

Performance of Sparse Decomposition Algorithms with Deterministic versus Random Dictionaries

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>



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



INRIA

centre de recherche

RENNES - BRETAGNE ATLANTIQUE

Summary

- Session 1:
 - ◆ role of sparsity for compression and inverse problems
- Session 2:
 - ◆ Review of main algorithms & complexities
 - ◆ Success guarantees for L1 minimization to solve under-determined inverse linear problems
- Session 3:
 - ◆ Robust guarantees & Restricted Isometry Property
 - ◆ Comparison of guarantees for different algorithms
 - ◆ Explicit guarantees for various inverse problems

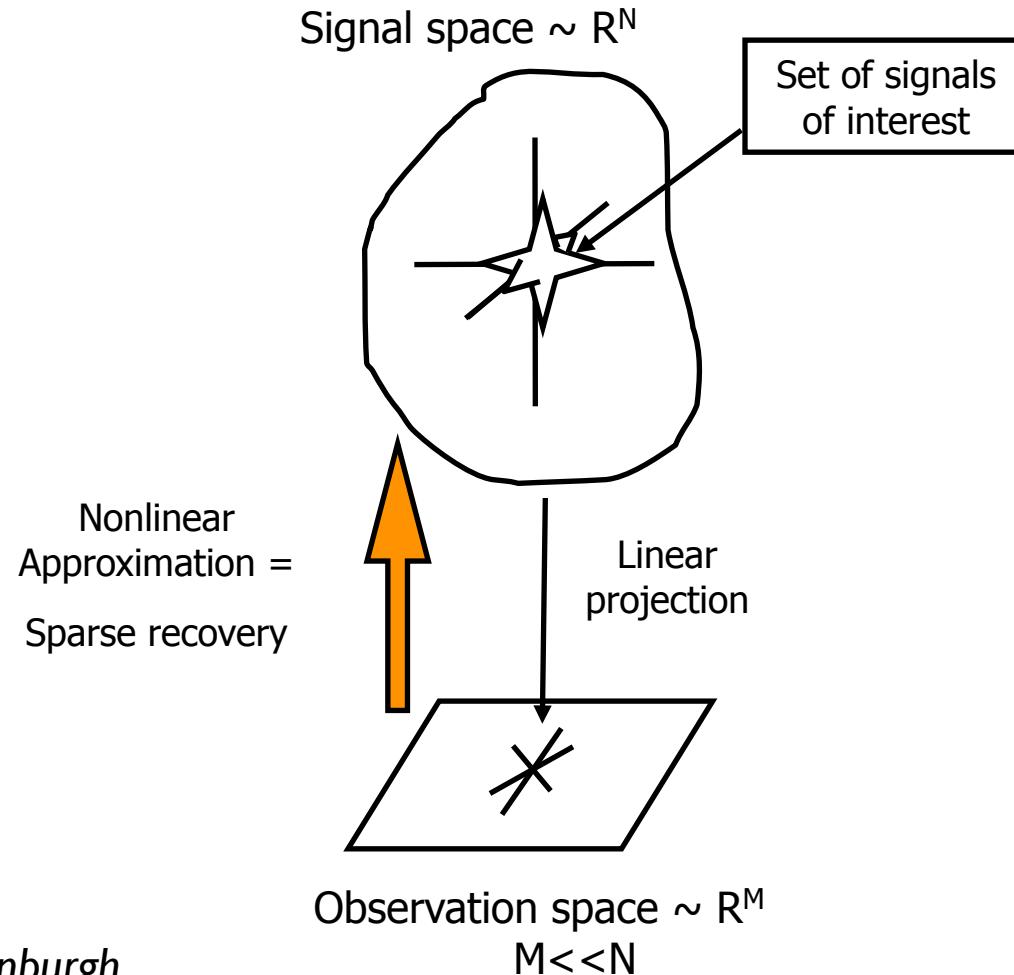


INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Inverse problems



Courtesy: M. Davies, U. Edinburgh

Stability and robustness



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



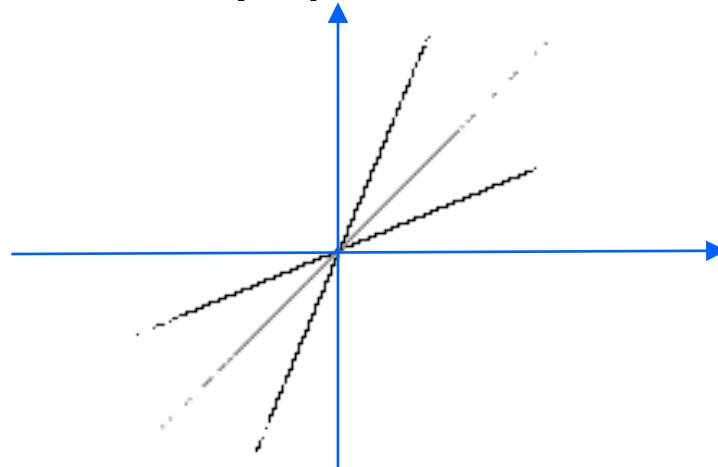
INRIA

centre de recherche

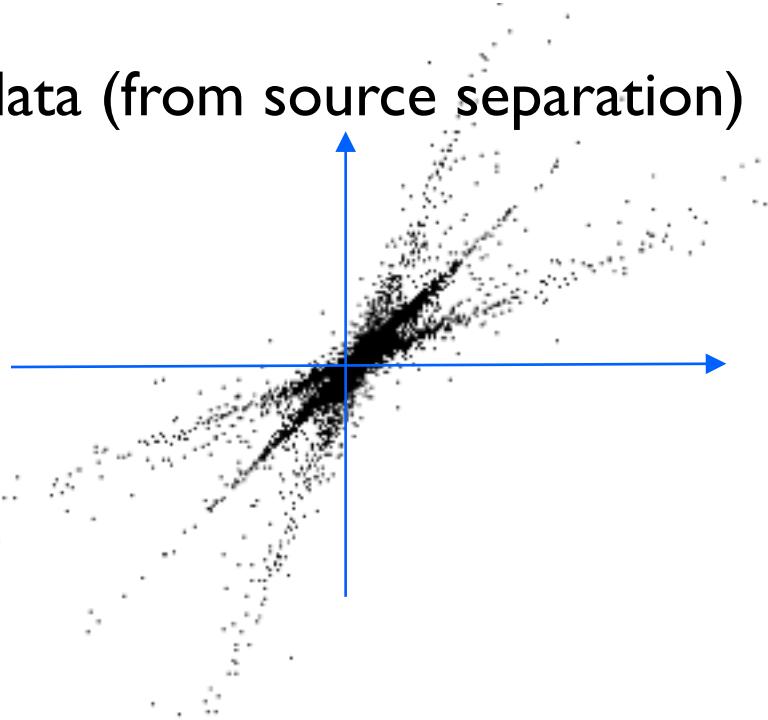
RENNES - BRETAGNE ATLANTIQUE

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*

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

[Cohen, Dahmen & De Vore 2006]



Stability for Lp minimization

- Assumption: «stable Null Space Property»

$\text{NSP}(k, \ell^p, t)$

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

- Conclusion: *instance optimality* for all x

$$\|x_p^\star(\mathbf{b}) - x\|_p^p \leq C(t) \cdot \sigma_k(x)_p^p$$

$$C(t) := 2 \frac{1+t}{1-t}$$

[Davies & Gribonval, SAMPTA 2009]



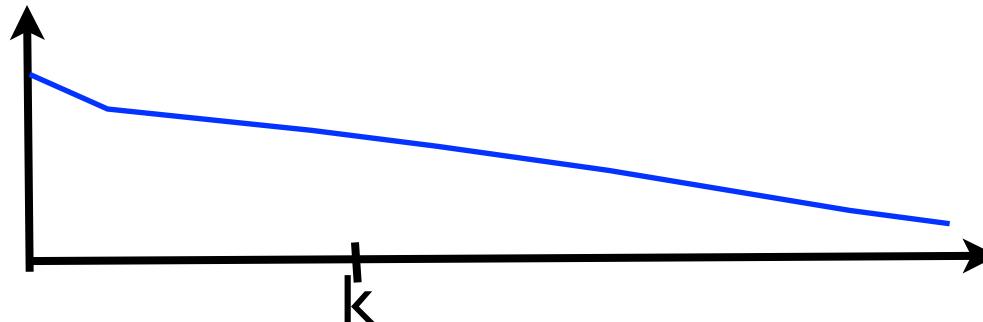
INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Reminder on NSP

- Geometry in coefficient space:
 - ◆ consider an element z of the Null Space of A
 - ◆ order its entries in decreasing order



- ◆ the mass of the largest k -terms should not exceed a fraction of that of the tail $\|z_{I_k}\|_p^p \leq t \cdot \|z_{I_k^c}\|_p^p$

All elements of the null space must be “flat”

Robustness

- Toy model = noiseless $\mathbf{b} = \mathbf{A}\mathbf{x}$
- Need to account for noise $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}$
 - ◆ measurement noise
 - ◆ modeling error
 - ◆ numerical inaccuracies ...
- Goal: predict robust estimation

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

- Tool: restricted isometry property

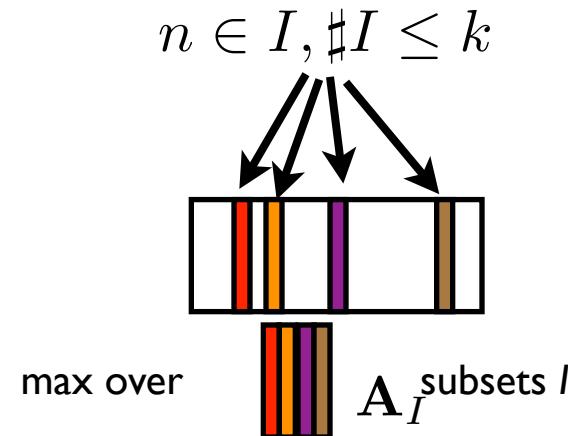
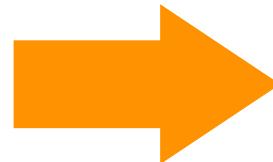


Restricted Isometry Property

- Definition



N columns



$$\frac{N!}{k!(N-k)!}$$

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

- Computation ?

- ◆ naively: combinatorial
- ◆ **open question:** NP ? NP-complete ?



Stability & robustness from RIP

RIP(k, δ)

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

[Candès 2008]



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

NSP(k, ℓ^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: Lp 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
- ♦ Needel & Tropp 2009, Blumensath & Davies 2009: RIP for greedy algorithms



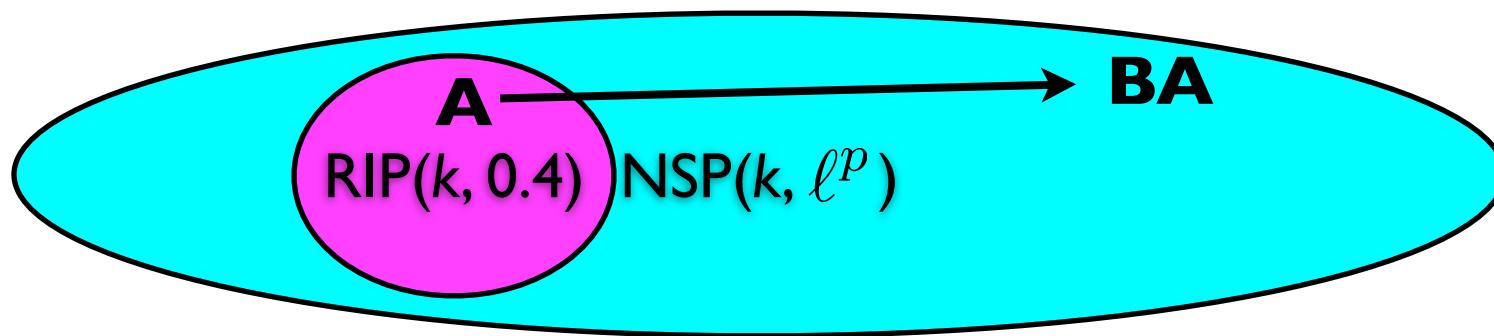
INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

 INRIA

centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Is the RIP a sharp condition ?

- The Null Space Property
 - ◆ “algebraic” + sharp property for L_p , only depends on $\mathcal{N}(A)$
 - ◆ invariant by linear transforms $A \rightarrow BA$
- 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]



Comparison between algorithms

- Recovery conditions based on number of nonzero components $\|x\|_0$ for $0 \leq q \leq p \leq 1$

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

Proof

- **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

[Gribonval & Nielsen, ACHA 2007]

Remaining agenda

- Recovery conditions based on number of nonzero components $\|x\|_0 \quad 0 \leq q \leq p \leq 1$

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

- **Question**
 - ◆ what is the order of magnitude of these numbers ?
 - ◆ how do we estimate them in practice ?
- A first element:
 - ◆ if \mathbf{A} is $m \times N$, then $k_0(\mathbf{A}) \leq \lfloor m/2 \rfloor$
 - ◆ for almost all matrices (in the sense of Lebesgue measure in \mathbb{R}^{mN}) this is an equality

Explicit guarantees in various inverse problems



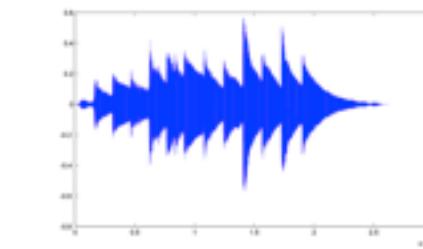
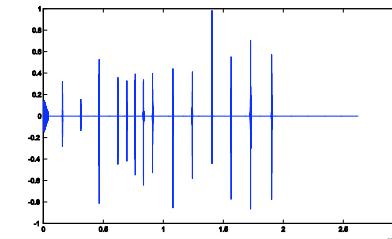
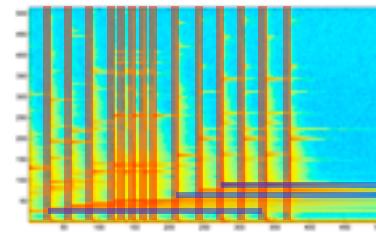
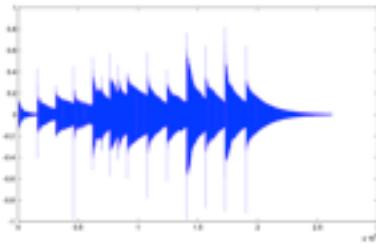
Scenarios

- 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 «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



Multiscale Time-Frequency Structures

- Audio = superimposition of structures
- Example : glockenspiel



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

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



Deterministic matrices and 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 $c \in \mathbb{R}^I$

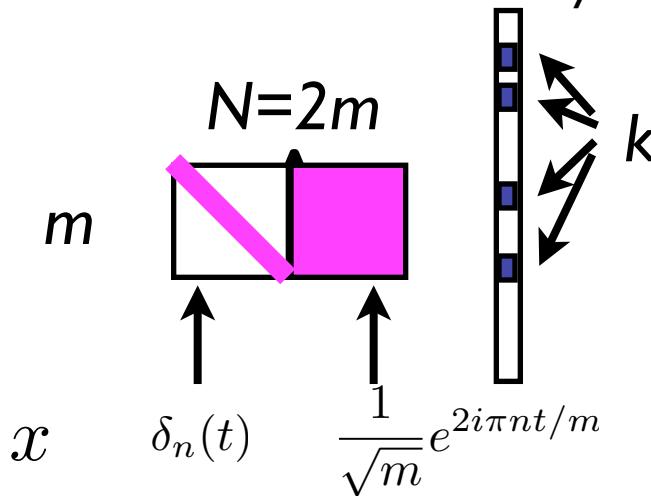
$$1 - (k - 1)\mu \leq \frac{\|\mathbf{A}_I c\|_2^2}{\|c\|_2^2} \leq 1 + (k - 1)\mu$$

- ◆ In other words $\delta_{2k} \leq (2k - 1)\mu$



Coherence vs RIP

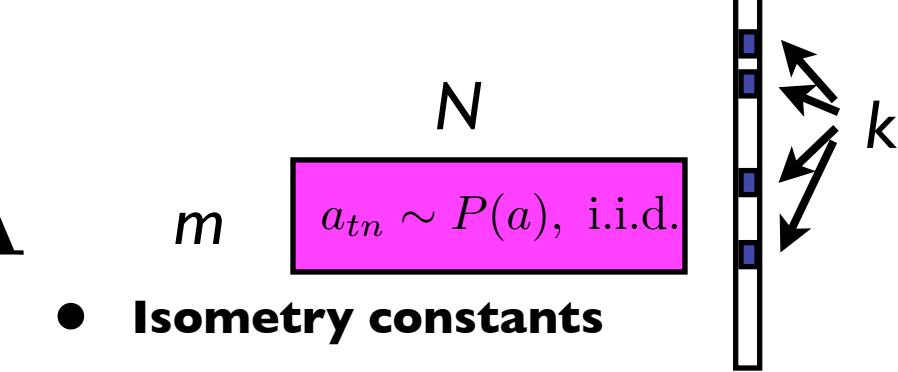
- **Deterministic** matrix, such as Dirac-Fourier dictionary



- **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

[Elad & Bruckstein 2002]

[Donoho & Tanner 2009]

Example: convolution operator

- Deconvolution problem with spikes

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

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

$$\mathbf{A}_n(i) = h(i - n) \quad \text{by convention } \|\mathbf{A}_n\|_2^2 = \sum_i h(i)^2 = 1$$

- ◆ 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 \mathbf{A} and adjoint.

Example: image inpainting

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



Inpainting



$$\mathbf{b} = \mathbf{M}\mathbf{y} = \mathbf{M}\Phi\mathbf{x}$$

Wavelets
 $y = \Phi x$



Compressed sensing

- Approach = acquire some data y with a limited number m of (linear) measures, modeled by a measurement matrix $\mathbf{b} \approx \mathbf{K}y$
- Key hypotheses
 - ◆ Sparse model: the data can be sparsely represented in a known dictionary $y \approx \Phi x \quad \sigma_k(x) \ll \|x\|$
 - ◆ The overall matrix $\mathbf{A} = \mathbf{K}\Phi$ leads to robust + stable sparse recovery, e.g. $\delta_{2k}(\mathbf{A}) \ll 1$
- Reconstruction = sparse recovery algorithm



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Key constraints to use Compressed Sensing

- Availability of sparse model= dictionary Φ
 - ◆ should fit well the **data**, not always granted. E.g.: cannot acquire white Gaussian noise!
 - ◆ require appropriate choice of dictionary, or **dictionary learning from training data**
- Measurement matrix K
 - ◆ must be associated with **physical sampling process** (hardware implementation ... designed aliasing?)
 - ◆ should guarantee **recovery** from $K\Phi$ through incoherence
 - ◆ should ideally enable fast algorithms through **fast computation** of $Ky, K^T b$



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

Remarks

- 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]?
 - ◆ Transmission is lossy
(robust to loss of a few measures)

Excessive pessimism ?



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

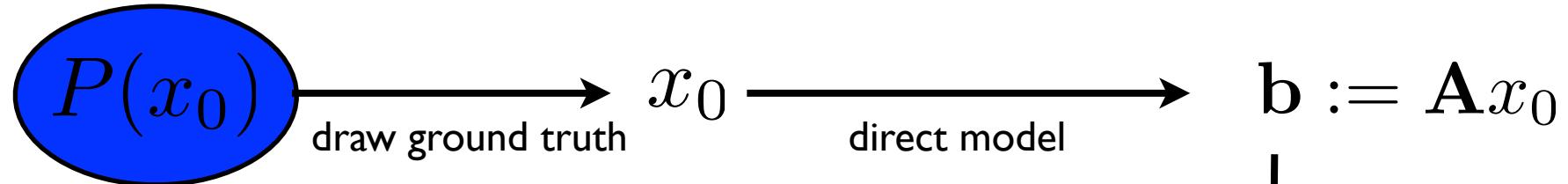


INRIA

centre de recherche

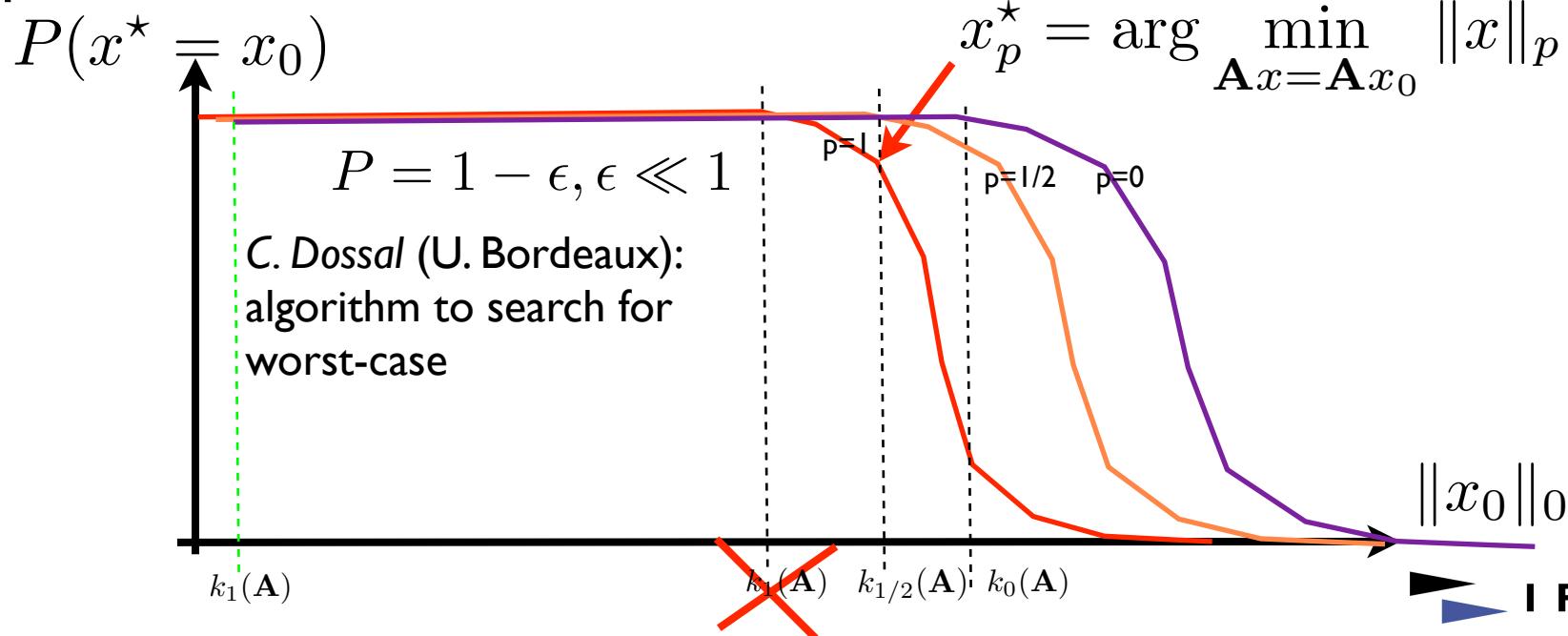
RENNES - BRETAGNE ATLANTIQUE

Average case analysis ?



Bayesian! Favorable priors?

Typical observation



Conclusions

- Sparsity helps solve ill-posed inverse problems (more unknowns than equations).
- If the solution is sufficiently sparse, any reasonable algorithm will find it (even simple thresholding!).
- Computational efficiency is still a challenge, but problem sizes up to 1000×10000 already tractable efficiently.
- Theoretical guarantees are mostly worst-case, empirical recovery goes far beyond but is not fully understood!
- Challenging practical issues include:
 - ◆ choosing / learning / designing dictionaries;
 - ◆ designing feasible compressed sensing hardware.



Thanks to

- F. Bimbot, G. Gonon, S. Krstulovic, B. Roy
- A. Ozerov, S. Lesage, B. Mailhé
- M. Nielsen, L. Borup (Aalborg Univ.)
- P. Vandergheynst, R. Figueras, P. Jost, K. Schnass (EPFL)
- H. Rauhut (U. Vienna)
- M. Davies (U. Edinburgh)
- and several other collaborators ...

The end

remi.gribonval@inria.fr

www.irisa.fr/metiss/gribonval



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche
RENNES - BRETAGNE ATLANTIQUE

LI vs Lp



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



centre de recherche

RENNES - BRETAGNE ATLANTIQUE

L_p better than L_1 (I)

- **Theorem 2** [G. Nielsen 2003]
 - ◆ Assumption 1: **sub-additivity** of sparsity measures f, g

$$f(a + b) \leq f(a) + f(b), \forall a, b$$
 - ◆ Assumption 2: the function $t \mapsto \frac{f(t)}{g(t)}$ is **non-increasing** ↘
 - ◆ Conclusion: $k_g(\mathbf{A}) \leq k_f(\mathbf{A}), \forall \mathbf{A}$

Minimizing $\|x\|_f$ can recover vectors which are less sparse than required for guaranteed success when minimizing $\|x\|_g$



L_p better than L₁ (2)

- **Example**

- ◆ sparsity measures $f(t) = t^p, g(t) = t^q, 0 \leq p \leq q \leq 1$

- ◆ sub-additivity

$$|a + b|^p \leq |a|^p + |b|^p, \forall a, b, 0 \leq p \leq 1$$

- ◆ function $\frac{f(t)}{g(t)} = t^{p-q}$ is non-increasing

- ◆ therefore

$$k_1(\mathbf{A}) \leq k_q(\mathbf{A}) \leq k_p(\mathbf{A}) \leq k_0(\mathbf{A}), \forall \mathbf{A}$$



L_p better than L_1 : proof

- 1) Since f/g non-decreasing:

$$z_1 \geq z_2 \geq 0$$



$$\frac{f(z_1)}{g(z_1)} \leq \frac{f(z_2)}{g(z_2)}$$

- 2) Similarly

$$z_1 \geq \dots \geq z_N \geq 0$$



$$\frac{\|z_{1:k}\|_f}{\|z_{1:k}\|_g} \leq \frac{\|z_{k+1:N}\|_f}{\|z_{k+1:N}\|_g}$$

$$I_k = 1 : k$$

$$\frac{\|z_{I_k}\|_f}{\|z_{I_k^c}\|_f} \leq \frac{\|z_{I_k}\|_g}{\|z_{I_k^c}\|_g}$$

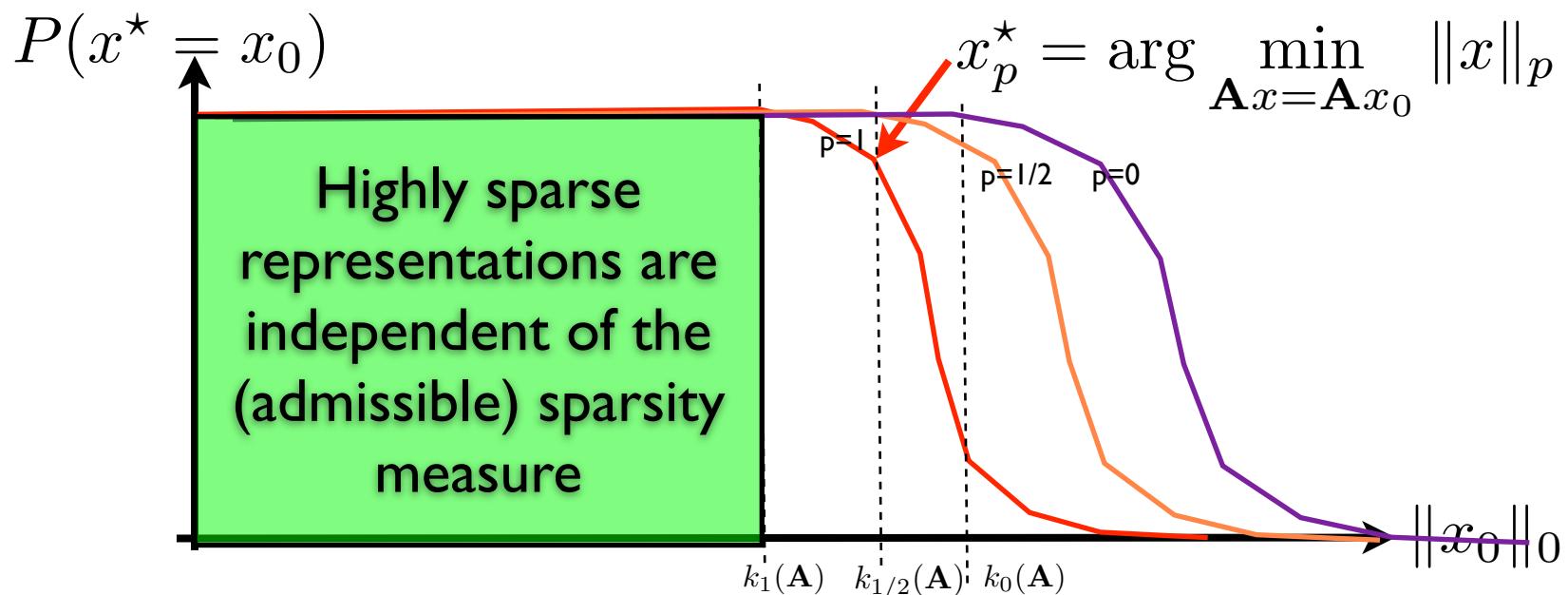
$$I_k^c = k + 1 : N$$

- 3) Conclusion : if $\text{NSP}(g, t, k)$ then $\text{NSP}(f, t, k)$



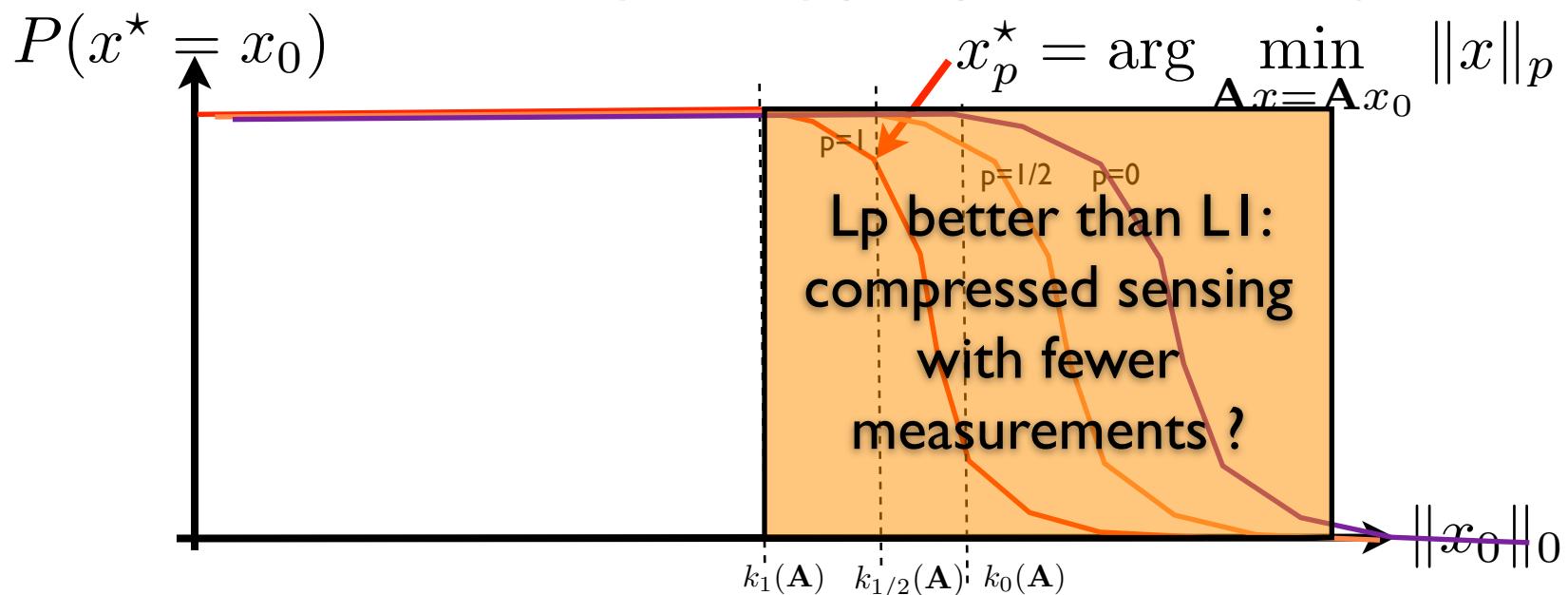
L_p better than L_1 (3)

- At sparsity levels where L_1 is guaranteed to “succeeds”, all L_p $p \leq 1$ is also guaranteed to succeed



L_p better than L_1 (4)

- + L_p $p < 1$ can succeed where L_1 fails
 - ◆ How much improvement ? Quantify $k_p(\mathbf{A})$?
- - L_p $p < 1$: nonconvex, has many local minima
 - ◆ Better recovery with L_p principle, what about *algorithms* ?



When does $\delta_{2k}(\mathbf{A}) < \delta$ imply $k \leq k_p(\mathbf{A})$?

