# M* proxy for the Sparse Recovery Threshold.

**Mathieu Barré**,
*D.I., Ecole normale supérieure.*

With Alexandre d'Aspremont (CNRS).

# Introduction

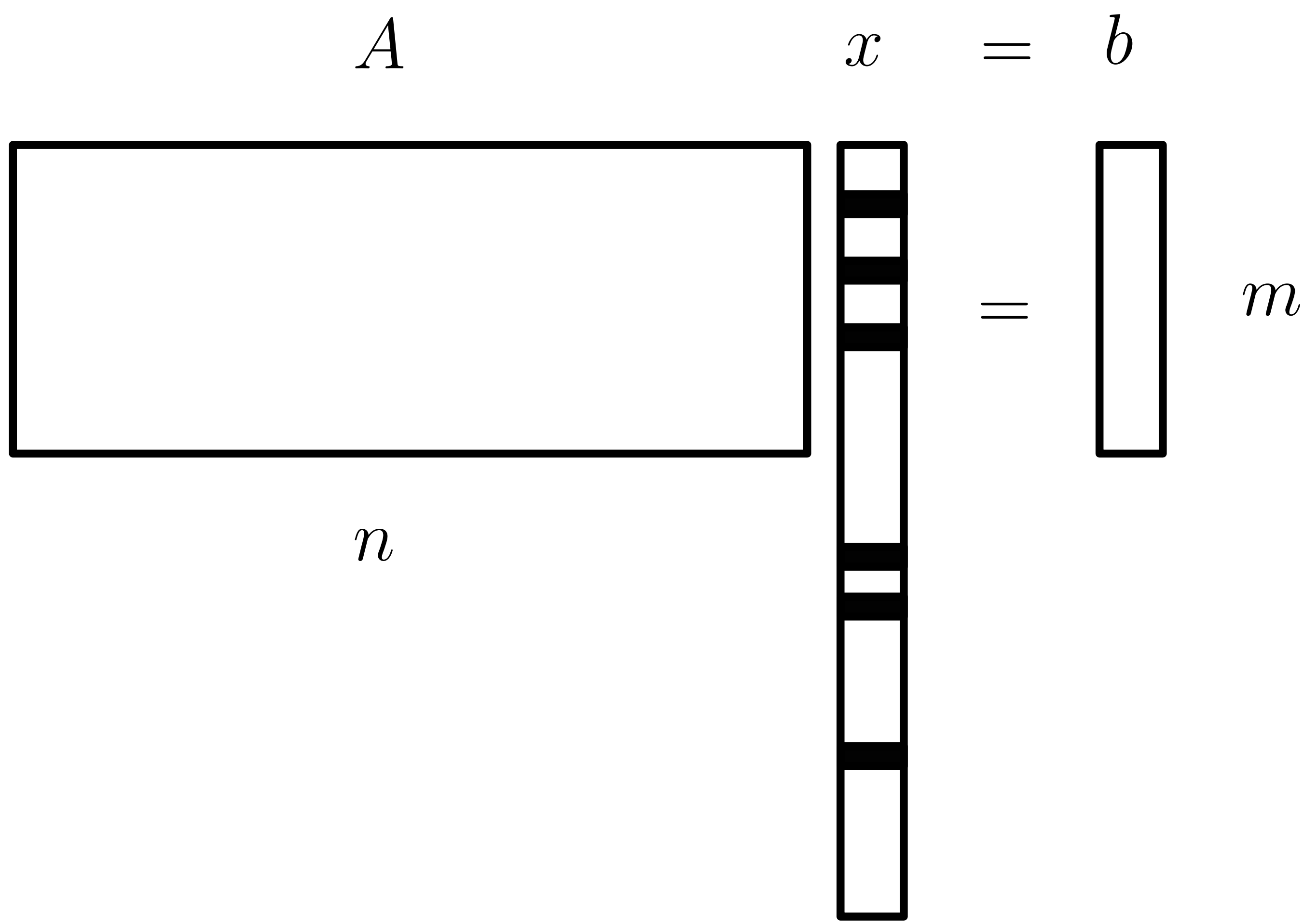
**Tractable sparse recovery guarantees.**

- The sparse recovery threshold of a given matrix is hard to compute.
- Probabilistic bounds on this threshold for classes of random matrices.

**Today:** Tractable proxy for sparse recovery threshold of deterministic matrices.

# Introduction

Given a signal sparse signal $x \in \mathbb{R}^n$ with $n$ potentially very large, we observe $b \in \mathbb{R}^m$ with $m << n$ defined as follow



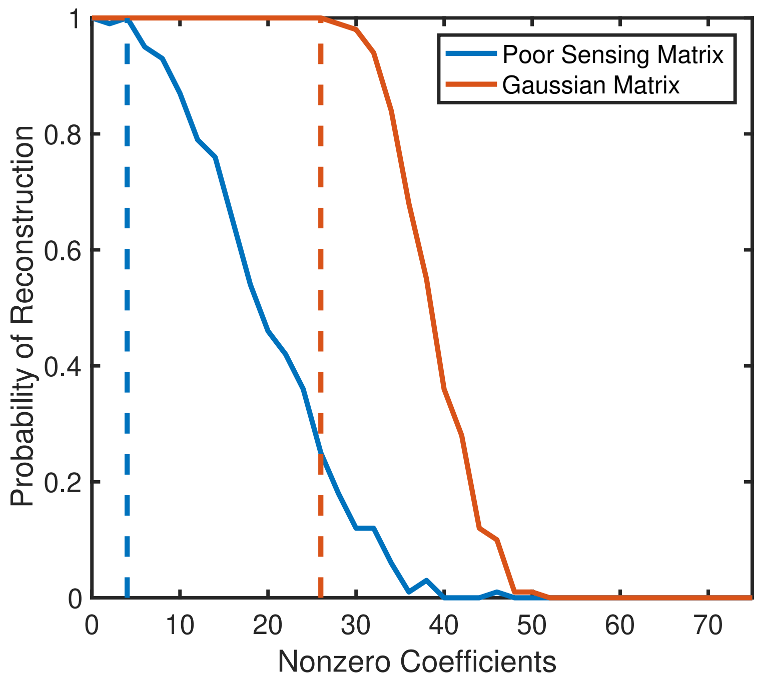

with $A \in \mathbb{R}^{m \times n}$ full rank.

# Introduction

Solving the $l_0$ problem is NP-hard. Relaxed as [Donoho and Tanner, 2005, Candès and Tao, 2005]

$$\begin{array}{ll} \text{minimize} & \|u\|_1 \\ \text{subject to} & Au = b \end{array} \qquad (l_1)$$

Define $\mathbf{k}(A)$ such that $(l_1)$ recovers any signal $x$ with $\mathbf{Card}(x) \leq \mathbf{k}(A)$.



# Outline

- **A lower bound for sparse recovery threshold from $\mathrm{M}^*$**
- $M^*$ regularization in Dictionary Learning
- Greedy $M^*$ strategy for generating MRI sampling schemes

# Bound on recovery threshold

**Theorem (Kashin and Temlyakov [2007, Th. 2.1])**

**Lower bound on k(A)** *Given a coding matrix* $A \in \mathbb{R}^{m \times n}$, $\mathbf{k}(A)$ *can be lower bounded as follow*

$$\mathbf{k}(A) \geq \frac{1}{S(A)^2}$$

*with* $S(A) = \sup_{u \neq 0} \frac{\|u\|_2}{\|u\|_1}$.

- $S(A) = \text{radius}\{\mathcal{K}(A)\}$ where $\mathcal{K}(A) \triangleq \{u \in \mathbb{R}^n : \|u\|_1 \leq 1,\ Au = 0\}$.
- Approximating the radius of a convex polytope is a hard problem.
- Semidefinite relaxations e.g. only certify recovery of cardinality $O(\sqrt{\mathbf{k}(A)})$.

# $M^*$ of a set

For $K \subset \mathbb{R}^n$ one defines $M^*(K)$ as

$$M^*(K) \triangleq \mathbf{E}\left[\sup_{x \in K} \langle g, x \rangle\right]$$

where $g \sim \mathcal{N}(0, \mathbf{I}_n)$

**Theorem (Pajor and Tomczak-Jaegermann [1986])**

**Low $M^*$** *Given a set $K \subset \mathbb{R}^n$ and an $q \times n$ matrix $G$ whose rows are independent isotropic Gaussian random vectors in $\mathbb{R}^n$, the radius of a section of $K$ by the nullspace of $A$ satisfies*

$$\mathbf{radius}(K \cap \mathcal{N}(G)) \leq \frac{c}{\sqrt{q}} M^*(K)$$

*with probability $1 - e^{-c'q}$, where $c, c' > 0$ are absolute constants.*

Apply this theorem to $\mathcal{K}(A)$ ?

# Recovery Threshold of Perturbed Matrix

- $\mathcal{K}(A) \cap \mathcal{N}(G) = \mathcal{K}(\begin{bmatrix} A \\ G \end{bmatrix})$ then taking $K = \mathcal{K}(A)$ in previous theorem leads to

**Proposition B. and d'Aspremont [2018]**

**Threshold of perturbed matrix** *Let $A \in \mathbb{R}^{m \times n}$ be a given matrix and $G \in \mathbb{R}^{q \times n}$ be a matrix with i.i.d. Gaussian coefficients. Suppose $u^{LP}$ solves*

$$\begin{array}{ll} \textit{minimize} & \|u\|_1 \\ \textit{subject to} & Au = Ax, \\ & Gu = Gx, \end{array}$$

*in the variable $u \in \mathbb{R}^n$, then with probability $1 - e^{-c'q}$*

$$\boxed{k(\begin{bmatrix} A \\ G \end{bmatrix}) \geq c \frac{q}{M^*(\mathcal{K}(A))^2}}$$

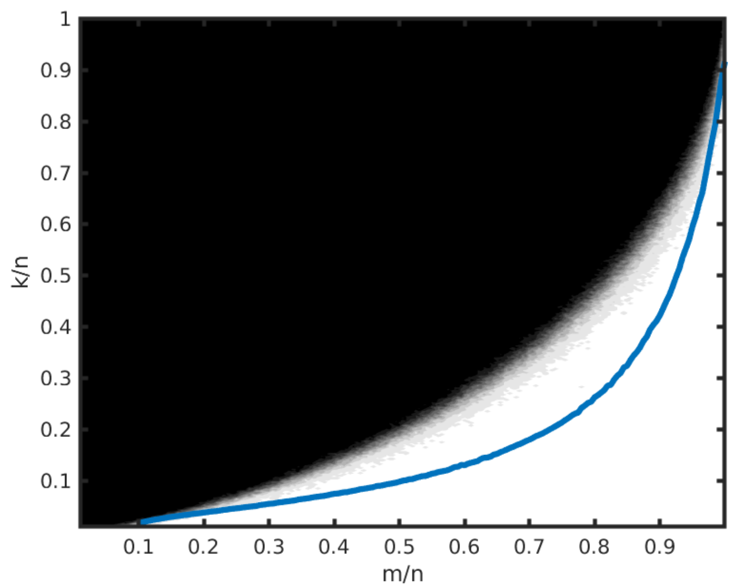
*with c,c' positive constants.*

# Estimating $M^*$

- The previous theorem gives a lower bound on the recovery threshold of a **perturbed** version of $A$, with $M^*(A)^{-2}$ a proxy for $\mathbf{k}(A)$.
- Approximating

$$M^*(A) = \mathbf{E}_g \left[ \sup_{\substack{Ax=0 \\ \|x\|_1 \leq 1}} \langle g, x \rangle \right]$$

by simulation involves solving multiple LPs.



# Outline

- A lower bound for sparse recovery threshold from $M^*$
- **$\mathbf{M}^*$ regularization in Dictionary Learning**
- Greedy $M^*$ strategy for generating MRI sampling scheme

## Dictionary Learning setting

Observations $Y \in \mathbb{R}^{n \times m}$, sparsity target $S$. Find an over-complete dictionary $D \in \mathbb{R}^{n \times p} (n < p << m)$ and representation $X \in \mathbb{R}^{p \times m}$ which minimize

$$\sum_i ||Y_i - DX_i||_2^2 = ||Y - DX||_F^2.$$

with $||X_i||_0 \leq S$.

- Classical normalization strategy is to impose $\|D_i\|_2 = 1$.
- Solved with alternating minimization ($e.g.$ **KSVD** [Elad and Aharon, 2006]).

# $M^*$ regularization

- Looking for a meaningful regularization strategy of dictionary learning.
- $M^*(D)$ used as proxy for the quality of $D$ as a sensing matrix.
- Optimize $||Y - DX||_F^2 + \lambda M^*(D)$ over $X$ and $D$ using alternate minimization with stochastic gradient descent on $D$.

**Lemma**

**SGD on $M^*$** *The regularized function*

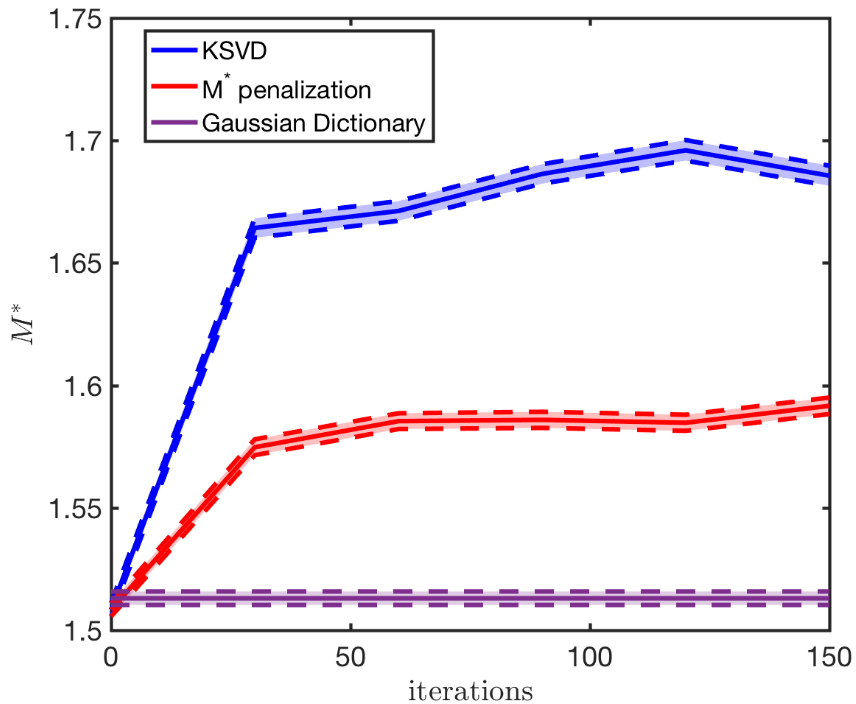
$$\nu_g(A) \triangleq \begin{array}{ll} \textit{sup.} & g^T x + \frac{\lambda}{2}||x||_2^2 + \frac{\sigma}{2}||r||_2^2 \\ \textit{s.t.} & Ax + r = 0 \\ & \|x\|_1 \leq 1 \end{array} \qquad \text{(P)}$$

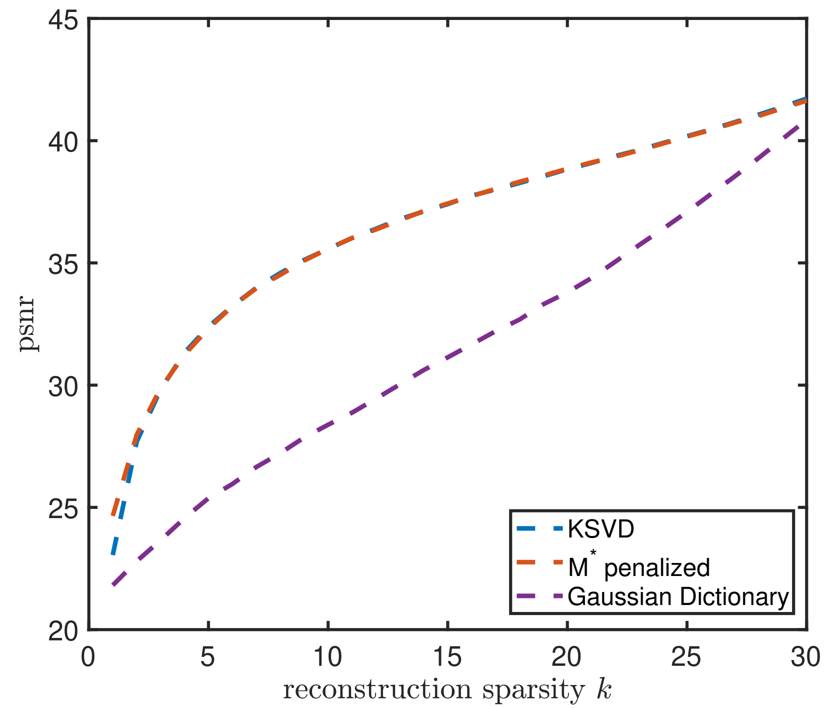
*in the variables* $x \in \mathbb{R}^n$ , $r \in \mathbb{R}^{n-m}$*, with* $\lambda, \sigma > 0$*, has gradient*

$$\nabla \nu_g(A) = y_g^*(A) x_g^{*T}(A),$$

*where* $x_g^*(A)$ *and* $y_g^*(A)$ *are the primal and dual solutions of problem* (P).

# Compression experiments



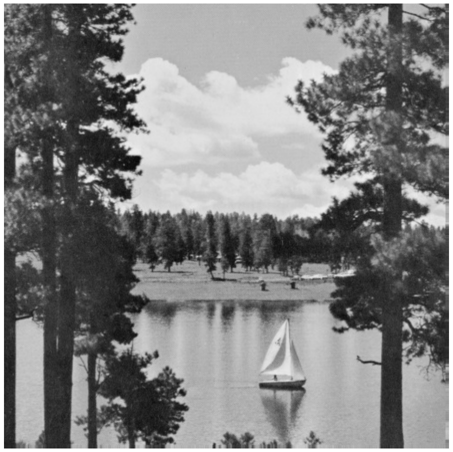

Lower $M^*$ means larger $K(D)$, i.e. better recovery performance.

# Inpainting setting

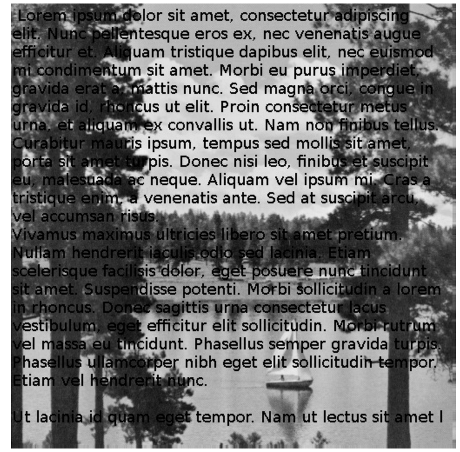
- Observations patches $Y \in \mathbb{R}^{n \times m}$ and mask $B \in \{0,1\}^{n \times m}$, only $B \odot Y$ is observed ($i.e$ black holes on training images).
- Minimize $\|B \odot (Y - DX)\|_F^2$ with respect to $X$ and $D$, same constraints $\|X_i\|_0 \leq S$ and $\|D_i\|_2 = 1$.
- Weighted KSVD [Mairal et al., 2008]. Modified KSVD that tackles the inpainting problem.
- Strong need to regularize, add $M^*$ penalization.

# Inpainting experiments

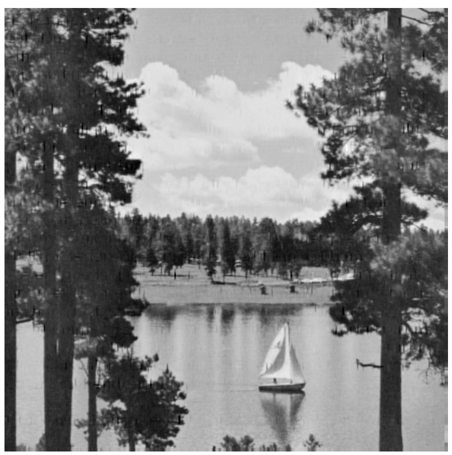
original



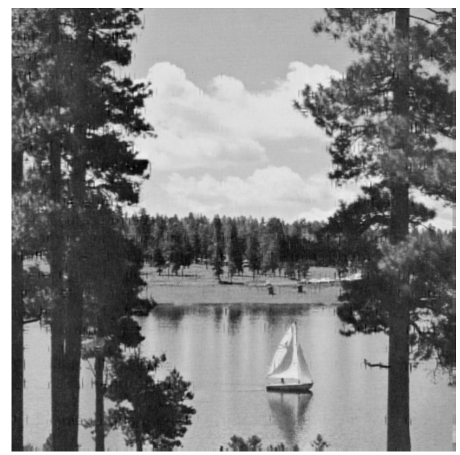
with mask




W-KSVD PSRN:30.7898 k = 15



Mstar PSNR:30.9293 k = 15



# Inpainting experiments

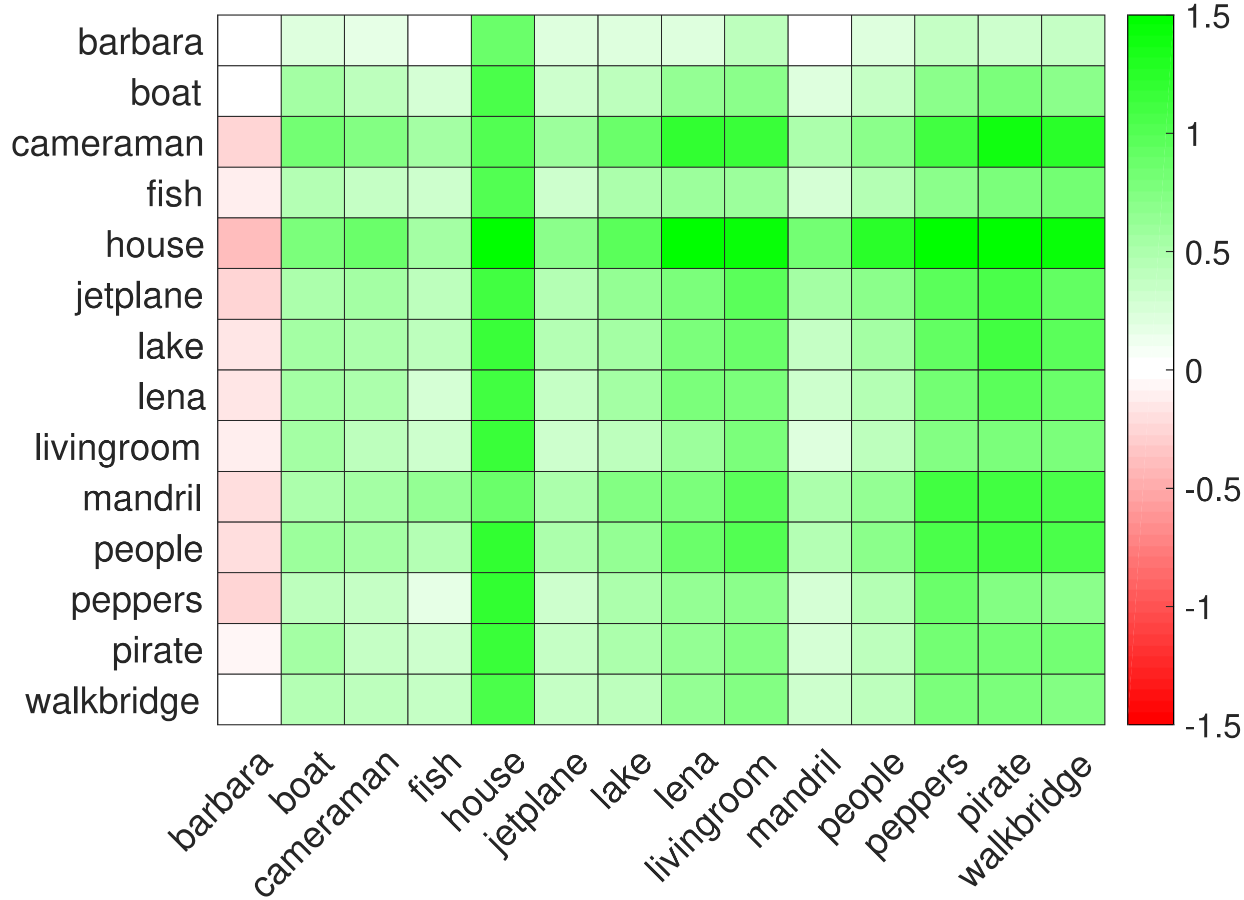



**Figure 1:** Heatmap of the average PSNR gap between the $M^*$ penalized method and wKSVD with on x-axis the training images and on y-axis the test images.

**Remark :** Penalizing the norm of the dictionary in some given random direction also works well on the Inpainting problem.

# Outline

- A lower bound for sparse recovery threshold from $M^*$
- $M^*$ regularization in Dictionary Learning
- **Greedy $\mathrm{M}^*$ strategy for generating MRI sampling scheme**

# MRI basic principle

- MRI recovers the density of matter in an object. Measures **Fourier transform** of the the density through a magnetic field.

- Classical signal processing theory (Nyquist-Shannon) applies. But sampling on a full regular grid in Fourier space is **very time consuming**.

- Compressed sensing allows to significantly reduce the number of measurements required to recover the original signal by **subsampling Fourier mesurements** [Lustig et al., 2008, Boyer et al., 2017].

# Density recovery

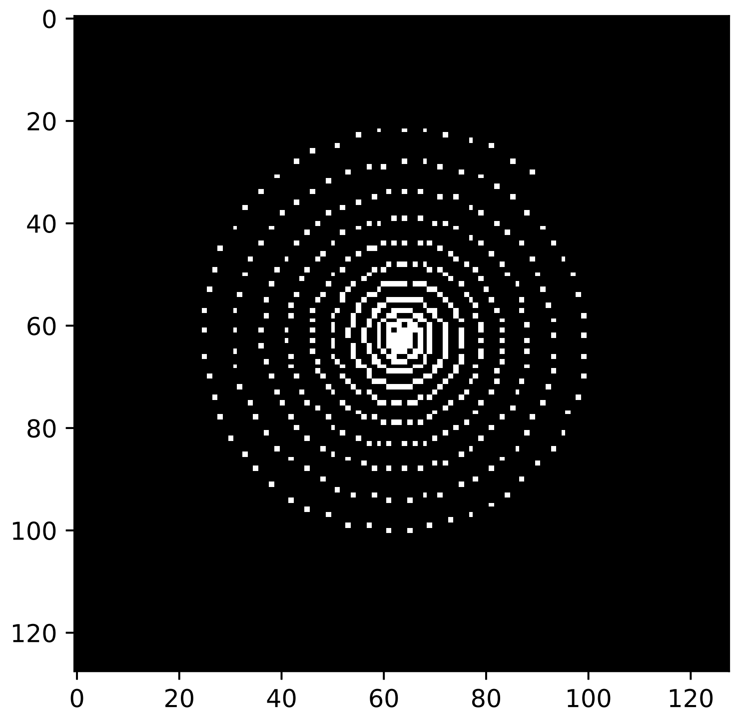
- Fourier coefficients can be measured along **particular trajectories** in Fourier space.
- Trajectories should be chosen to be **sampled efficiently by MRI machines.** [Boyer et al., 2017]
- Reconstructing the original image from the observations $y$ involves solving

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & FH^*x = y \end{array}$$

where $F$ is the Fourier operator on the sub-sampled frequencies and $H^*$ is the inverse wavelet transform operator.

# Generating good sampling scheme

Here, use spirals as trajectories in Fourier space.




Here: use a greedy strategy based on the $M^*$ to choose a subset of spirals that has good generic reconstruction properties.

# Experiments

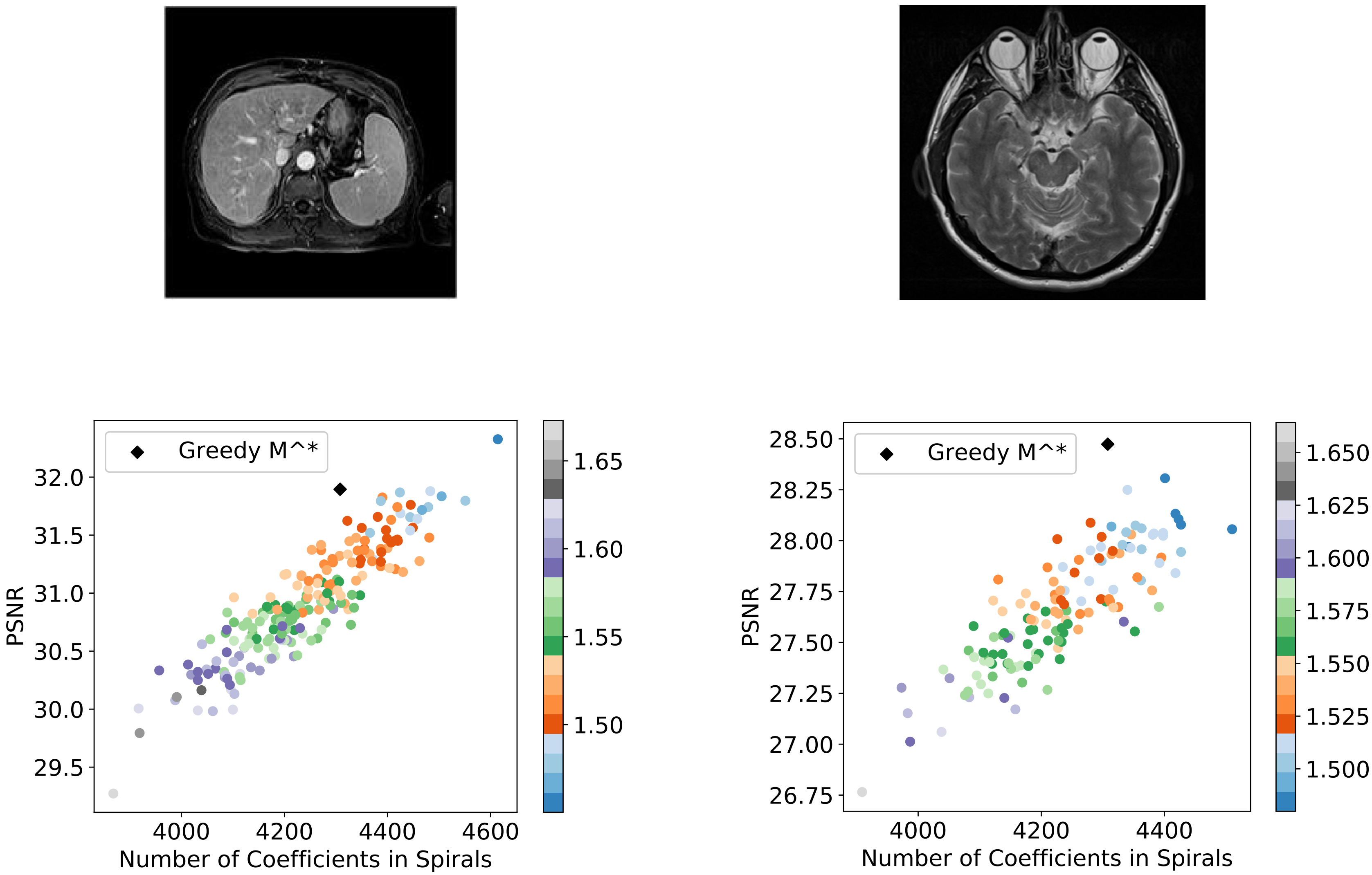



**Figure 2:** Comparaison between Greedy M* strategy to select the right spirals compared to selecting it at random.

# Conclusion

- Tractable proxy for the sparse recovery threshold associate to $(l_1)$.
- Measure generalization properties of a dictionary
- Use $M^*$ to select good sampling schemes in MRI.

**Open problems.**

- Dictionary Learning applications where $M^*$ regularization might be really helpful, $e.g.$with images very different from natural images ?
- More realistic setting for MRI experiments.
- New $M^*$ applications.

*

References

Mathieu B. and Alexandre d'Aspremont. M*-regularized dictionary learning. *arXiv preprint arXiv:1810.02748*, 2018.

Claire Boyer, Jérémie Bigot, and Pierre Weiss. Compressed sensing with structured sparsity and structured acquisition. *Applied and Computational Harmonic Analysis*, 2017.

E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

D. L. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. of the National Academy of Sciences*, 102(27):9446–9451, 2005.

M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

B.S. Kashin and V.N. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82(5):748–755, 2007.

Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE signal processing magazine*, 25(2): 72–82, 2008.

J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on image processing*, 17(1):53–69, 2008.

A. Pajor and N. Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional banach spaces. *Proceedings of the American Mathematical Society*, 97(4):637–642, 1986.