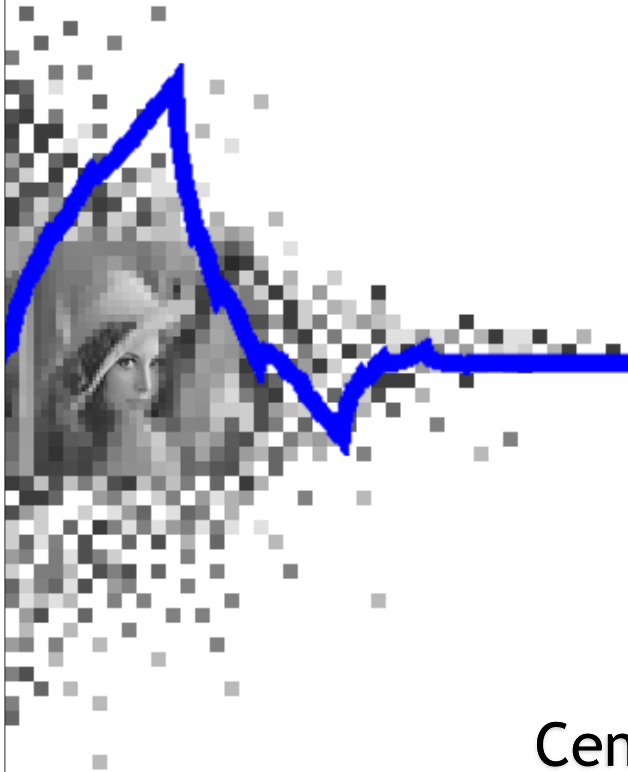
Aisenstadt Chair Course
CRM September 2009



Part IV
Compressed Sensing

Stéphane Mallat

Centre de Mathématiques Appliquées
Ecole Polytechnique

Compressed Sensing

- Acquire few measurements and reconstruct a high resolution signal, if the signal has a sparse representation in a dictionary.

- A super-resolution problem, where the measurement operator can be chosen.

- **Key idea:** use random measurement operators to construct incoherent transformed dictionaries.

-

Representation from Linear Sampling

- ***Linear sampling:*** an analog signal $\bar{f}(x)$ is projected on a basis $\{\bar{\phi}_n\}_{n<N}$ of an approximation space $\mathbf{V}_N$, which specifies a linear approximation:

$$P_{\mathbf{U}}\bar{f}(x) = \sum_{n<N} \langle \bar{f}, \bar{\phi}_n \rangle \, \tilde{\bar{\phi}}_n$$

- Uniform sampling: $\bar{\phi}_n(x) = \bar{\phi}(x - Tn)$

- ***Sparse representation of*** the discrete signal

$$f[n] = \langle \bar{f}, \bar{\phi}_n \rangle \in \mathbf{R}^N$$

in a basis $\{g_p\}_{p\in\Gamma}$ with the *M* largest coefficients

$$\{\langle f, g_p \rangle\}_{p\in\Lambda} \text{ with } |\Lambda| = M \ .$$

Compressive Sensing

- Sparse random analog measurements

$$Y[q] = \bar{U}\bar{f}[q] + W[q] = \langle \bar{f}, \bar{u}_q \rangle + W[q]$$

where $\bar{u}_q(x)$ are realizations of a random process.

- Discretization: if $\bar{u}_q(x) \in \mathbf{V}_N$ then

$$Y[q] = U f[q] + W[q] = \langle f, u_q \rangle + W[q]$$

where $f[n] = \langle \bar{f}, \bar{\phi}_n \rangle$ and $u_q[n] \in \mathbf{R}^N$ is a random vector.

- If f is sparse in $\{g_p\}_{p \in \Gamma}$ can we recover f from Y?

Compressive Sensing Recovery

- From measurements with a random operator

$$Y[q] = Uf[q] + W[q]$$

sparse super-resolution estimation:

$$\tilde{F} = \sum_{p \in \tilde{\Lambda}} \tilde{a}[p]\, g_p$$

where $\tilde{a}[p]$ has a support $\tilde{\Lambda}$ computed with a sparse decomposition of Y in $\mathcal{D}_U = \{U g_p\}_{p \in \Gamma}$ by minimizing:

$$\frac{1}{2} \| \sum_{p \in \Gamma} a[p]\, U g_p - Y \|^2 + T \sum_{p \in \Gamma} |a[p]|$$

or with an orthogonal matching pursuit.

Restricted Isometry and Incoherence

- Riesz basis condition for recovery stability

$$(1-\delta_\Lambda) \sum_{p\in\Gamma} |a[p]|^2 \leq \Big\| \sum_{p\in\Lambda} a[p]\, U g_p \Big\|^2 \leq (1+\delta_\Lambda) \sum_{p\in\Gamma} |a[p]|^2 \ .$$

- Restricted isometry condition: $\delta_\Lambda \leq \delta_M(\mathcal{D}_U) < 1$ if $|\Lambda| \leq M$

$$(1-\delta_M) \sum_{p\in\Gamma} |a[p]|^2 \leq \Big\| \sum_{p\in\Lambda} a[p]\, U g_p \Big\|^2 \leq (1+\delta_M) \sum_{p\in\Gamma} |a[p]|^2 \ .$$

any such family $\{U g_p\}_{p\in\Lambda}$ is "nearly" orthogonal.

- Relation to incoherence:

$$\delta_M(\mathcal{D}_U) \leq (M-1)\,\mu(\mathcal{D}_U) \quad \text{with} \quad \mu(\mathcal{D}_U) = \max_{p\neq q} \langle U g_p, U g_q \rangle .$$

Exact Recovery

- **Theorem:**

$$\text{If } \quad f = \sum_{p \in \Lambda} a[p]\, g_p \quad \text{with} \quad |\Lambda| = M \quad \text{and} \quad \delta_{3M} < 1/3$$

$$\text{then} \quad a = \arg\min_b \frac{1}{2} \| \sum_{p \in \Gamma} b[p]\, U g_p - Y \|^2 + T\, \|b\|_1$$

- Sparse signals are exactly recovered.

Gaussian Random Matrices

- We want to have $\{Ug_p\}_{p\in\Gamma}$ nearly uniformly distributed over the unit sphere of $\mathbf{R}^Q$ so that $\{Ug_p\}_{p\in\Lambda}$ is as orthogonal as possible even for Λ not small.

- The distribution of a Gaussian white noise of variance 1 is a uniform measure in the neighborhood of the unit sphere of $\mathbf{R}^Q$.

- If $\{g_p\}_{p\in\Gamma}$ is an orthonormal basis of $\mathbf{R}^N$ and if U is a matrix of Q by N values taken by independant Gaussian random variables (white noise) then $\{Ug_p\}_{p\in\Gamma}$ are values taken by Q independant Gaussian random variables.

RIP Stability for Gaussian Matrices

- **Theorem:** If U is a Gaussian random matrix then for any $\delta < 1$ there exists $\beta > 0$ such that

$$\delta_M(\mathcal{D}_U) \leq \delta \quad \text{if} \quad M \leq \frac{\beta\, Q}{\log(N/Q)}.$$

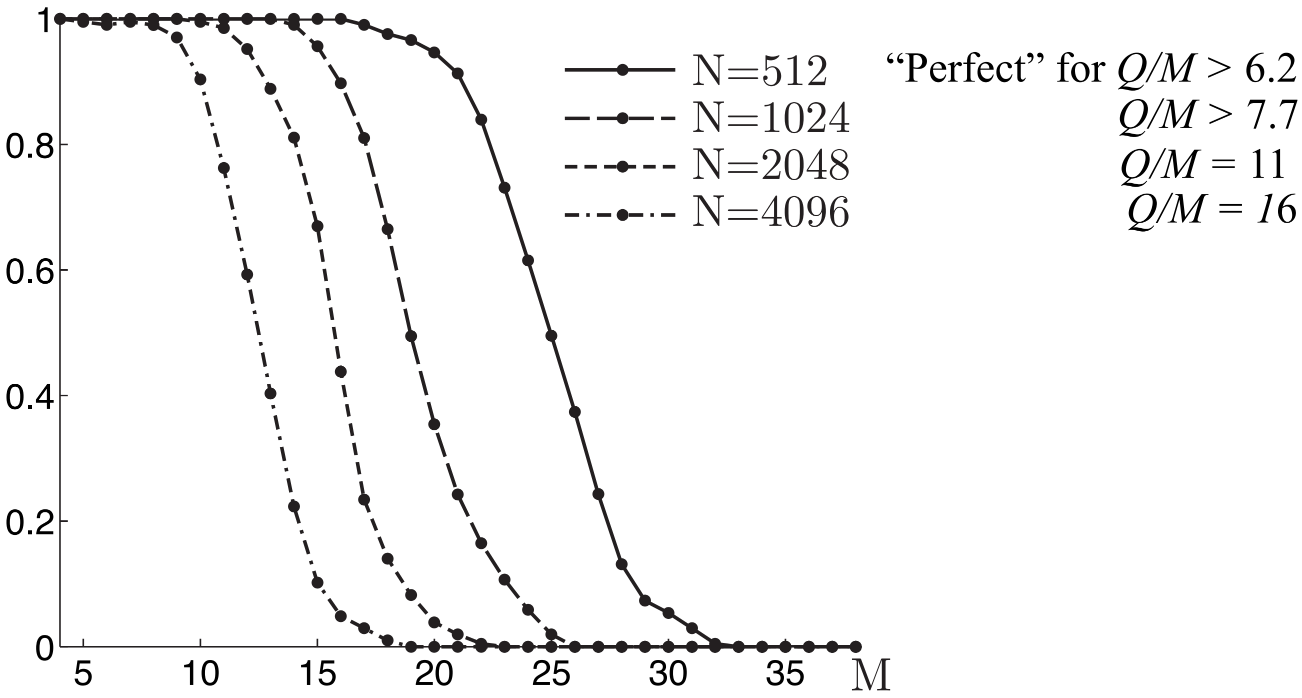
- Valid for random Bernouilli matrices (random 1 and -1).

- We need $Q \sim C\, M \log(N/M)$ measurements to recover M values and M unknown indices among N.
- Coding would require of the order of $M \log(N/M)$ bits.

Perfect Recovery Constants

- Monte-Carlo experiments for recovering signals with M non-zero coefficients out of N with Q random Gaussian measurements: $Q \sim C\, M \log(N/M)$

Ratio of recovered signals with $Q = 100$



Other Random Operators

- Storing a random Gaussian matrix U and computing Uh requires $O(N^2)$ memory and calculations, too much.

- RIP theorem valid for Bernouilli matrices (random 1 and -1). Still too much memory and computations.

- Similar RIP theorem valid for a random projector in an orthonormal basis $\{g'_m\}_m$ which is highly incoherent with the sparsity basis $\{g_p\}_p$. May require only *O(N)* memory and *O(N log N)* computations.

Random Sparse Spike Inversion

- Measurements

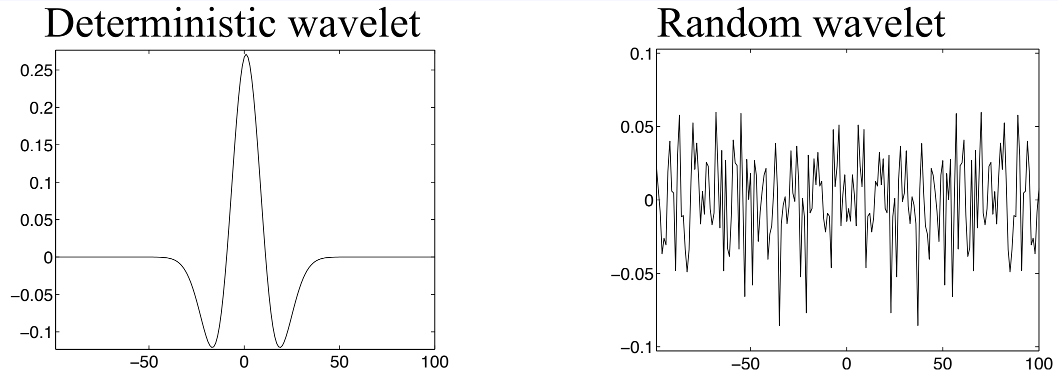
$$Y = u * f + W \quad \text{with} \quad f[n] = \sum_{p \in \Lambda} a[p] \, \delta[n-p] \ .$$

- Random wavelet makes a random sampling of the Fourier coefficients of f : $\hat{u}[k]$ is the indicator of random set of frequencies.

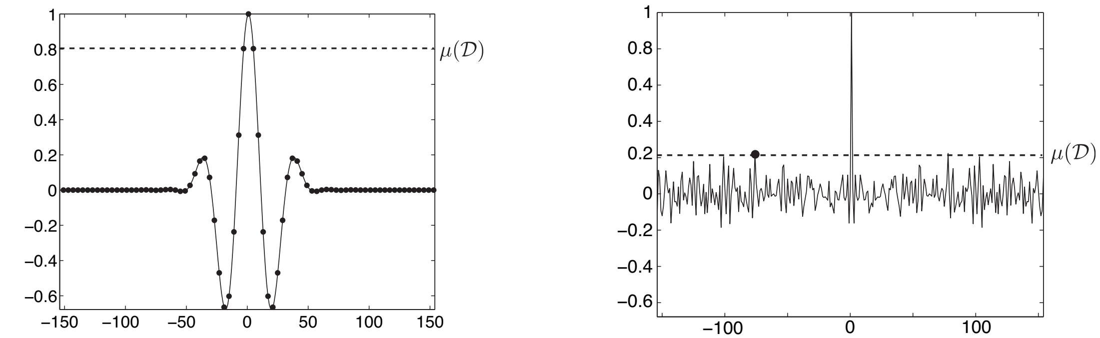
- Fourier and Dirac bases have a low coherence.

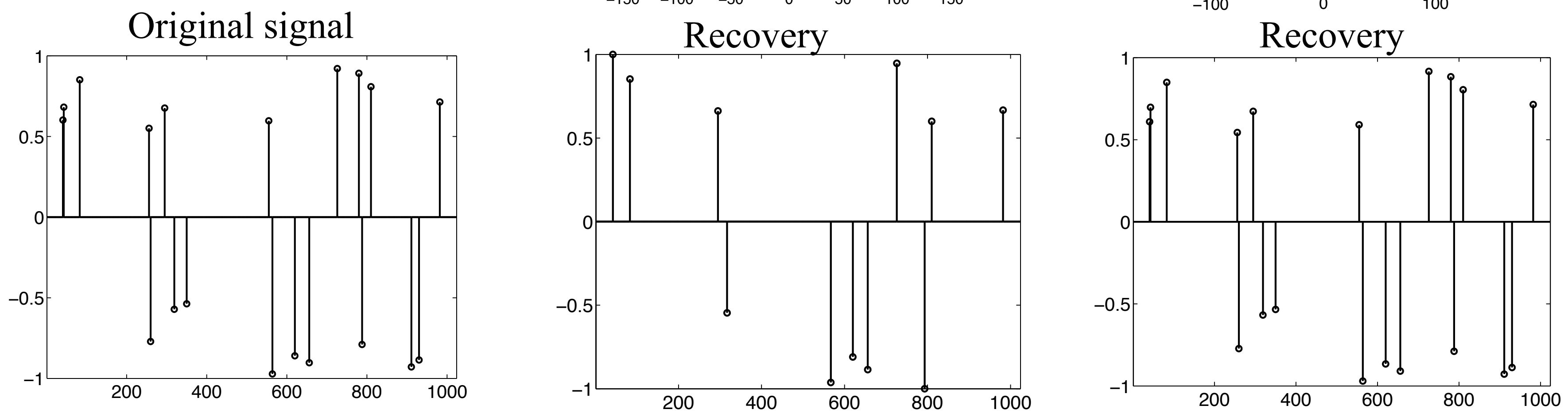
Random Sparse Spike Inversion

Seismic wavelet $u[n]$



Incoherence $u * \tilde{u}[p]$





Non-Linear Approximation Error

- $\{\langle f, g_{m_k} \rangle\}_k$ sorted with decreasing amplitude

$$|\langle f, g_{m_{k+1}} \rangle| \leq |\langle f, g_{m_k} \rangle|.$$

- Non-linear approximation in an orthonormal basis:

$$f_M = \sum_{k=1}^{M} \langle f, g_{m_k} \rangle \, g_{m_k}$$

and

$$\|f - f_M\|^2 = \sum_{k=M+1}^{N} |\langle f, g_{m_k} \rangle|^2 \; .$$

If $|\langle f, g_{m_k} \rangle| = O(k^{-\alpha})$ then $\|f - f_M\|^2 = O(M^{1-2\alpha})$.

Stability and Recovery Error

- **Theorem:** There exists C such that if $\delta_{3M} < 1/3$ and

$$\tilde{a} = \arg\min_b \frac{1}{2} \|Y - \sum_{p \in \Gamma} b[p] \, U g_p\|^2 + T \, \|b\|_1$$

then

$$\|f - \sum_{p \in \Gamma} \tilde{a}[p] \, g_p\|^2 \leq \frac{C}{\sqrt{M}} \sum_{k=M}^{N-1} |\langle f, g_{m_k} \rangle| + C \, \|W\|$$

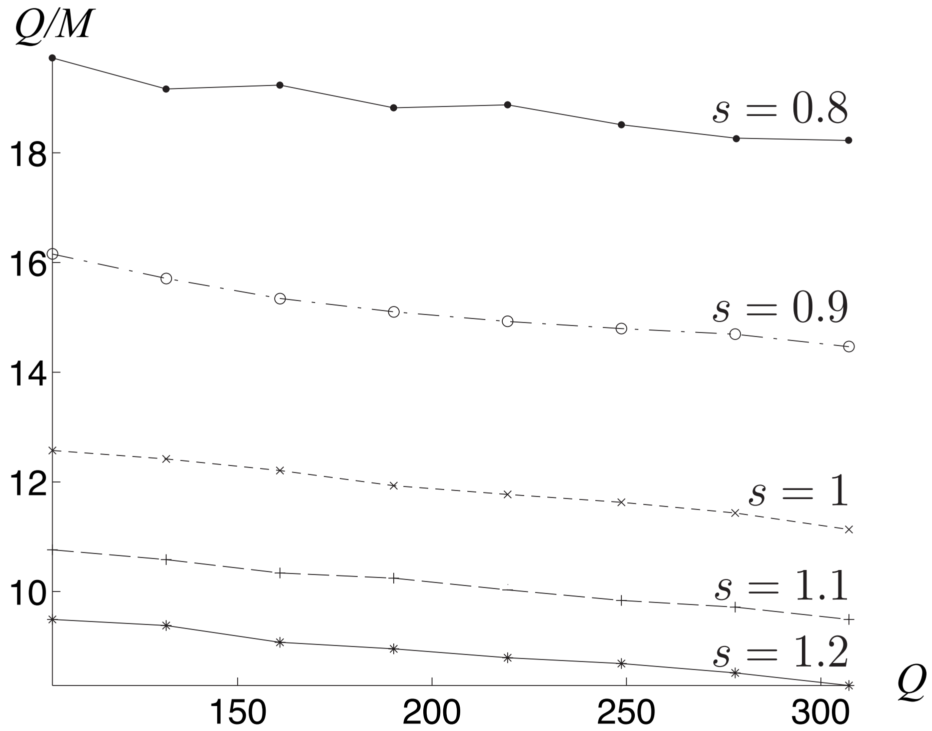
and if $|\langle f, g_{m_k} \rangle| = O(k^{-s})$ with $s > 1$ and $\|W\| = 0$

$$\|f - \sum_{p \in \Gamma} \tilde{a}[p] \, g_p\|^2 = O(M^{-2s+1}) \; .$$

- Requires $Q \sim C' \, M \, \log(N/M)$ random measurements.

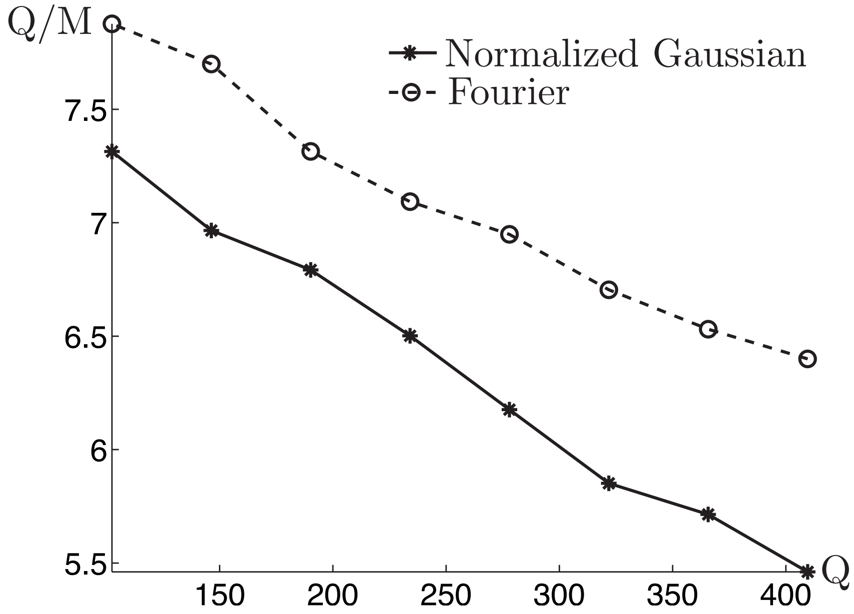
Approximation Constants

- For Q random measurements, M is the number basis coefficients defining an approximation having the same error. The ratio Q/M is computed with a Monte-Carlo experiment for different decay exponents s for N=1024.



Recovery Efficiency

Compressive sensing efficiency in random Gaussian and Fourier dictionaries.



Comparison of Basis Pursuit, Matching Pursuit and Orthogonal Matching Pursuits.

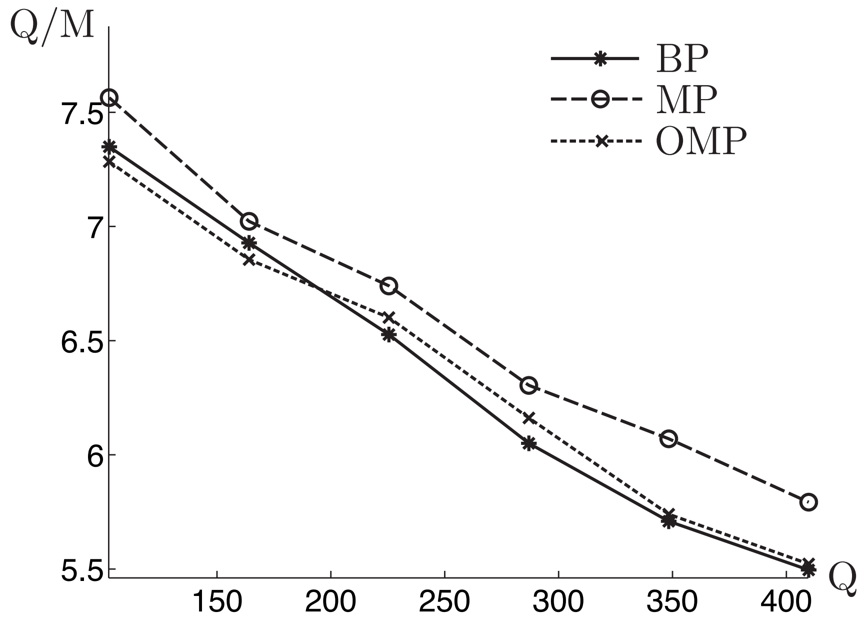


Image Compressive Sampling

- Wavelet coefficients of images often have the decay exponent $s = 1$ of bounded variation images.
- Q measurements with a linear uniform sampling satisfy $Q/M < 5$, relatively to a non-linear approximation with M coefficients.
- Direct compressive is worst with typically $Q/M > 7$.
- Improvements by using some prior information on coefficients. Grouping wavelet coefficients scale by scale improves image approximations.

Compressive Sensing Applications

- **Robustness:** compressive sensing is a “democratic” acquisition process where all samples are equally important. A missing sample introduce an error that is diluted across the signal.

- **Analog to Digital Converters**: for signals that have a sparse Fourier transform, with a random time sampling. For utrawide band signals having a Nyquist rate which is too large.

- **Single pixel camera:** random Bernouilli (1 or -1) mesurements of images with a single pixel at very high sampling rate.

- **Medical Resonance Imaging:** randomize as much as possible the Fourier sampling of images obtained with MRI, and use their wavelet sparsity to improve their resolution.

Conclusion to Compressive Sensing

- Sparse super-resolution becomes stable with randomized measurements. Large potential applications.
- Asymptotic performance equivalent to a non-linear approximation with a known signal.
- The devil is in the constants, compared to a linear uniform sampling.
- Technological difficulties for the signal recovery: large amount of memory and computations are required.