DE LA RECHERCHE À L'INDUSTRIE

Sparsifying Transform Learning for Blind Compressed Sensing In MRI

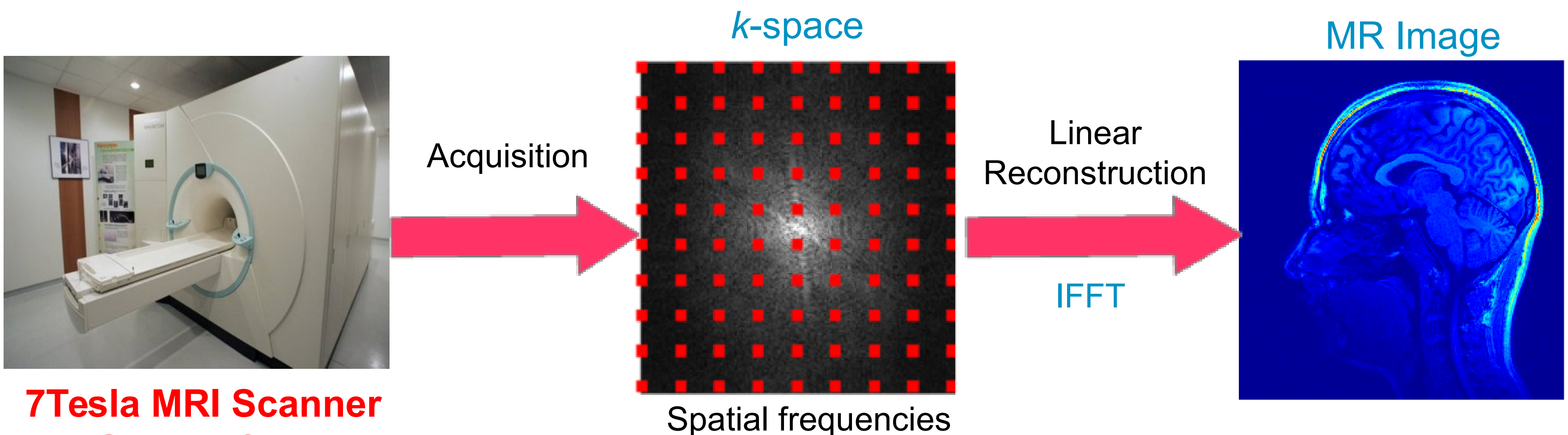
Philippe Ciuciu & Carole Lazarus
Joint work with Pierre Weiss & Maximilan März

03/21/2017

Short Reminder on Compressed Sensing for Magnetic Resonance Imaging

CARTESIAN SAMPLING SCHEME IN MRI

- Data: samples in k-space i.e. of spatial Fourier transform of object:



7Tesla MRI Scanner @NeuroSpin

Specificities in MRI:

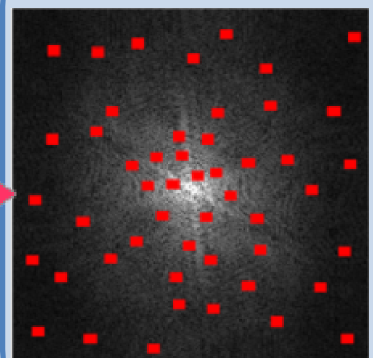
- k-space samples collected sequentially in MRI
- Acquisition rate limited by MR physics, gradient constraints and physiological constraints on RF energy deposition

COMPRESSED SENSING IN MRI

- Objective: Accelerate the data acquisition in MRI

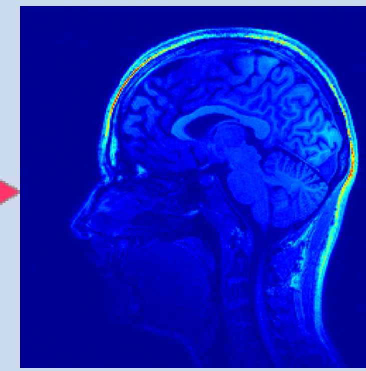
7Tesla MRI Scanner @NeuroSpin

Compressed Sensing *[Lustig et al., MRM 2007]*



Random undersampling

Nonlinear reconstruction



Sparsity

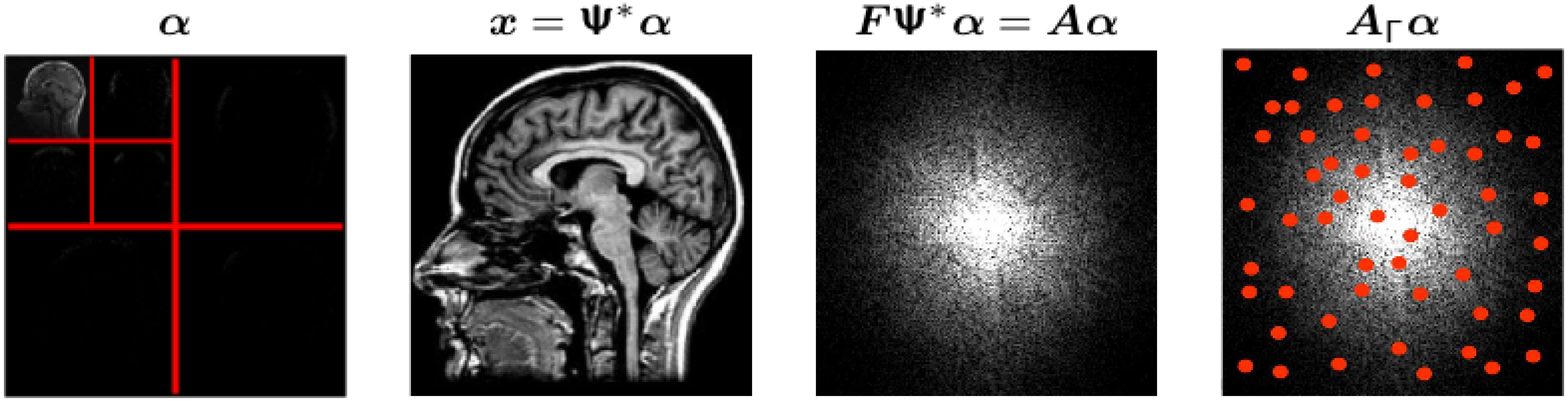
Three ingredients:
- Sparsity of image in transform domain or dictionary
- Measurement process incoherent with the sparsifying transform
- Nonlinear reconstruction algorithm

CS-MRI MATHEMATICAL FORMULATION (1/2)

- Compressed sensing theory:
  - $\boldsymbol{x}$ is sparse in a given basis (e.g. wavelets): $\boldsymbol{x} = \boldsymbol{\Psi}^* \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{C}^n$ is s-sparse.
  - Acquisition matrix: $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{\Psi}^*$.

Let $\Gamma \subseteq \{1, \cdots, n\}$ and $\boldsymbol{A}_\Gamma = (\boldsymbol{a}_i^*)_{i\in\Gamma}$. We acquire a measurement vector:

$$\boldsymbol{y} = \boldsymbol{F}_\Gamma \boldsymbol{x} + \boldsymbol{b} = \boldsymbol{A}_\Gamma \boldsymbol{\alpha} + \boldsymbol{b}$$



Nonlinear reconstruction (synthesis formulation, e.g. FISTA):

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\in\mathbb{C}^n}{\arg\min} \|\boldsymbol{y} - \boldsymbol{A}_\Gamma \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

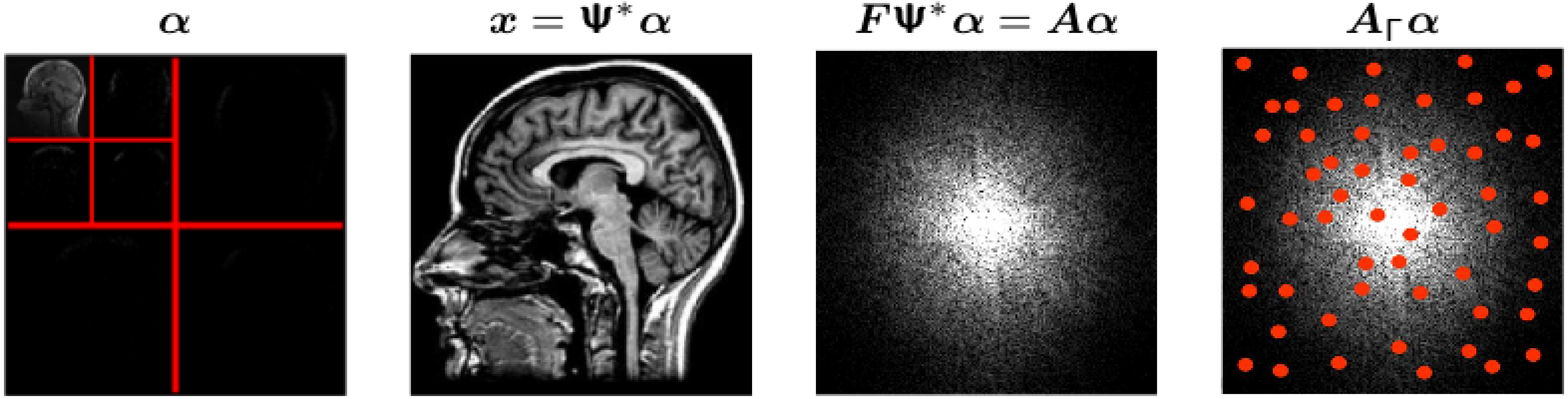
$$\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}^* \widehat{\boldsymbol{\alpha}}$$

CS-MRI MATHEMATICAL FORMULATION (2/2)

- Compressed sensing theory:
  - $\boldsymbol{x}$ is sparse in a given basis (e.g. wavelets): $\boldsymbol{x} = \Psi^* \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{C}^n$ is s-sparse.
  - Acquisition matrix: $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{\Psi}^*$.

Let $\Gamma \subseteq \{1, \cdots, n\}$ and $\boldsymbol{A}_\Gamma = (\boldsymbol{a}_i^*)_{i \in \Gamma}$. We acquire a measurement vector:

$$\boldsymbol{y} = \boldsymbol{F}_\Gamma \boldsymbol{x} + \boldsymbol{b} = \boldsymbol{A}_\Gamma \boldsymbol{\alpha} + \boldsymbol{b}$$



Nonlinear reconstruction (analysis formulation, e.g. MM, ADMM):

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{C}^n}{\arg\min} \|\boldsymbol{y} - \boldsymbol{F}_\Gamma \boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{\Psi}\boldsymbol{x}\|_1$$

TWO COMPLEMENTARY DIRECTIONS IN CS-MRI

- Design of new sampling patterns (k-space trajectories)
  - **SPARKLING: Segmented Projection Algorithm for Random K-space samplING**

- Better or adaptive sparse modeling
  - **Either fixed such as curvelets, starlets, shearlets …**
  - **Or learned from (trained) or compressed data**

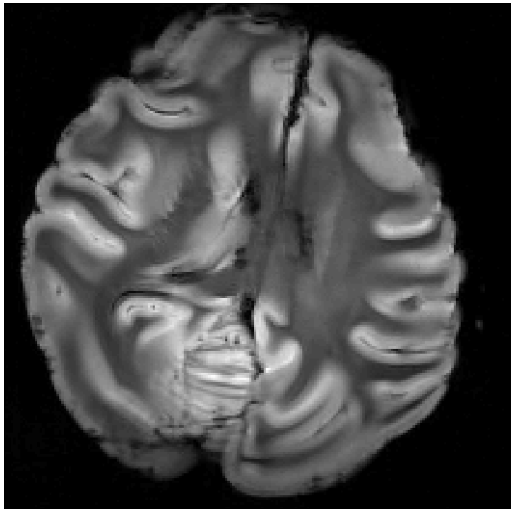
TWO COMPLEMENTARY DIRECTIONS IN CS-MRI

- Design of new sampling patterns (k-space trajectories)
  - **SPARKLING: Segmented Projection Algorithm for Random K-space samplING**

DIRECTION 1: NEW K-SPACE SAMPLING STRATEGIES

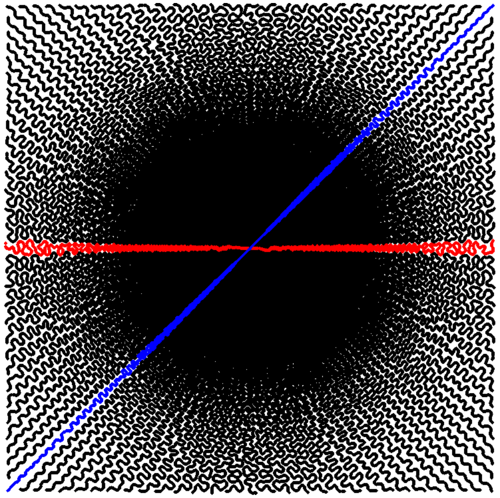
- Design of new sampling patterns (k-space trajectories)

Sequence parameters:
- N=512
- FOV=200x200 mm²
- TR=550 ms and TE=30 ms
- α=25°
- BW=32.55 Hz/px
- **Tobs=30.72 ms**
- Slice thickness: 3 mm
- Single channel receiver coil
- Ex-vivo baboon brain



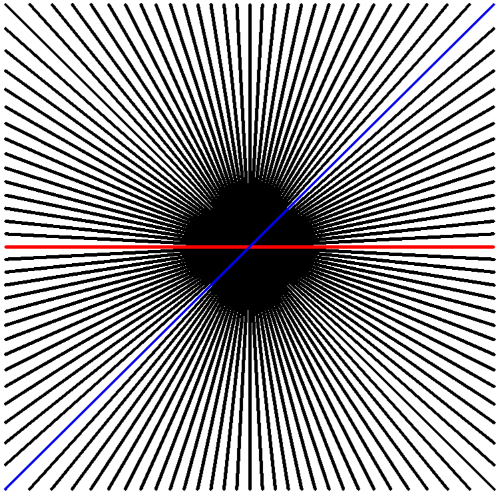
REFERENCE
BABOON BRAIN

SPARKLING



vs.

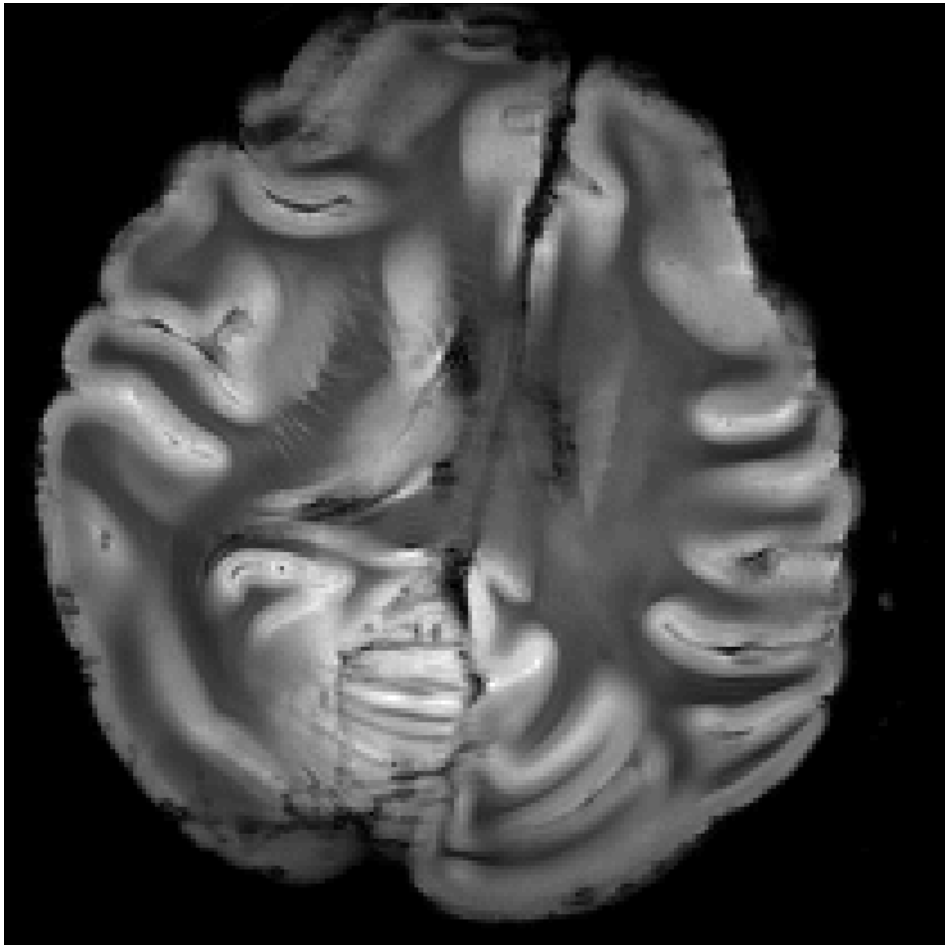
RADIAL

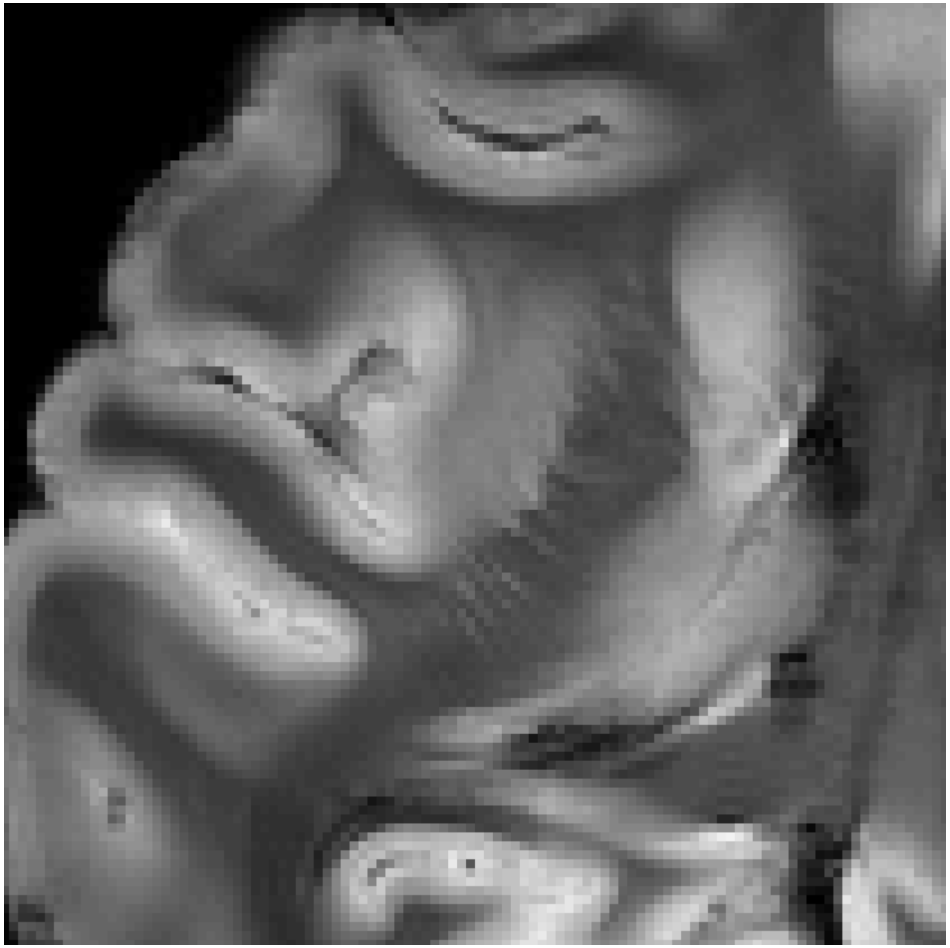


[Lazarus et al., ISMRM 2017]

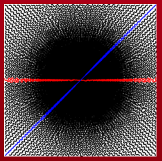
- nc = 60 ⇔ **8.5x acceleration in time**
- ns = 4096 ⇔ Dt = 7.5 µs

REFERENCE WITH NEX=32



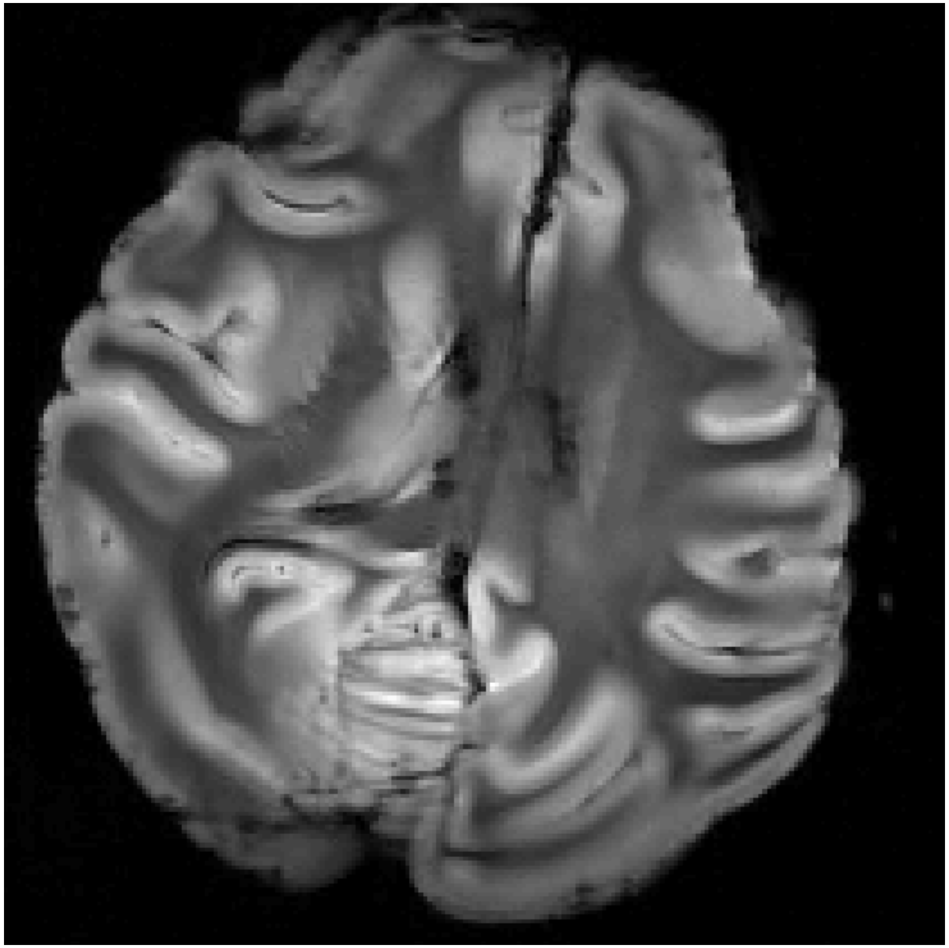


SPARKLING + FISTA - NEX=32



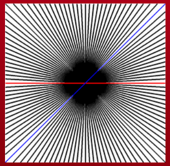
nc=60 and ns=4096 - FISTA

Acceleration factor = 8.5



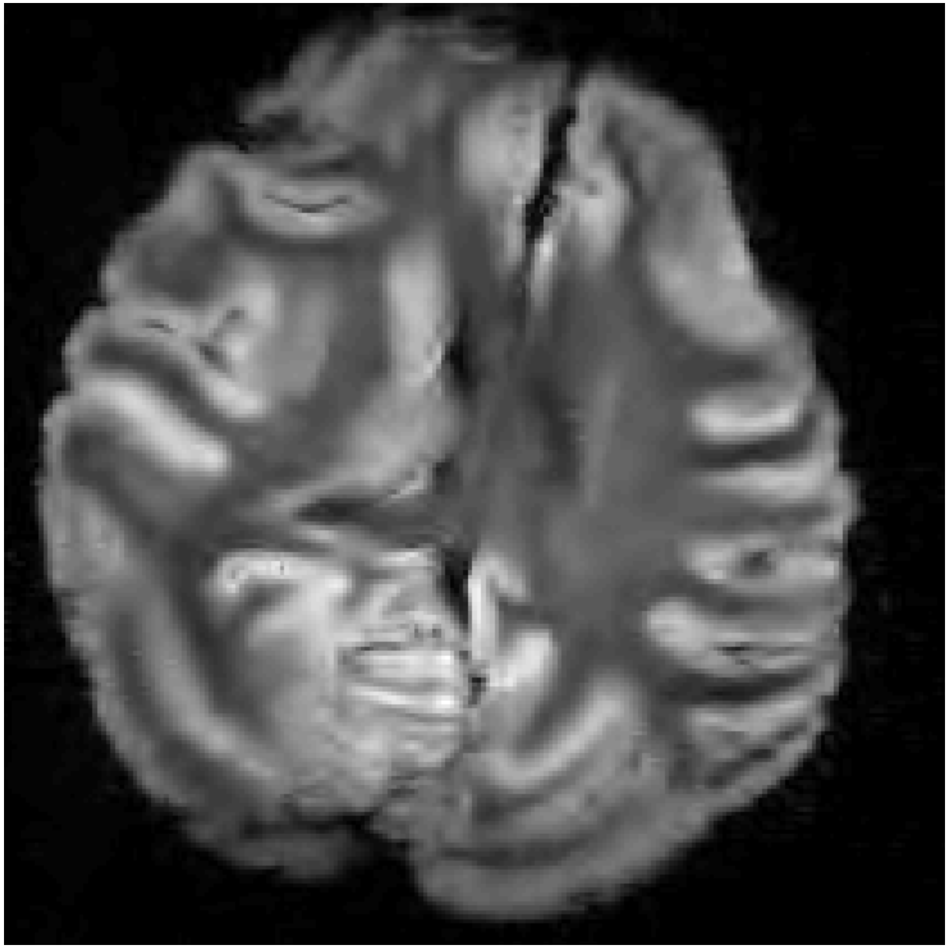
$\lambda = 1.10^{-6}$

III. RADIAL TO CORNERS + FISTA - NEX=32



nc=60 and ns=4096 - FISTA

Acceleration factor = 8.5



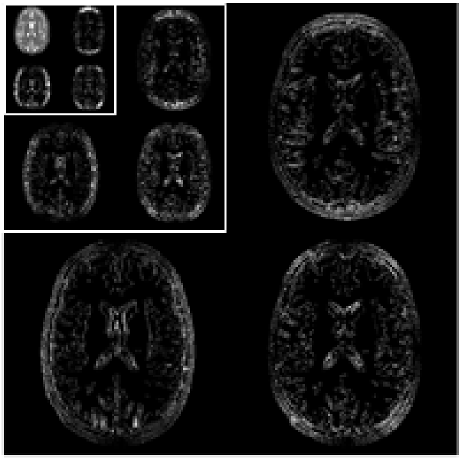
λ=2.10^{-5}

DIRECTION 2: RICHER DICTIONARIES

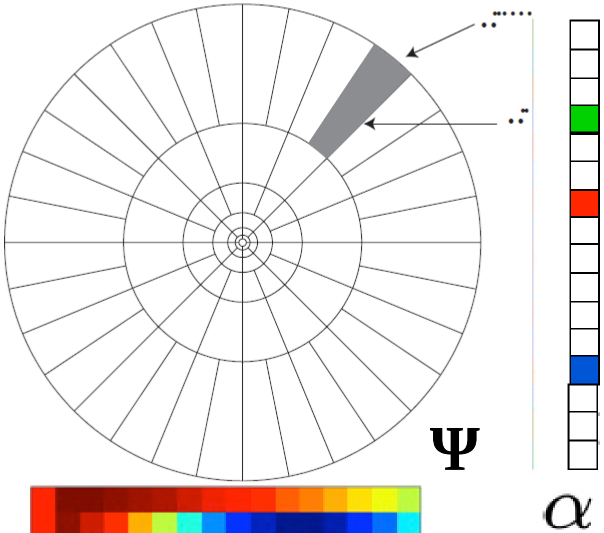
- Better or adaptive sparse modeling
  - **Either fixed such as curvelets, starlets, shearlets …**
  - **Or learned from (trained) or compressed data**

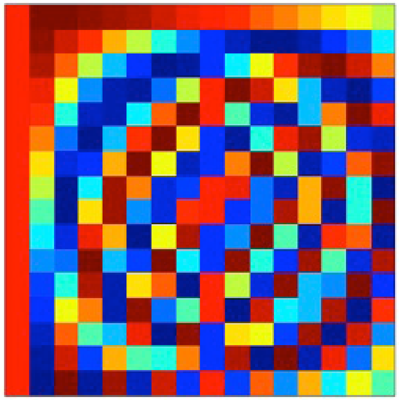
is not sparse…
… but compressible!

Wavelet decomposition of MRI scan is sparse!



3 levels of
decomposition





Redundant transforms **Ψ** induce sparser decompositions **α**

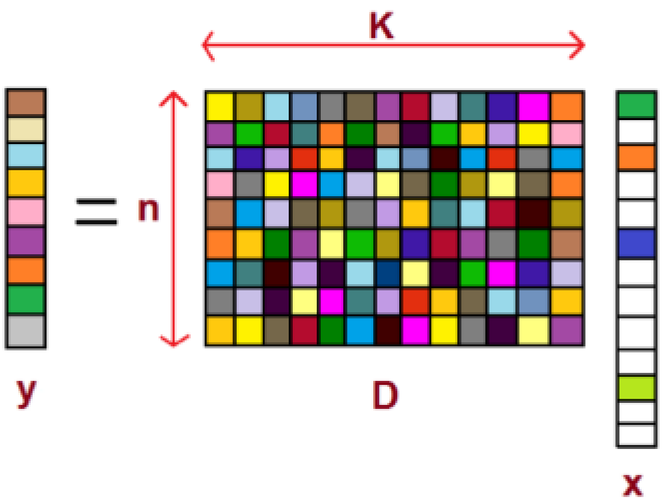
Curvelets
Starlets, shearlets

Dictionary Learning for Blind CS-MRI

[Ravishankar & Bresler, 2011, 13, 15 & 16]:
IEEE Trans Med Imaging, SIAM Imaging Sci,
IEEEComput Imaging

SYNTHESIS MODEL FOR SPARSE REPRESENTATION

- Given a signal $\boldsymbol{y} \in \mathbb{R}^n$ ($\mathbb{C}^n$ in MRI) and dictionary $\boldsymbol{D} \in \mathbb{R}^{n \times K}$, we assume $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 \ll K \Rightarrow$ a union of subspaces model.



- Real-world signals modeled as $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}$ is a deviation term.
- Given $\boldsymbol{D}$, sparsity level s, the *synthesis sparse coding* problem is:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \leqslant s$$

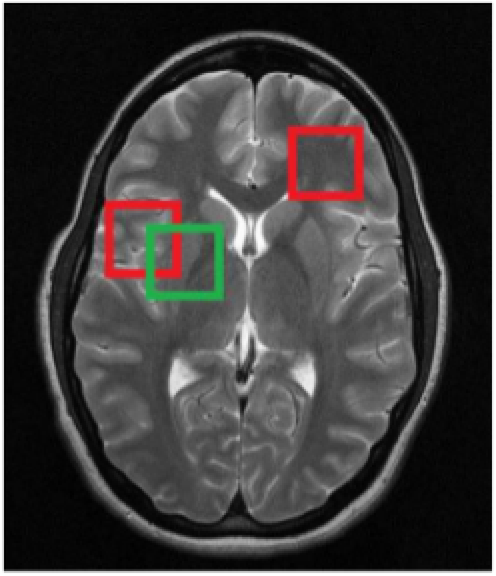
- This problem is **NP-hard**.

SYNTHESIS DICTIONARY LEARNING

- The DL problem (NP-hard):

$$\min_{\boldsymbol{D},\boldsymbol{B}} \sum_{j=1}^{N} \|\boldsymbol{R}_j \boldsymbol{x} - \boldsymbol{D}\boldsymbol{b}_j\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{d}_k\|_2 = 1 \forall k, \ \|\boldsymbol{b}_j\|_0 \leqslant s, \forall j. \tag{1}$$

- $\boldsymbol{R}_j \boldsymbol{x} \in \mathbb{C}^n$: $\sqrt{n} \times \sqrt{n}$ patch indexed by location in image
- $\boldsymbol{R}_j$ extracts patch with upper left corner located in pixel j.
- $\boldsymbol{D} \in \mathbb{C}^{n \times K}$: patch based dictionary.
- $\boldsymbol{b}_j$: sparse, $\boldsymbol{x}_j \approx \boldsymbol{D}\boldsymbol{b}_j$.
- s: sparsity level, $\boldsymbol{B} = \left[\boldsymbol{b}_1 | \boldsymbol{b}_2 | \cdots | \boldsymbol{b}_N\right]$.
- DL minimizes fit error of all patches using sparse representations w.r.t. $\boldsymbol{D}$.



SYNTHESIS-BASED BLIND COMPRESSED SENSING

[Ravishankar and Bresler, IEEE TMI 2011]

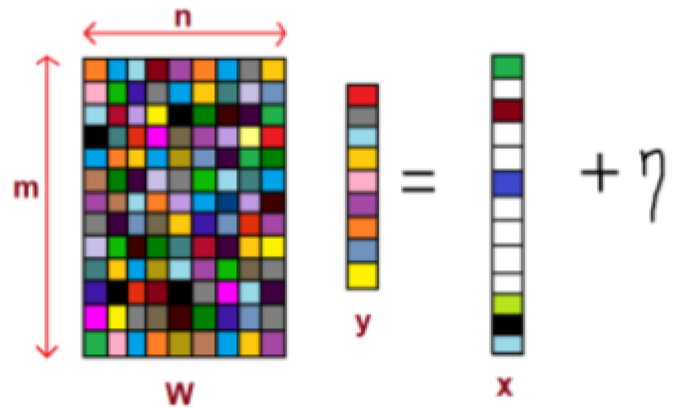
$$(P0): \quad \min_{\boldsymbol{x},\boldsymbol{D},\boldsymbol{B}} \overbrace{\sum_{j=1}^{N} \|\boldsymbol{R}_j \boldsymbol{x} - \boldsymbol{D}\boldsymbol{b}_j\|_2^2}^{\text{Sparse fitting Regularizer}} + \nu \overbrace{\|\boldsymbol{A}_\Gamma \boldsymbol{x} - \boldsymbol{y}\|_2^2}^{\text{Data Fidelity}}$$

$$\text{s.t. } \|d_k\|_2 = 1, \forall k, \quad \|\boldsymbol{b}_j\|_0 \leqslant s, \forall j.$$

- $\boldsymbol{B} \triangleq [\boldsymbol{b}_1|\boldsymbol{b}_2|\cdots|\boldsymbol{b}_N] \in \mathbb{C}^{n\times N}$.
- $(P0)$ learns $\boldsymbol{D} \in \mathbb{C}^{n\times K}$ and reconstructs $\boldsymbol{x}$ from only undersampled dsata $\boldsymbol{y} \Rightarrow$ **dictionary adaptive to underlying image.**
- $(P0)$ is NP-hard, non-convex even if the ℓ_0 semi-norm is relaxed to the ℓ_1 norm.
- **DLMRI[1] solves $(P0)$ for MRI and works better than non-adaptive CS.**
- Synthesis BCS algorithms have no guarantees and are expensive.

ALTERNATIVE: SPARSIFYING TRANSFORM MODEL

- Given a signal $\boldsymbol{y} \in \mathbb{R}^n$ ($\mathbb{C}^n$ in MRI) and transform $\boldsymbol{W} \in \mathbb{R}^{m \times n}$, we model $\boldsymbol{W}\boldsymbol{y} = \boldsymbol{x} + \eta$ with $\|\boldsymbol{x}\|_0 \ll m$ and η an error term.



- Natural signals are naturally sparse in Wavelets, DCT.
- Given $\boldsymbol{W}$, sparsity level s, the *transform sparse coding* problem is:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{W}\boldsymbol{y} - \boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \leqslant s$$

- $\widehat{\boldsymbol{x}} = H_s(\boldsymbol{W}\boldsymbol{y})$ computed by thresholding $\boldsymbol{W}\boldsymbol{y}$ to the s largest magnitude elements. Sparse coding is cheap. Signal recovered as $\boldsymbol{W}^\dagger \widehat{\boldsymbol{x}}$.
- Sparsifying transforms exploited for compression (JPEG 2000), ...

ALTERNATIVE: SPARSIFYING TRANSFORM LEARNING

- **Square Transform Models**
  - Unstructured transform Learning [Ravishankar and Bresler, 2013]
  - Doubly sparse transform learning
  - Online learning for big data [Ravishankar et al., 2015]
  - ...
- **Overcomplete Transform Models**
  - Unstructured overcomplete transform learning
  - Learning structured overcomplete transforms with block cosparsity (OCTOBOS) [Wen et al., 2015]
- Applications: Sparse representations, Image & Video denoising, Classification, Blind Compressed Sensing (BCS) for imaging.

SQUARE TRANSFORM LEARNING FORMULATION

$$(P1): \quad \min_{\boldsymbol{W},\boldsymbol{B}} \overbrace{\sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2}^{\text{Sparsification Error}} + \lambda \overbrace{\left(0.5\|\boldsymbol{W}\|_F^2 - \log|\det \boldsymbol{W}|\right)}^{\text{Regularizer}}$$

$$\text{s.t. } \|\boldsymbol{b}_j\|_0 \leqslant s, \forall j.$$

- Sparsification Error – measures deviation of data in transform domain from perfect sparsity.
- Regularizer enables complete control over conditioning & scaling of $\boldsymbol{W}$.
- If $\exists(\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{B}})$ such that the condition number $\kappa(\widehat{\boldsymbol{W}}) = 1$, $\widehat{\boldsymbol{W}}\boldsymbol{R}_j\boldsymbol{x} = \widehat{\boldsymbol{b}}_j$, $\|\widehat{\boldsymbol{b}}_j\|_0 \leqslant s, \forall j \Rightarrow$ **globally identifiable by solving $(P1)$.**
- $(P1)$ favors both a low sparsification error and good conditioning.
- The solution to $(P1)$ is unitary as $\lambda \to +\infty$.

TRANSFORM-BASED BLIND COMPRESSED SENSING (BCS)

$$(P2): \quad \min_{\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{B}} \overbrace{\sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2}^{\text{Sparsification Error}} + \nu \overbrace{\|\boldsymbol{A}_\Gamma \boldsymbol{x} - \boldsymbol{y}\|_2^2}^{\text{Data Fidelity}} + \lambda \overbrace{v(\boldsymbol{W})}^{\text{Regularizer}}$$

$$\text{s.t.} \quad \sum_{j=1}^{N} \|\boldsymbol{b}_j\|_0 \leqslant s, \quad \|\boldsymbol{x}\|_2 \leqslant C.$$

- $(P2)$ learns $\boldsymbol{W}$, reconstructs $\boldsymbol{x}$ from only undersampled data $\boldsymbol{y}$ $\Rightarrow$.transform adaptive to underlying image.
- Regularizer $v(\boldsymbol{W}) \triangleq 0.5\|\boldsymbol{W}\|_F^2 - \log|\det \boldsymbol{W}|$ controls scaling and κ of $\boldsymbol{W}$.
- $\|\boldsymbol{x}\|_2 \leqslant C$ is an energy/range constraint, with $C > 0$.

TRANSFORM-BASED BCS: IDENTIFIABILITY & UNIQUENESS

Proposition 1

Let $\boldsymbol{x} \in \mathbb{C}^p$ and let $\boldsymbol{y} = \boldsymbol{A}_\Gamma \boldsymbol{x}$ with $\boldsymbol{A} \in \mathbb{C}^{m \times p}$ Suppose:

- $\|\boldsymbol{x}\|_2 \leqslant C$
- $\boldsymbol{W} \in \mathbb{C}^{n \times n}$ is a unitary transform
- $\sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x}\|_0 \leqslant s$

Further, let $\boldsymbol{B}$ denote the matrix that has $\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x}$ as its columns. Then, $(\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{B})$ is a global minimizer of Problem $(P2)$, i.e., it is **identifiable** by solving $(P2)$.

- Given minimizer $(\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{B})$ of $(P2)$, $(\boldsymbol{x}, \boldsymbol{\Theta W}, \boldsymbol{\Theta B})$ is another **equivalent minimizer** $\forall \boldsymbol{\Theta}$ s.t. $\boldsymbol{\Theta}^{\mathrm{H}}\boldsymbol{\Theta} = \boldsymbol{I}$, $\sum_j \|\boldsymbol{\Theta b}_j\|_0 \leqslant s$. Th optimal $\boldsymbol{x}$ is invariant to such transformations of $(\boldsymbol{W}, \boldsymbol{B})$.

BLOCK COORDINATE DESCENT (BCD) ALGORITHM FOR (P2)

- $(P2)$ solved by alternating between updating $\boldsymbol{W}, \boldsymbol{B}, \boldsymbol{x}$.
- Alternate a few times between the $\boldsymbol{W}$ and $\boldsymbol{B}$ updates, before performing an image update.

- **Sparse Coding Step** solves $(P2)$ for $\boldsymbol{B}$ with fixed $\boldsymbol{x}, \boldsymbol{W}$.

$$\min_{\boldsymbol{B}} \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2 \quad \text{s.t.} \quad \sum_{j=1}^{N} \|\boldsymbol{b}_j\|_0 \leqslant s. \tag{2}$$

- **Cheap Solution:** Let $\boldsymbol{Z} \in \mathbb{C}^{n \times N}$ be the matrix with $\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x}$ as its columns. $\widehat{\boldsymbol{B}} = H_s(\boldsymbol{Z})$ computed exactly by zeroing out all but the s largest magnitude coefficients in $\boldsymbol{Z}$.

BCD ALGORITHM FOR (P2): TRANSFORM UPDATE STEP

- **Transform Update Step**: Solves $(P2)$ for $\boldsymbol{W}$ with fixed $\boldsymbol{x}, \boldsymbol{B}$.

$$\min_{\boldsymbol{W}} \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2 + 0.5\lambda\|\boldsymbol{W}\|_F^2 - \lambda \log|\det \boldsymbol{W}| \qquad (3)$$

  - Let $\boldsymbol{X} \in \mathbb{C}^{n \times N}$ be the matrix with $\boldsymbol{R}_j\boldsymbol{x}$ as its columns.
  - **Closed-form solution:**

$$\widehat{\boldsymbol{W}} = 0.5\boldsymbol{R}\left(\boldsymbol{\Sigma} + \left(\boldsymbol{\Sigma}^2 + 2\lambda\boldsymbol{I}\right)^{1/2}\right)\boldsymbol{V}^{\mathrm{H}}\boldsymbol{L}^{-1} \qquad (4)$$

    where $\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} + 0.5\lambda\boldsymbol{I} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{H}}$, and $\boldsymbol{L}^{-1}\boldsymbol{X}\boldsymbol{B}^{\mathrm{H}}$ has a full SVD of $\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{R}^{\mathrm{H}}$.
  - Solution is unique if and only if $\boldsymbol{X}\boldsymbol{B}^{\mathrm{H}}$ is non-singular.

BCD ALGORITHM FOR (P2): IMAGE UPDATE STEP

- **Image Update Step**: Solves $(P2)$ for $\boldsymbol{x}$ with fixed $\boldsymbol{W}, \boldsymbol{B}$.

$$\min_{\boldsymbol{x}} \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2 + \nu\|\boldsymbol{A}_\Gamma\boldsymbol{x} - \boldsymbol{y}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_2 \leqslant C. \tag{5}$$

- Least squares problem with ℓ_2 norm constraint.
- Solution is unique as long as the set of overlapping patches covers all all image pixels.
- **Solve Least squares Lagrangian formulation:**

$$\min_{\boldsymbol{x}} \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2 + \nu\|\boldsymbol{A}_\Gamma\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \mu\left(\|\boldsymbol{x}\|_2^2 - C\right) \tag{6}$$

- The optimal multiplier $\hat{\mu} \in \mathbb{R}_+$ is the smallest real such that $\|\boldsymbol{x}\|_2 \leqslant C$. $\hat{\mu}$ and $\widehat{\boldsymbol{w}}$ can be found cheaply.

BCD CONVERGENCE GUARANTEES - NOTATION

- Define the barrier function $\psi_s(\boldsymbol{B})$ as

$$\psi_s(\boldsymbol{B}) = \begin{cases} 0, & \text{if } \sum_{j=1}^{N} \|\boldsymbol{b}_j\|_0 \leqslant s \\ +\infty, & \text{otherwise} \end{cases}$$

- $\chi_C(\boldsymbol{x})$ is the barrier function corresponding to $\|\boldsymbol{x}\|_2 \leqslant C$.
- $(P2)$ is equivalent to the problem of minimizing the objective

$$g(\boldsymbol{W}, \boldsymbol{B}, \boldsymbol{x}) = \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{R}_j\boldsymbol{x} - \boldsymbol{b}_j\|_2^2 + \nu\|\boldsymbol{A}_\Gamma\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda v(\boldsymbol{W}) + \psi(\boldsymbol{B}) + \chi_C(\boldsymbol{x})$$

- For $\boldsymbol{H} \in \mathbb{C}^{p \times q}$, $\rho_j(\boldsymbol{H})$ is the magnitude of the j^{th} largest element (magnitude-wise) of $\boldsymbol{H}$.
- $\boldsymbol{X} \in \mathbb{C}^{n \times N}$ denotes a matrix with $\boldsymbol{R}_j\boldsymbol{x}$, $1 \leqslant j \leqslant N$, as its columns.

TRANSFORM BCS CONVERGENCE GUARANTEES (1/2)

Theorem 1

For the sequence $\{\boldsymbol{W}^t, \boldsymbol{B}^t, \boldsymbol{x}^t\}$ generated by the BCD Algorithm with initial $(\boldsymbol{W}^0, \boldsymbol{B}^0, \boldsymbol{x}^0)$ we have:

- $\{g(\boldsymbol{W}^t, \boldsymbol{B}^t, \boldsymbol{x}^t)\} \to g^* = g(\boldsymbol{W}^0, \boldsymbol{B}^0, \boldsymbol{x}^0)$.
- $\{\boldsymbol{W}^t, \boldsymbol{B}^t, \boldsymbol{x}^t\}$ is bounded, and all its accumulation points are equivalent, i.e. they achieve the same value g^* of the objective.
- $\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\|_2 \to 0$ as $t \to \infty$.
- Every accumulation point $(\boldsymbol{W}, \boldsymbol{B}, \boldsymbol{x})$ is a **critical point** of g satisfying the following partial global optimality conditions:

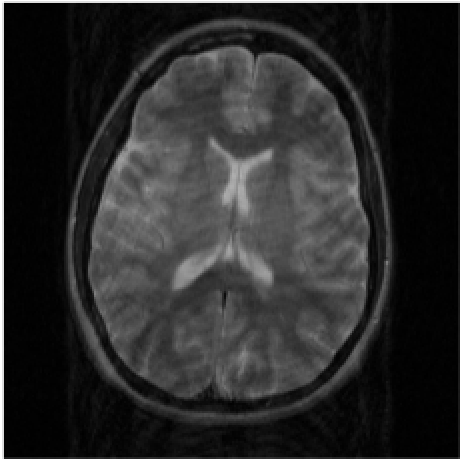
$$\boldsymbol{x} \in \arg\min_{\tilde{\boldsymbol{x}}} g(\boldsymbol{W}, \boldsymbol{B}, \tilde{\boldsymbol{x}}) \tag{7}$$

$$\boldsymbol{W} \in \arg\min_{\tilde{\boldsymbol{W}}} g(\tilde{\boldsymbol{W}}, \boldsymbol{B}, \boldsymbol{x}), \quad \boldsymbol{B} \in \arg\min_{\tilde{\boldsymbol{B}}} g(\boldsymbol{W}, \tilde{\boldsymbol{B}}, \boldsymbol{x}) \tag{8}$$

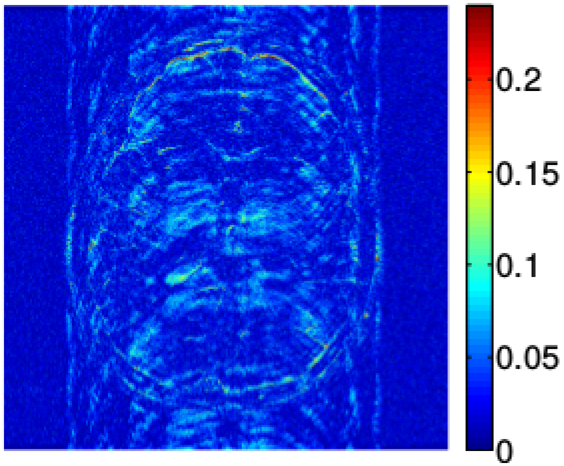
COMPUTATIONAL ADVANTAGES

- Cost per iteration of transform BCS: $O(p^4 NL)$
  - N overlapping patches of size $p \times p$, $\boldsymbol{W} \in \mathbb{C}^{n \times n}$, $n \triangleq p^2$.
  - L # inner alternations between transform updates & sparse coding.
- Cost per iteration of synthesis BCS DLMRI: $O(p^6 NJ)$
  - $\boldsymbol{D} \in \mathbb{C}^{n \times K}$, $n \triangleq p^2$, $K \propto n$, sparsity $s \propto n$.
  - J # inner iterations of dictionary learning using K-SVD [Aharon et al., 2006].
- In practice, transform BCS converges quickly and is much cheaper for large p.
- In 3D or 4D imaging, $n = p^3$ or p^4, and the gain in computations is about a factor of n in order.

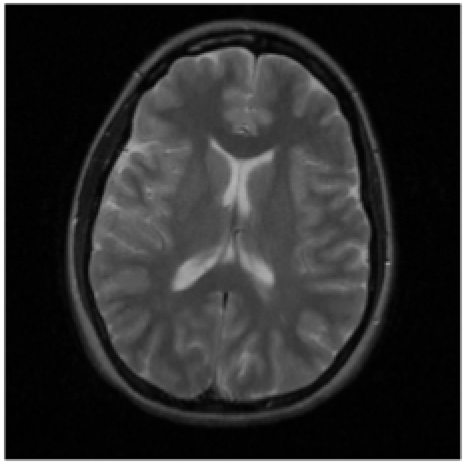
CONVERGENCE & LEARNING – 4x UNDERSAMPLING (S=3.4 %)



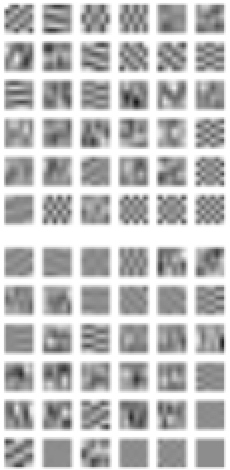
Zero-filling (28.94 dB)



Zero-filling Error

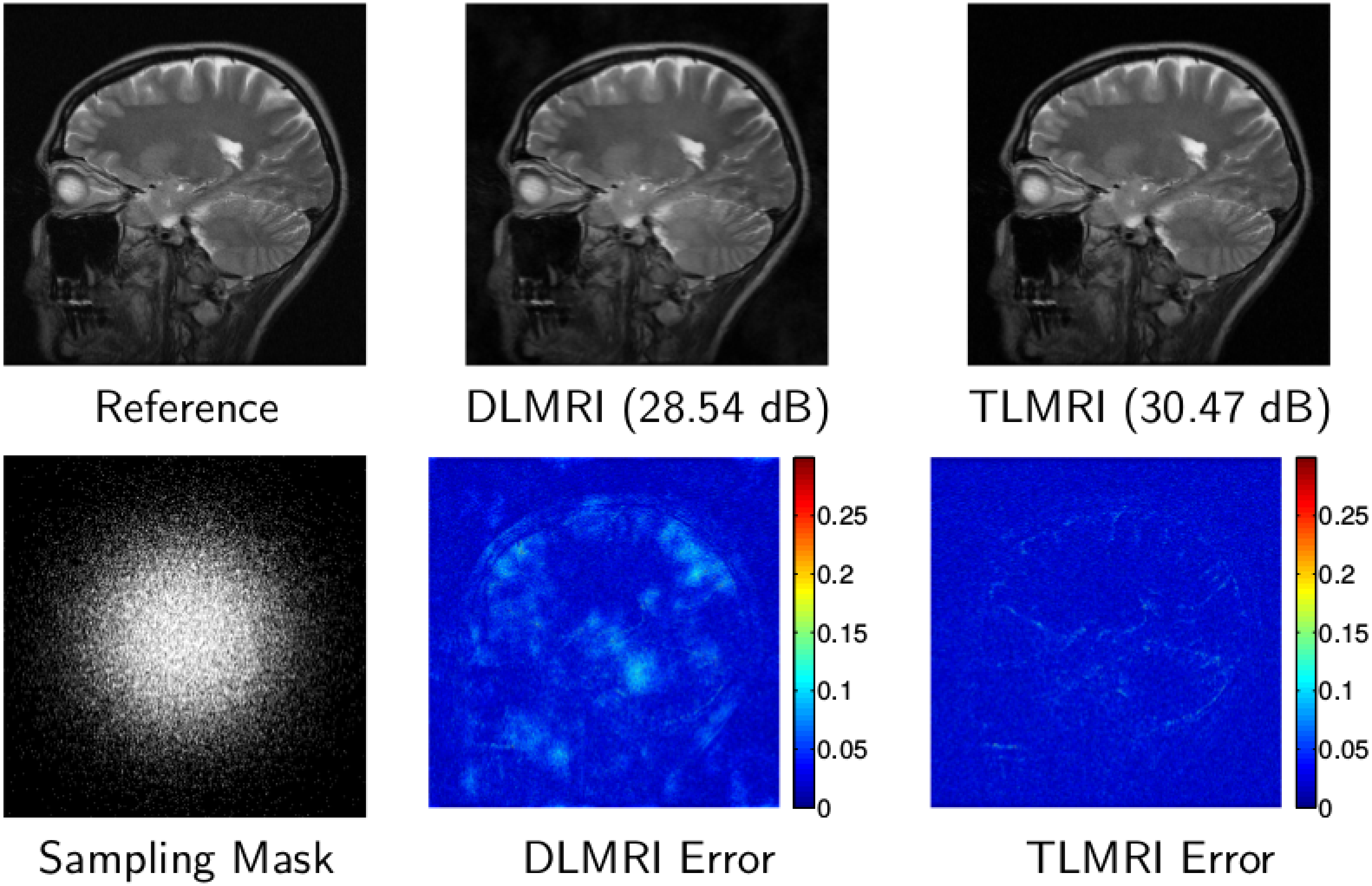


TLMRI (32.66 dB)



real (top), imaginary (bottom) parts of learnt 36×36 W

EXAMPLE – 2D RANDOM 5x UNDERSAMPLING



Comparison of FISTA and TLMRI in the non-Cartesian setting

INTRODUCTION

- **Objective:**

  Compare Transform Learning (TLMRI[1] adapted to NFFT) reconstructions with FISTA-Symmlet reconstructions (currently used), in the case of prospective SPARKLING data.

- **TLMRI parameters:**
  - **λ**: parameter to set the weight on the negative log-determinat+Frobenius norm terms in the problem formulation

    range tested → 0.1, **0.2** (**default**), 0.3
  - **v**: weight on the data fidelity term in the problem formulation

    range tested → 10^5, $\mathbf{10^6}$ (**default**), 10^7, 10^8
  - **s**: sparsity fraction (i.e., fraction of non-zeros in the sparse code matrix)

    range tested → constant s of 0.03, **0.045 (default)**, 0.06 and one s increasing from 0.013 to 0.055
  - **n**: Patch size, i.e., Total number of pixels in a square patch

    range tested → 25, **36 (default)**, 49, 64, 81
- **FISTA parameters:**
  - **λ**: regularization parameter in front of L_1-norm

    → range tested: from 10^{-7} to 10^{-4}

[1]: Ravishankar, Saiprasad and Yoram Bresler. "Efficient Blind Compressed Sensing Using Sparsifying Transforms with Convergence Guarantees and Application to MRI." SIAM Imaging Sciences (2015): 8(4); 2519-2557.

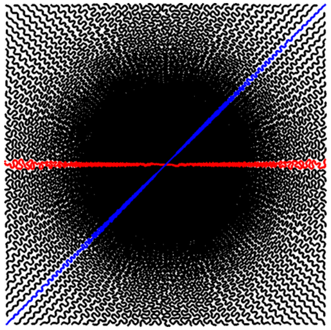
REMINDER: SPARKLING TRAJECTORIES

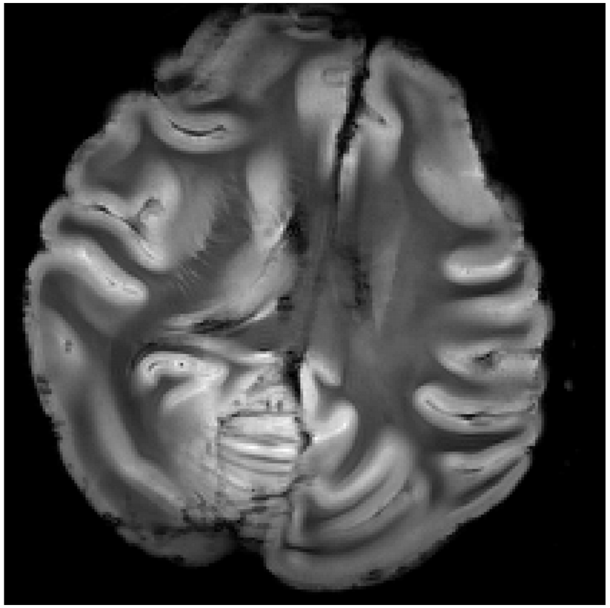
- **Data used for recontructions**
  - N=512
  - Prospective SPARKLING data yielding a 8.5-fold acceleration in time

    (nc=60 segments and ns=4096 samples per segment)
  - 2 cases tested:
    1) **Low SNR: NEX=1**
    2) **High SNR: NEX=32**

- **Image quality metrics:**
  - SSIM
  - NRMSE
  - Reference is the full Cartesian image with very high SNR (NEX=32)

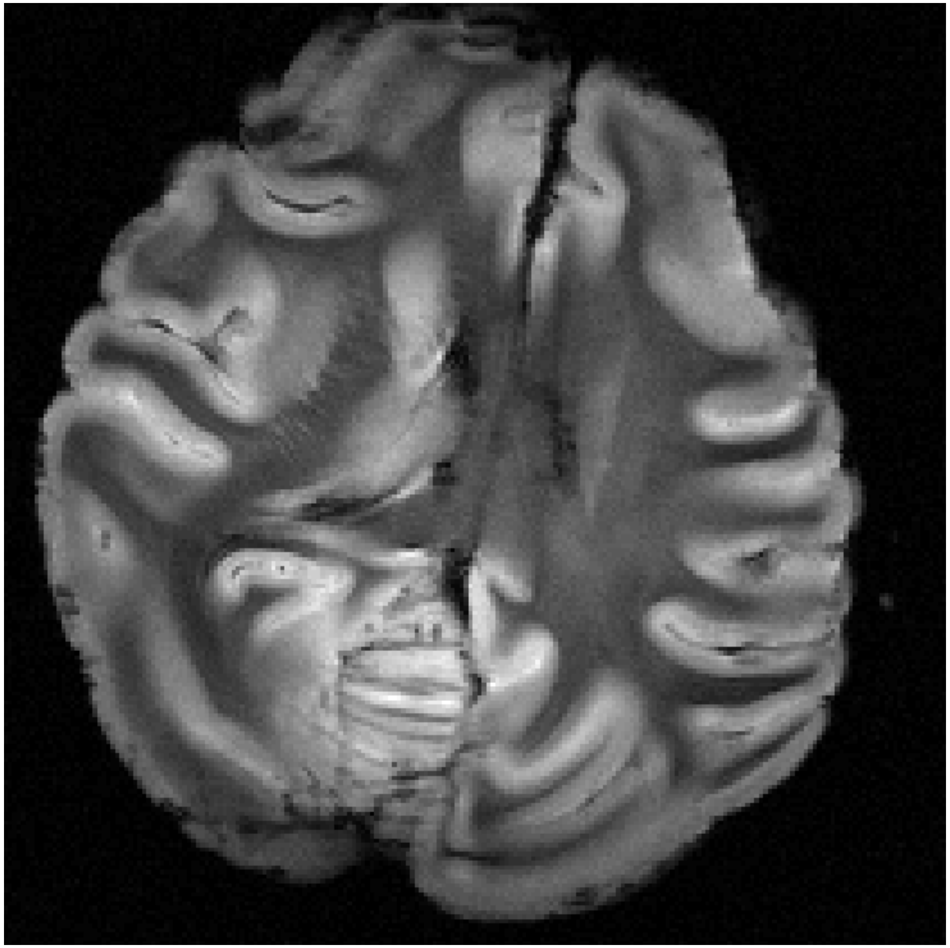
Sequence parameters:
- N=512
- FOV=200x200 mm²
- TR=550 ms and TE=30 ms
- α=25°
- BW=32.55 Hz/px
- **Tobs=30.72 ms**
- Slice thickness: 3 mm
- Single channel receiver coil
- Ex-vivo baboon brain



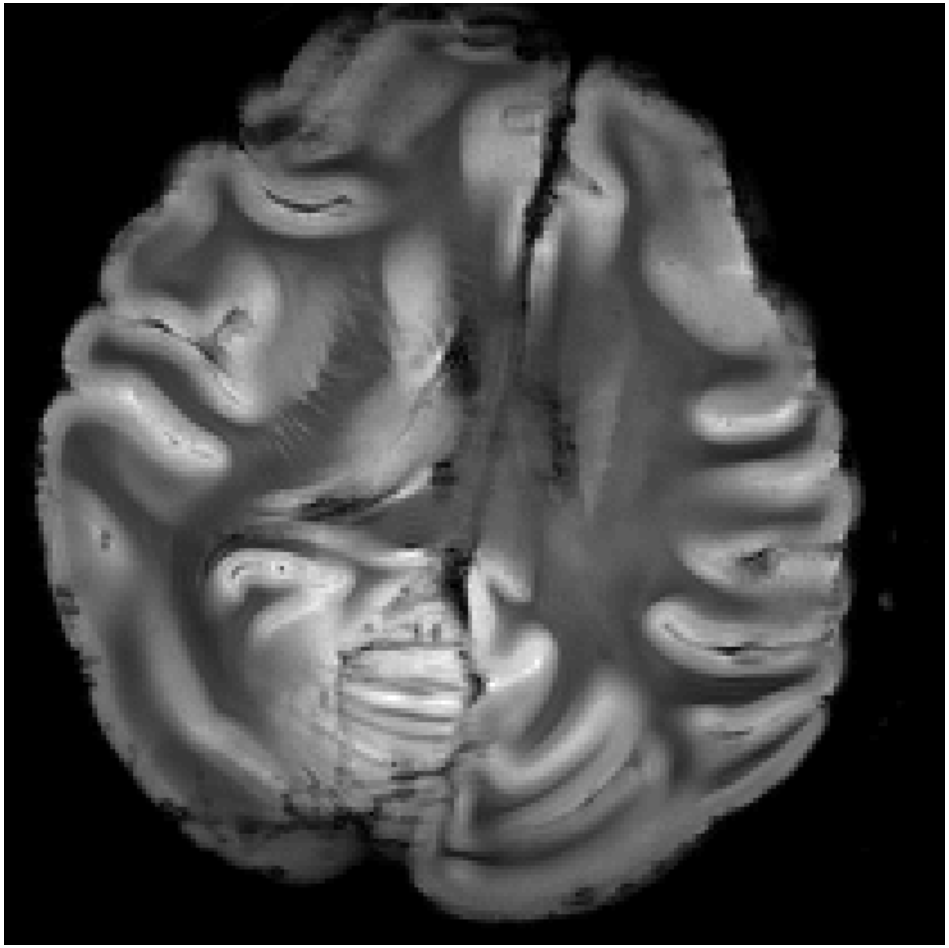


0. FULL CARTASIAN IMAGES FOR NEX=1 AND 32

NEX=1



NEX=32

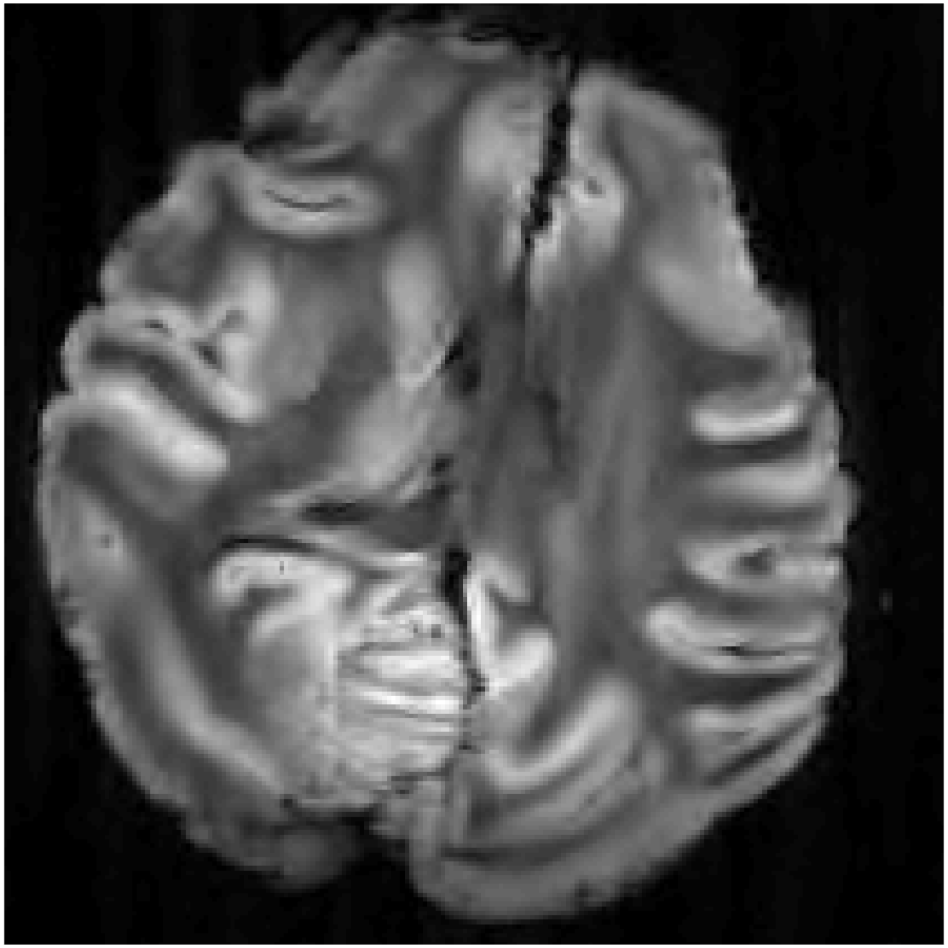


I.
Visual comparison of best image reconstructions

I. BEST RECONSTRUCTIONS: NEX=1

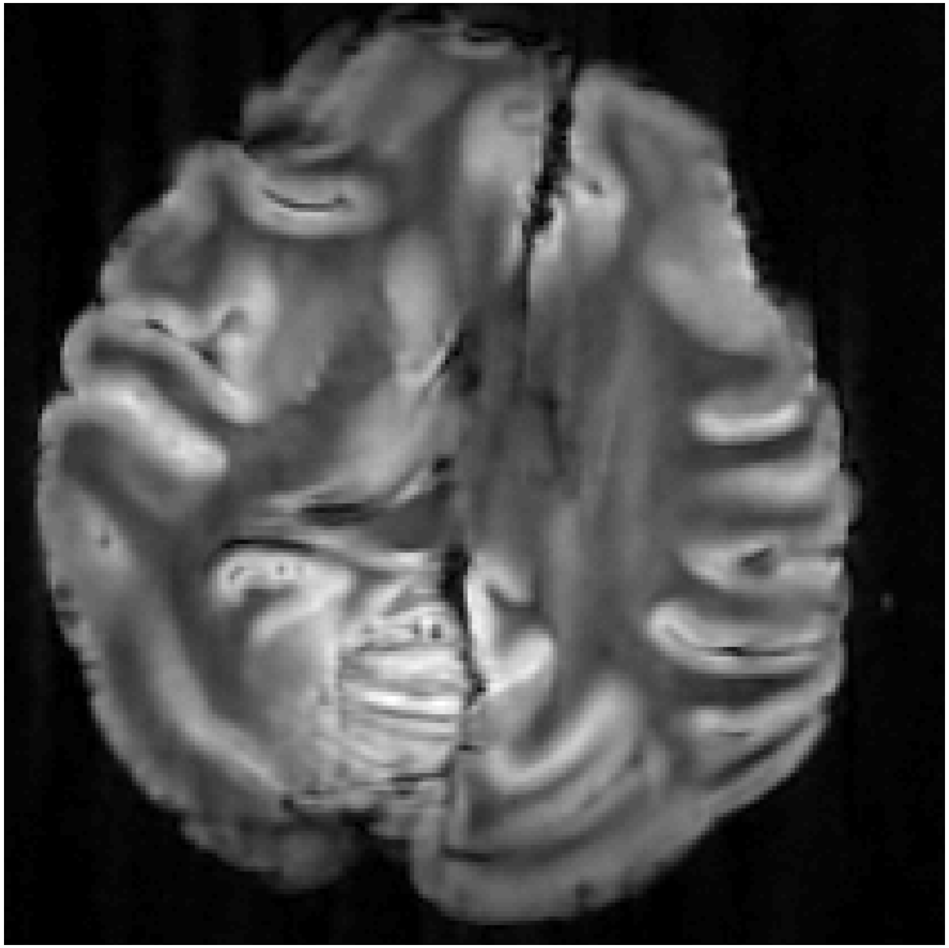
Acceleration factor = 8.5

FISTA-Symmlet



SSIM=0.862
NRMSE=0.133

TLMRI

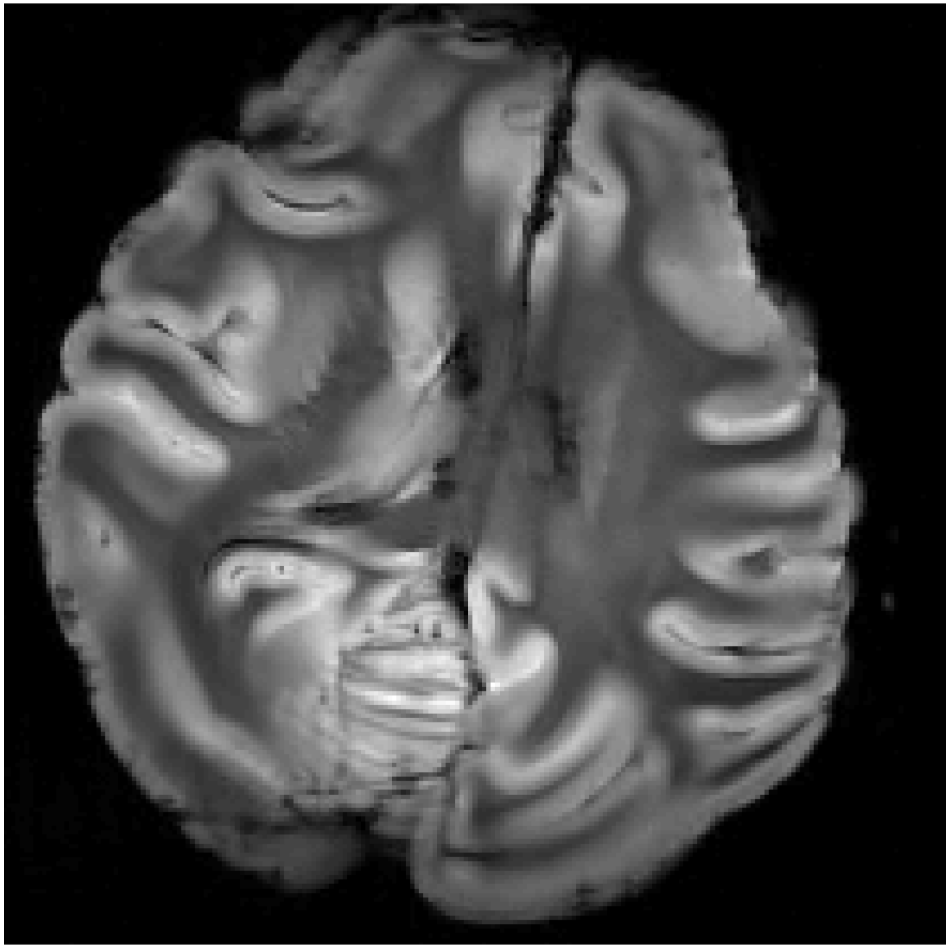


SSIM=0.874
NRMSE=0.136

I. BEST RECONSTRUCTIONS: NEX=32

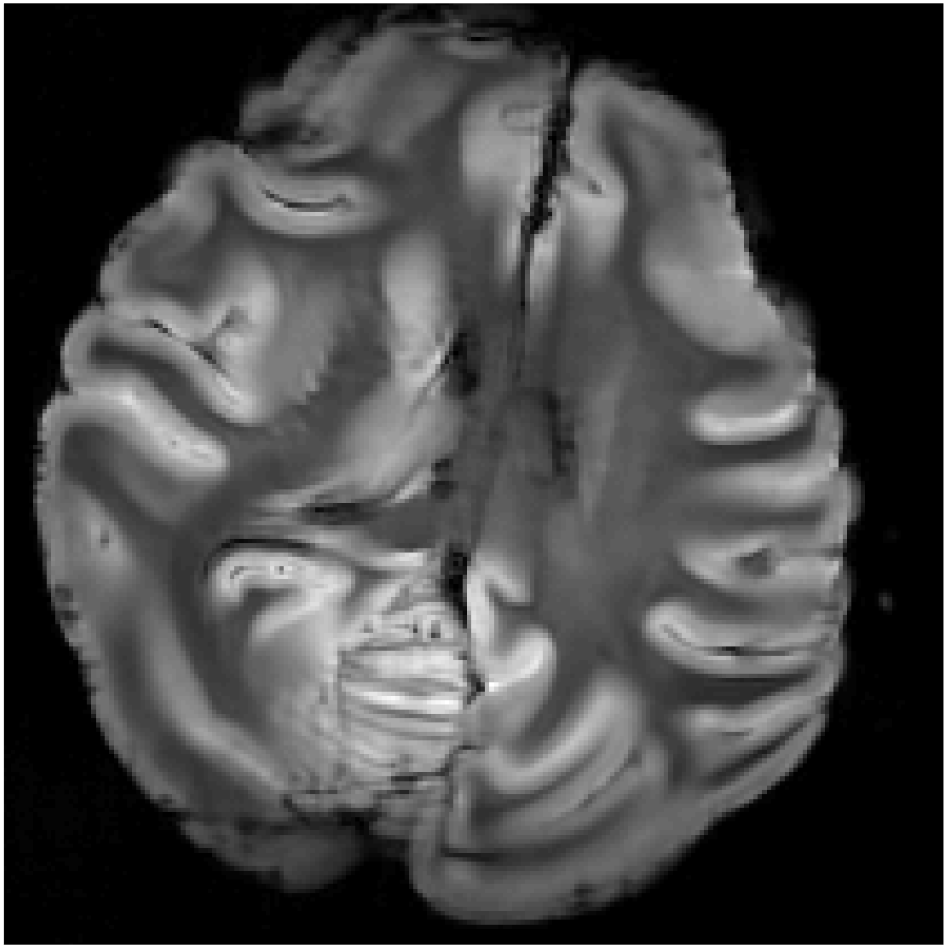
Acceleration factor = 8.5

FISTA-Symmlet



SSIM=0.935
NRMSE=0.078

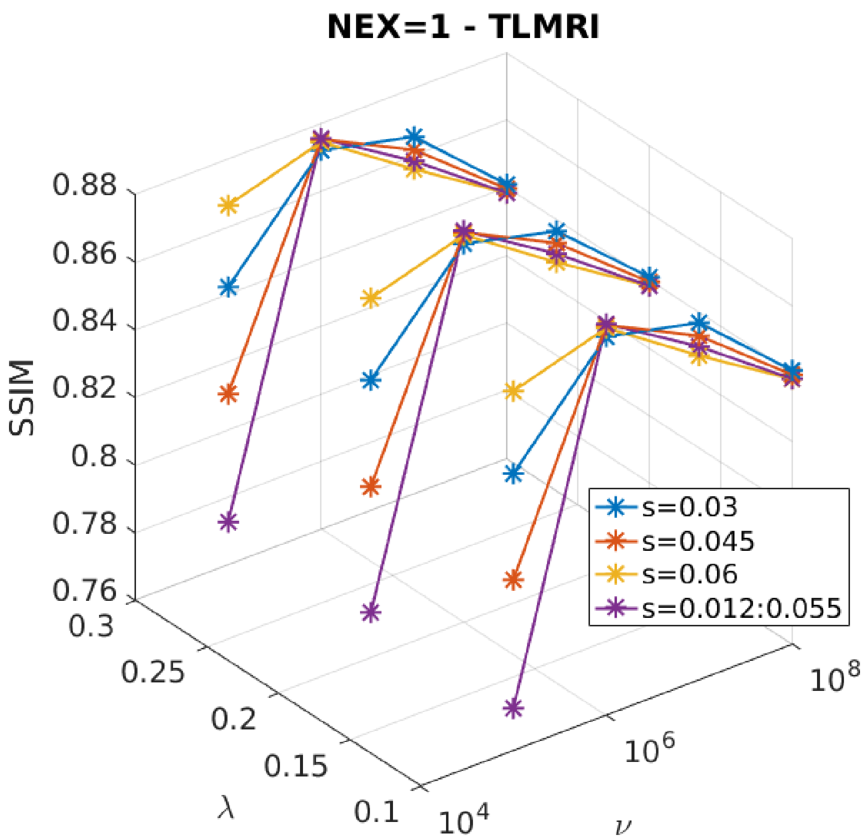
TLMRI

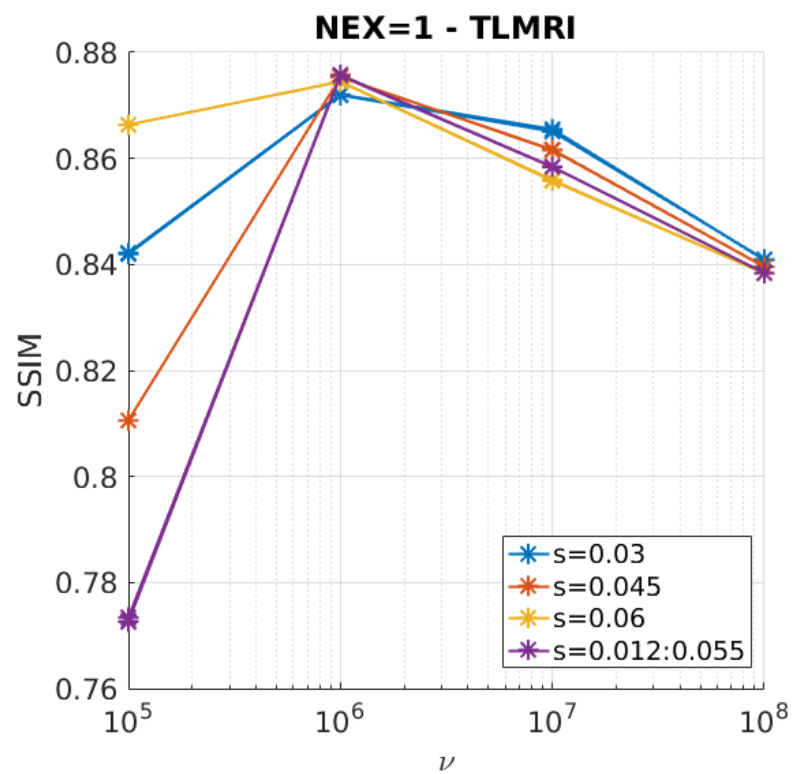


SSIM=0.937
NRMSE=0.079

II. Parameters' influence on TLMRI quality scores for NEX=1

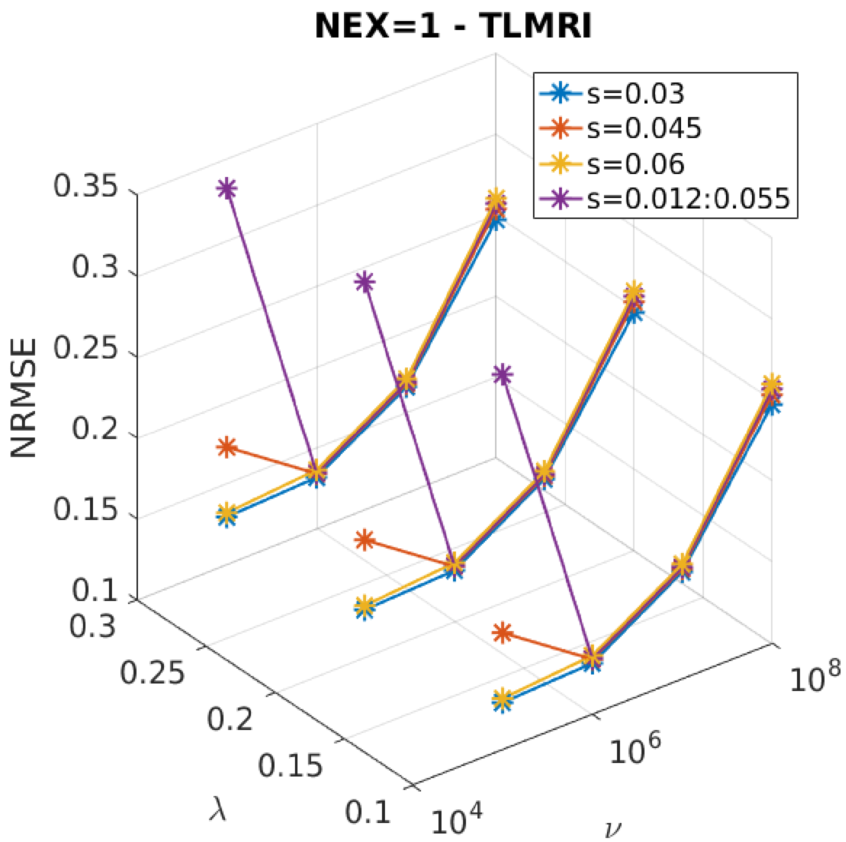
II. NEX=1 – SSIM vs. λ, ν and s

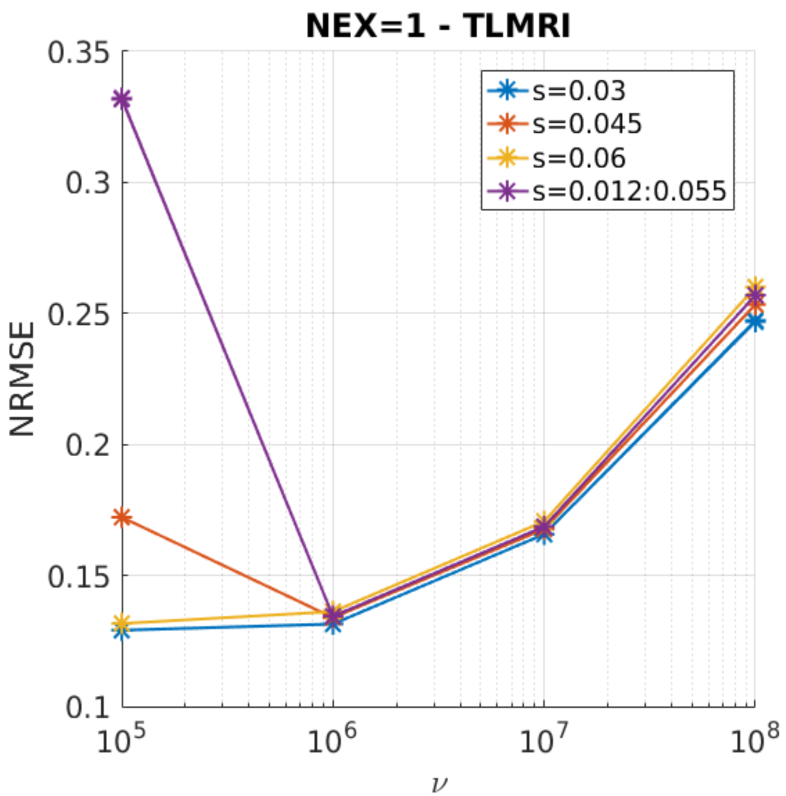




- Optimal value reached for $\nu=10^6$.
- The sparsity fraction s has a significant impact on SSIM scores.
- λ does not influence the SSIM a lot.

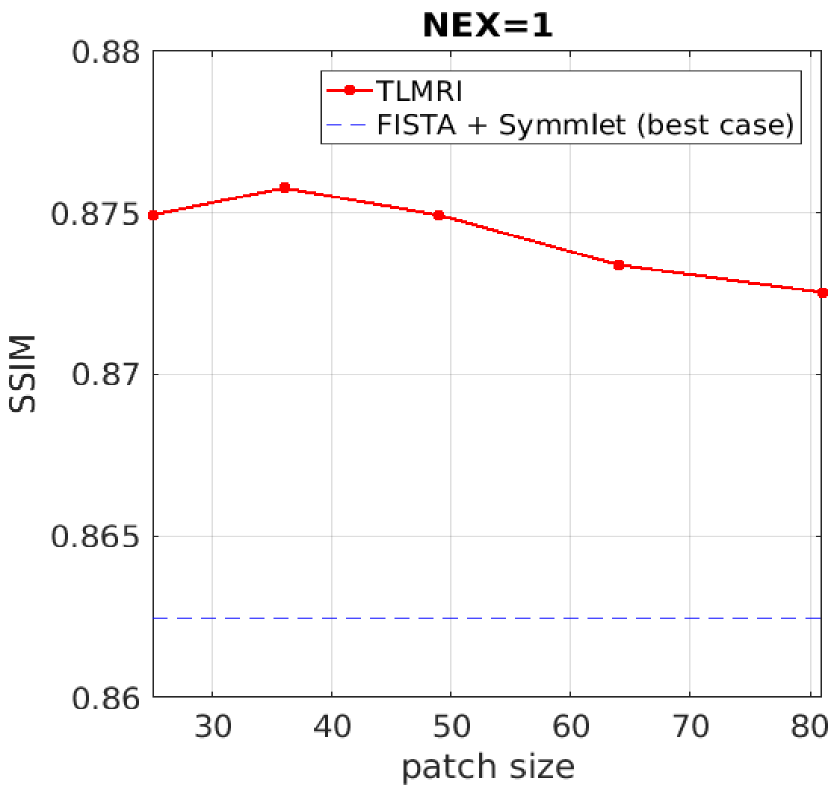
II. NEX=1 – NRMSE vs. λ, ν and s





- Likewise, for NRMSE, there is an optimal value reached for $\nu=10^6$.
- The sparsity fraction does matter.
- λ does not significantly influence the NRMSE.

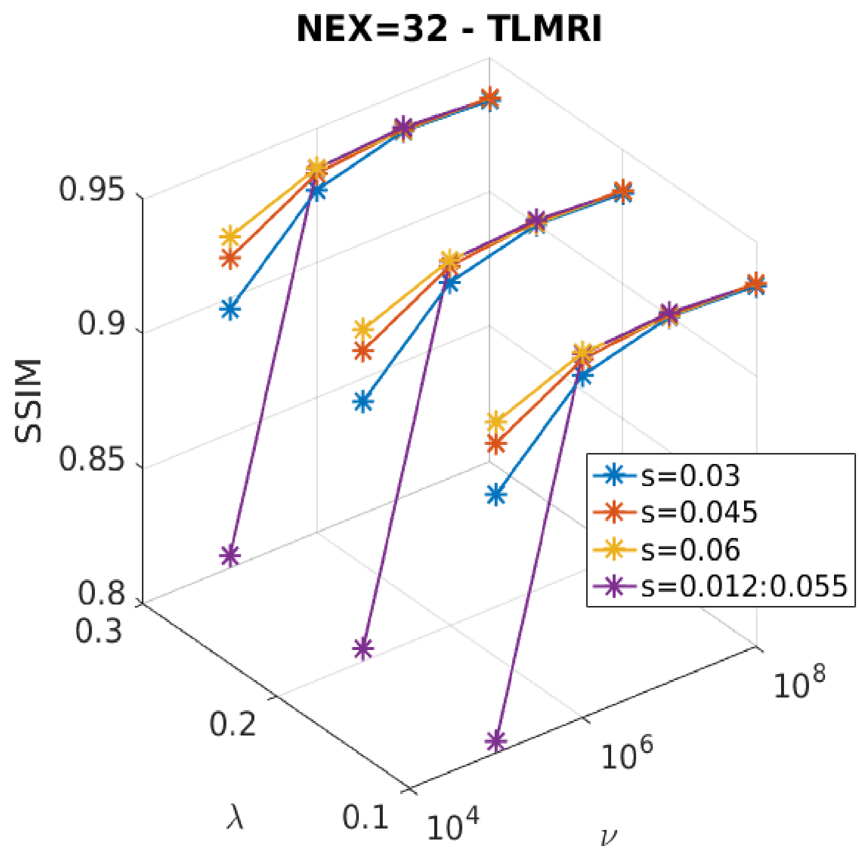
II. NEX=1 – SSIM vs. patch size

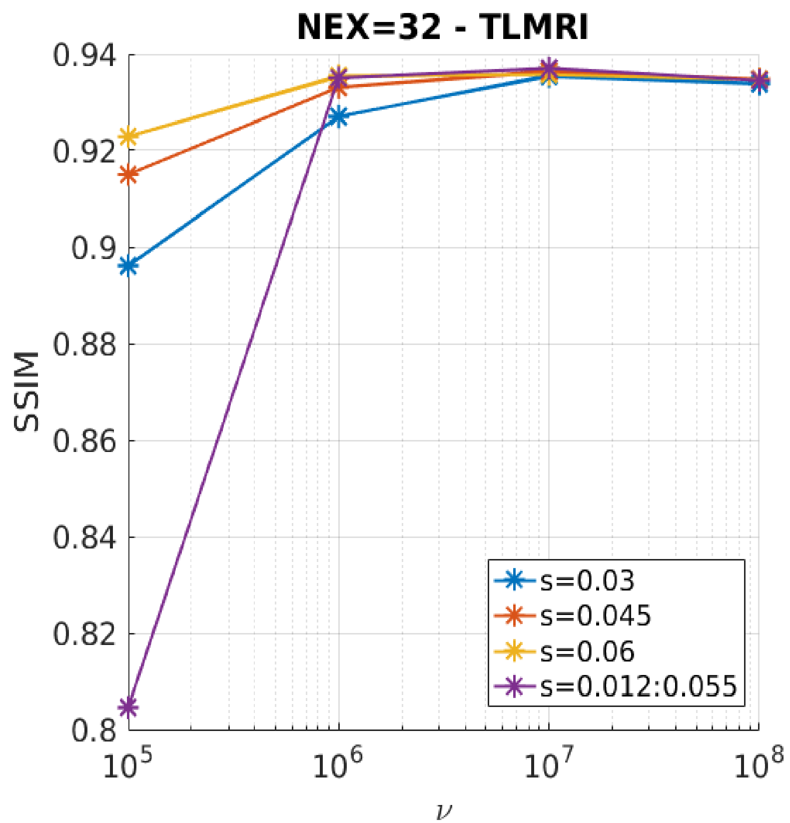


- For NEX=1, patch size does not influence a lot the SSIM scores.

III.
Parameters' influence on TLMRI quality scores for NEX=32

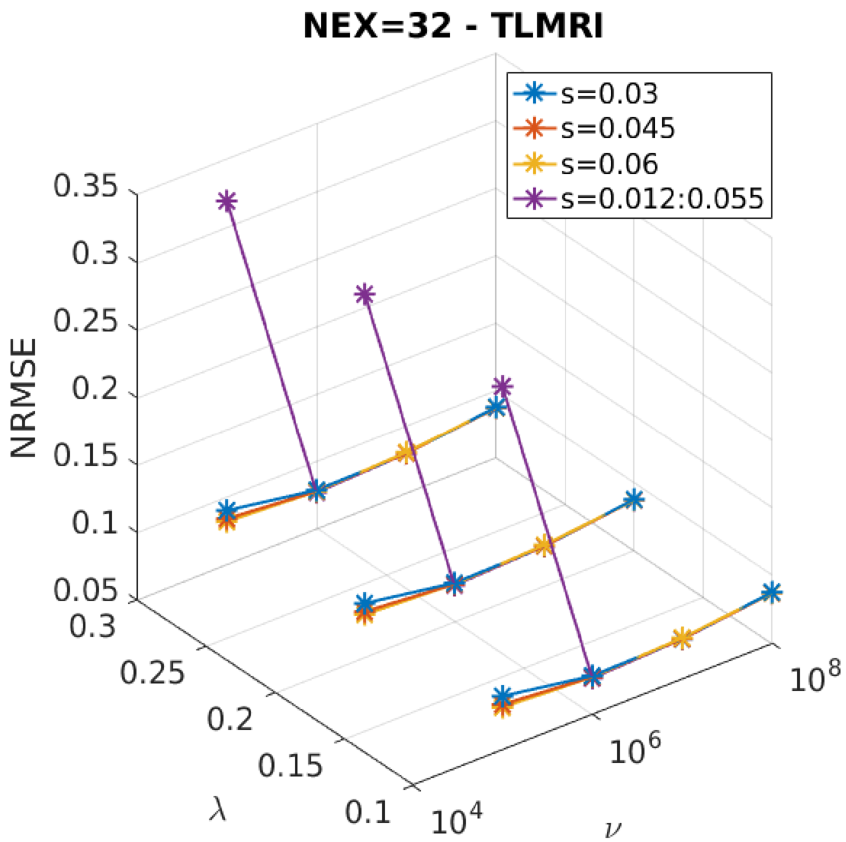
III. NEX=32 – SSIM vs. λ, ν and s

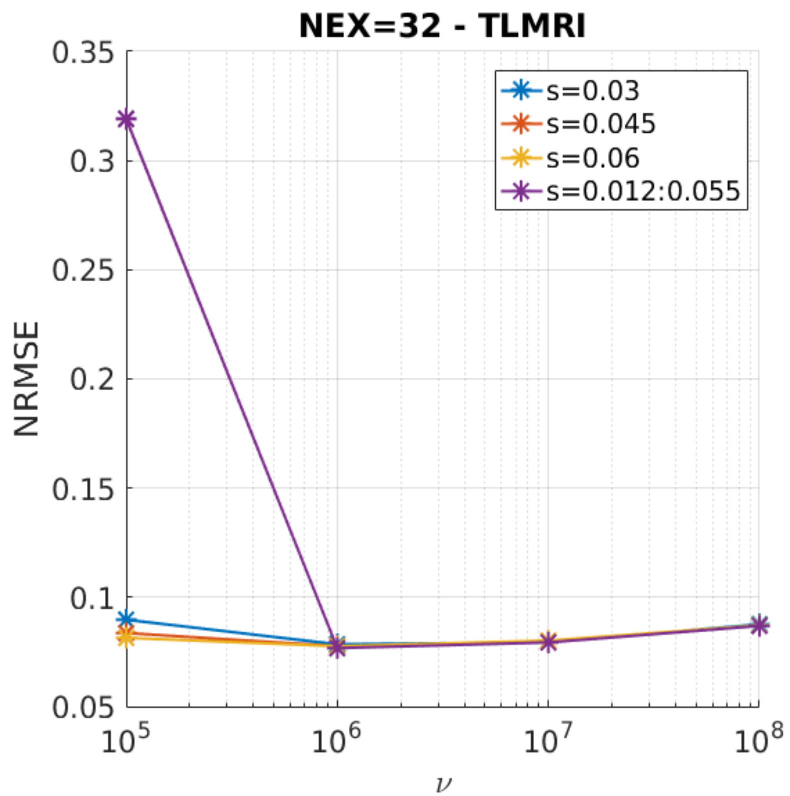




- Optimal value reached for $\nu=10^7$.
- The sparsity fraction s has a significant impact on SSIM scores
- λ does not influence the SSIM a lot.

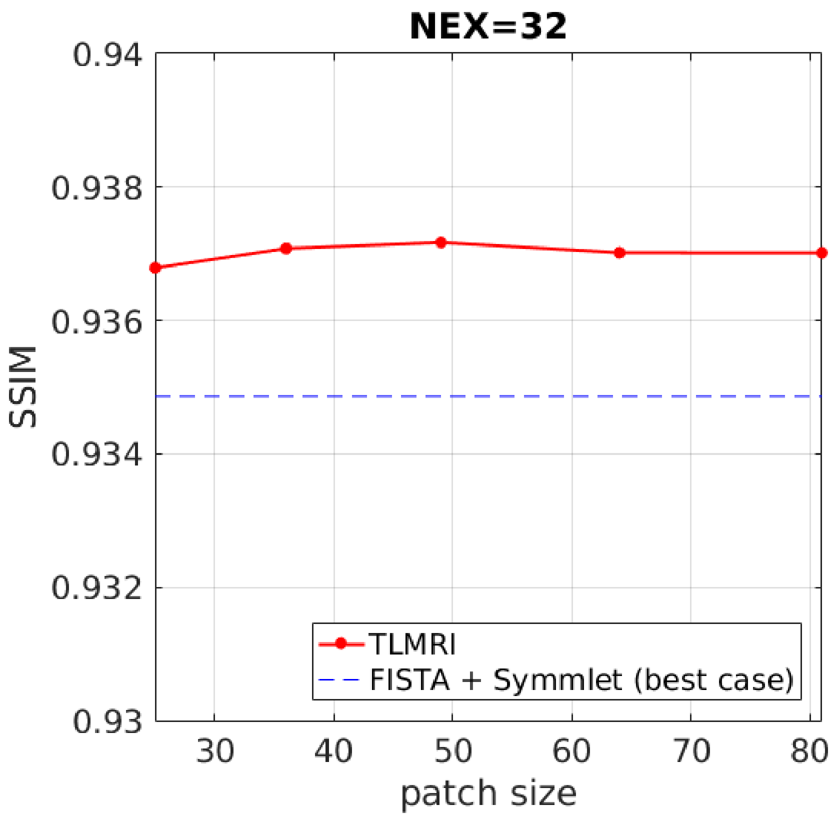
III. NEX=32 – NRMSE vs. λ, ν and s





- Optimal value reached for $\nu=10^6$.
- The sparsity fraction s has an impact on SSIM scores, but less than for NEX=1
- λ does not influence the SSIM a lot.

III. NEX=32 – SSIM vs. patch size



- For NEX=32, patch size does not significantly influence SSIM scores.

CONCLUSIONS

- **TLMRI vs. FISTA-Symmlet**
  - There is a small but noticeable gain in image quality using TLMRI instead of FISTA-Symmlet for the low SNR case of NEX=1.
  - This advantage is hardly visible on the high SNR case of NEX=32.

- **Parameters tuning for TLMRI**
  - Parameters ν (data fidelity parameter) and s (sparsity fraction) have to be tuned carefully since they influence a lot the image quality. Optimal values for the tested cases were close to default settings of TLMRI.
  - Patch sizes n and λ have a small (or no) impact on the image quality.

- **Algorithm durations**
  - One iteration in TLMRI lasts between 60 and 100 s (depending on convergence of PCG).
  - One iteration of FISTA lasts about 0.75 seconds.
  - ….

Thanks for your attention!

Any questions?

16

BIBLIOGRAPHY

[Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006).
rmk-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.
IEEE Transactions on signal processing, 54(11):4311–4322.

[Lustig et al., 2007] Lustig, M., Donoho, D. L., and Pauly, J. M. (2007).
Sparse MRI: The application of compressed sensing for rapid MR imaging.
Magnetic Resonance in Medicine, 58(6):1182–1195.

[Ravishankar and Bresler, 2011] Ravishankar, S. and Bresler, Y. (2011).
Mr image reconstruction from highly undersampled k-space data by dictionary learning.
IEEE transactions on medical imaging, 30(5):1028–1041.

[Ravishankar and Bresler, 2013] Ravishankar, S. and Bresler, Y. (2013).
Learning sparsifying transforms.
IEEE Transactions on Signal Processing, 61(5):1072–1086.

[Ravishankar et al., 2015] Ravishankar, S., Wen, B., and Bresler, Y. (2015).
Online sparsifying transform learning—part i: Algorithms.
IEEE Journal of Selected Topics in Signal Processing, 9(4):625–636.

[Wen et al., 2015] Wen, B., Ravishankar, S., and Bresler, Y. (2015).
Structured overcomplete sparsifying transform learning with convergence guarantees and applications.
International Journal of Computer Vision, 114(2-3):137–167.