Inverse problems, Dictionary based Signal Models and Compressed Sensing

Rémi Gribonval
METISS project-team (audio signal processing, speech recognition, 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 de Peyresq
Peyresq, Juillet 2009

UNIVERSITÉ DE RENNES 1 · INSA · CNRS · INRIA RENNES

Agenda

- Recovery conditions based on number of nonzero components $\|x\|_0$

$$k_{*\mathrm{MP}}(\mathbf{A}) \leq k_1(\mathbf{A}) \leq k_p(\mathbf{A}) \leq k_q(\mathbf{A}) \leq k_0(\mathbf{A}), \forall \mathbf{A}$$

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

Scenarios

- Range of "choices" for the matrix **A**
  - imposed by physics of inverse problems (ex: convolution operator)
  - chosen signal dictionary for sparse modeling (ex: union of wavelets + curvelets + spikes)
  - designed Compressed Sensing matrix (ex: random Gaussian matrix)
- Estimation of the recovery regimes
  - coherence for deterministic matrices
  - typical results for random matrices

Deterministic matrices and coherence

- **Lemma**
  - Assume normalized columns $\|\mathbf{A}_i\|_2$
  - Define coherence $\mu = \max_{i \neq j} |\mathbf{A}_i^T \mathbf{A}_j|$

  - Consider index set I of size $\sharp I \leq k$
  - 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 $\delta_{2k} \leq (2k-1)\mu$

Consequence

- Since $\delta_{2k} \leq \mu \cdot (2k-1)$ we obtain $\delta_{2k} \leq \delta$ as soon as

$$k < (1 + \delta/\mu)/2$$

- Combining with best known RIP condition for stable L1 recovery $\delta \approx 0.4531$

$$k_1(\mathbf{A}) \geq \left\lfloor \left(1 + 0.4531/\mu\right)/2 \right\rfloor$$

- In fact, can prove with other techniques that

$$k_0(\mathbf{A}) \geq k_1(\mathbf{A}) \geq \left\lfloor \left(1 + 1/\mu\right)/2 \right\rfloor$$

[G. Nielsen 2003]

Observation

- Assume the *m* x *N* matrix A has normalized columns and contains an orthonormal basis
- Then its coherence is at least

$$\mu \geq \frac{1}{\sqrt{m}}$$

- The bounds are therefore, at best, of the order

$$\left\lfloor \left(1+\sqrt{m}\right)/2 \right\rfloor \leq k_1(\mathbf{A}) \leq k_0(\mathbf{A}) \leq \left\lfloor m/2 \right\rfloor$$

Example : Dirac-Fourier dictionary

- Fourier matrix in dimension $m = r^2$

$$\mathbf{F}_m = \frac{1}{\sqrt{m}} \cdot (\exp(-2i\pi kn))_{0 \leq k,n < m}$$

- Dirac comb is *r*-sparse $c[n] = \sum_{\ell=0}^{r-1} \delta_{\ell r}[n]$
- Poisson formula $\mathbf{F}_m c = c$
- Dictionary $\mathbf{A} = [\mathbf{Id}_m, -\mathbf{F}_m]$
  - null space element $z = [c, c]$ has *r* nonzero entries, all of equal magnitude.
  - for *k* = *r*+*1*, and *I* a set with *k* nonzero entries of z:

$$\|z_I\|_1 = r + 1 > \|z_{I^c}\|_1 = r - 1$$

  - It follows that

$$k_1(\mathbf{A}) \leq r = \sqrt{m} = \frac{1}{\mu}$$

Example: convolution operator

- Deconvolution problem $y = h \star s + e$
  - ✦ re-expressed in matrix-vector form as $\mathbf{b} = \mathbf{A}x + \mathbf{e}$
  - ✦ **A** = Toeplitz or circulant matrix $[\mathbf{A}_1, \ldots, \mathbf{A}_N]$
  $$\mathbf{A}_n(i) = h(i-n)$$
  - ✦ convention
  $$\|\mathbf{A}_n\|_2^2 = \sum_i h(i)^2 = 1$$
  - ✦ coherence: given by 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 results
    - ✤ worst case = close spikes, usually difficult and not robust
    - ✤ results assuming distance between spikes *[Dossal]*

Example: source separation

- Time-domain model $\mathbf{b}(t) = \mathbf{A}x(t), \forall t$
- Time-frequency domain model (STFT)

$$\mathbf{B}(t, f) = \mathbf{A}X(t, f), \forall t, f$$

- Minimum Lp solution *[Bofill & Zibulevsky, Vincent]*

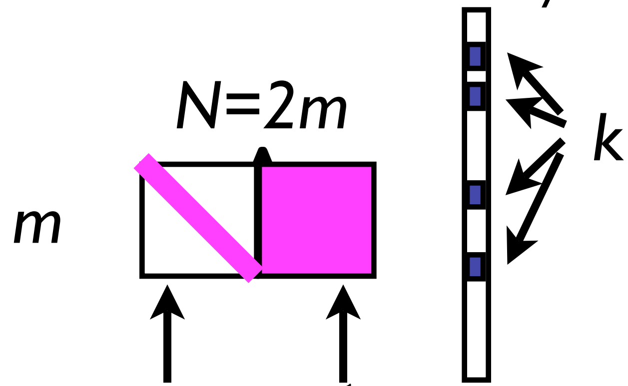
$$\hat{X}(t, f) = \arg\min \|X(t, f)\|_p$$

- Reconstruction (inverse STFT) $\hat{x}(t)$
- 2x3 case (stereophonic, three sources)
  - 1-dimensional null space, compute NSP constants
  - instance optimality guarantees: $\|\hat{X}(t, f) - X(t, f)\| \leq C\sigma_1(X(t, f))$

IRISA

Random matrix scenario

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



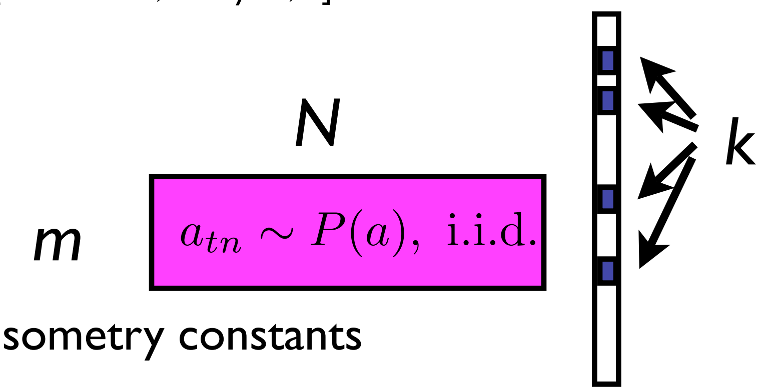
x $\delta_n(t)$ $\frac{1}{\sqrt{m}} e^{2i\pi nt/m}$

- Coherence

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

A

- "**Generic**" (random) dictionary
[Candès & al, Vershynin, ...]



- Isometry constants

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

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

Recovery regimes

$$k_1(\mathbf{A}) \approx 0.914\sqrt{m}$$
$$k_{*\mathrm{MP}}(\mathbf{A}) \geq 0.5\sqrt{m}$$

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

[Elad & Bruckstein 2002] *[Donoho & Tanner 2009]*

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 some dictionary $y \approx \mathbf{\Phi}x$ $\sigma_k(x) \ll \|x\|$
  - ✦ The overall matrix $\mathbf{A} = \mathbf{K}\mathbf{\Phi}$ leads to robust + stable sparse recovery, e.g. $\delta_{2k}(\mathbf{A}) \ll 1$
- Reconstruction = sparse recovery algorithm

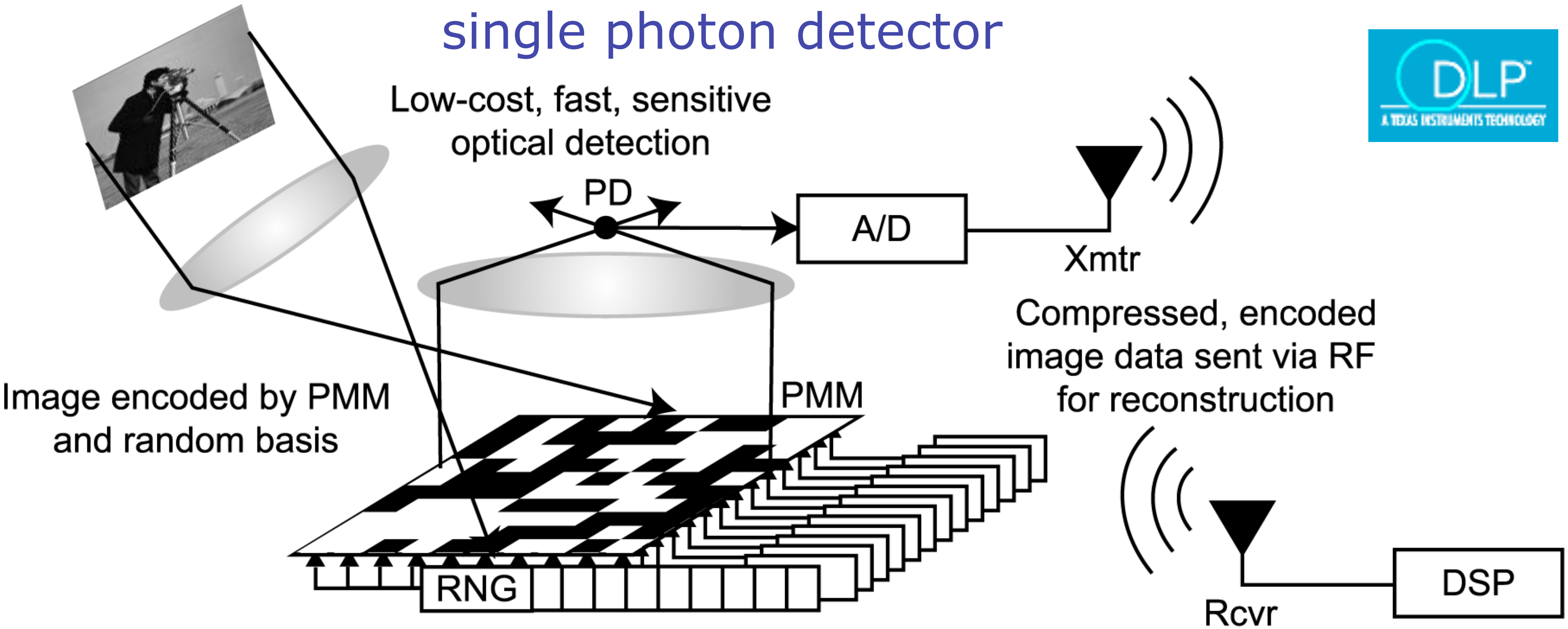
IRISA

Compressed Sensing

- Sparse model: Φ (synthesis .... or analysis)
  - should fit well the **data**, not always granted. E.g.: cannot aquire white Gaussian noise!
  - require appropriate *choice* of dictionary, or **dictionary learning from training data**
- Measurement matrix $\mathbf{K}$
  - must be associated with **physical sampling process** (hardware implementation)
  - should guarantee **recovery** from $\mathbf{K}\Phi$
  - should ideally enable fast algorithms through **fast computation** of $\mathbf{K}y, \mathbf{K}^T\mathbf{b}$

Example : Rice University Single Pixel Camera

single photon detector



Random pattern on DMD array

image reconstruction

IRISA

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

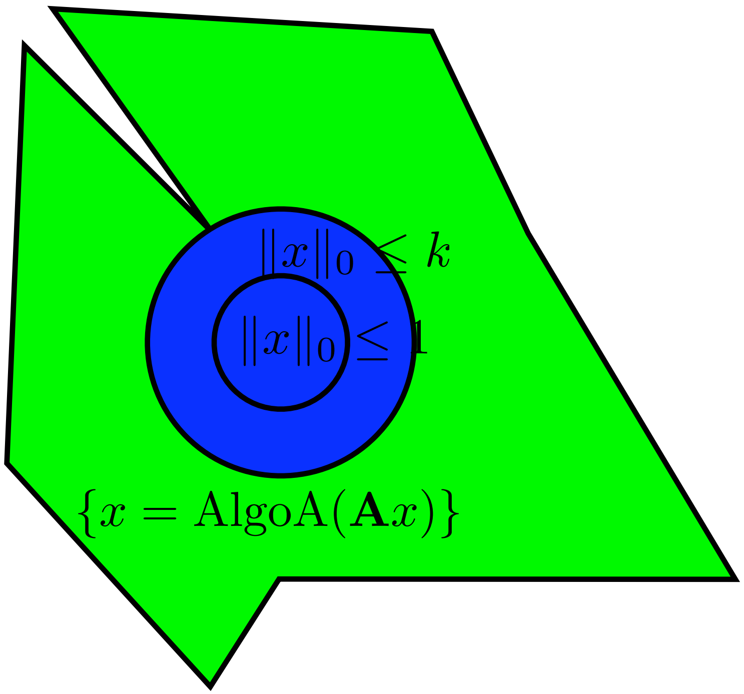
  (robust to loss of a few measures)

Excessive pessimism ?

Recovery analysis $\mathrm{b} = \mathbf{A}x$

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

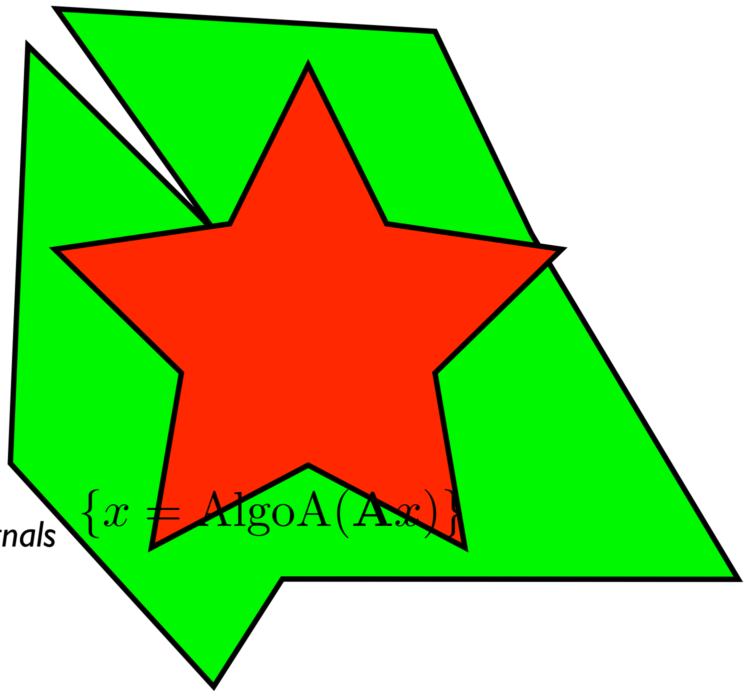
= too pessimistic!



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

- Recoverable set for a given "inversion" algorithm
- Level sets of L0-norm
- Worst case
  = too pessimistic!
- Finer "structures" of x
  $\mathrm{support}(x), \mathrm{sign}(x)$

Borup, G. & Nielsen ACHA 2008,
***A** = Wavelets U Gabor, recovery of infinite supports for analog signals*

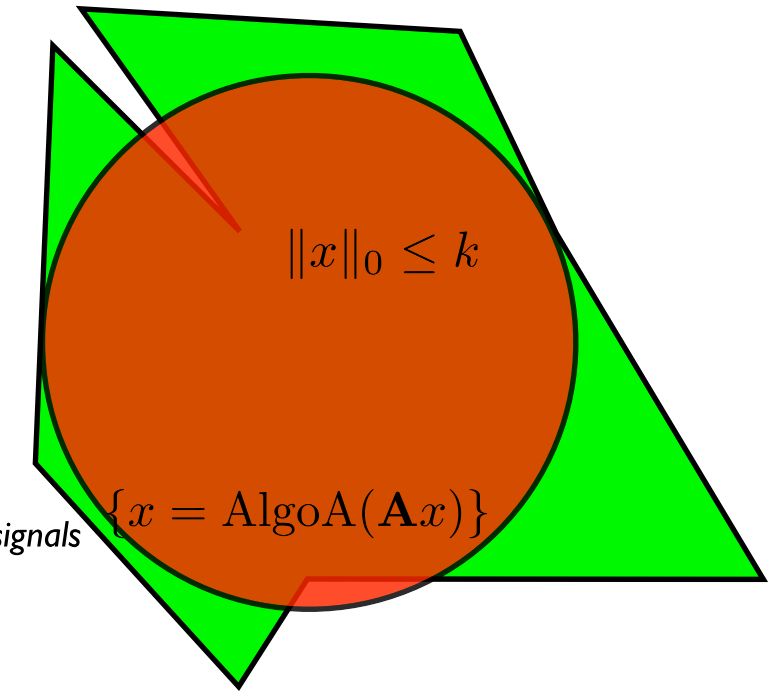


IRISA

Recovery analysis $\mathrm{b} = \mathbf{A}x$

- Recoverable set for a given “inversion” algorithm
- Level sets of L0-norm
- Worst case
  = too pessimistic!
- Finer “structures” of x
  $\mathrm{support}(x), \mathrm{sign}(x)$

Borup, G. & Nielsen ACHA 2008,
A *= Wavelets U Gabor, recovery of infinite supports for analog signals*



- 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 ?

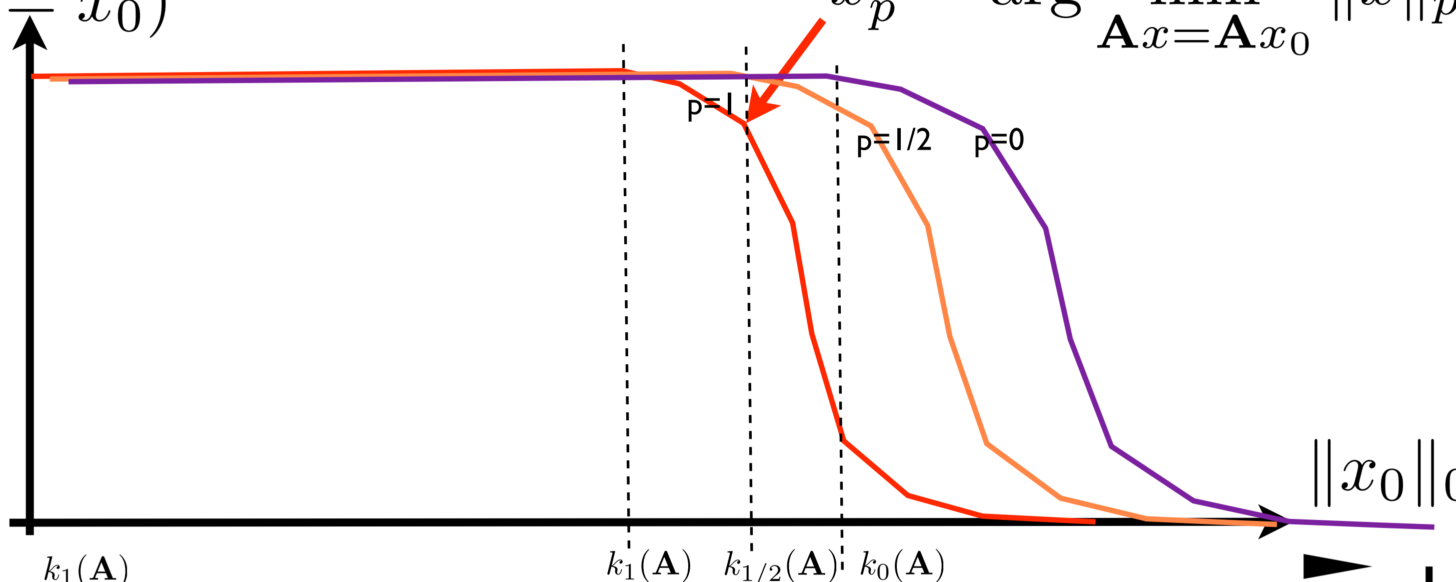
$$x_0 \xrightarrow{\text{direct model}} \mathbf{b} := \mathbf{A}x_0$$

inverse problem

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

Typical observation

$P(x^\star = x_0)$



Average case analysis ?

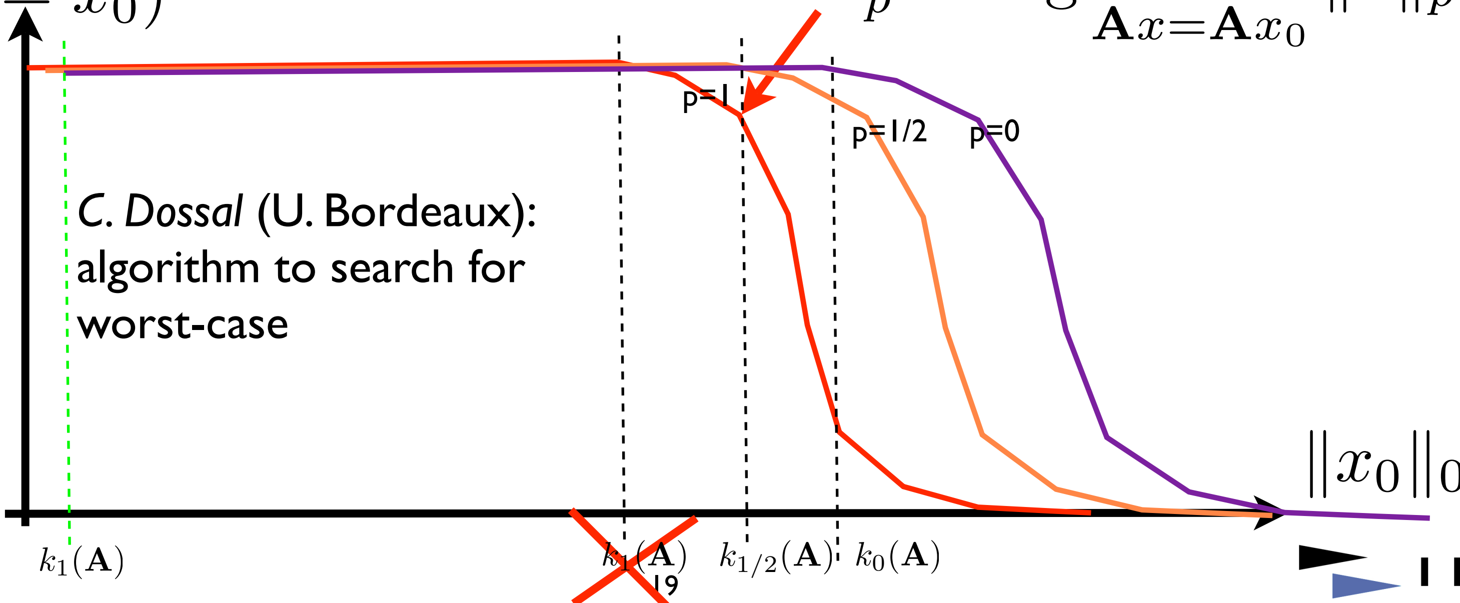
$$x_0 \xrightarrow{\text{direct model}} \mathbf{b} := \mathbf{A}x_0$$

inverse problem

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

Typical observation

$P(x^\star = x_0)$



C. Dossal (U. Bordeaux): algorithm to search for worst-case

Average case analysis ?

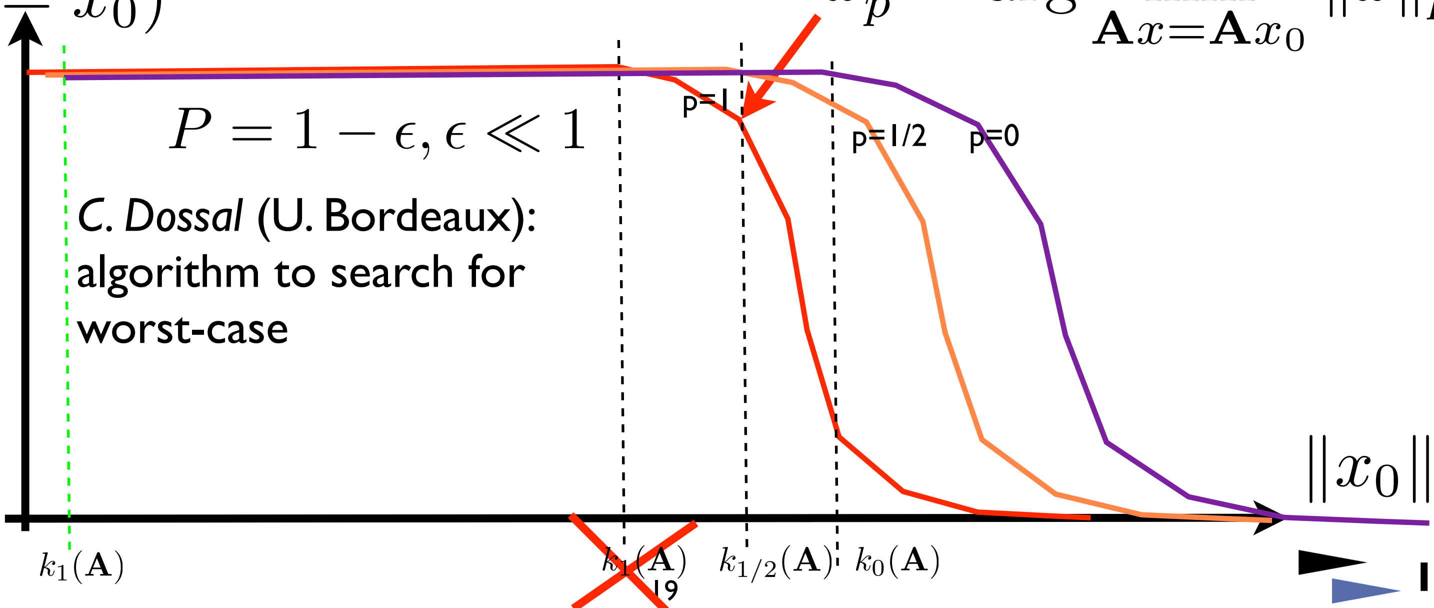
$$x_0 \xrightarrow{\text{direct model}} \mathbf{b} := \mathbf{A}x_0$$

inverse problem

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

Typical observation

$P(x^\star = x_0)$



C. Dossal (U. Bordeaux): algorithm to search for worst-case

IRISA

Average case analysis ?

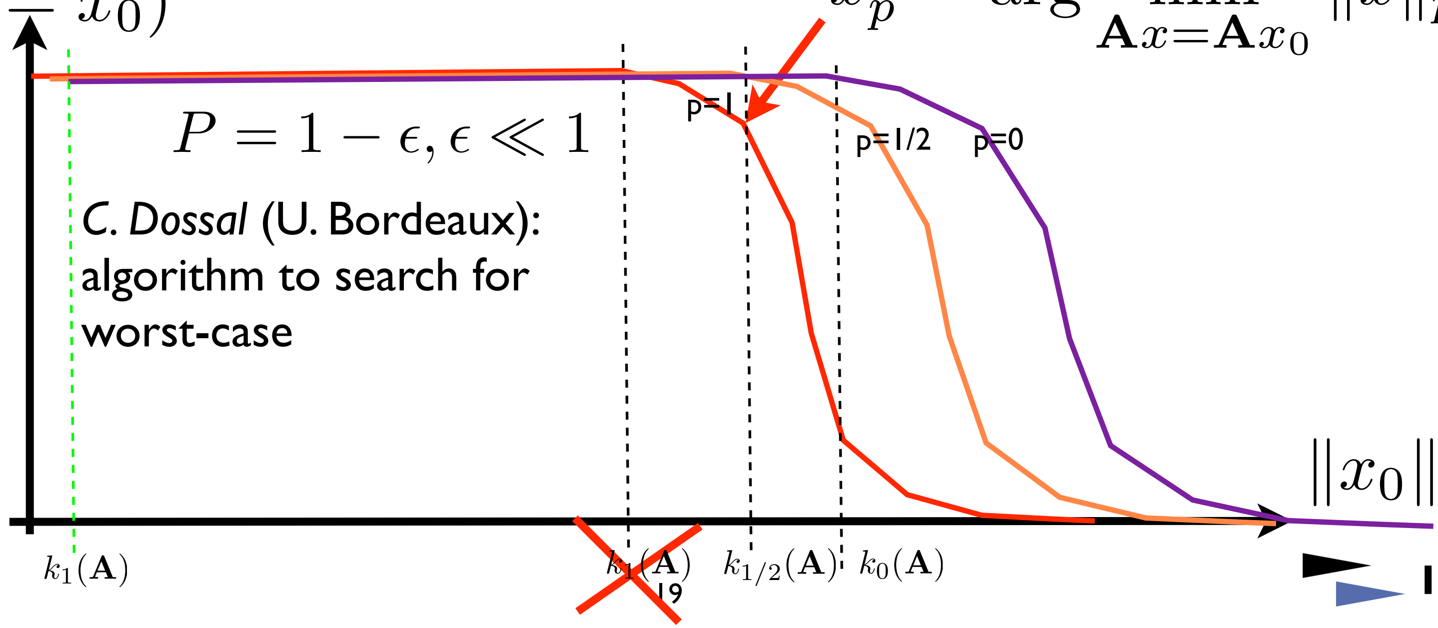
$P(x_0)$ → draw ground truth → x_0 → direct model → $\mathbf{b} := \mathbf{A}x_0$

↓ inverse problem

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

Bayesian! Favorable priors?

Typical observation



The Bayesian bit: L1 minimization and the Laplacian distribution

IRISA

Bayesian modeling

- Observation : $\mathbf{b} = \mathbf{A}x$
- "True" Bayesian model $P(x_k) \propto \exp(-f(|x_k|))$
- Maximum likelihood estimation

$$\max_x \prod_k P(x_k) \Leftrightarrow \min_x \sum_k f(|x_k|)$$

- L1 minimization equivalent to MAP with Laplacian model

$$\hat{P}(x_k) \propto \exp(-|x_k|)$$

- Does L1 minimization fit Laplacian data ?

L1 minimization for Laplacian data ...

- Gaussian matrix

$$\mathbf{A} \in \mathbb{R}^{m \times N} \quad N = 128 \quad 1 \leq m \leq 100$$

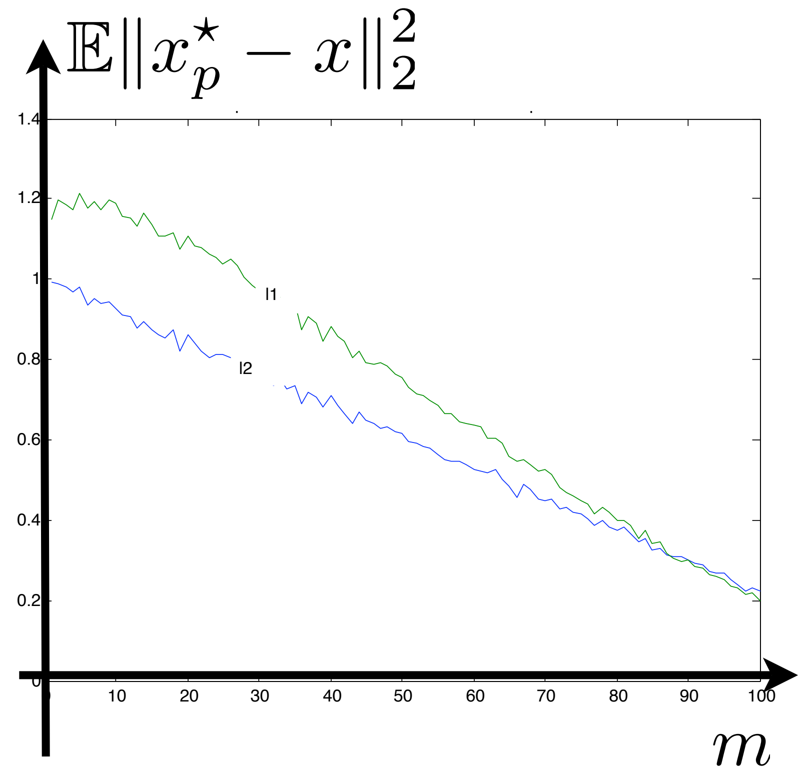
- Laplacian data, 500 draws

$$x \in \mathbb{R}^N \longrightarrow \mathbf{b} = \mathbf{A}x$$

- Reconstruction L1 or L2

$$x_p^\star := \arg\min \|x\|_p, \ \ p = 1, 2$$

= ML with Laplacian / Gaussian prior



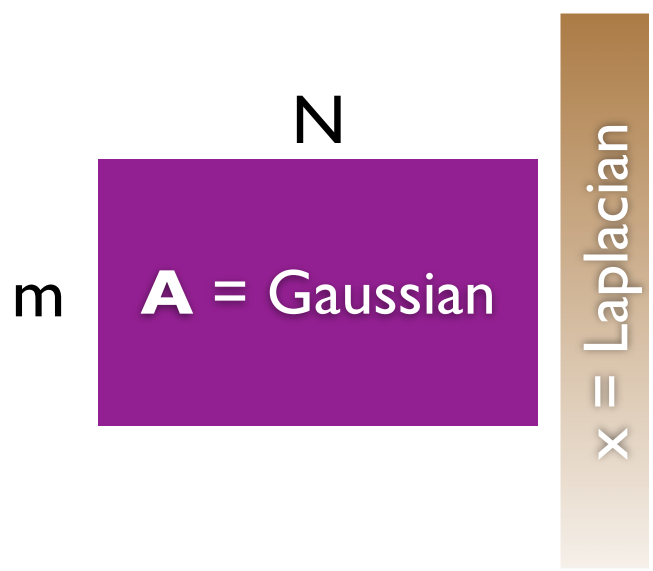
cf also Seeger and Nickish, ICML 2008

MAP is bad when the model fits the data!
Mikolova 2007, Inverse Problems and Imaging

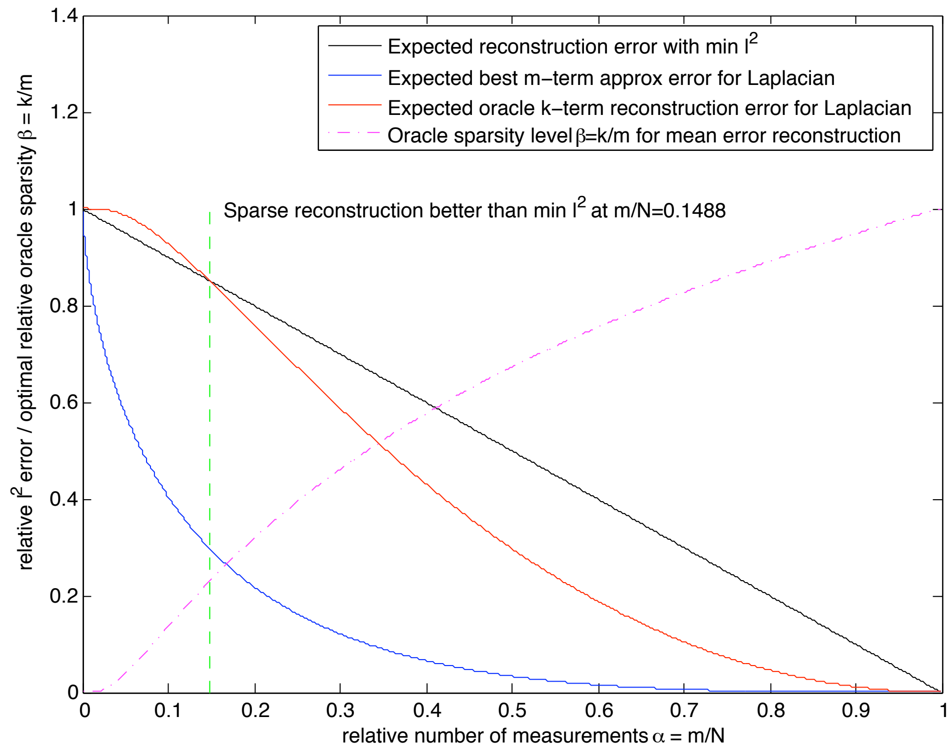
IRISA

Sparse recovery for Laplacian data ?

- Asymptotic analysis with “oracle” sparse estimation



$$N \longrightarrow \infty$$



work in progress, G. & Davies

IRISA

Thanks to

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

The end

remi.gribonval@inria.fr
www.irisa.fr/metiss/gribonval