Collaborative Sparsity and Compressive MRI

Robert Crandall

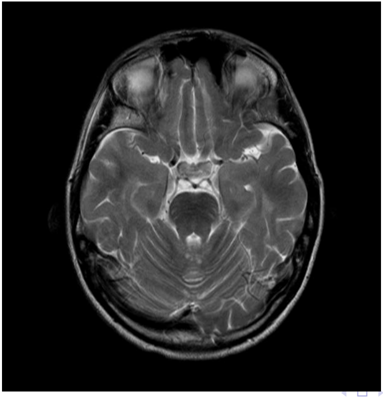
February 14, 2013

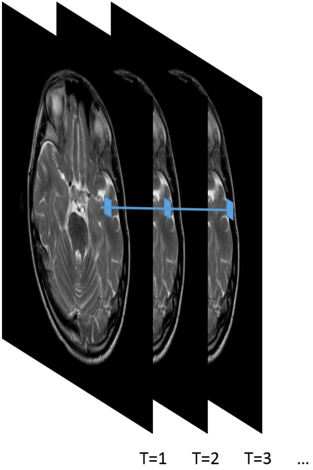
Table of Contents

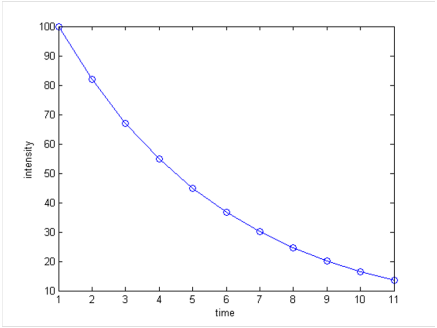
MRI (Very) Basics

T2-weighted Magnetic Resonance Imaging (MRI). Acquire Fast Spin Echo (FSE) sequence: intensity values at each pixel decay over time, with decay rate depending on a spin-spin relaxation time T_2. Can model this as a simple exponential decay:

$$I(t, x) \approx I_0(x) e^{-t/T_2(x)}$$




T=1
T=2
T=3
...


intensity
time
100
90
80
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
11

T2 Estimation

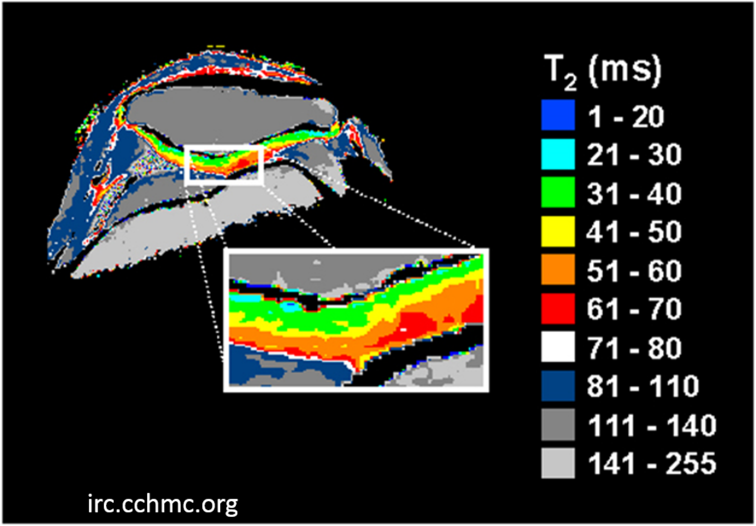
Images from T_2 weighted MRI useful by themselves, but would be great to have actual T_2 values!

The value of T_2 at each pixel is a useful diagnostic tool: differentiates tissue! Goal: produce a single image containing the T_2 values at each pixel.

Estimating T_2 straightforward in principle: acquire a sequence, fit an exponential curve to the sequence at each pixel to estimate T_2.

T2 maps

"Increased T2 relaxation time is an early marker of cartilage injury."



The Need For Undersampling

MRI is costly and time consuming. By speeding up acquisitions we can reduce cost, improve patient comfort, and limit motion artifacts.

To reduce acquisition times, we can take fewer measurements. In MRI, measurements are taken in Fourier space (”*k-space*”)

Radial Undersampling

In this talk just consider radial sampling: undersampling means measuring fewer radial lines (in Fourier domain)

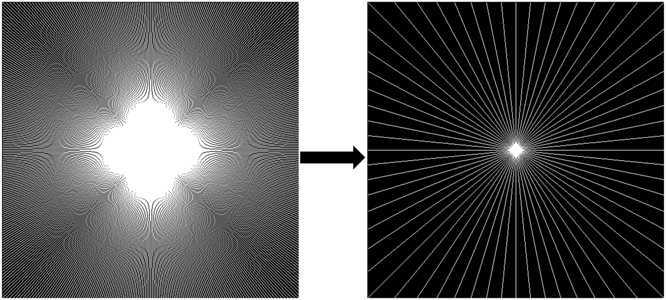
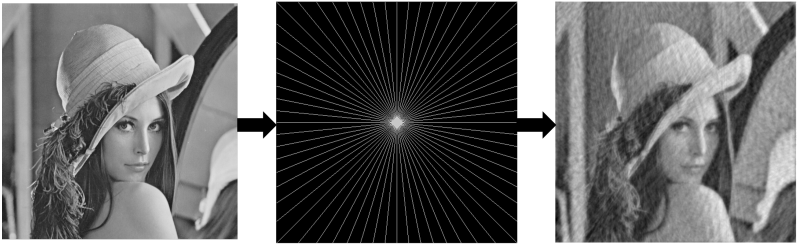


Figure : Undersampled DFT

Radial Undersampling

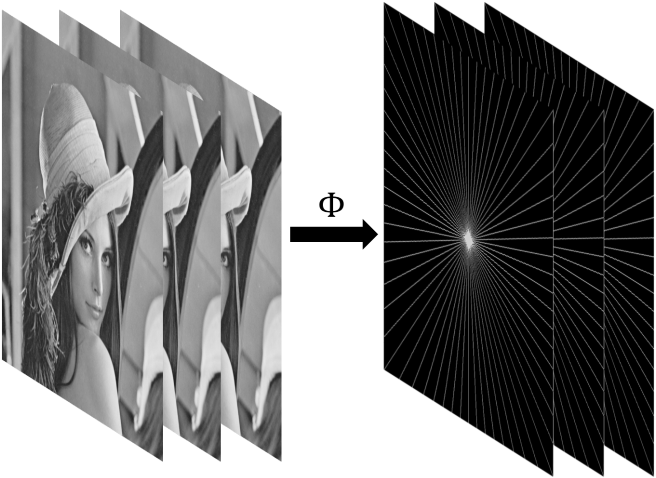
Information is lost in the undersampling process:



"Naive" solution minimizes $||y - \Phi x||_2$, found by inverting DFT

Measuring Image Sequence

For T_2 estimation we have a sequence of images, all containing "similar" information, so measure different radial lines on each:



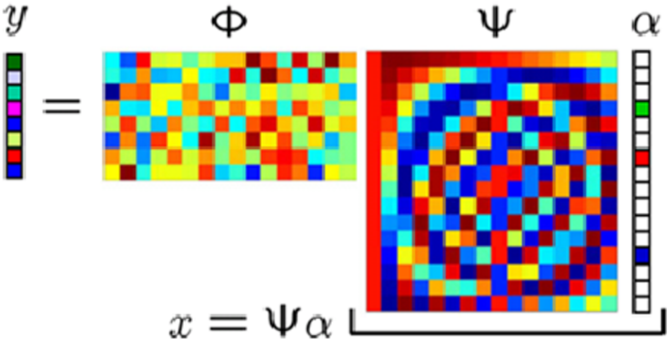
We need to reconstruct image x from this undersampled DFT:

$$y = \Phi x$$

$\Phi \in \mathbb{R}^{m \times n}$, $m < n$: *underdetermined* system, infinitely many solutions! This is where *compressed sensing* comes in...

Compressed Sensing

Assume image x has a sparse representation α in a "dictionary" Ψ (often a basis, but not necessarily)



(image from Sourabh Bhattacharya's page at Iowa State)

Compressed Sensing

Use *compressed sensing* methodology. Assume x is *sparse* in a basis Ψ. Seek sparsest solution consistent with the measurements:

$$\underset{x}{\operatorname{argmin}}\ \lambda_{l^0}||\Psi x||_0 + ||y - \Phi x||_2$$

This is NP-hard... $O(2^n)$ computations. Convex relaxation:

$$\underset{x}{\operatorname{argmin}}\ \lambda_{l^1}||\Psi x||_1 + ||y - \Phi x||_2$$

The Restricted Isometry Property

Why should we expect convex relaxation to work? Introduce Restriced Isometry Property (RIP) for a matrix A. A satisfies RIP-(s,δ_S) if for any s-sparse vector x,

$$(1-\delta_S)||x||_2^2 \leq ||Ax||_2^2 \leq (1+\delta_S)||x||_2^2$$

Says that any s-column submatrix of A is "nearly" orthonormal. Alternative formulation: measure *coherence*

$$\mu(A) = \max_{i \neq j} \frac{\langle a_i, a_j \rangle}{||a_i|| \cdot ||a_j||}$$

Smaller coherence $\implies$ smaller RIP constant

The Restricted Isometry Property

So what? Well... look at l^1 and l^0 problems in noiseless case.

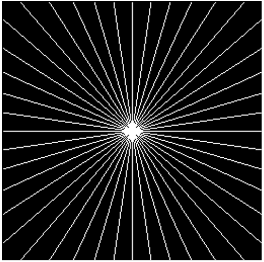
$$x_{l^0} = \underset{x}{\text{argmin}} \, ||\Psi x||_0 \text{ s.t. } y = \Phi x$$

$$x_{l^1} = \underset{x}{\text{argmin}} \, ||\Psi x||_1 \text{ s.t. } y = \Phi x$$

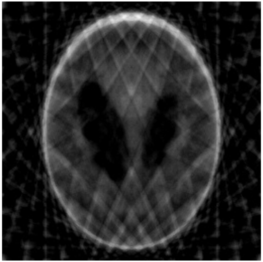
Theorem: if $\delta_{2S} < 1$, solution x_{l^0} is unique. If $\delta_{2S} < \sqrt{2} - 1$, $x_{l^0} = x_{l^1}$.
Proof of first part: $\delta_{2S} < 1 \implies$ no 2s-sparse vectors in $null(\Phi)$.
Thus $\Phi(x_1 - x_2) = 0$ iff $x_1 = x_2$ for s-sparse x_1, x_2.

Can solve combinatorially hard l^0 problem using convex optimization if we have RIP! Also have robustness to error when $= \Phi x + e$.

Some early CS results show perfect recovery of simple images from highly undersampled measurements.



Sample 22 radial lines



Naïve reconstruction

CS by l1 minimization

Summary so far...

Our goal:

- Measure a sequence of images with undersampled FT
- Model sequence (in image domain) as

$$I(x,t) = I_0 e^{-t/T_2(x)}$$

- Reconstruct sequence of images using CS techniques
- Estimate T_2 map from image sequence

Nonlinear problem

Straightforward approach:

$$\underset{T_2, I_0}{\text{argmin}} \, \|y - \Phi(I_0 e^{-t/T_2})\|_2^2 + P(I_0, T_2)$$

P is penalty function, e.g. l^1 norms of ΨI_0, ΨT_2

Suffers from problems of *scale mismatch* between I_0, T_2. To get around this, let's *linearize* by representing exponential decay using PCA.

Modeling T2 Decay with PCA

Model each pixel as a simple exponential decay:

$$I(x,t) = I_0(x)e^{-t/T_2(x)}$$

For a given range of expected T_2 values, use training data to learn PC basis:

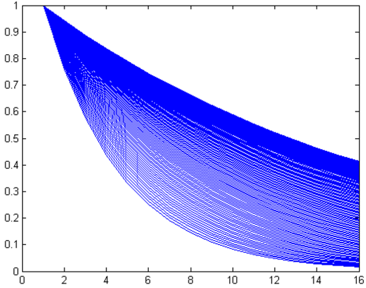
$$D = U\Sigma V^*$$

U is PC basis. Compute PC coefficients as

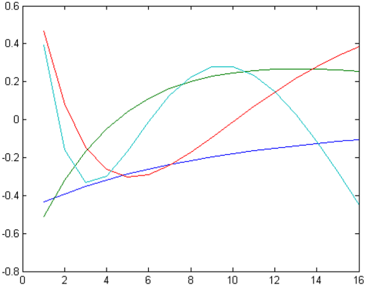
$$M(x) = U^* I(x,t)$$

Reconstruct using $I(x,t) = UM(x)$. In practice we truncate U to just a few columns; temporal sparsity assumption

Can model T_2 decays in a given range as linear combination of just a few basis functions:



Training Set



Truncated PCA basis

4 of 16 coefficients reconstruct training set with < 0.04% error

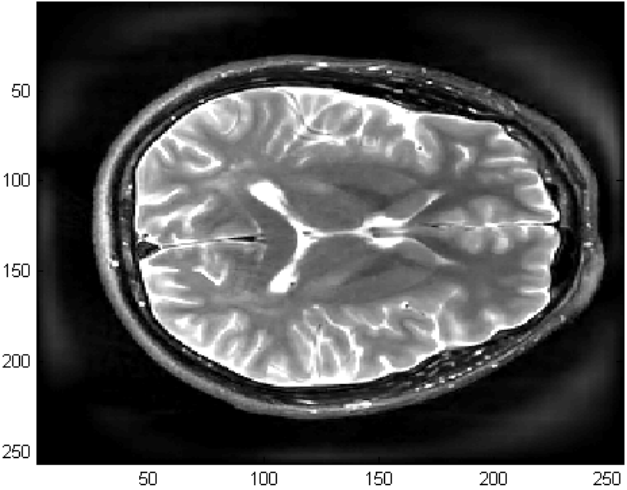
REPCOM Implementation

Reconstruction of Pricincipal Component Coefficient Maps (REPCOM, Huang et al 2012): solves for principal component representation M by minimizing

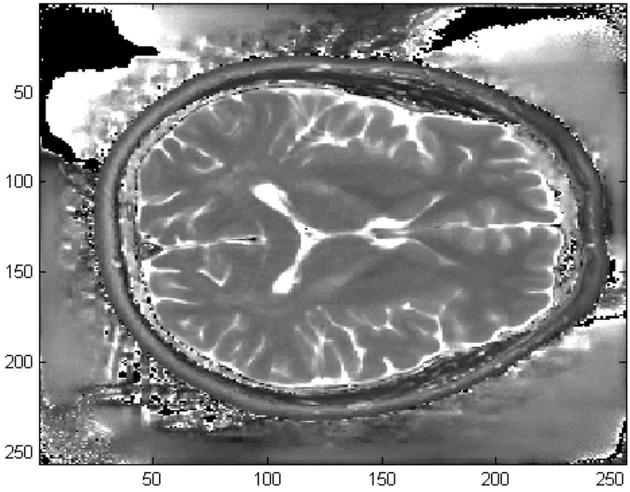
$$C(M) = ||y - \Phi(UM)||_2 + \lambda_{l1}||\Psi M||_1 + \lambda_{TV} TV(M)$$

using conjugate gradient optimization. Ψ is an orthonormal wavelet basis promoting sparsity. Estimate T_2 maps using least-squares exponential fit at each pixel of $I = UM$.

Image Reconstruction with REPCOM



T2 Map with REPCOM



L1 to L0

In solving l^1 minimization we implicitly assume RIP. But Fourier and Wavelet bases are not perfectly incoherent... not ideal in terms of RIP! l^1 and l^0 solutions are different. Recent work suggests possible improvements by using l^0 minimization.

- l^0 potentially better for problems with poor RIP
- NP-hard $\implies$ can only hope to find *locally* optimal solutions
- Harder to give error guarantees, and often hard to show convergence of algorithms

IHT with Orthonormal Wavelet

Example of l^0 approach: let Ψ be orthonormal wavelet. To find local minimizer of

$$||y - \Phi(UM)||_2^2 + \lambda||\Psi M||_0$$

or equivalently

$$||y - \Phi U \Psi^* \alpha||_2^2 + \lambda||\alpha||_0 \text{ s.t. } \alpha = \Psi M$$

use gradient descent:

$$\nabla(\alpha) = -2\Psi U^* \Phi^* (y - \Phi U \Psi^* \alpha)$$

use Iterative Hard Thresholding (IHT):

$$\alpha^{n+1} = H(\alpha^n + \mu \nabla(\alpha^n))$$

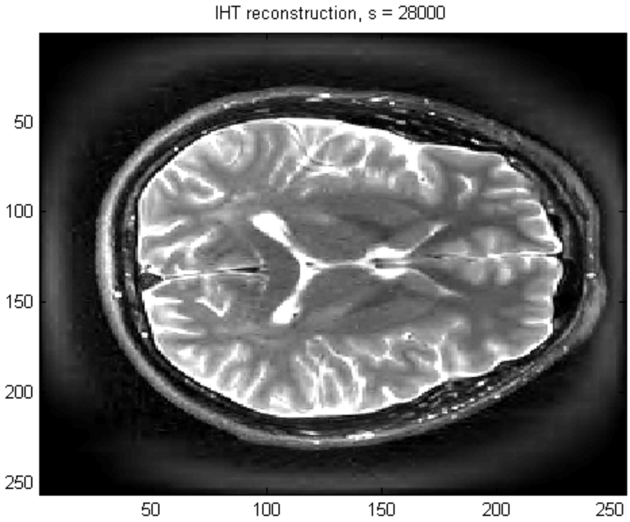
$$M^{n+1} = \Psi^* \alpha^{n+1}$$

IHT with Orthonormal Wavelet

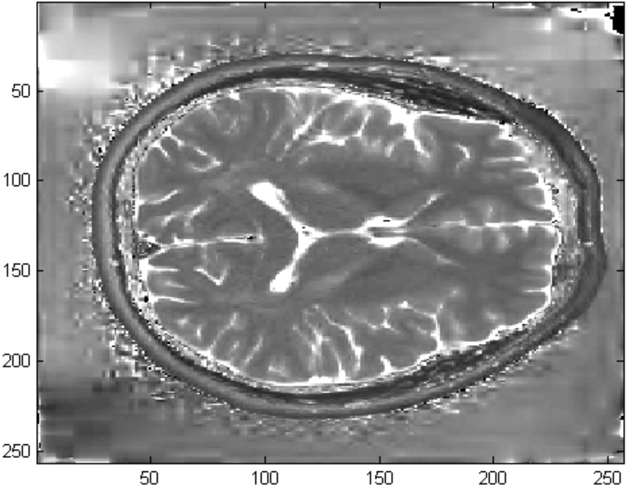
IHT...

- Easy to implement
- Converges to local optimum of l^0 problem
- Provably "good" when we have RIP
- Can use fast transforms (FFT, DWT, etc.)

Image Reconstruction with IHT (orthonormal wavelet)

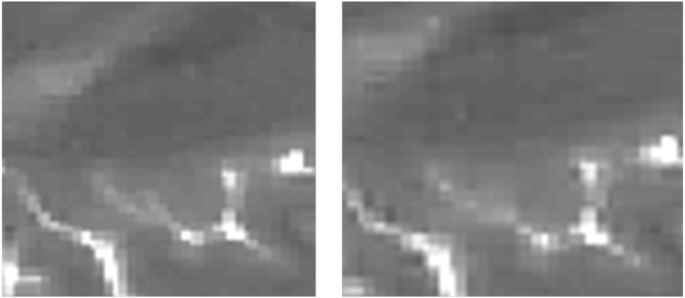


T2 Map with IHT (orthonormal wavelet)



T2 Map Comparisons

IHT approach not quite as good as REPCOM...



REPCOM   IHT

Beyond Global Sparsity...

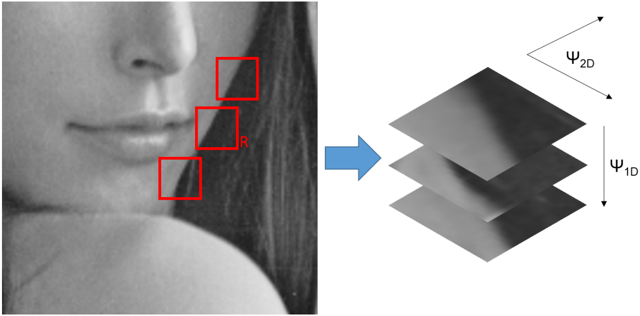
To do better... look at *adaptive* transforms, i.e. sparsity transforms that depend on the image being recovered. Makes analysis more difficult, but potential for better reconstructions.

BM3D

State-of-the-art in image denoising: Block-Matching 3D. Assumes *local* sparsity, and *collaborative* sparsity between similar blocks. Gives highly redundant frame representation.

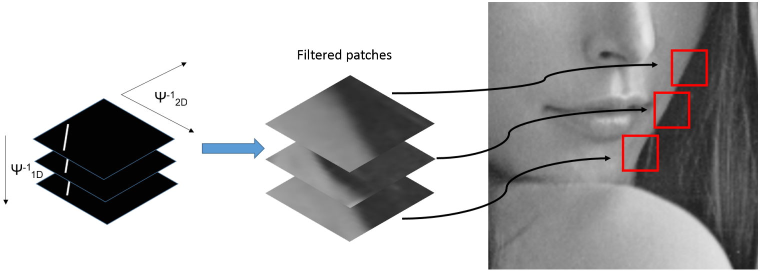
BM3D

For a given *reference block*, find similar blocks (e.g. similar by l^2 distance). Stack these into a 3D *group*.



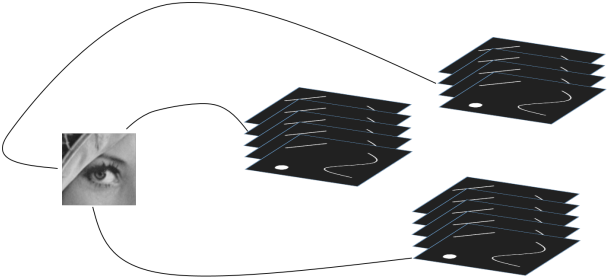
Use 3D sparsifying transform on groups: take advantage of local image sparsity AND correlation between patches

Inverting block transform: return filtered blocks to original locations. Use weighted average where blocks overlap (e.g. weights proportional to inverse of variance in a patch). This is a *left inverse* of Ψ as long as weights are normalized.



Frame Interpretation of BM3D

BM3D transformation is a *frame*: highly redundant representation that can be stably inverted. Regrouping of blocks is a left inverse: if we skip the thresholding step we get back what we started with



Analysis vs. Synthesis Approach

How do we formulate optimization problem in redundant case?

$$\alpha_{\text{synthesis}} = \operatorname*{argmin}_{\alpha} \lambda ||\alpha||_0 + ||y - \Phi \Psi^* \alpha||_2^2$$

$$x_{\text{analysis}} = \operatorname*{argmin}_{x} \lambda ||\Psi x||_0 + ||y - \Phi x||_2^2$$

Distinct problems! There are many α satisfying $x = \Psi^* \alpha$ for any x. Which is better? Note that we can't solve the analysis approach directly with IHT-type algorithms.

Iterating BM3D

First cut at using BM3D: initialize solution $M = 0$, residual $r = 0$

- Gradient step: $M = M + U^*\Phi^* r$
- Filtering: $M = BM3D(M, \sigma)$
- Update residual: $r = y - \Phi U M$

Filtering enforces sparsity of the *adaptive* transform coefficients $\Psi_i M$, but we observe convergence in practice.

Not exactly gradient descent since we apply inverse of redundant transform, not adjoint...

Iterating BM3D

256 radial lines

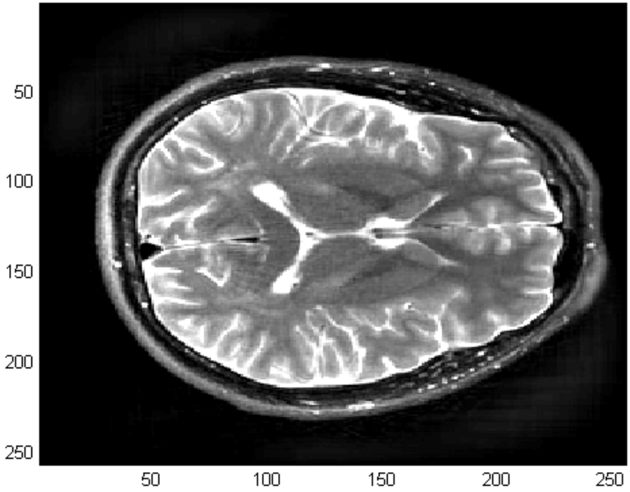
Iterating BM3D

128 radial lines

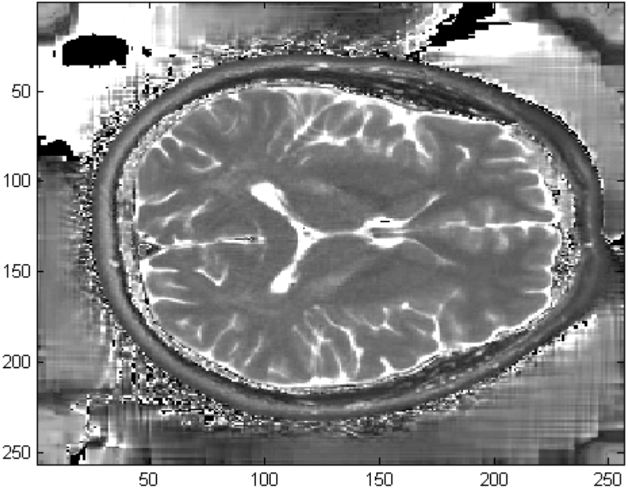
Iterating BM3D

64 radial lines

Image Reconstruction with BM3D

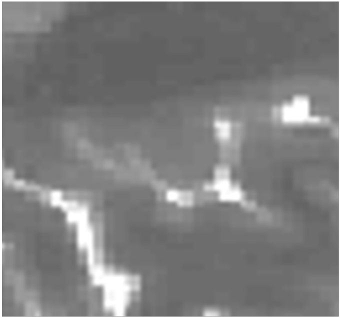


T2 Map with BM3D

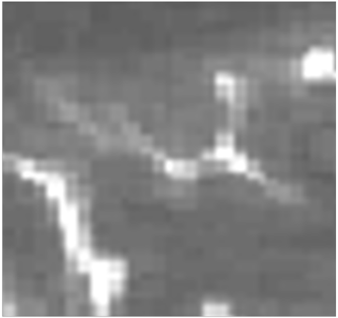


T2 Map with BM3D

...



REPCOM



BM3D

What now?

- Pick "optimal" σ for this applicaiton (BM3D hard thresholding level)
- Formulate optimization problem... analysis, synthesis or balanced approach?
- Proof of convergence for chosen problem?

Thanks!