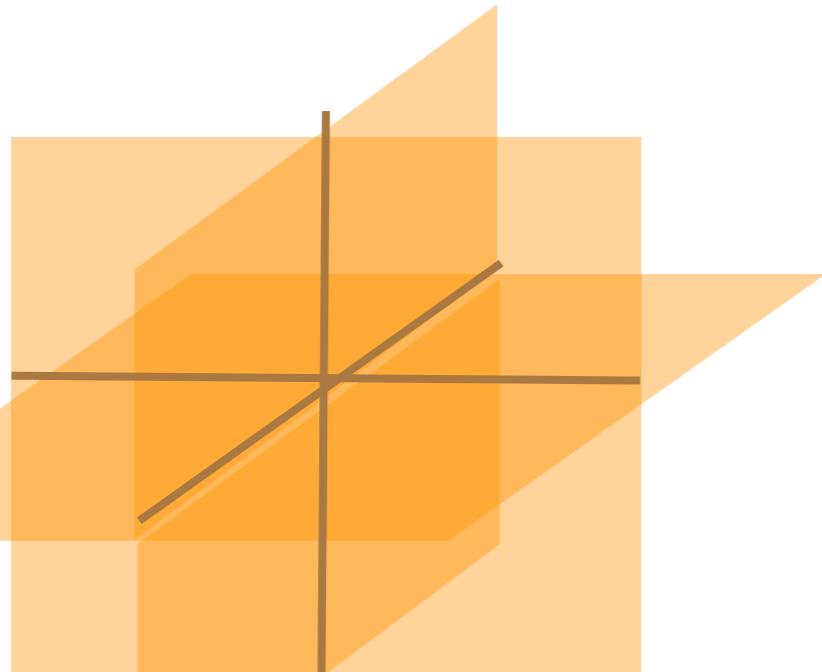
# Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

- 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 ...



# Comparison between algorithms

- **Theorem:** recovery conditions based on number of nonzero components  $\|x\|_0$  for  $0 \leq q \leq p \leq 1$  satisfy

$$k^*_{\text{MP}}(\mathbf{A}) \leq k_1(\mathbf{A}) \leq k_p(\mathbf{A}) \leq k_q(\mathbf{A}) \leq k_0(\mathbf{A}), \forall \mathbf{A}$$

- **Warning :**
  - ✓ there often exists vectors beyond these critical sparsity levels, which are recovered
  - ✓ there often exists vectors beyond these critical sparsity levels, where the successful algorithm is not the one we would expect
- **How do we estimate these sparsity thresholds ?**

[Gribonval & Nielsen, ACHA 2007]

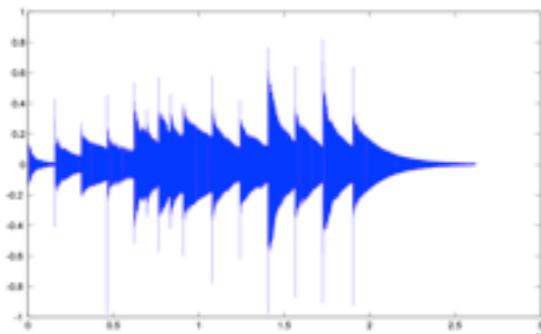
# Explicit guarantees ?

# Signal Processing Scenarios

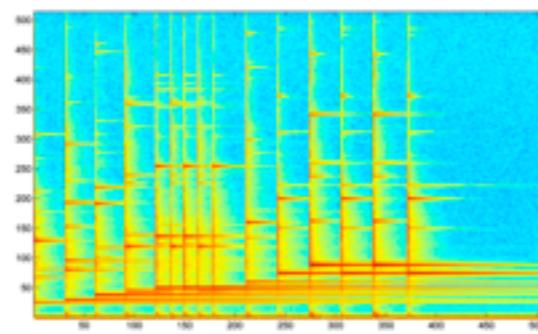
- Range of “choices” for the matrix **A**
  - ✓ Dictionary modeling structures of signals
    - ◆ Constrained choice = to fit the data.
    - ◆ Ex: *union of wavelets + curvelets + spikes*
  - ✓ «Transfer function» from physics of inverse problem
    - ◆ Constrained choice = to fit the direct problem.
    - ◆ Ex: *convolution operator / transmission channel*
  - ✓ Designed / chosen «Compressed Sensing» matrix
    - ◆ «Free» design = to maximize recovery performance vs cost of measures
    - ◆ Ex: *random Gaussian matrix... or coded aperture, etc.*
- Estimation of the recovery regimes
  - ✓ coherence for deterministic matrices
  - ✓ typical results for random matrices

# Example 1: Multilayer Audio Decomposition

- Audio = superimposition of structures



$$\mathbf{b} = \{b(t)\}_t$$



$$x = \{x(s, \tau, f)\}_{s, \tau, f}$$

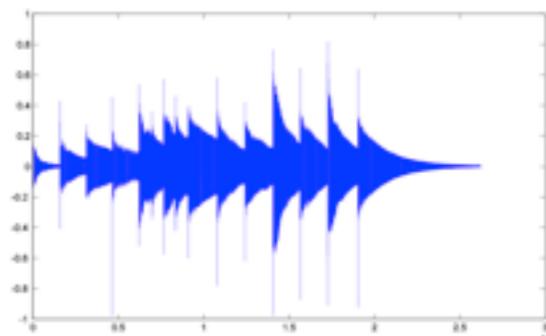
- ✓ transients = short, small scale
- ✓ harmonic part = long, large scale

- Model = Gabor atoms  $g_{s, \tau, f}(t) := \frac{1}{\sqrt{s}} w\left(\frac{t - \tau}{s}\right) e^{2i\pi ft}$
- Dictionary matrix:  $\mathbf{A}_n = \{g_{s_n, \tau_n, f_n}(t)\}_t$

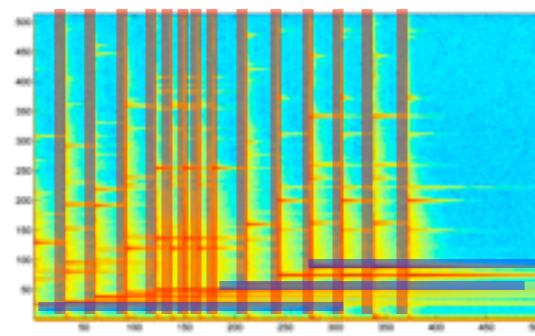
$$\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_N]$$

# Example 1: Multilayer Audio Decomposition

- Audio = superimposition of structures



$$\mathbf{b} = \{b(t)\}_t$$



$$x = \{x(s, \tau, f)\}_{s, \tau, f}$$

- ✓ transients = short, small scale
- ✓ harmonic part = long, large scale

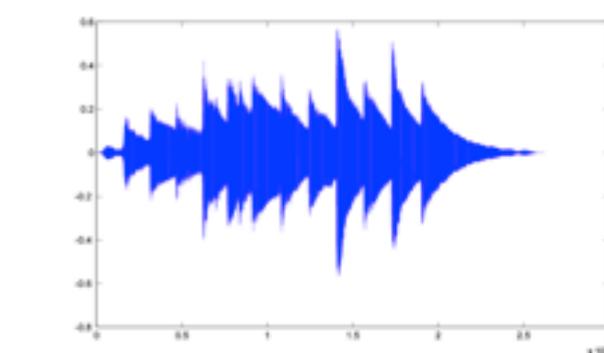
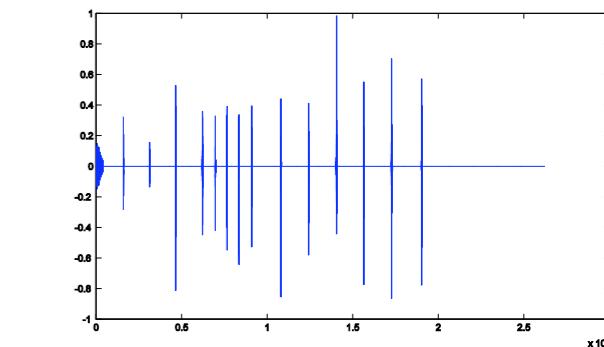
- Model = Gabor atoms

$$g_{s, \tau, f}(t) := \frac{1}{\sqrt{s}} w\left(\frac{t - \tau}{s}\right) e^{2i\pi ft}$$

- Dictionary matrix:

$$\mathbf{A}_n = \{g_{s_n, \tau_n, f_n}(t)\}_t$$

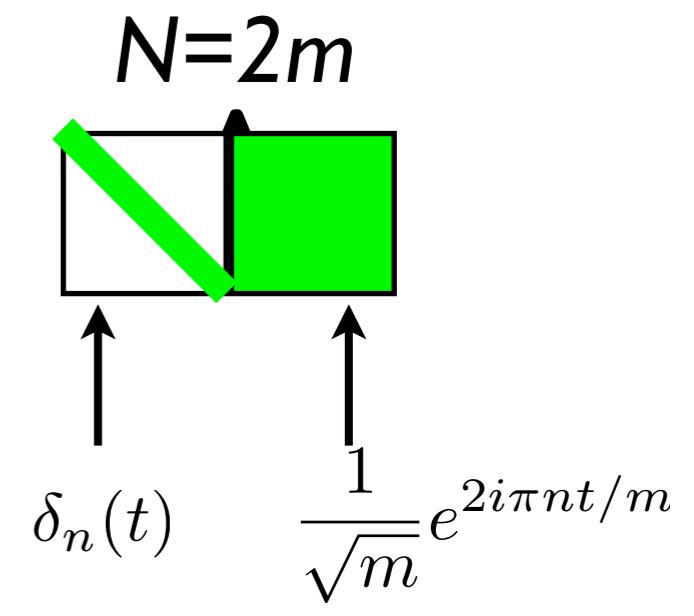
$$\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_N]$$



# Caricature of two-scale Gabor dictionary

- Dirac-Fourier dictionary

$$\mathbf{A} = \mathbf{m}$$



- Coherence

$$\mu = 1/\sqrt{m}$$

- Sparsity thresholds

$$k^*_{\text{MP}}(\mathbf{A}) \geq 0.5\sqrt{m}$$

## Example 2: convolution operator

- Deconvolution problem with spikes

$$b = h \star x + e$$

- ✓ Matrix-vector form  $\mathbf{b} = \mathbf{A}x + \mathbf{e}$  with  $\mathbf{A} =$  Toeplitz or circulant matrix  $[\mathbf{A}_1, \dots, \mathbf{A}_N]$

$$\mathbf{A}_n(i) = h(i - n)$$

by convention

$$\|\mathbf{A}_n\|_2^2 = \sum_i h(i)^2 = 1$$

- ✓ Coherence = autocorrelation, can be large

$$\mu = \max_{n \neq n'} \mathbf{A}_n^T \mathbf{A}_{n'} = \max_{\ell \neq 0} h \star \tilde{h}(\ell)$$

- ✓ Recovery guarantees

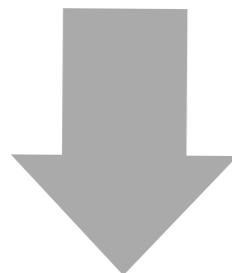
- ◆ Worst case = close spikes, usually difficult and not robust
- ◆ Stronger guarantees assuming distance between spikes [Dossal 2005]

- ✓ Algorithms: exploit fast convolution to apply  $\mathbf{A}$  and adjoint.

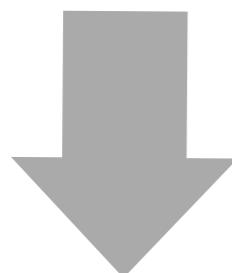
# Coherence vs RIP vs NSP

# Tools that we can use

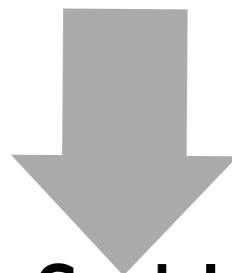
Incoherence



RIP( $k$ )



NSP( $k, p=1$ )



$$k < \frac{1}{2}(1 + 1/\mu)$$

$$\delta_{2k} < \delta$$

$$\|z_{I_k}\|_1 < \|z\|_1/2 \quad \forall z \in \text{Ker } \mathbf{A} \setminus \{0\}$$

**Stable LI Null Space Property (NSP)**

Stable  $L_p$  recovery for all  $k$ -sparse vectors, for all  $p \leq 1$

# How small is the coherence ?

- **Orthonormal basis**  $\mathbf{A} = \mathbf{B}$   $\mathbf{B}^T \mathbf{B} = \text{Id}_m$   
 $\mu(\mathbf{A}) = 0$
- **Orthonormal basis + some unit vectors**  $\mathbf{A} = [\mathbf{B}; \mathbf{C}]$   
 $\mu(\mathbf{A}) \geq 1/\sqrt{m}$   $(1 + 1/\mu)/2 \lesssim \sqrt{m}$
- **Welch bound:**  $\mathbf{A} \in \mathbb{R}^{m \times N}$  **with unit norm columns**

$$\mu(\mathbf{A}) \geq \sqrt{\frac{N-m}{m(N-1)}}$$

# RIP from coherence

- **Lemma**

- ✓ Assume normalized columns  $\|\mathbf{A}_i\|_2 = 1$

- ✓ Define **coherence**  $\mu = \max_{i \neq j} |\mathbf{A}_i^T \mathbf{A}_j|$

- ✓ Consider index set  $I$  of size  $\#I \leq k$

- ✓ Then for any coefficient vector  $c \in \mathbb{R}^I$

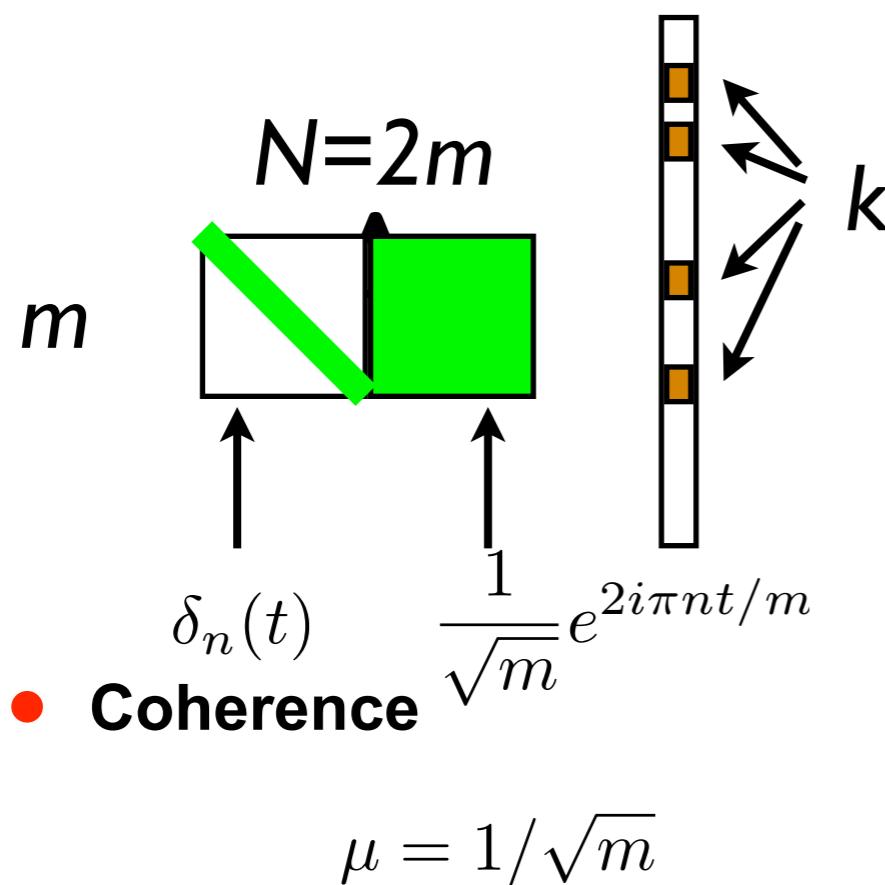
$$1 - (k - 1)\mu \leq \frac{\|\mathbf{A}_I c\|_2^2}{\|c\|_2^2} \leq 1 + (k - 1)\mu$$

- ✓ In other words

$$\delta_{2k} \leq (2k - 1)\mu$$

# Coherence vs RIP

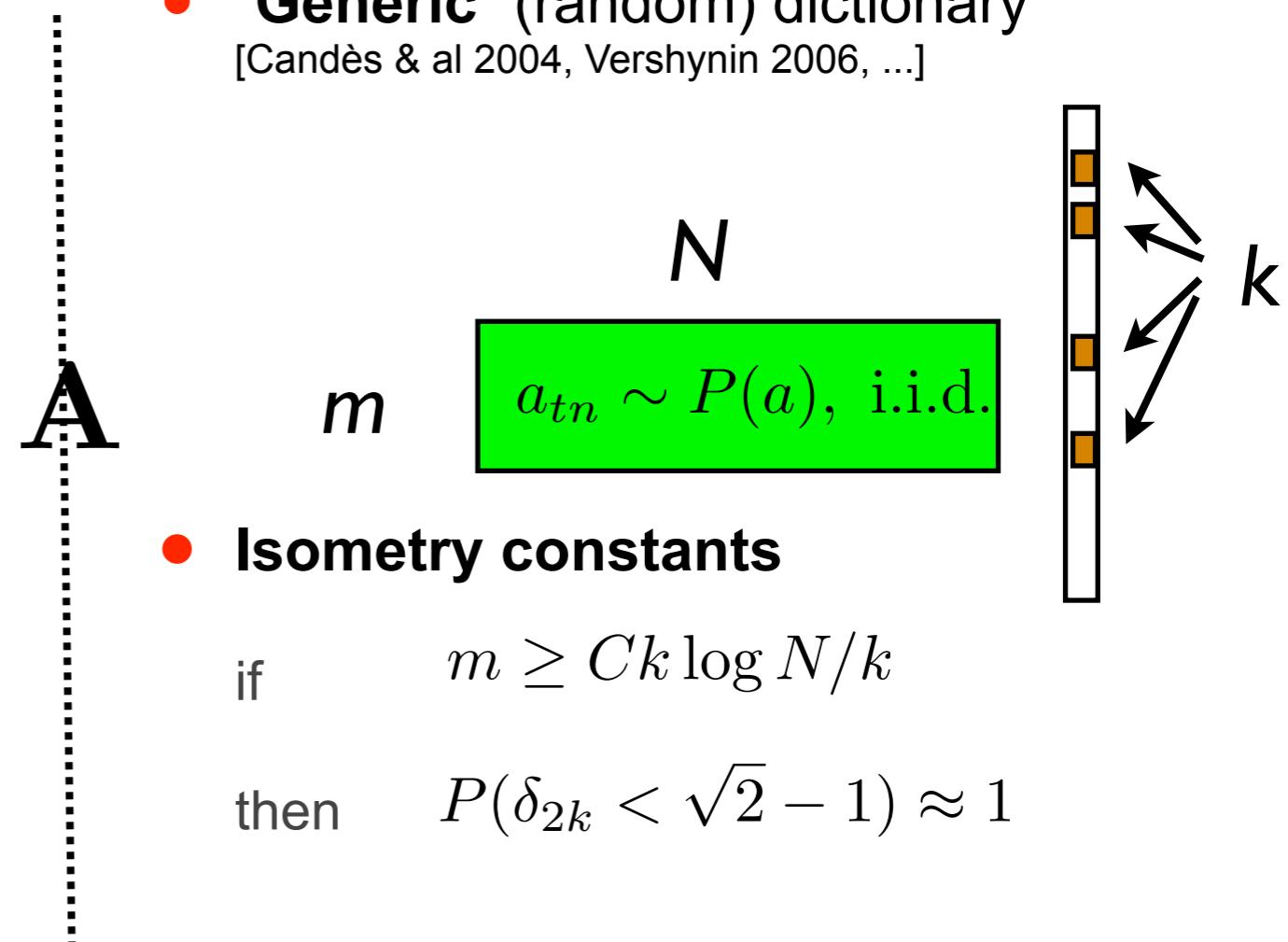
- Deterministic matrix, such as Dirac-Fourier dictionary



- Coherence

$$\mu = 1/\sqrt{m}$$

- “Generic” (random) dictionary [Candès & al 2004, Vershynin 2006, ...]



- Isometry constants

if  $m \geq Ck \log N/k$

then  $P(\delta_{2k} < \sqrt{2} - 1) \approx 1$

## Recovery regimes

$$k_1(\mathbf{A}) \approx 0.914\sqrt{m}$$

$$k^*_{\text{MP}}(\mathbf{A}) \geq 0.5\sqrt{m}$$

$$k_1(\mathbf{A}) \approx \frac{m}{2e \log N/m}$$

with high probability for Gaussian  $\mathbf{A}$

[Elad & Bruckstein 2002]

[Donoho & Tanner 2009]

# Compressed sensing

- Approach = acquire some data  $y$  with a limited number  $m$  of (linear) measures, modeled by a **measurement matrix**  $M$ :  $b \approx My$
- Key hypotheses
  - ✓ Sparse model: the data can be sparsely represented in a known dictionary  $\Phi$ :  $y \approx \Phi x$ , with  $\sigma_k(x) \ll \|x\|$
  - ✓ Overall matrix  $A = M\Phi$  is «incoherent»:  $\delta_{2k}(A) \ll 1$  leading to robust + stable sparse recovery
- Reconstruction = sparse recovery algorithm

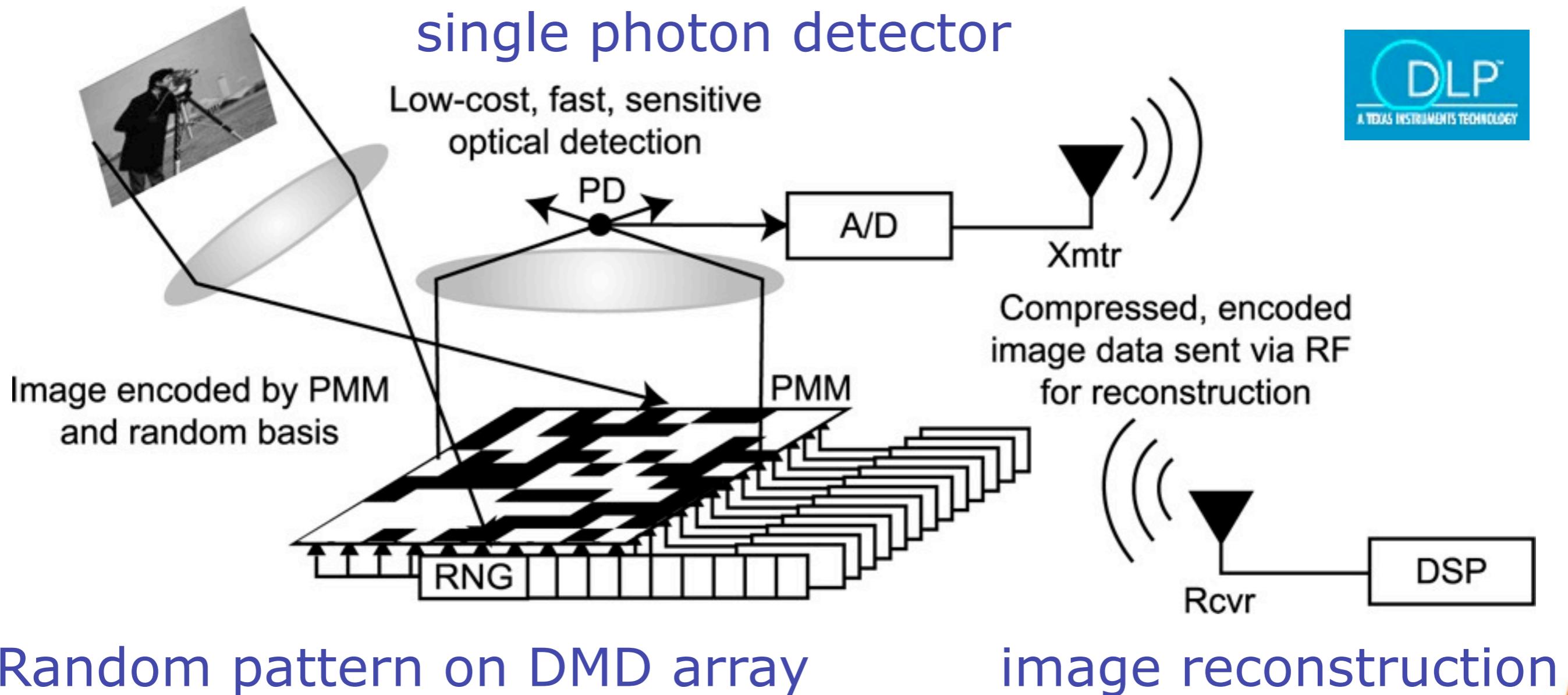
# Compressed Sensing: key requirements

- Sparse model= dictionary  $\Phi$ 
  - ✓ need to «fit» the **data**
    - ◆ does not always exist: e.g. white Gaussian noise cannot be sparsified!
  - ✓ dictionary design:
    - ◆ expert knowledge
    - ◆ **dictionary selection** from a library (wavelets, curvelets, Gabor, ...)
    - ◆ **dictionary learning**
- Measurement matrix  $M$ 
  - ✓ **physically feasible:** hardware implementation!
  - ✓ **recovery guarantees:** incoherence of  $M\Phi$
- Efficiency:
  - ✓ **fast computation** of  $M\Phi y, (M\Phi)^T b$

# Compressed Sensing: when is it worth it ?

- **Worthless** if high-res. sensing+storage = cheap  
*i.e., not for your personal digital camera!*
- **Worth it whenever**
  - ✓ High-res. = impossible
    - ◆ no miniature sensor, e.g, certain wavelength
  - ✓ Cost of each measure is high
    - ◆ Time constraints [fMRI]
    - ◆ Economic constraints [well drilling]
    - ◆ Intelligence constraints [furtive measures]?
    - ◆ Constraints on data flow
  - ✓ Transmission is lossy  
(CS=robust to loss of a few measures)

# Example : single-pixel camera, Rice University



# Stability & robustness from RIP

RIP( $k, \delta$ )

$$\delta_{2k}(\mathbf{A}) \leq \delta$$

[Candès 2008]



$$t := \sqrt{2}\delta/(1 - \delta)$$

NSP( $k, \ell^1, t$ )

$$\|z_{I_k}\|_1 \leq t \cdot \|z_{I_k^c}\|_1 \quad \text{when} \quad z \in \mathcal{N}(\mathbf{A}), z \neq 0$$

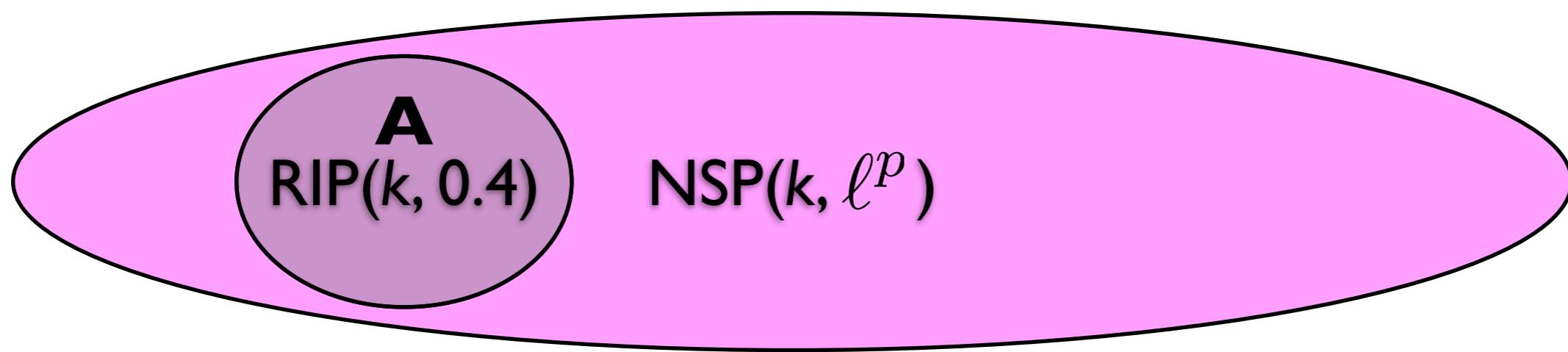
- Result: **stable + robust L1-recovery under assumption that**

$$\delta_{2k}(\mathbf{A}) < \sqrt{2} - 1 \approx 0.414$$

- ✓ *Foucart-Lai 2008: L<sub>p</sub> with p<1, and  $\delta_{2k}(\mathbf{A}) < 0.4531$*
- ✓ *Chartrand 2007, Saab & Yilmaz 2008: other RIP condition for p<1*
- ✓ *G., Figueras & Vandergheynst 2006: robustness with f-norms*
- ✓ *Needell & Tropp 2009, Blumensath & Davies 2009: RIP for greedy algorithms*

# Is the RIP a sharp condition ?

- The Null Space Property
  - ✓ “algebraic” + sharp property for  $L_p$ , only depends on  $\text{Ker } A$

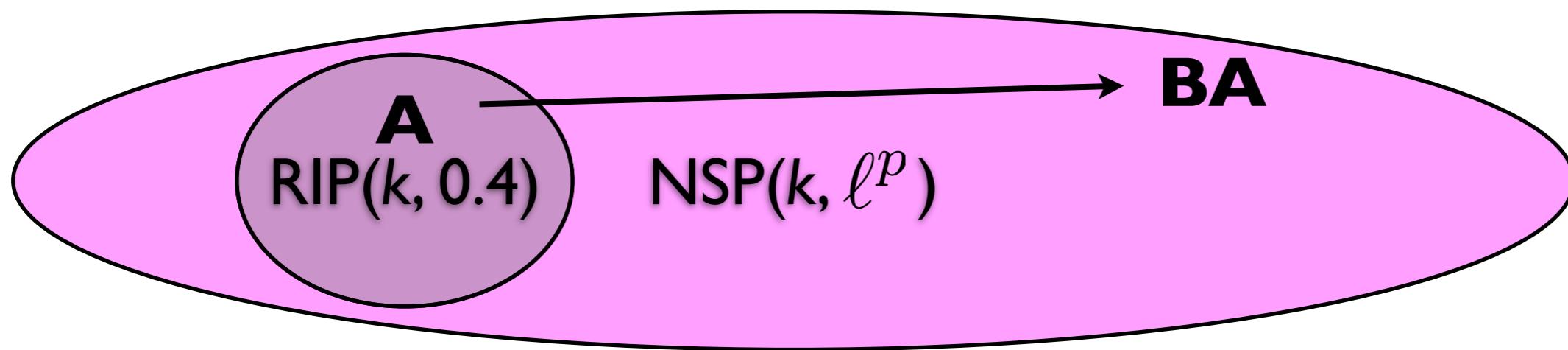


[Davies & Gribonval, IEEE Inf.Th. 2009]

# Is the RIP a sharp condition ?

- The Null Space Property

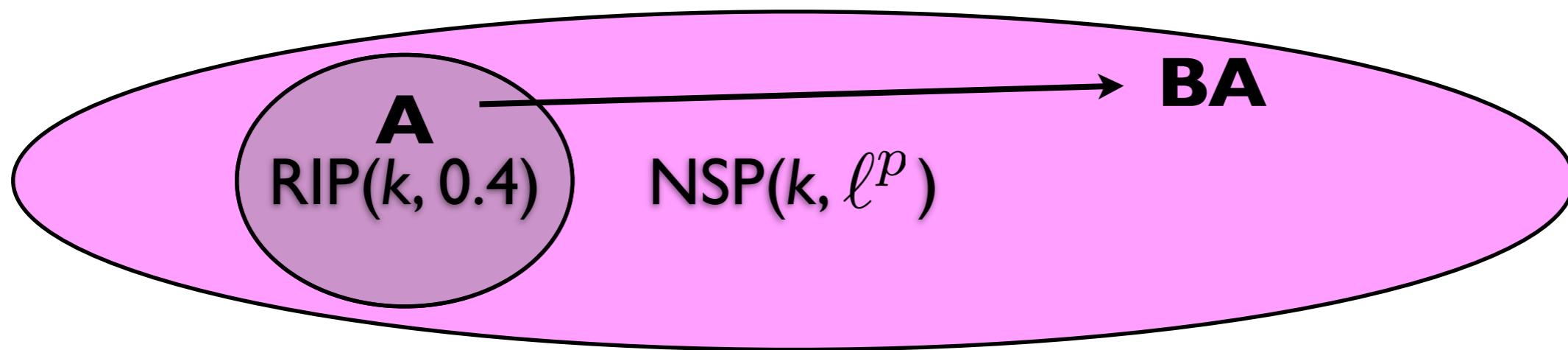
- ✓ “algebraic” + sharp property for  $L_p$ , only depends on  $\text{Ker } A$
- ✓ invariant by linear transforms  $A \rightarrow BA$



[Davies & Gribonval, IEEE Inf.Th. 2009]

# Is the RIP a sharp condition ?

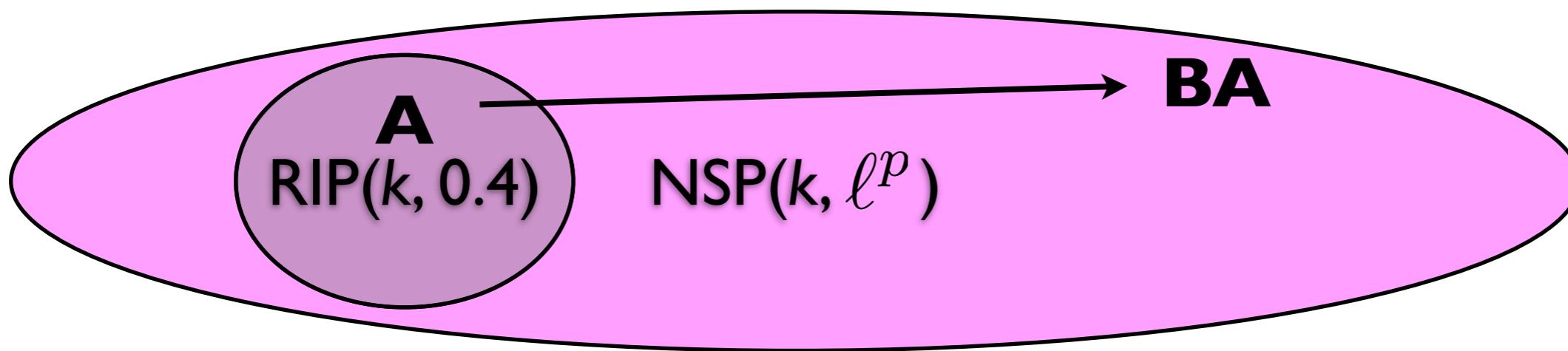
- The Null Space Property
  - ✓ “algebraic” + sharp property for  $L_p$ , only depends on  $\text{Ker } A$
  - ✓ invariant by linear transforms  $A \rightarrow BA$
- The RIP( $k, \delta$ ) condition



[Davies & Gribonval, IEEE Inf.Th. 2009]

# Is the RIP a sharp condition ?

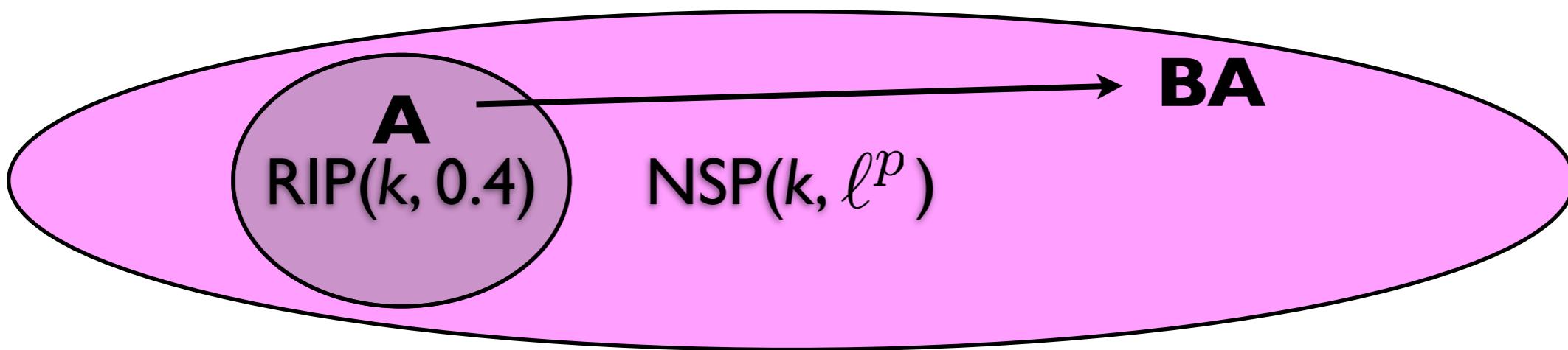
- The Null Space Property
  - ✓ “algebraic” + sharp property for  $L_p$ , only depends on  $\text{Ker } A$
  - ✓ invariant by linear transforms  $A \rightarrow BA$
- The RIP( $k, \delta$ ) condition
  - ✓ “metric” ... and not invariant by linear transforms



[Davies & Gribonval, IEEE Inf.Th. 2009]

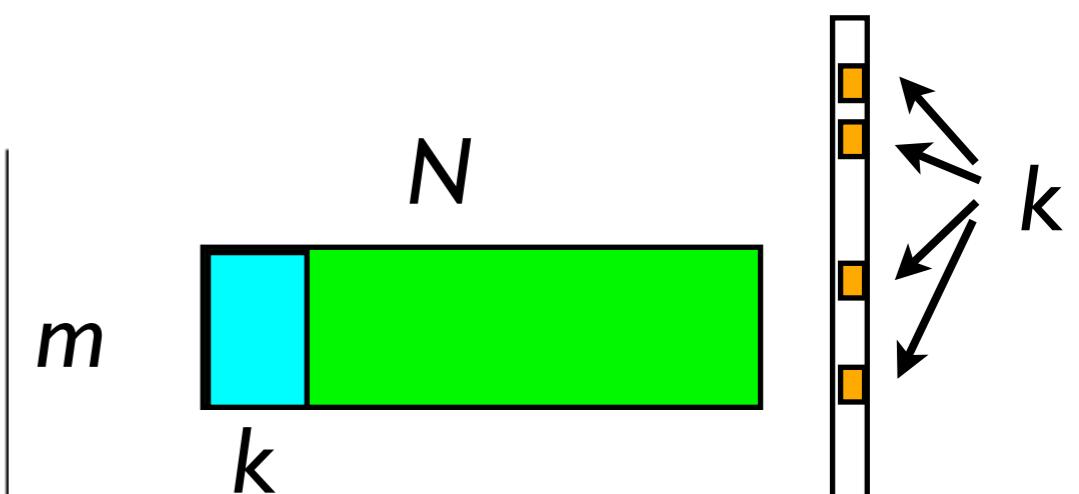
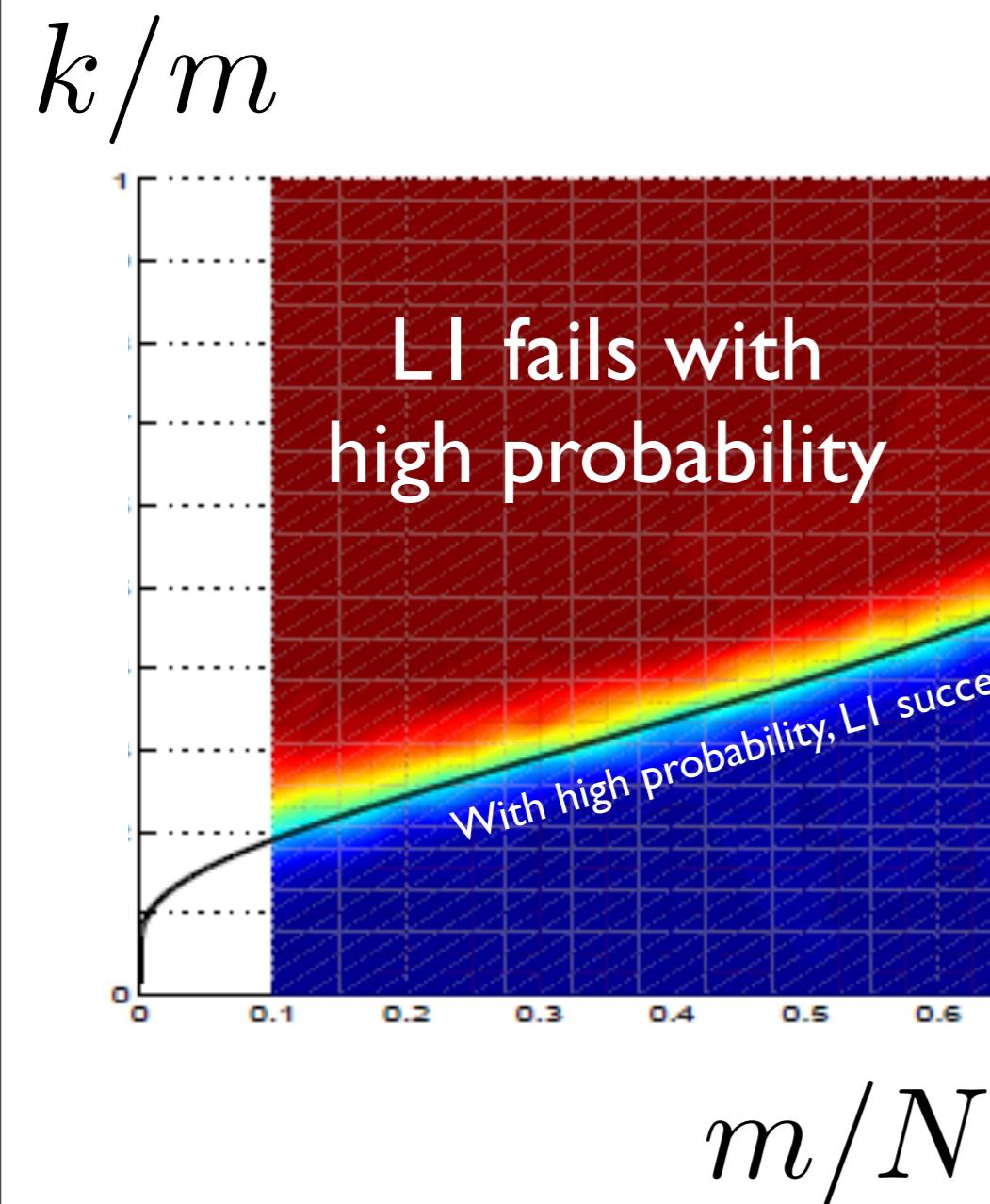
# Is the RIP a sharp condition ?

- The Null Space Property
  - ✓ “algebraic” + sharp property for  $L_p$ , only depends on  $\text{Ker } \mathbf{A}$
  - ✓ invariant by linear transforms  $\mathbf{A} \rightarrow \mathbf{B}\mathbf{A}$
- The  $\text{RIP}(k, \delta)$  condition
  - ✓ “metric” ... and not invariant by linear transforms
  - ✓ predicts performance + **robustness of several algorithms**



[Davies & Gribonval, IEEE Inf.Th. 2009]

# Phase transitions for Gaussian A



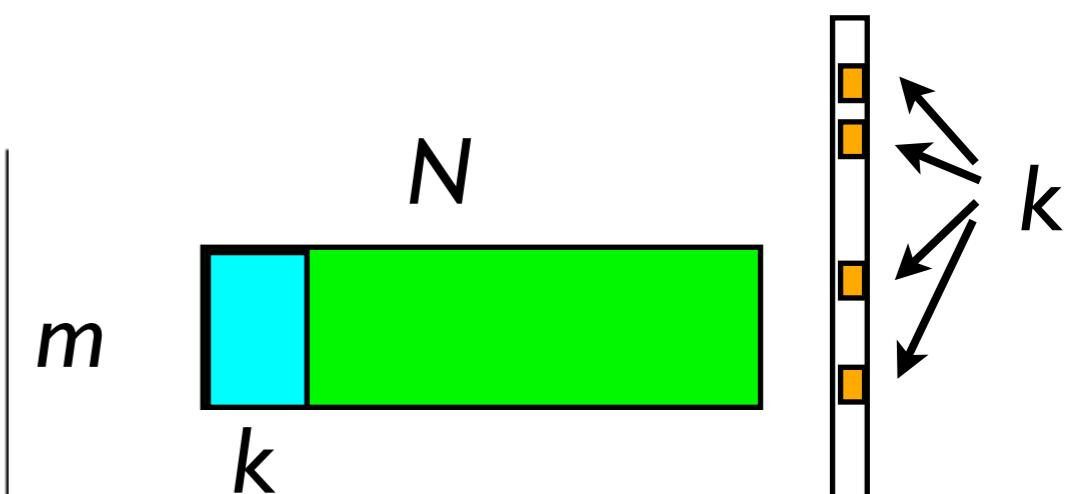
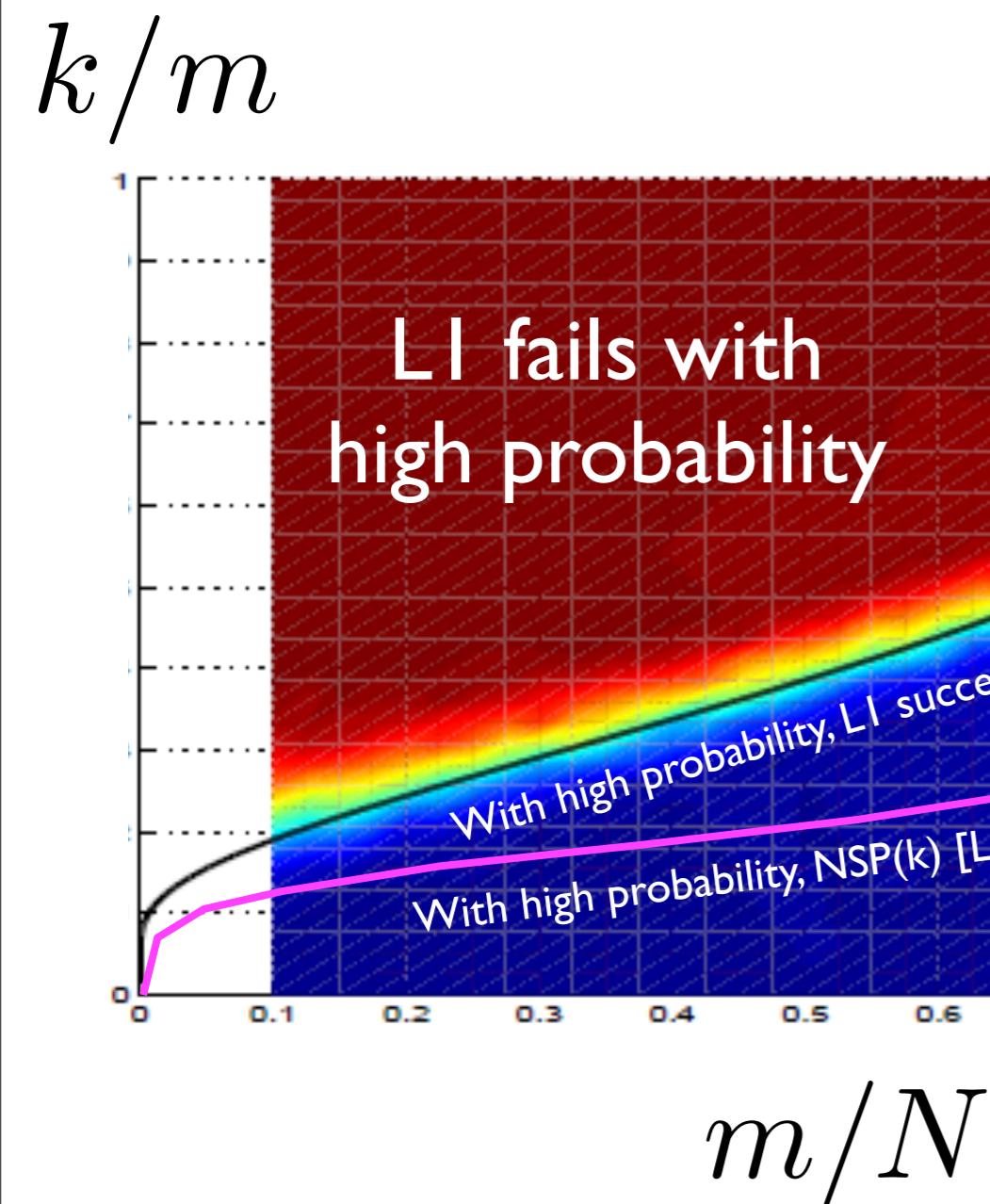
$$m \geq Ck \log N/k$$

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

$$k_1(\mathbf{A}) \approx \frac{m}{2e \log N/m}$$

[Donoho & Tanner 2009]

# Phase transitions for Gaussian A



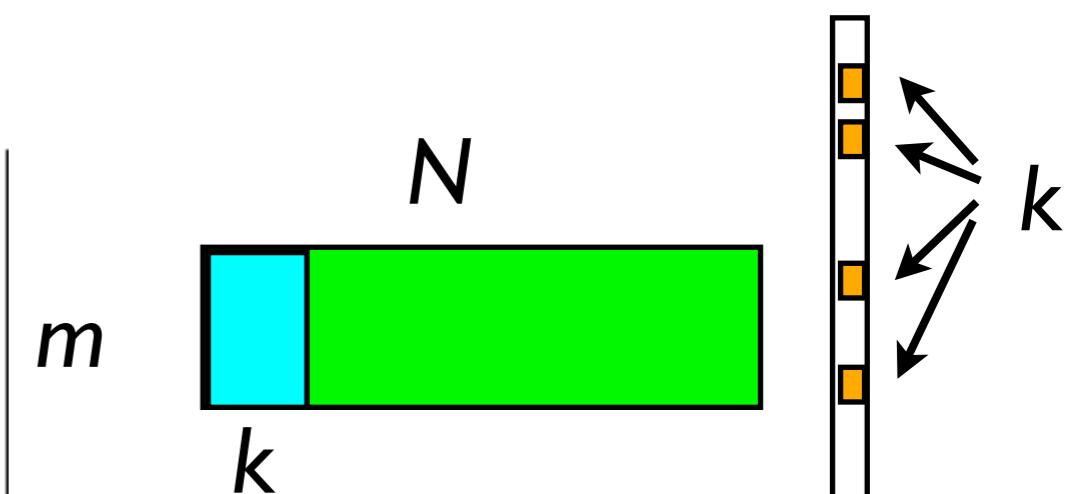
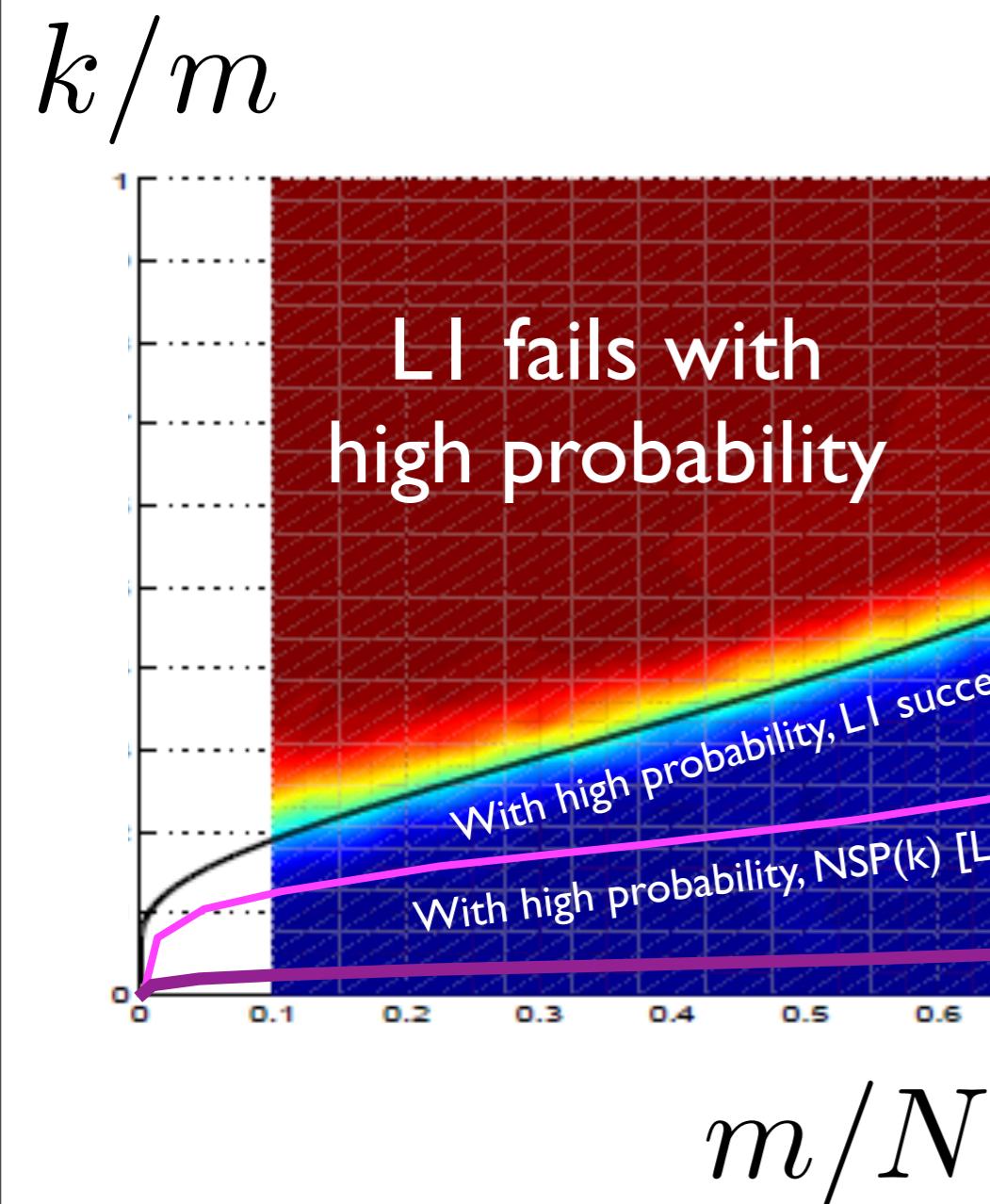
$$m \geq Ck \log N/k$$

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

$$k_1(\mathbf{A}) \approx \frac{m}{2e \log N/m}$$

[Donoho & Tanner 2009]

# Phase transitions for Gaussian A



$$m \geq Ck \log N/k$$

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

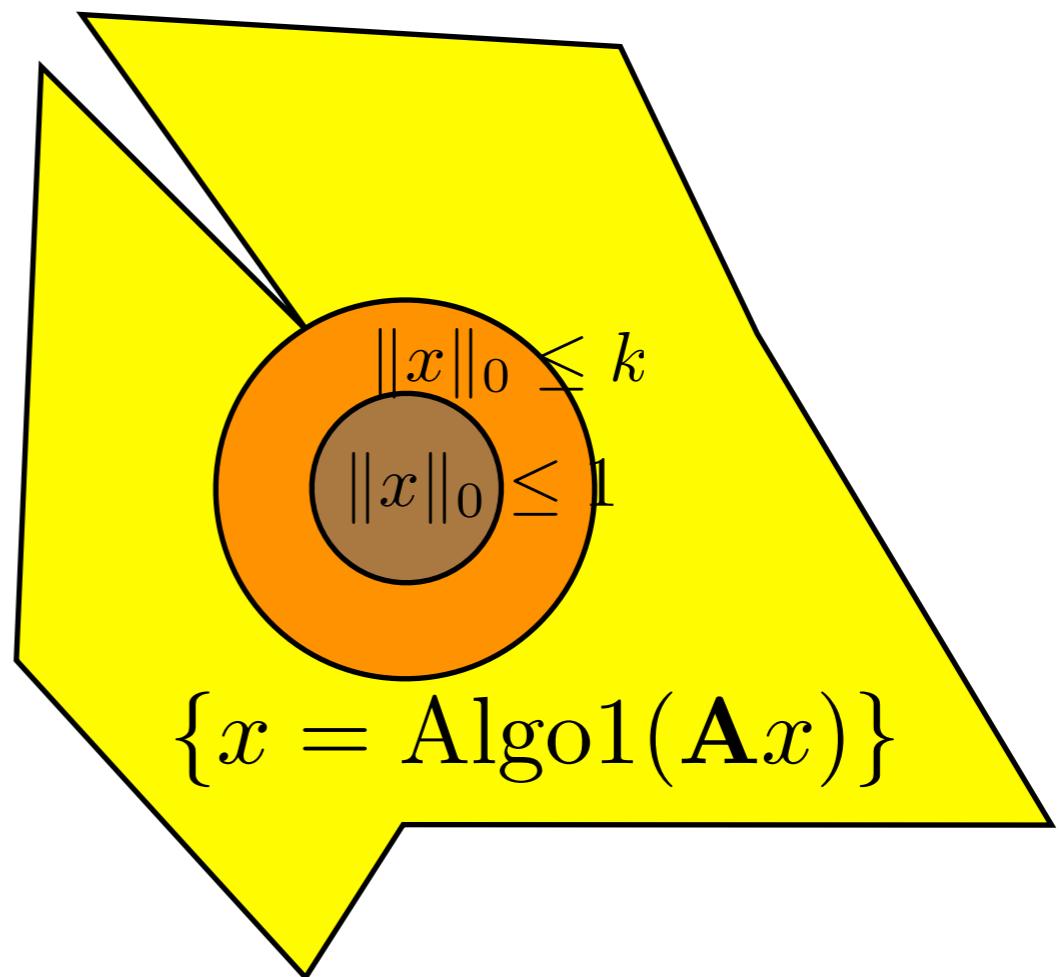
$$k_1(\mathbf{A}) \approx \frac{m}{2e \log N/m}$$

[Donoho & Tanner 2009]

# Excessive pessimism ?

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

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
- Worst case
  - ✓ = too pessimistic!



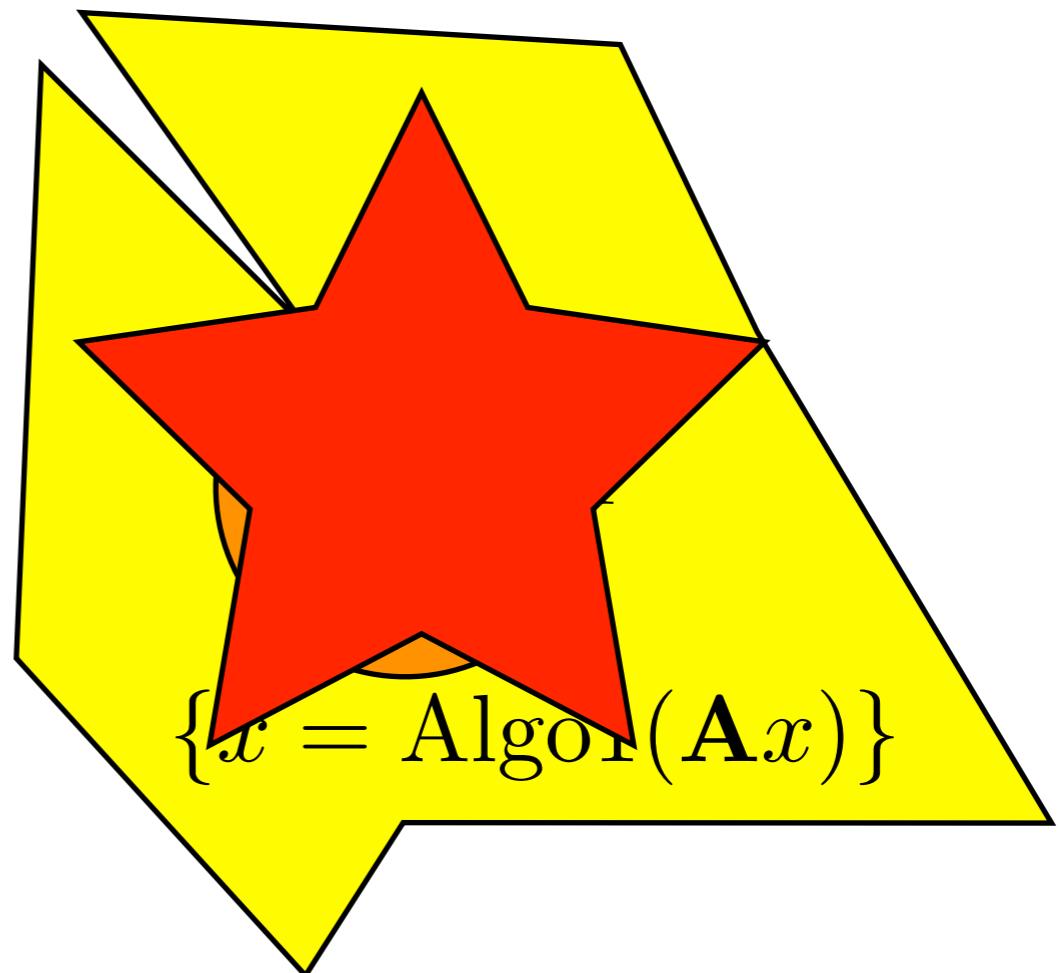
# Recovery analysis for inverse problem $\mathbf{b} = \mathbf{A}x$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
- Worst case
  - ✓ = too pessimistic!
- Finer “structures” of  $x$

Borup, G. & Nielsen ACHA 2008,  $\mathbf{A}$  = Wavelets U Gabor,  
recovery of infinite supports for analog signals

$\text{support}(x)$ ,  $\text{sign}(x)$

Fuchs 2005; Zhao & Yu 2006; Zou 2006; Yuan & Lin 2007;  
Wainwright 2009; Duarte & al 2009



# Recovery analysis

- Recoverable set for a given “inversion” algorithm  $\mathbf{b} = \mathbf{A}\mathbf{x}$
- Level sets of L0-norm
- Worst case  
= too pessimistic!
- Finer “structures” of  $\mathbf{x}$

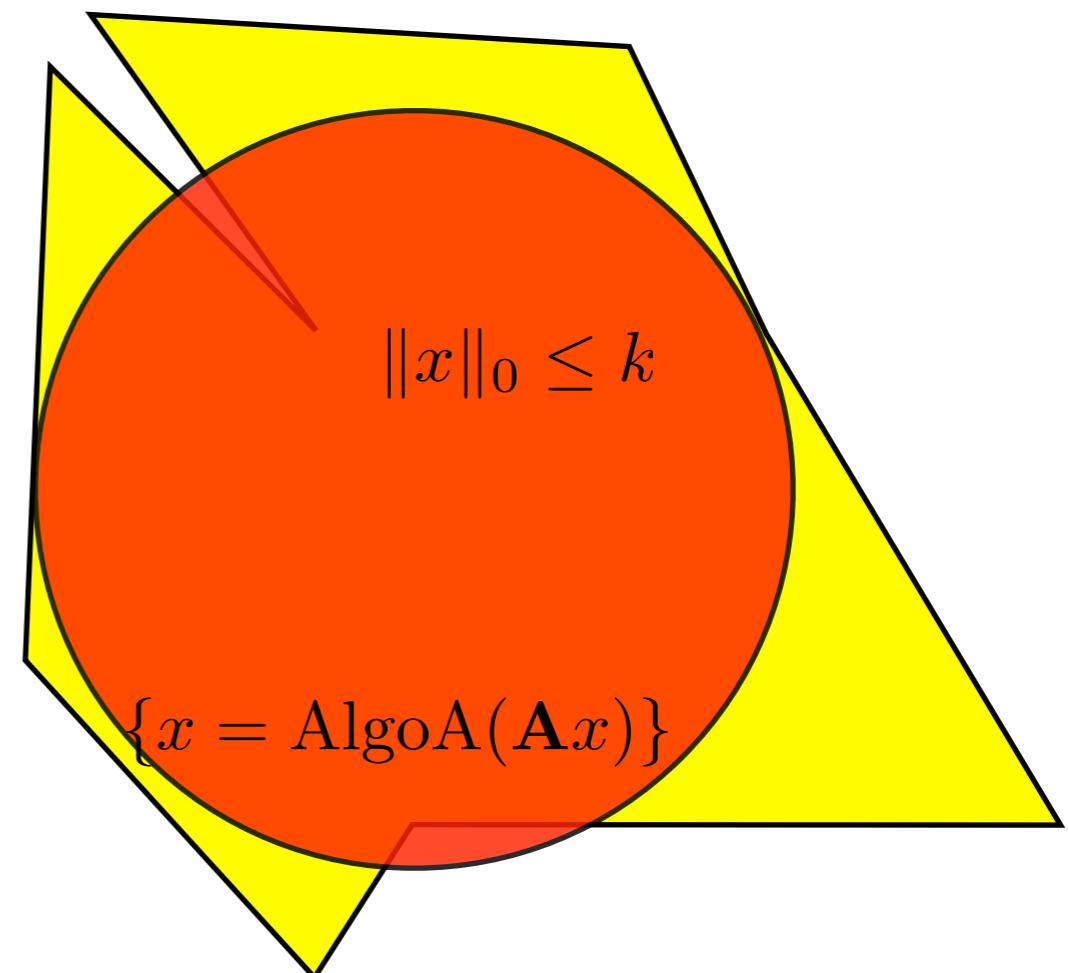
Borup, G. & Nielsen ACHA 2008,  $\mathbf{A}$  = Wavelets U Gabor,  
recovery of infinite supports for analog signals

$\text{support}(\mathbf{x})$ ,  $\text{sign}(\mathbf{x})$

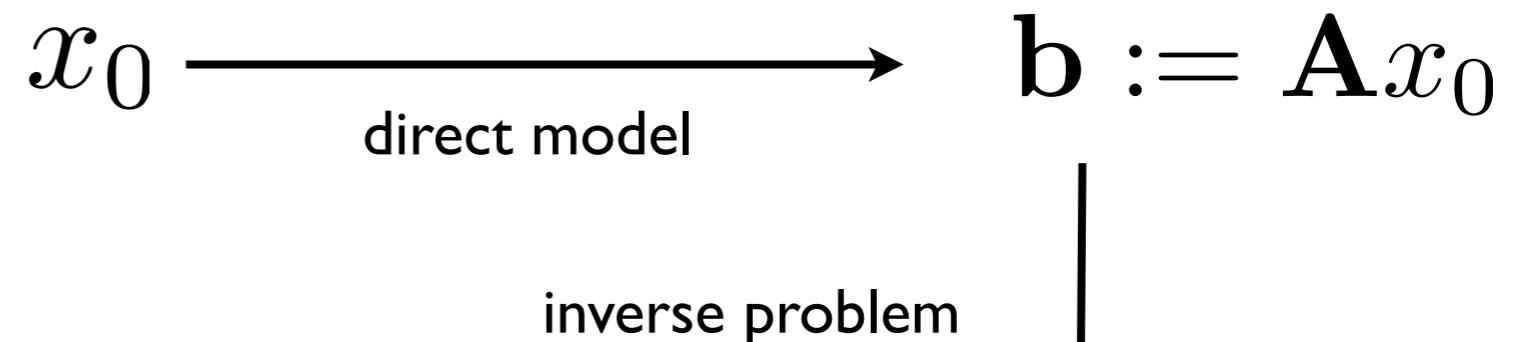
Fuchs 2005; Zhao & Yu 2006; Zou 2006; Yuan & Lin 2007;  
Wainwright 2009;

- Average/typical case

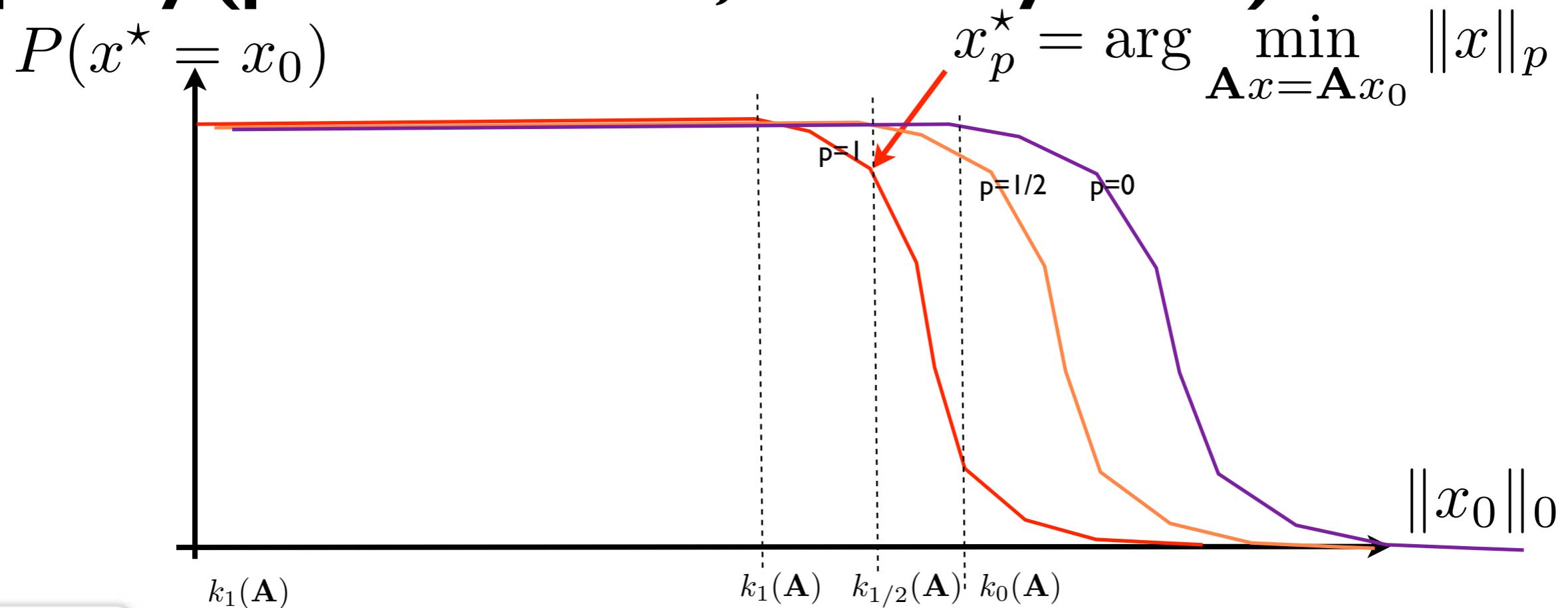
G., Rauhut,, Schnass & Vandergheynst, JFAA 2008,  
“Atoms of all channels, unite! Average case analysis of multichannel  
sparse recovery using greedy algorithms”.



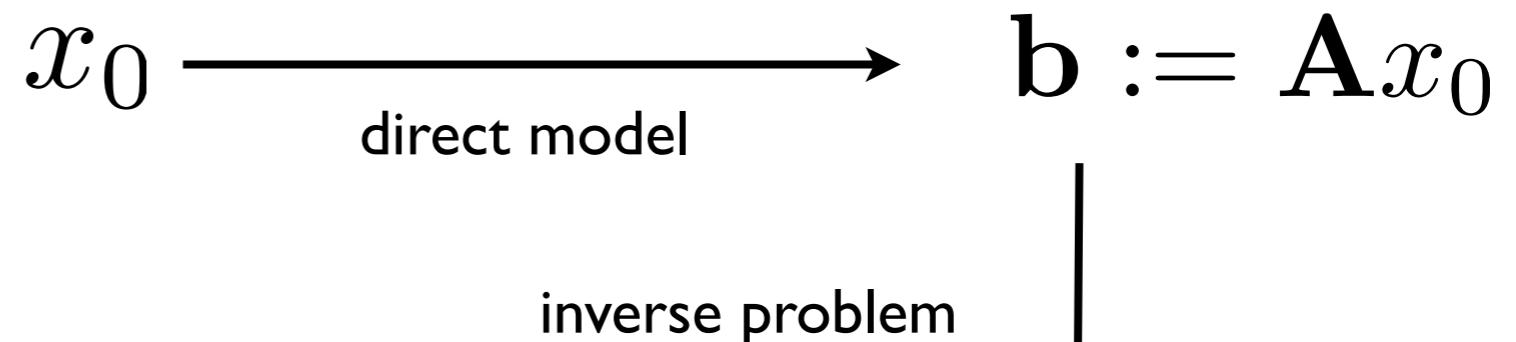
# Average case analysis ?



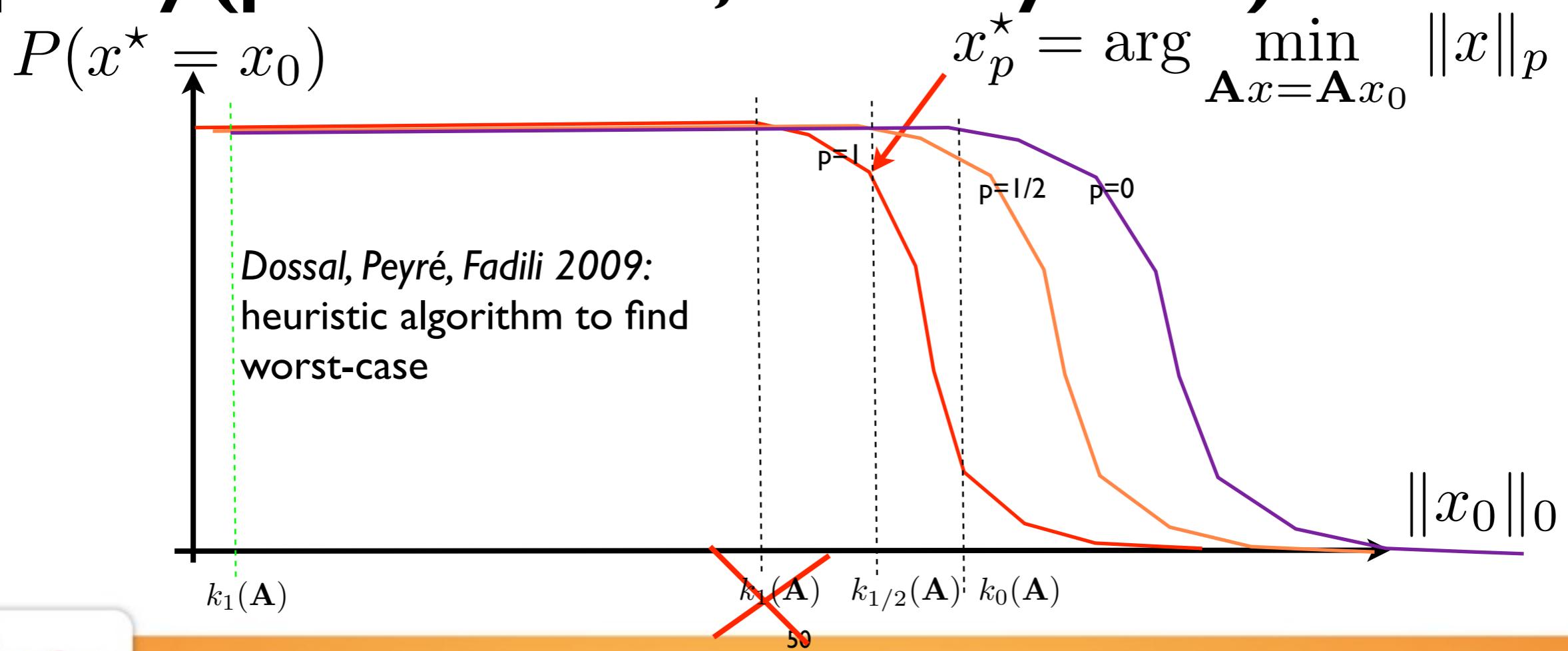
**Typically (pseudo-curves, drawn by hand!):**



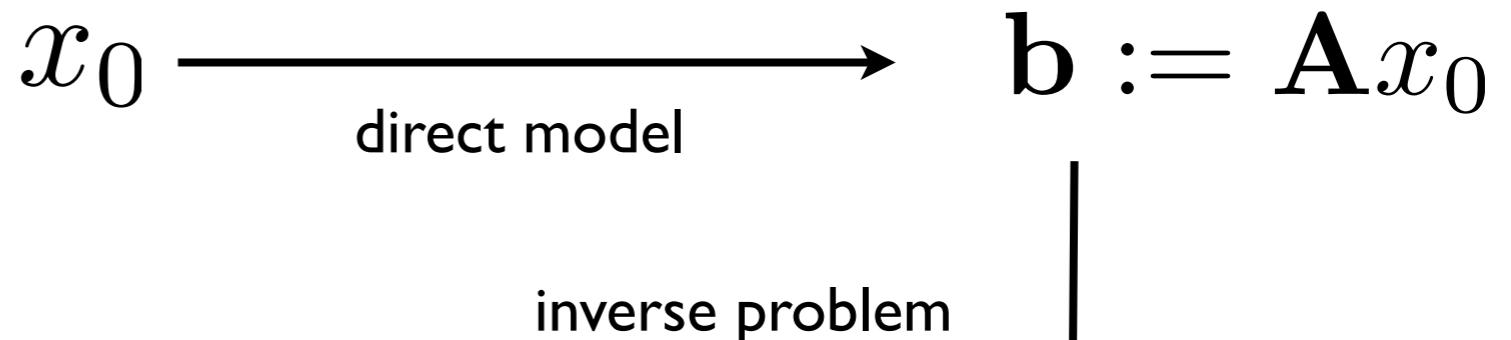
# Average case analysis ?



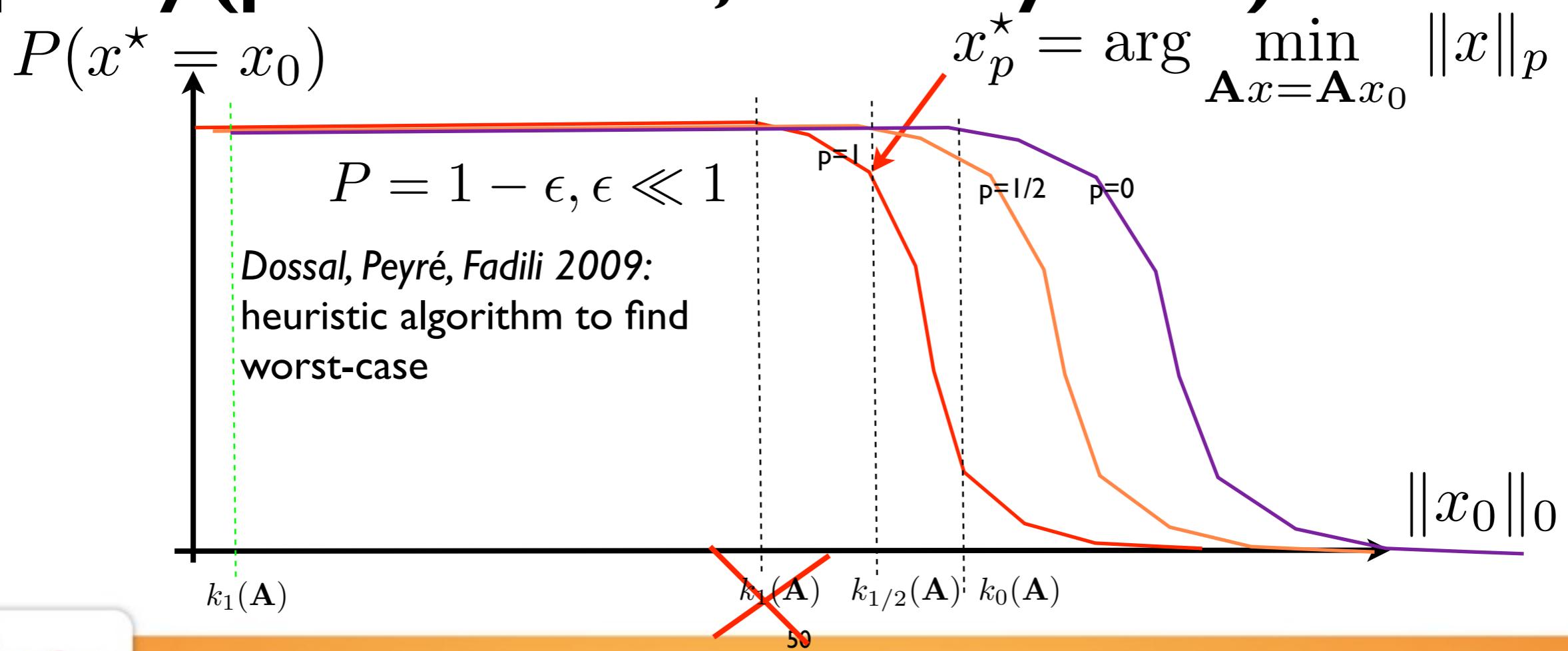
**Typically (pseudo-curves, drawn by hand!):**



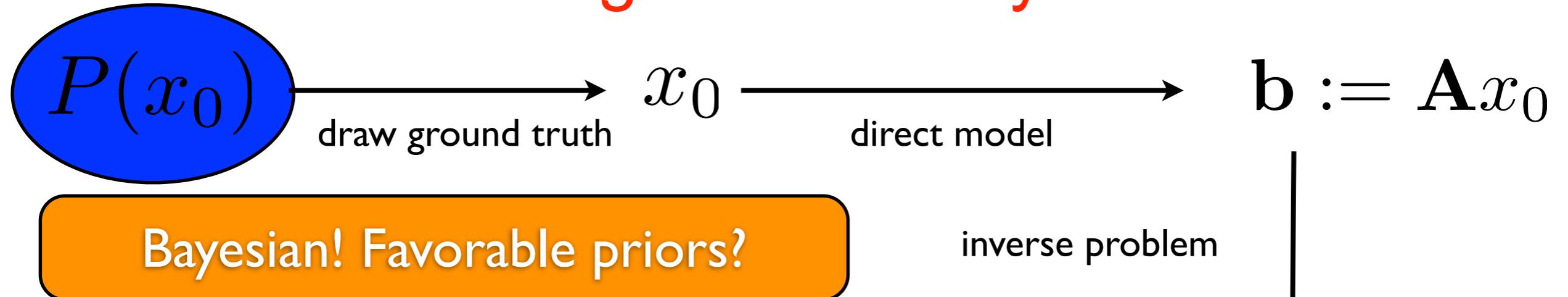
# Average case analysis ?



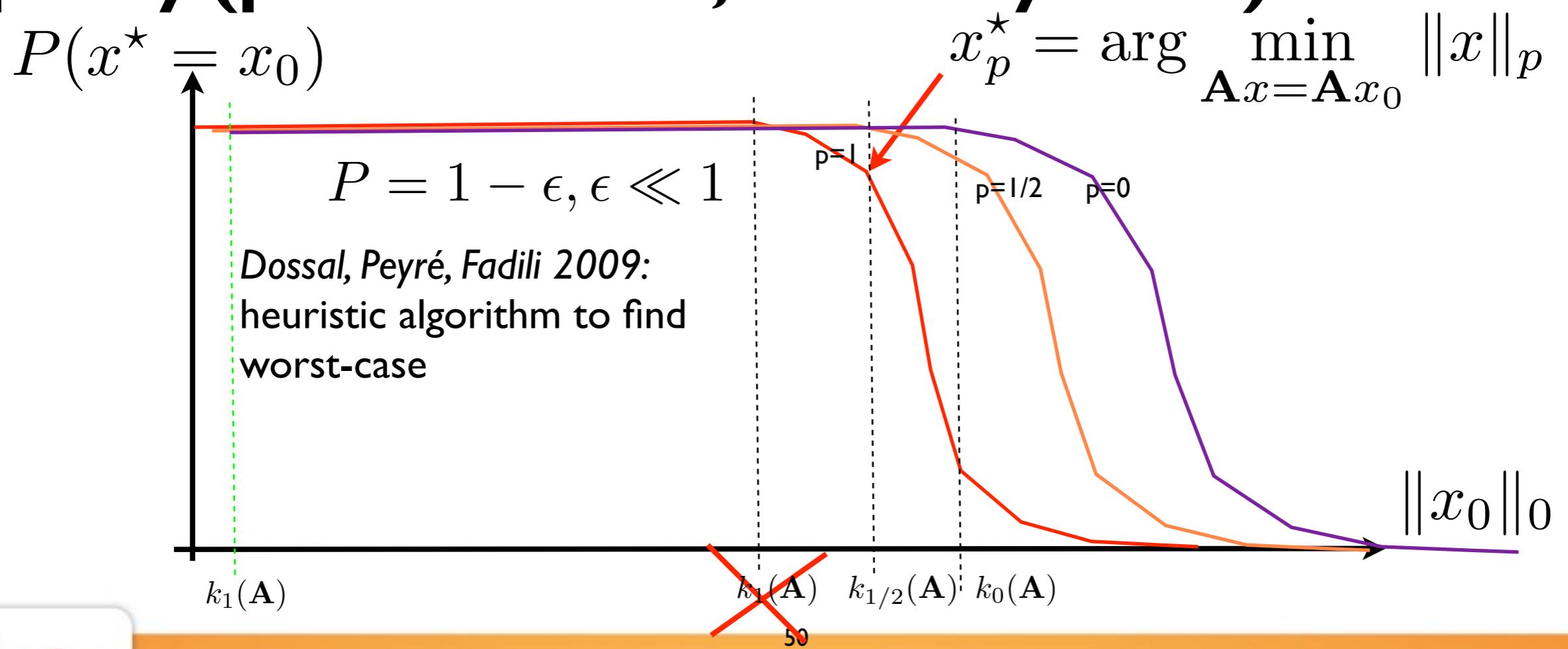
**Typically (pseudo-curves, drawn by hand!):**



# Average case analysis ?



**Typically (pseudo-curves, drawn by hand!):**



# Conclusions

- Sparsity: prior to solve **ill-posed inverse problems**
- If solution sufficiently sparse, **reasonable algorithms are guaranteed to find it** (even one step thresholding!).
- **Computational efficiency still a challenge**
  - ✓ problem sizes up to  $1000 \times 10000$  already efficiently tractable.
- Theoretical guarantees are mostly worst-case
  - ✓ Empirical recovery goes far beyond but is not fully understood.
- Challenging practical issues include:
  - ✓ choosing / learning / designing dictionaries;
  - ✓ exploiting structures beyond sparsity;
  - ✓ designing feasible compressed sensing hardware.

# Hot Topics, not covered in this tutorial

- Structured sparsity: group LASSO, etc.
- Combinatorial algorithms: submodular functions, etc.
- Approximate Message Passing algorithms
- Analysis vs synthesis sparsity
- Dictionary learning
- Low-rank matrices & sparsity

