Step-by-Step Reconstruction Using Learned Dictionaries

Jon Tamir, PhD

www.jtsense.com

Electrical and Computer Engineering
The University of Texas at Austin

ISMRM 2020 Virtual Conference

Declaration of Financial Interests or Relationships

Speaker Name: Jonathan I Tamir

I have no financial interests or relationships to disclose with regard to the subject matter of this presentation.

Links

- Compressed sensing MRI overview:

  https://www.ismrm.org/19/program_files/WE22.htm

- Hands-on examples:

  https://github.com/utcsilab/dictionary_learning_ismrm_2020

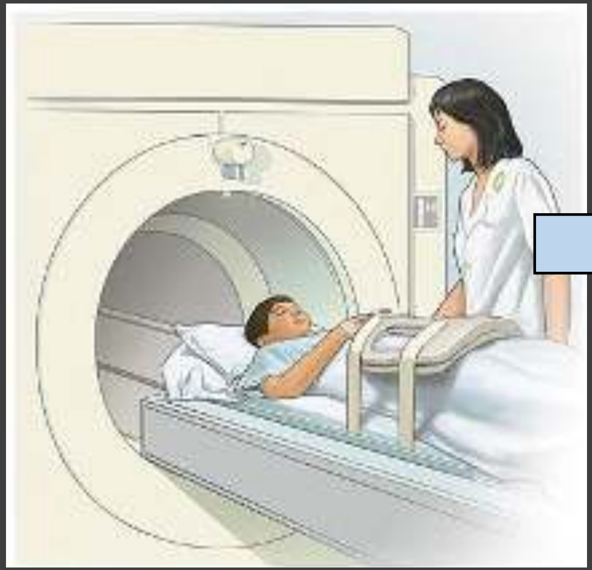
CS-MRI Talk

Hands-on code

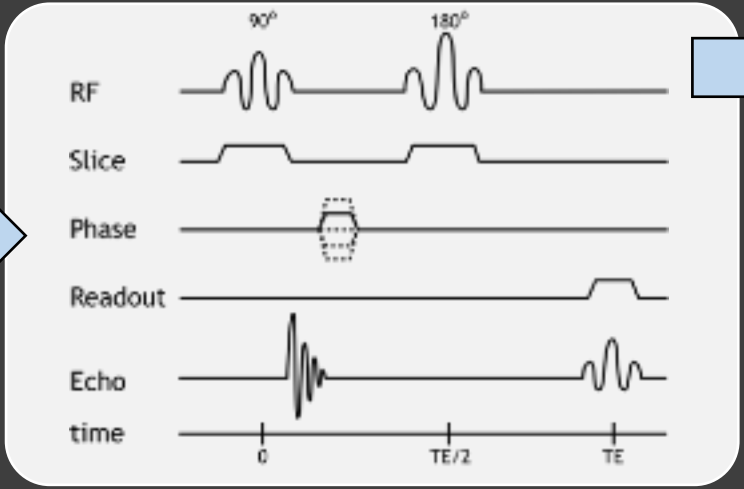
F Ong et al., SigPy: https://sigpy.readthedocs.io/en/latest/

MRI Background

Patient in MRI scanner



Pulse sequence controls MRI signal



Measurements are collected

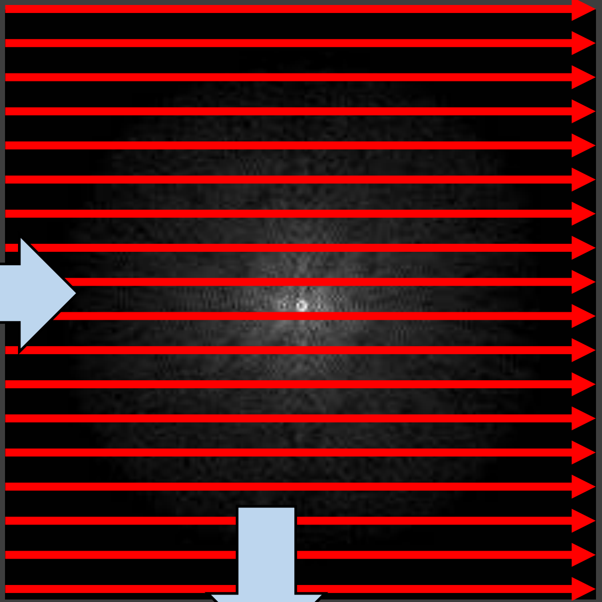
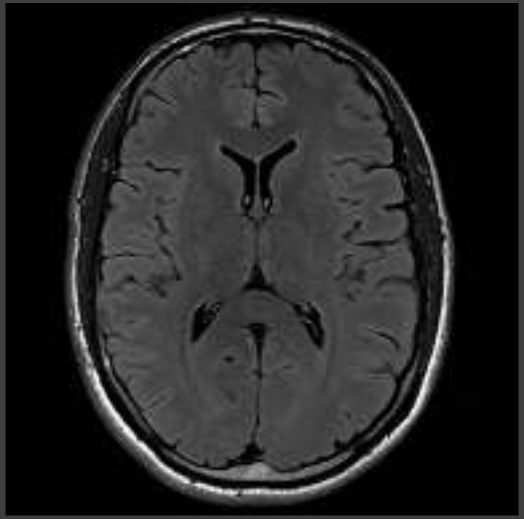


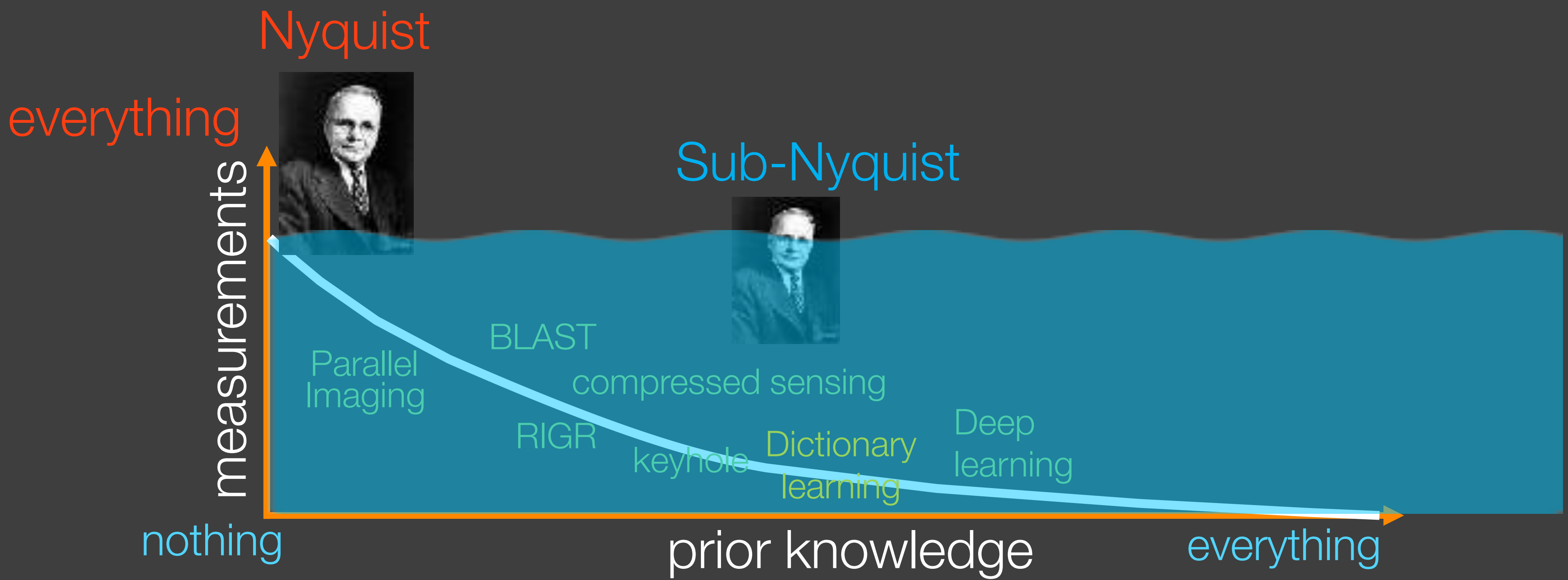
Image is reconstructed



https://www.mountnittany.org/assets/images/krames/Image_201608102119_2786.jpg
http://xrayphysics.com/se_diag_wgrad.png

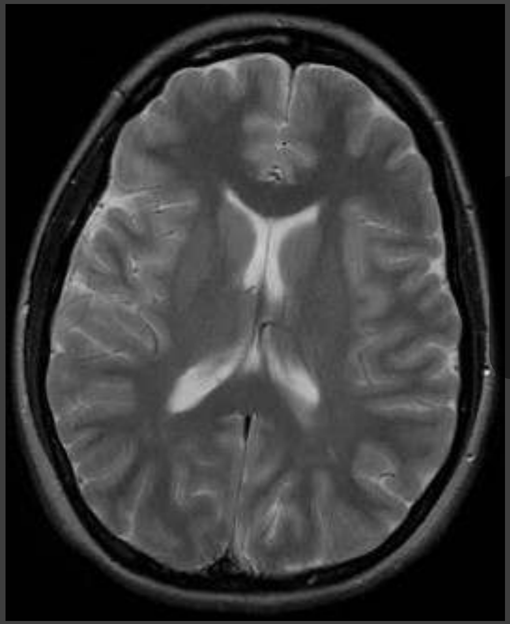
Data redundancy

Redundancy reduces sampling requirements
(The more you know, the less you need)

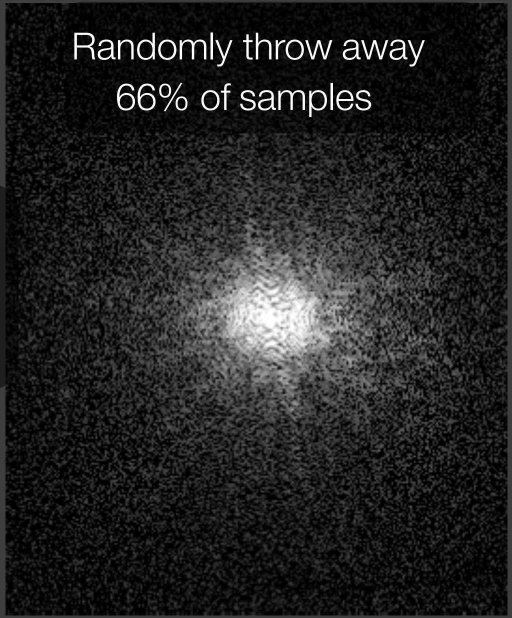


From M. Lustig

Compressed sensing MRI

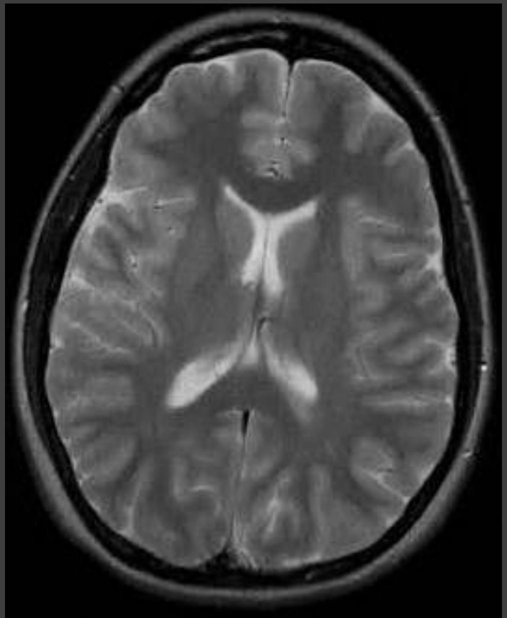


Fourier transform



standard recon

compressed sensing



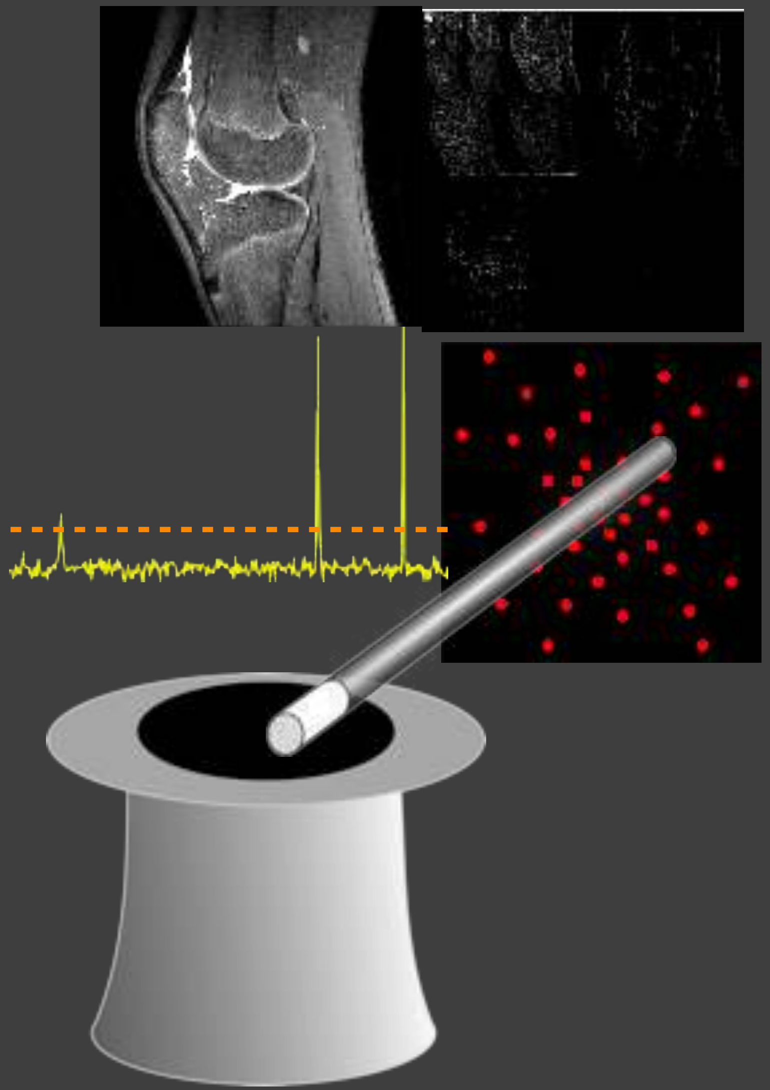
From Miki Lustig

Compressed sensing recipe

1. Sparse signal model

2. Incoherent sensing operator

3. Non-linear reconstruction algorithm



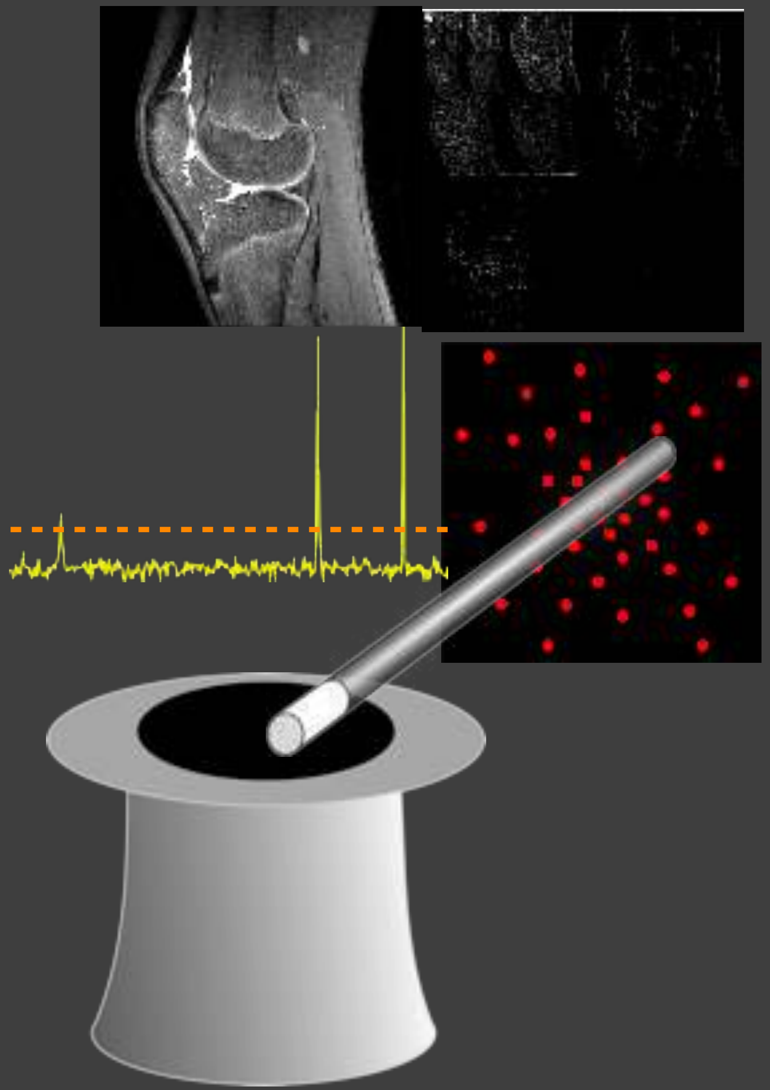
EJ Candes et al., IEEE SPM, 2008
M Lustig et al., IEEE SPM, 2008

Compressed sensing recipe

1. Sparse signal model

2. Incoherent sensing operator

3. Non-linear reconstruction algorithm



EJ Candes et al., IEEE SPM, 2008
M Lustig et al., IEEE SPM, 2008

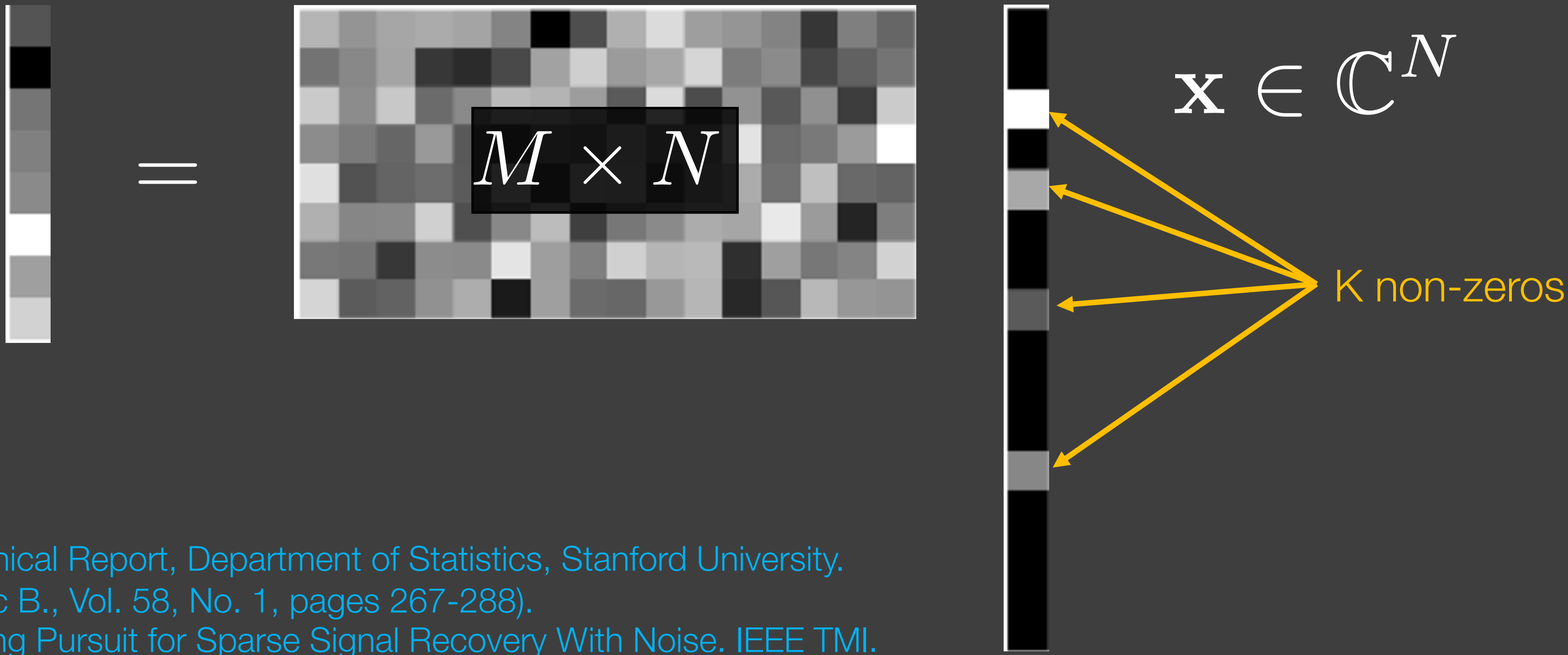
Sparse signal modeling

- Assumption: x is a K-sparse signal (K << N)
  - Make M (K < M < N) incoherent linear measurements

$\mathbf{E} \in \mathbb{C}^{M \times N}$

$\mathbf{y} \in \mathbb{C}^{M}$

$\mathbf{x} \in \mathbb{C}^{N}$



S Chen, D Donoho. "Basis Pursuit," Technical Report, Department of Statistics, Stanford University.
Tibshirani, R. (1996). J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288).
Cai and Wang (2011). Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. IEEE TMI.

Sparse signal modeling

- Assumption: x is a K-sparse signal (K << N)
  - Make M (K < M < N) incoherent linear measurements
  - Enforce sparsity during reconstruction

Greedy methods

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{E}\mathbf{x}||_2^2$$
$$\text{subject to } ||\mathbf{x}||_0 \leq K$$

Relaxation methods

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{E}\mathbf{x}||_2 + \lambda||\mathbf{x}||_1$$

S Chen, D Donoho. "Basis Pursuit," Technical Report, Department of Statistics, Stanford University.
Tibshirani, R. (1996). J. Royal. Statist. Soc B., Vol. 58, No. 1, pages 267-288).
Cai and Wang (2011). Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. IEEE TMI.

Sparse signal modeling

- What if the signal is not sparse?

not sparse

not sparse

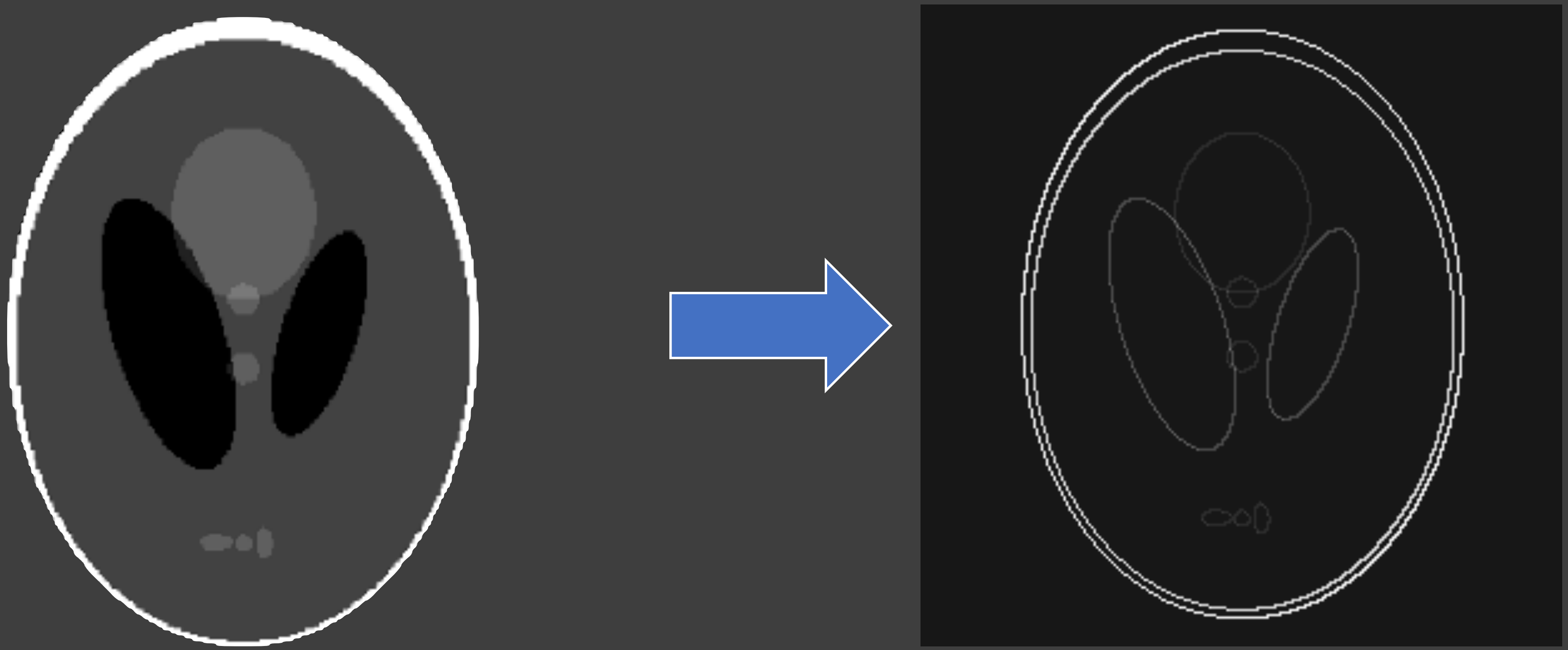
not sparse

not sparse

Transform sparsity

- Most medical images are sparse in an alternative representation

not sparse sparse edges

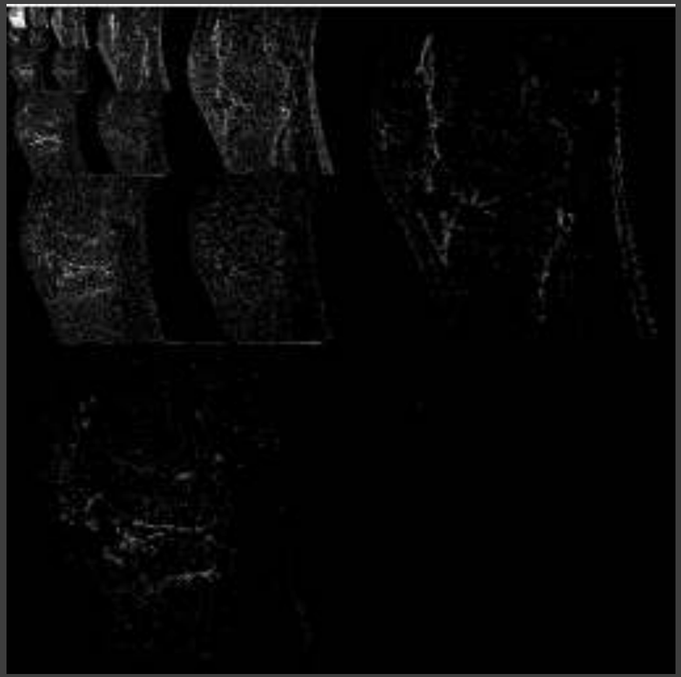


Transform sparsity

- Most medical images are sparse in an alternative representation

not sparse

sparse wavelet



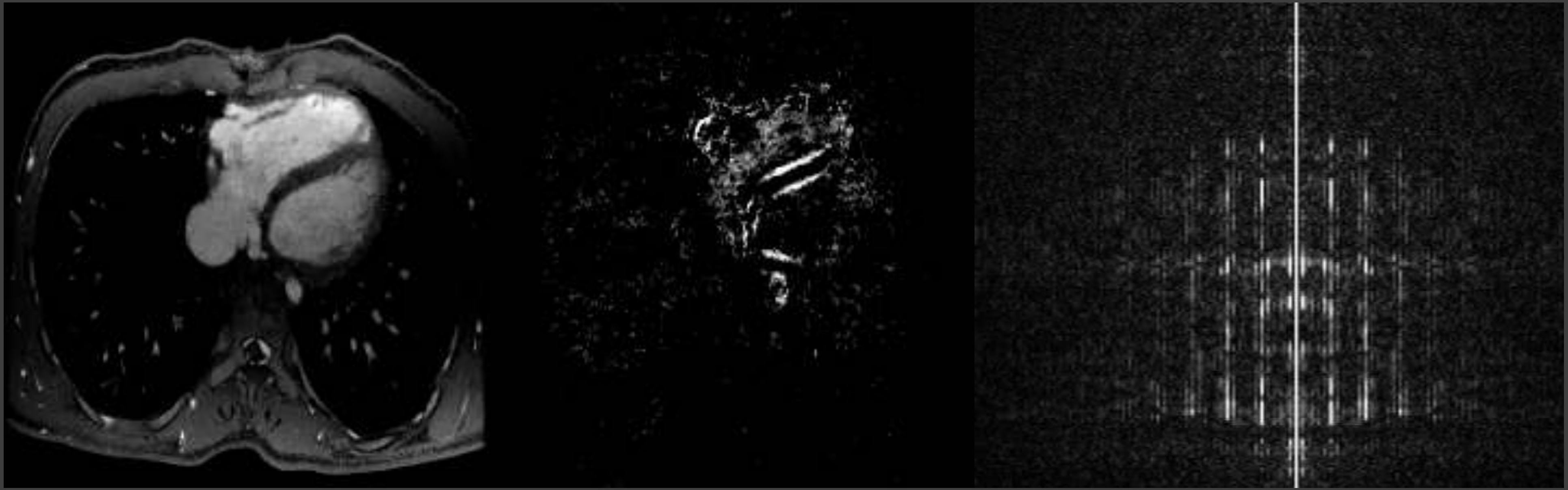
Transform sparsity

- Most medical images are sparse in an alternative representation

not sparse

sparse temporal finite differences

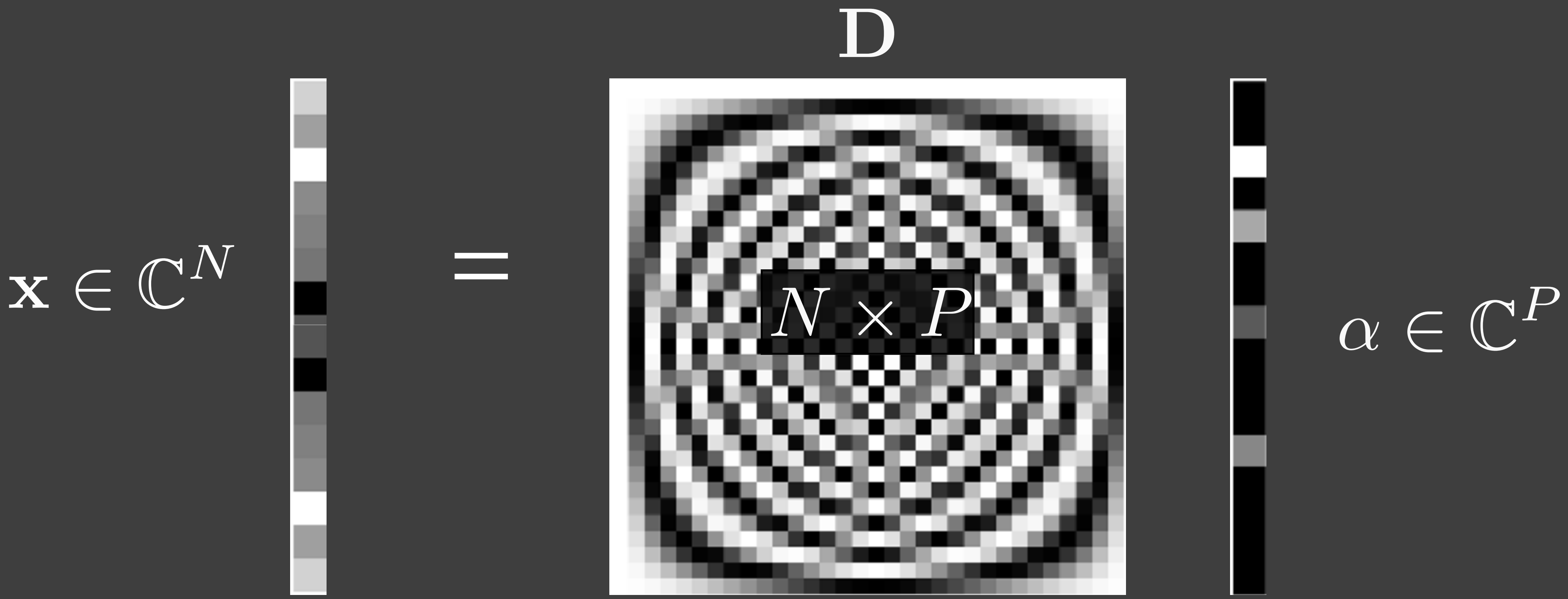
sparse temporal frequency



From Joseph Cheng

Transform sparsity

- Represent the original signal as sparse in a transform domain



Transform sparsity

- Represent the original signal as sparse in a transform domain
- Enforce sparsity on transformed coefficients

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{ED}\alpha||_2^2 + \lambda||\alpha||_1$$

$$\min_{\alpha} \frac{1}{2}||\mathbf{y} - \mathbf{ED}\alpha||_2^2 \quad \text{subject to } ||\alpha||_0 \leq K$$

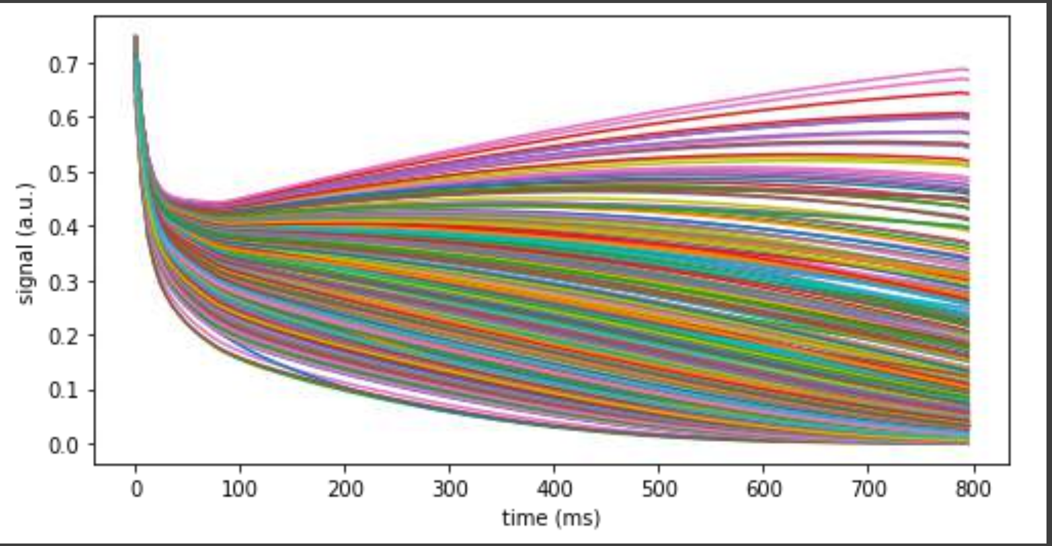
Transform sparsity

- Problem:
  - Sparsity is only as good as our transform
  - Need to choose the right transform for each signal
- Solution?
  - Learn the transform! → Dictionary Learning

Dictionary learning

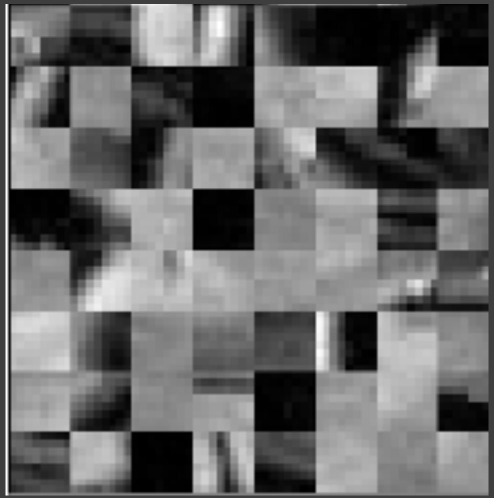
Given training data:

Temporal relaxation curves



Images

Patches

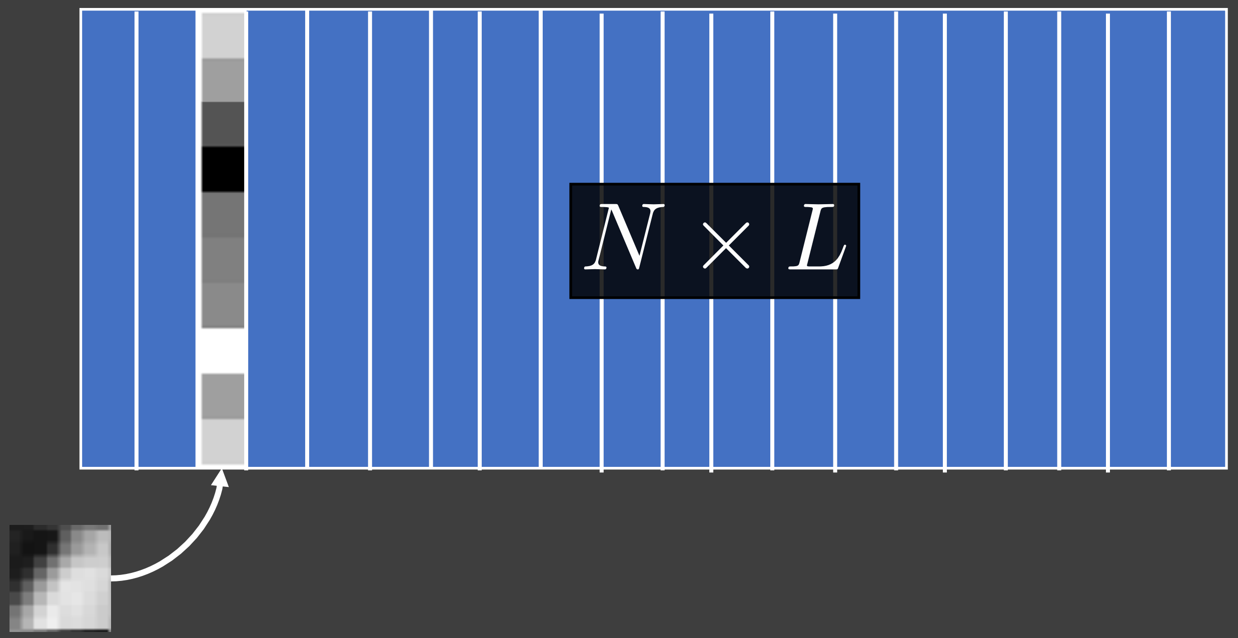


Dictionary learning

Given training data: form a data matrix

Data matrix

$\mathbf{X}$



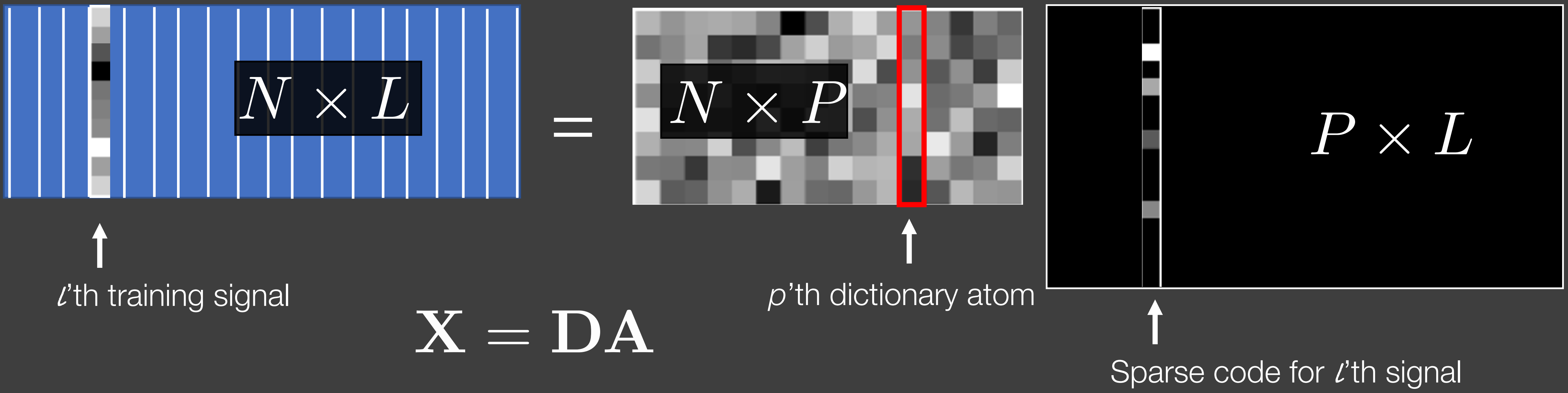
L training examples
Each length-N

ℓ'th training signal

Dictionary learning

Given training data: form a data matrix

- Jointly learn the dictionary and the sparse representation



$$\mathbf{X} = \mathbf{DA}$$

Dictionary learning

Given training data: form a data matrix

- Jointly learn the dictionary and the sparse representation

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2}||\mathbf{X} - \mathbf{DA}||_2^2 \quad \text{subject to } ||\alpha_l||_0 \leq K \quad l = 1, \ldots, L$$

$$||\mathbf{d}_p||_2 \leq 1 \quad p = 1, \ldots, P$$

Dictionary learning

Given training data: form a data matrix

- Jointly learn the dictionary and the sparse representation
- Approach: alternating minimization

Step 1: Update coefficients (sparse coding)

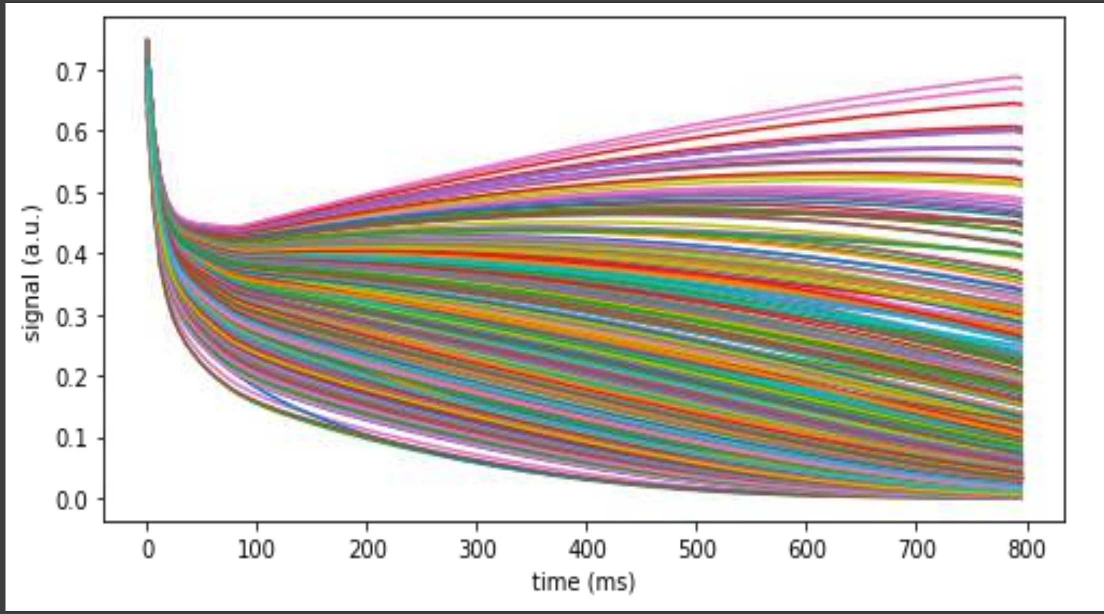
$$\min_{\mathbf{A}} \frac{1}{2}||\mathbf{X} - \mathbf{DA}||_2^2 \quad \text{subject to } ||\alpha_l||_0 \leq K \quad l = 1, \ldots, L$$

Step 2: Update dictionary

$$\min_{\mathbf{D}} \frac{1}{2}||\mathbf{X} - \mathbf{DA}||_2^2 \quad \text{subject to } ||\mathbf{d}_p||_2 \leq 1 \quad p = 1, \ldots, P$$

Dictionary learning: example

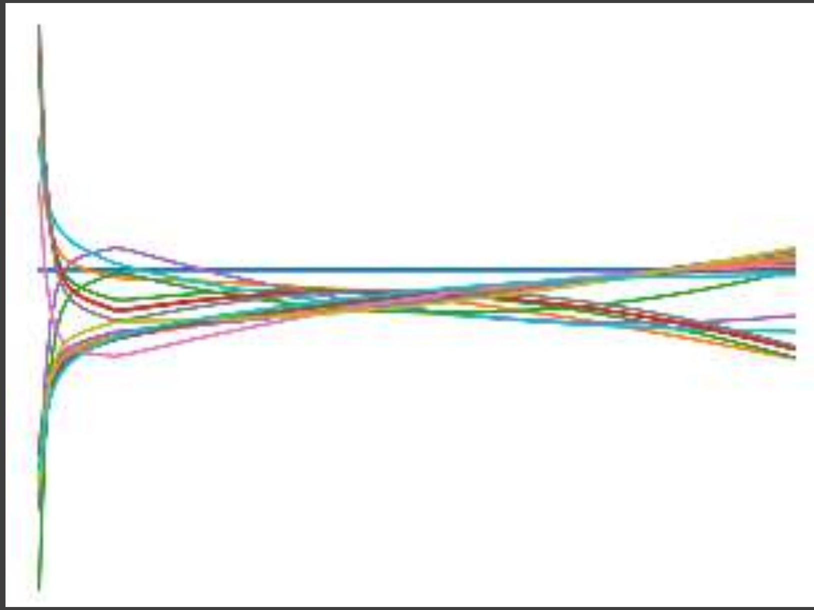
Temporal relaxation curves



N = 130

$\mathbf{X}$

=



P=20

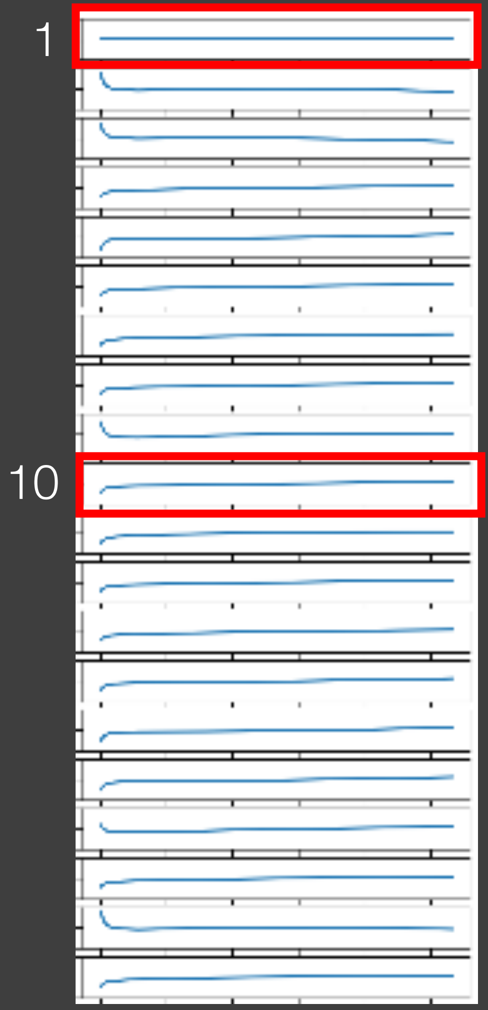
$\mathbf{D}$

K=2

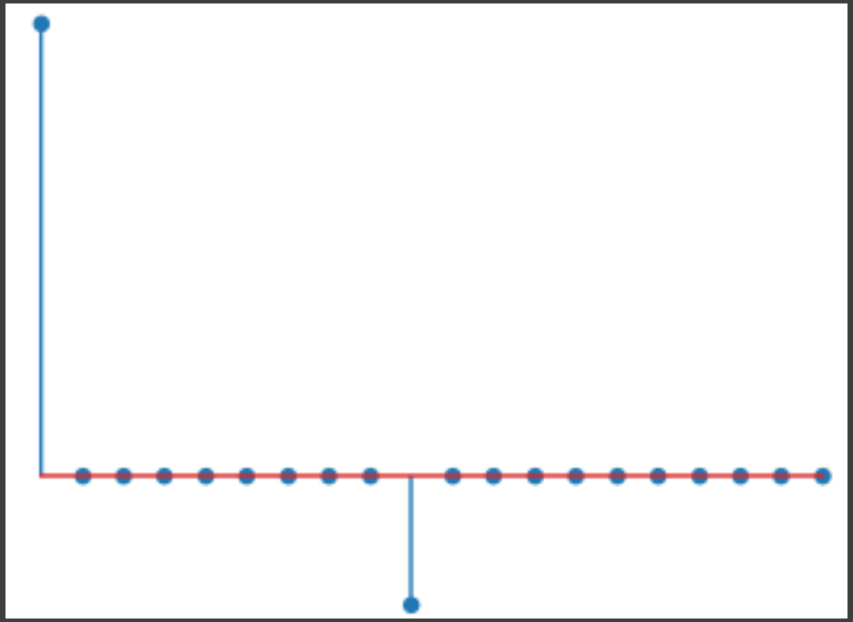
α

M. Doneva et al., MRM 2010. doi: 10.1002/mrm.22483
J.I. Tamir et al., MRM 2016. doi: 10.1002/mrm.26102

Dictionary learning: example

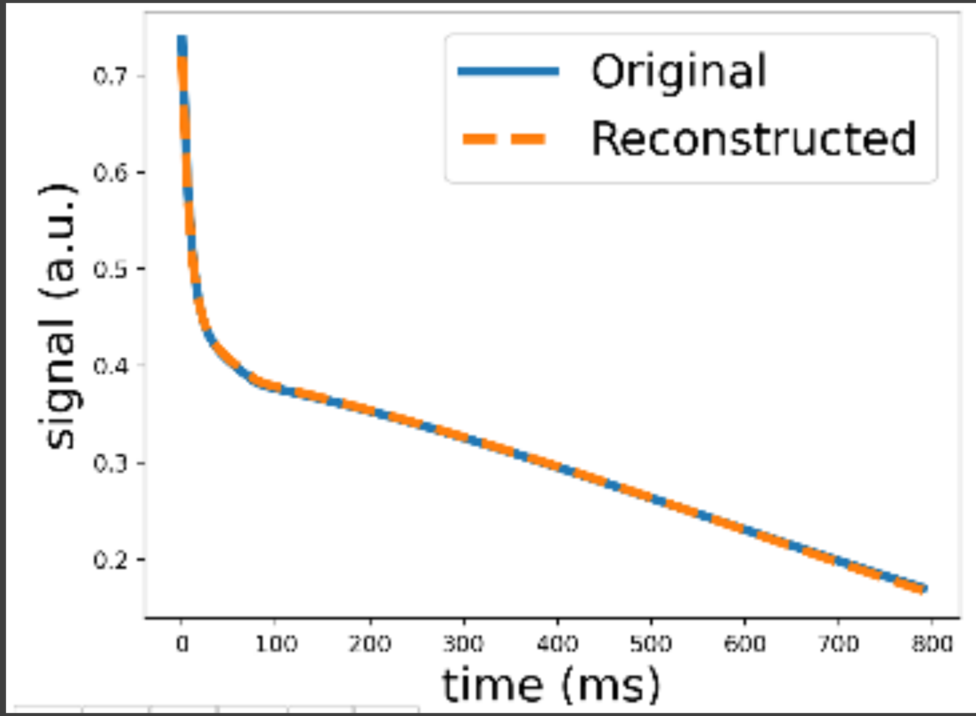


$\mathbf{D}$



α

$=$

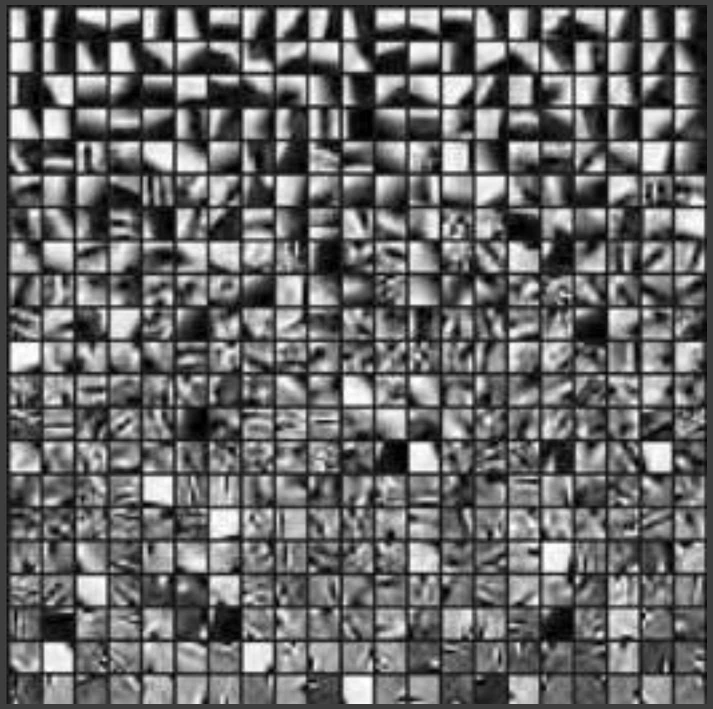


$$\mathbf{d}_1\alpha_1 + \mathbf{d}_{10}\alpha_{10} = \hat{\mathbf{x}}$$

Dictionary learning: example

Image patches

$=$



K=7

N = 8 x 8

$\mathbf{X}$

$\mathbf{D}\alpha$

P=441

Aharon et al., IEEE TSP, 2006, doi: 10.1109/TSP.2006.881199

From image space to k-space

- We can learn the dictionary directly from training examples…
  - At “inference time”, solve the sparse coding problem

- But we can also learn the dictionary directly from the MRI data!

M. Doneva et al., MRM 2010. doi: 10.1002/mrm.22483
Ravishankar et al., IEEE TMI, 2011. doi: 10.1109/TMI.2010.2090538

Dictionary learning MRI

$$\min_{\mathbf{x},\mathbf{D},\mathbf{A}} \frac{1}{2}||\mathbf{y}-\mathbf{E}\mathbf{x}||_2^2 + \frac{\lambda}{2}||\mathbf{x}-\mathbf{R}\left(\mathbf{D}\mathbf{A}\right)||_2^2 \quad \text{subject to}$$

$$||\alpha_l||_0 \leq K$$

$$||\mathbf{d}_p||_2 \leq 1$$

$$l = 1, ..., L$$
$$p = 1, ..., P$$

Data consistency

Dictionary fit

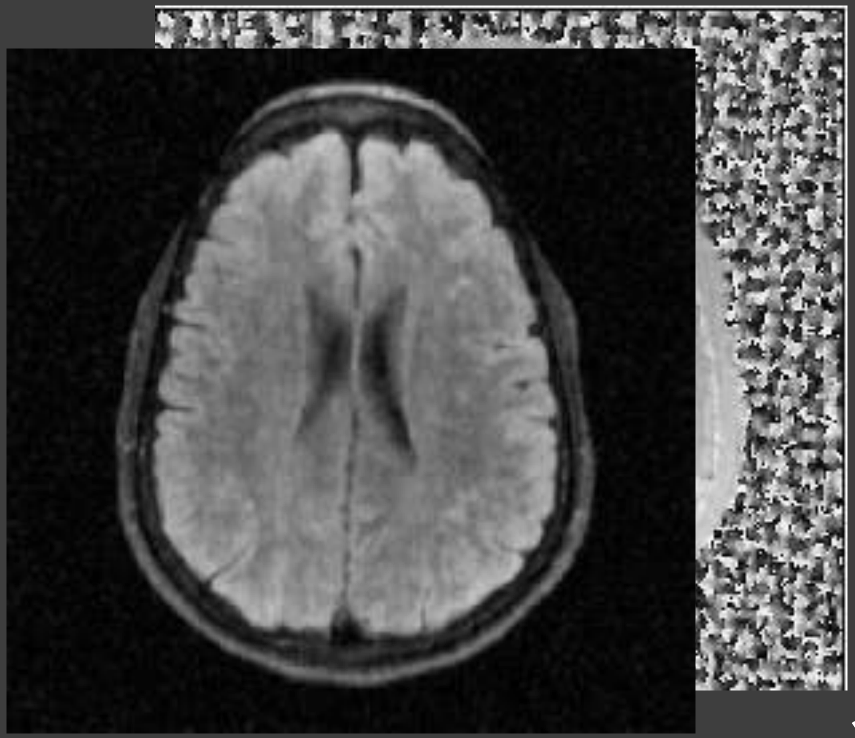
Converts from data matrix (patches) to image

Solution: → Alternating minimization (again!)

M. Doneva et al., MRM 2010. doi: 10.1002/mrm.22483
Ravishankar et al., IEEE TMI, 2011. doi: 10.1109/TMI.2010.2090538

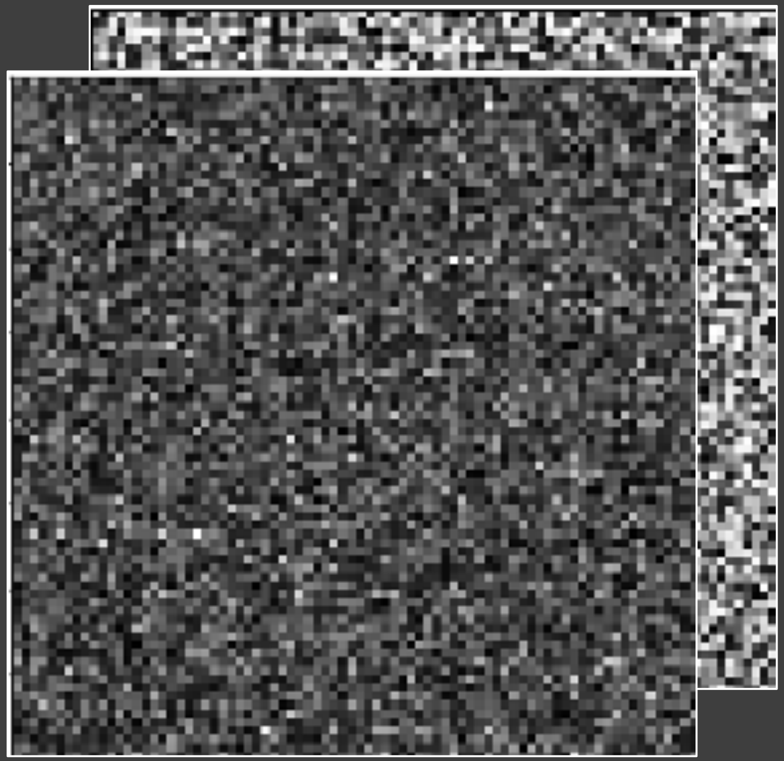
Dictionary learning MRI

Step 0. Initialize image and dictionary



$\mathbf{x}$

(mag + phase)



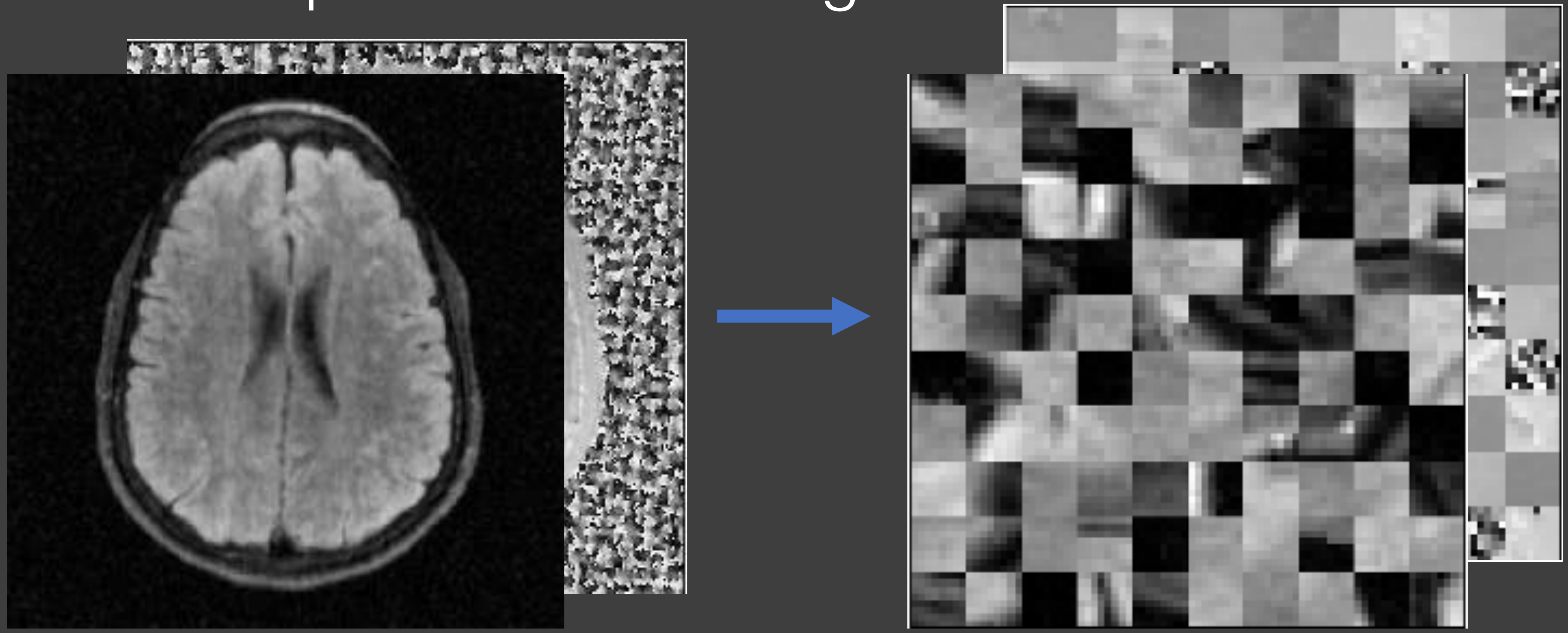
$\mathbf{D}$

(mag + phase)

Dictionary learning MRI

For iterations 1:T

Step 1. Extract patches from image



$\mathbf{x}$

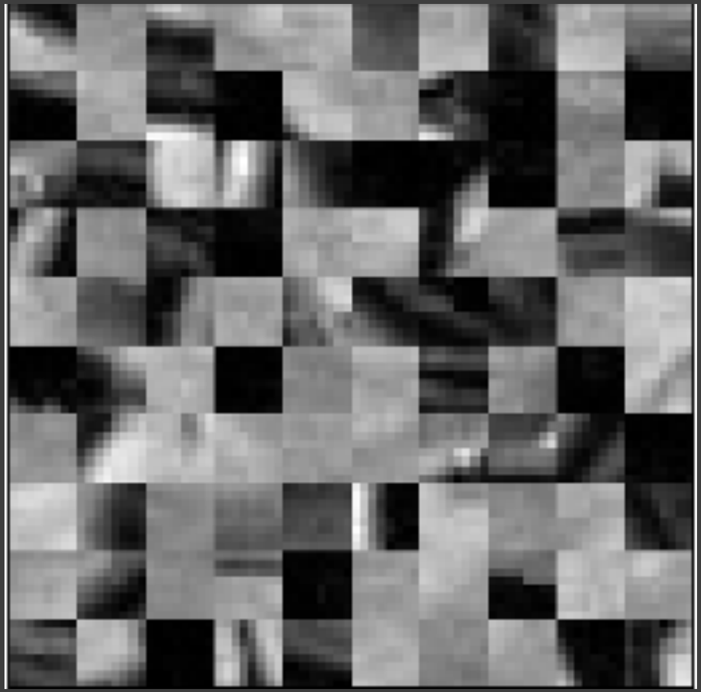
$\mathbf{X} = \mathbf{R}^T \mathbf{x}$

(mag + phase)

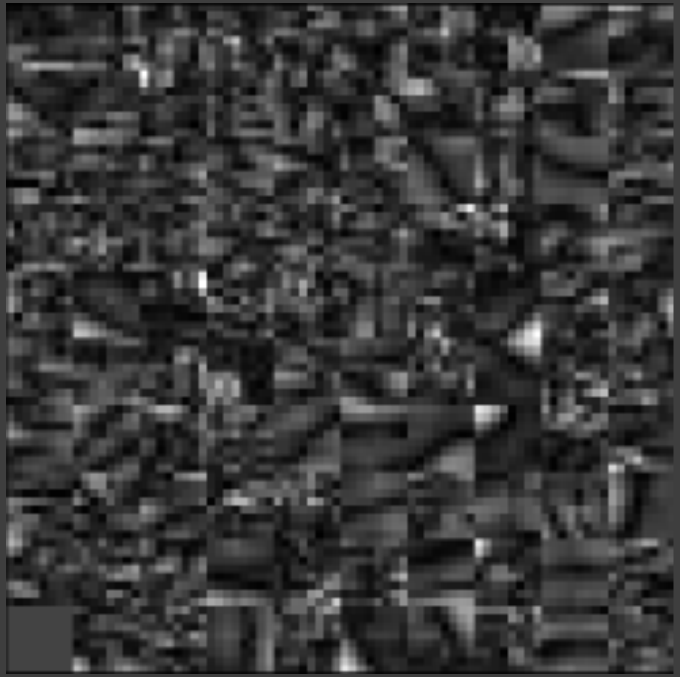
Dictionary learning MRI

For iterations 1:T

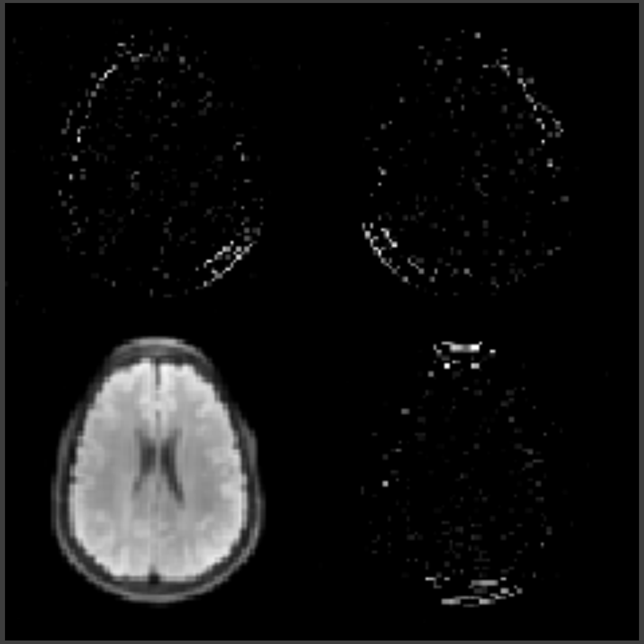
Step 2. Fit dictionary and sparse codes to patches



$\mathbf{X}$



$\mathbf{D}$

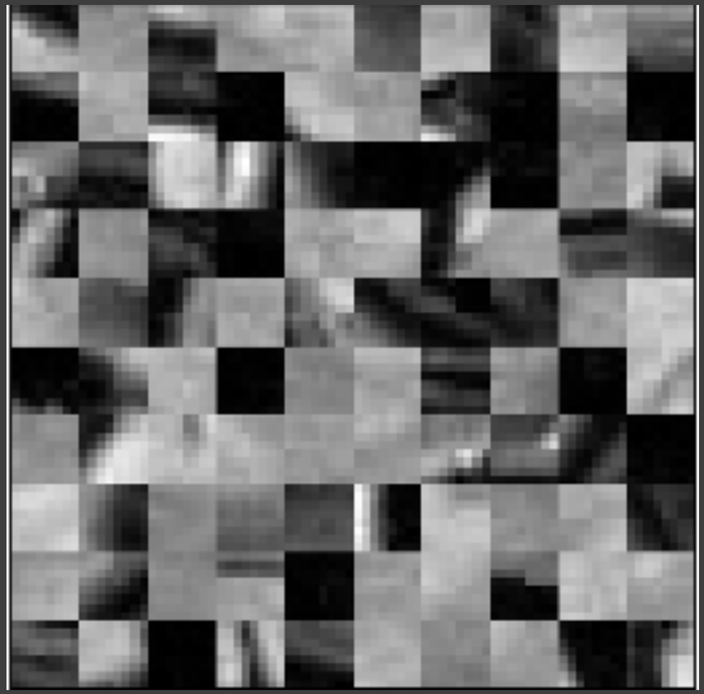


$\mathbf{A}$

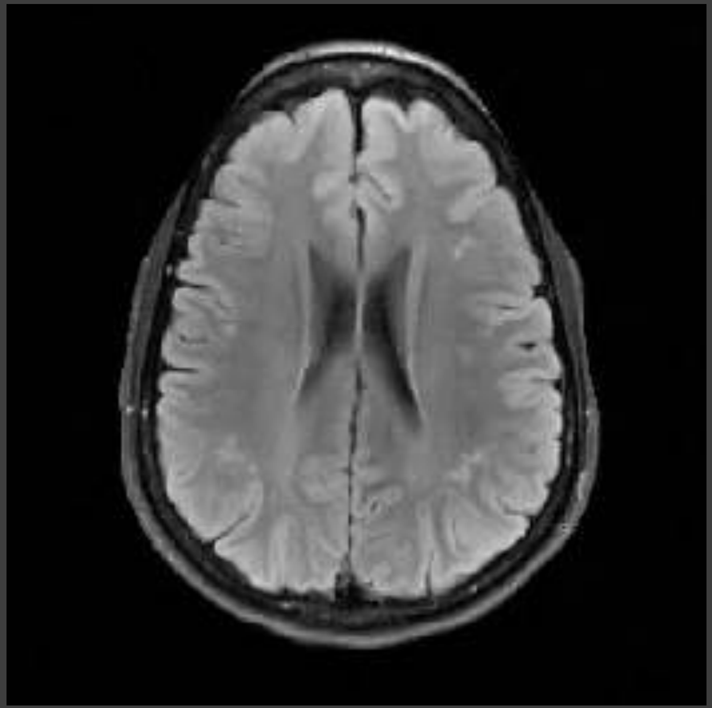
Dictionary learning MRI

For iterations 1:T

Step 3. Reshape and average patches into image



$$\mathbf{DA}$$

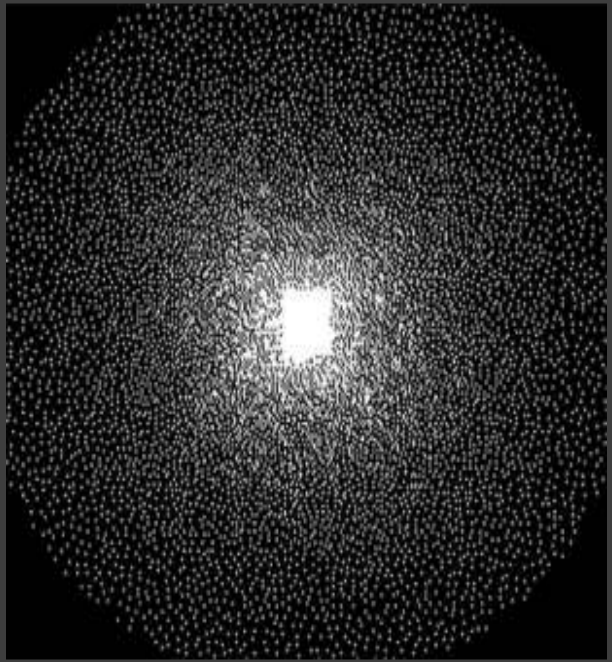


$$\mathbf{x}_D = \mathbf{RDA}$$

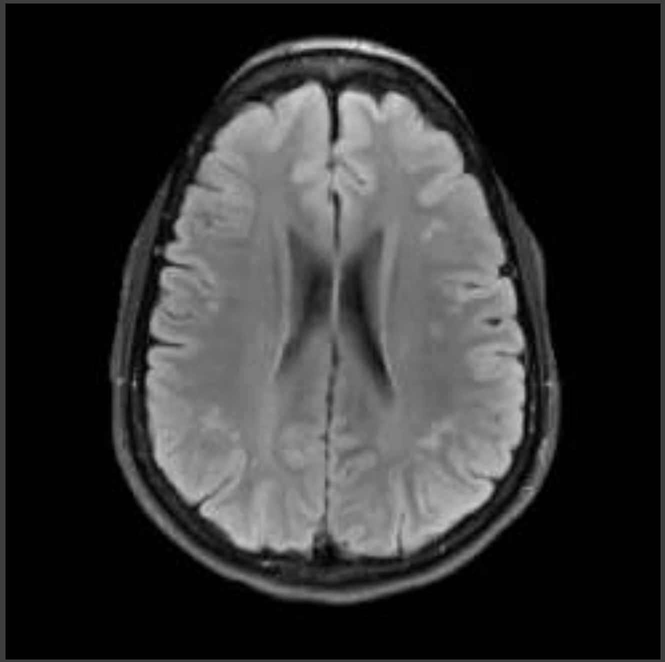
Dictionary learning MRI

For iterations 1:T

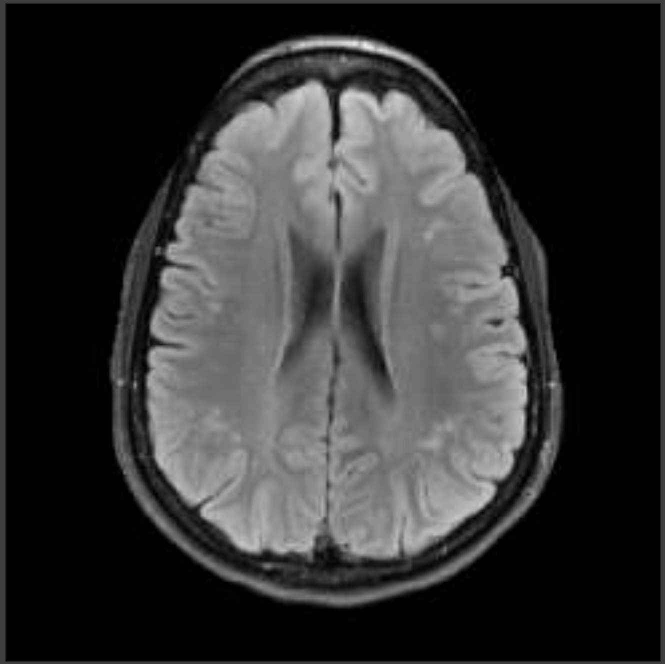
Step 4. Enforce data consistency with dictionary fit



$\mathbf{y}$

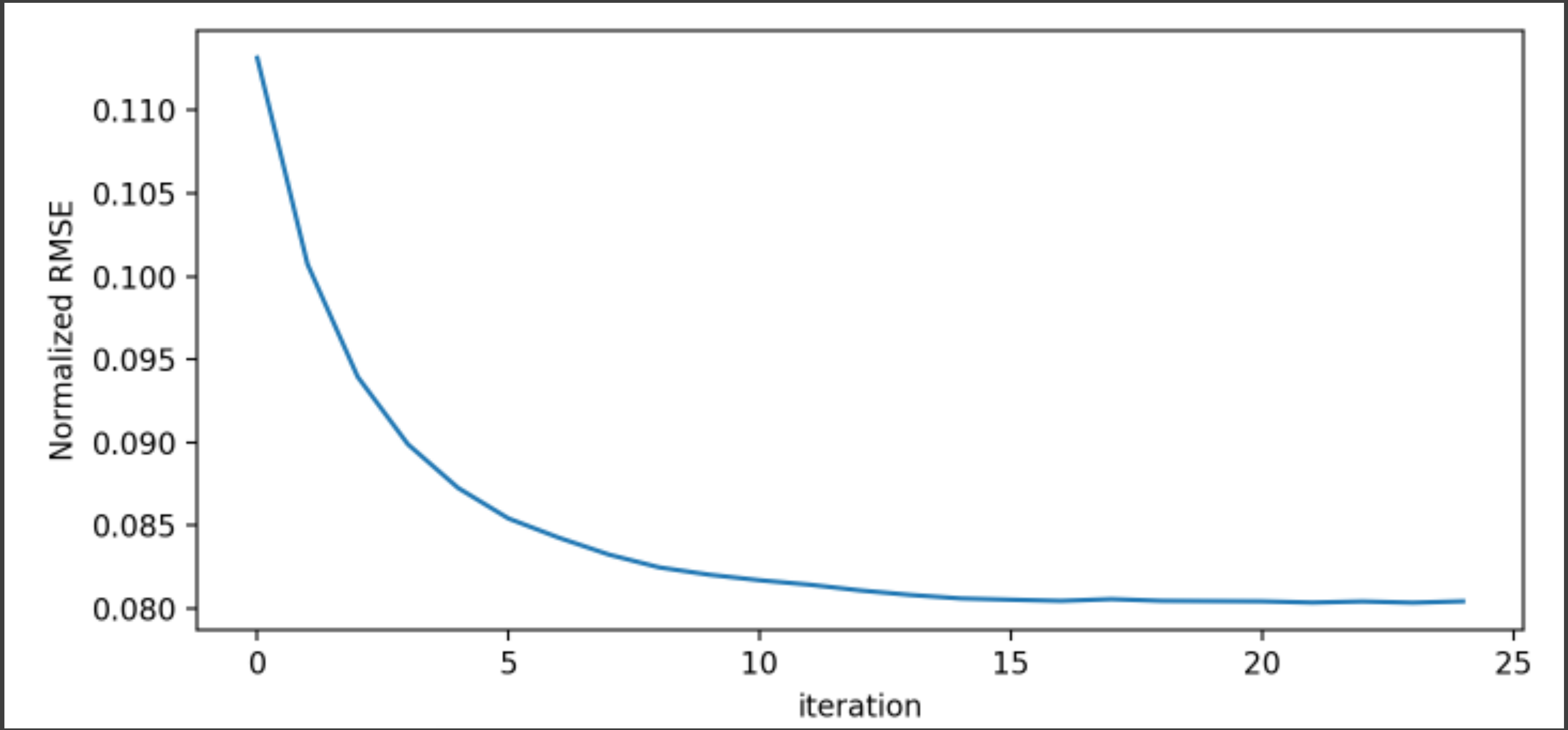


$\mathbf{x}_D$



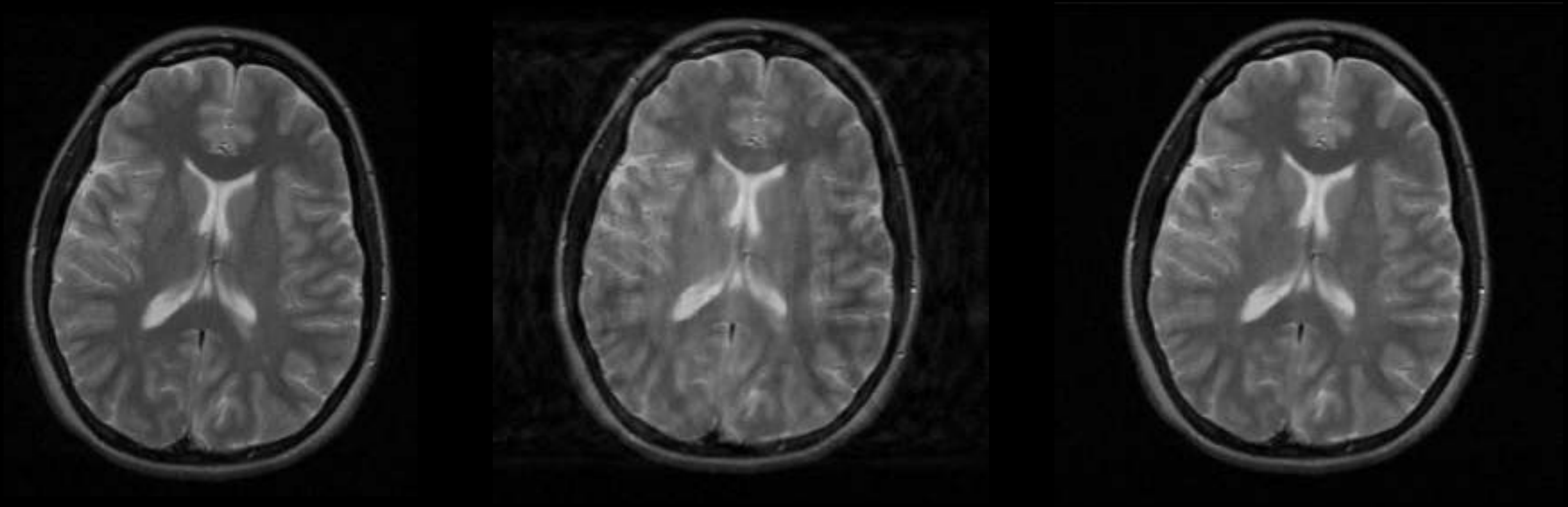
$\mathbf{x}$

Dictionary learning MRI



Error decreases with number of iterations

Dictionary learning MRI



Ground-truth CS-MRI Dictionary Learning

Ravishankar et al., IEEE TMI, 2011. doi: 10.1109/TMI.2010.2090538

5x under-sampling, single-coil

The Devil's in the details…

- How should I remove "bad" atoms?
- How big should the dictionary be?
- How sparse should the signal be?
- Should I update the dictionary all at once?
- How should I initialize the dictionary?
- How should I enforce sparsity?

Hands-on dictionary learning

- Tutorial code for dictionary learning:

  https://github.com/utcsilab/dictionary_learning_ismrm_2020

Tutorial Code

- Based on SigPy, a Python toolbox for iterative signal processing

  https://sigpy.readthedocs.io/en/latest/

SigPy

From local to global modeling

- Dictionary learning: local representation
  - The dictionary is built to represent local patches
  - Each patch is a sparse combination of the dictionary
  - The patches are reshaped and averaged to form the image

- Can we create a sparse global model?

Ref: [1] Zeiler et al. CVPR 2010, [2] Quan et al, ISBI 2016, [3] Elad et al., TMI 2006

From local to global modeling

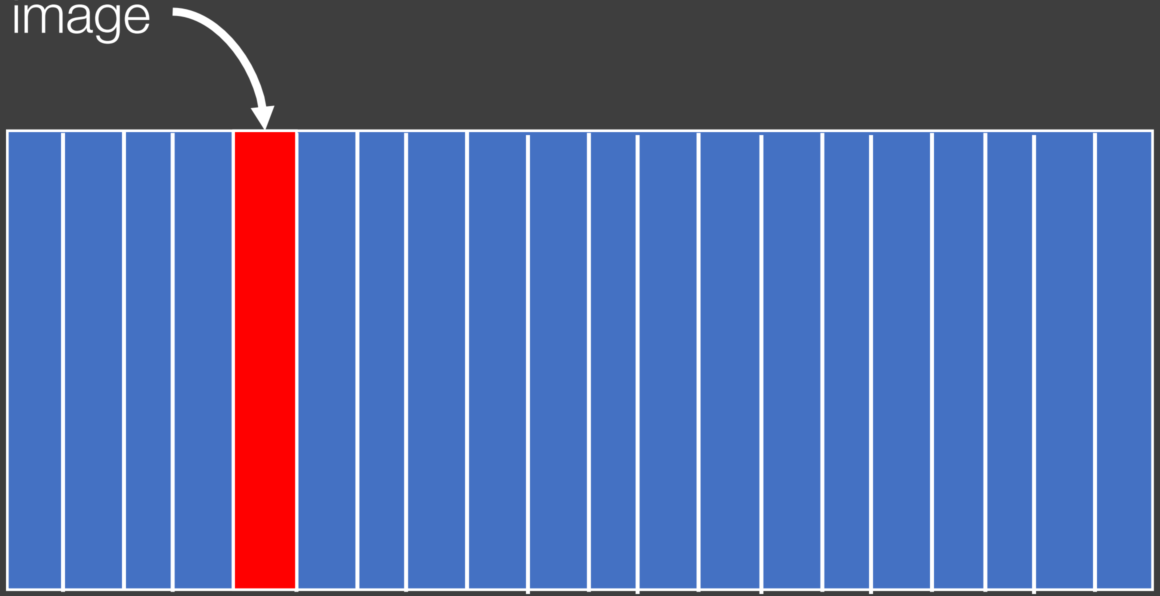
- Can we create a sparse global model?
  - Could we represent the image as a sparse combination of *images?*
  - Our dictionary is N x P, where N is the number of pixels!!

    → Not feasible!

Images

Vectorize

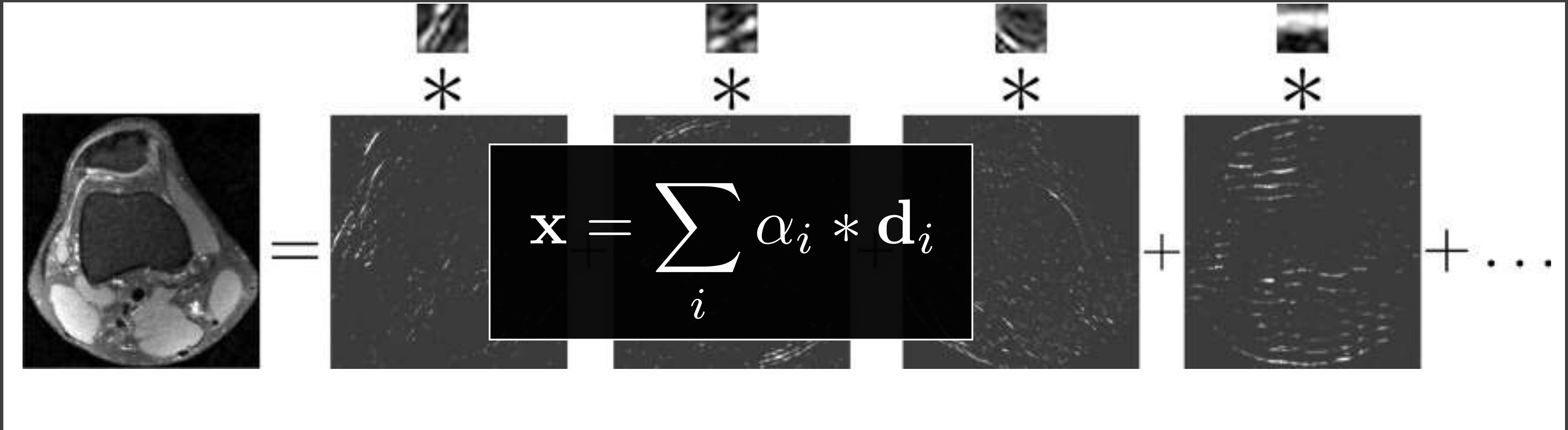
One image



$\mathbf{X}$

Alternative: convolutional sparse coding

- Represent the full image as a sum of dictionary filters convolved with sparse coefficients



$$\mathbf{x} = \sum_i \alpha_i * \mathbf{d}_i$$

From Frank Ong

Ref: [1] Zeiler et al. CVPR 2010, [2] Quan et al, ISBI 2016, [3] Elad et al., TMI 2006, [4] F Ong et al., ISMRM 18

Convolutional sparse coding (CSC)

- In k-space, over multiple data sets:

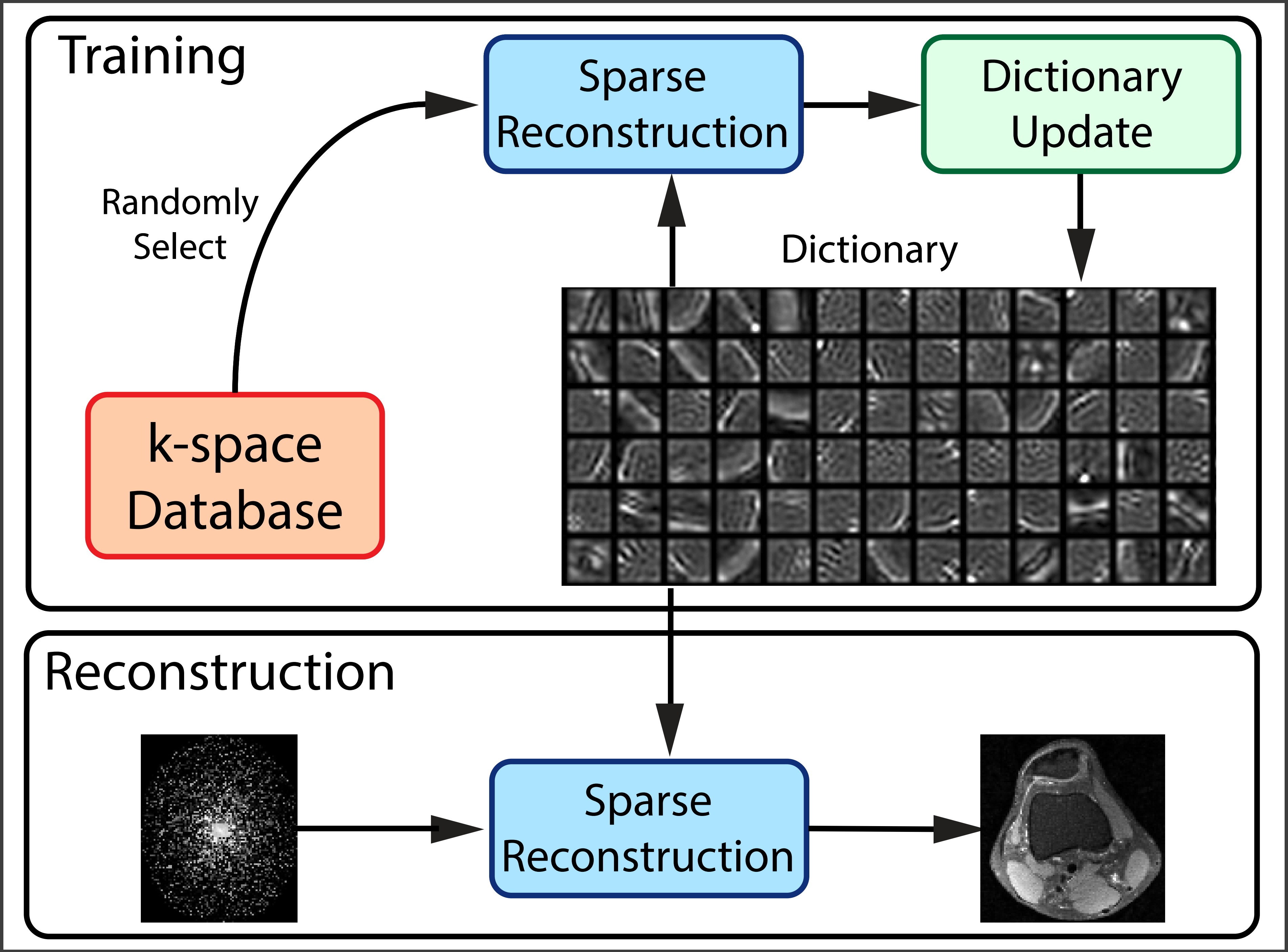
$$\min_{\alpha_j, \mathbf{D}} \sum_j \frac{1}{2} ||\mathbf{y_j} - \mathbf{E}_j \sum_i \alpha_{ij} * \mathbf{d}_i||_2^2 + \lambda ||\alpha_{ij}||_1$$

Same filters for all datasets

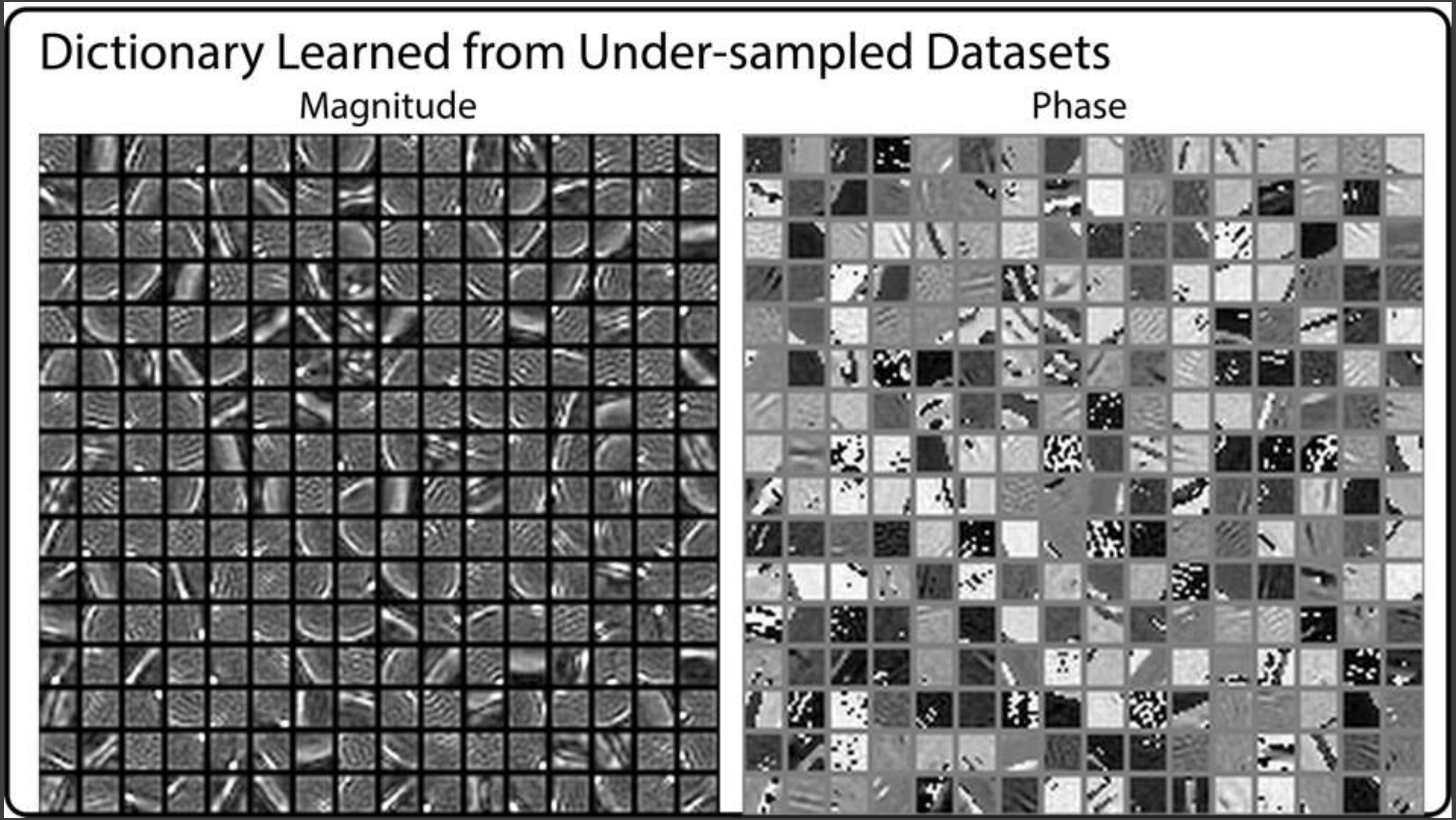
j'th k-space dataset

i'th sparse channel for j'th dataset

Ref: [1] Zeiler et al. CVPR 2010, [2] Quan et al, ISBI 2016, [3] Elad et al., TMI 2006, [4] F Ong et al., ISMRM 18

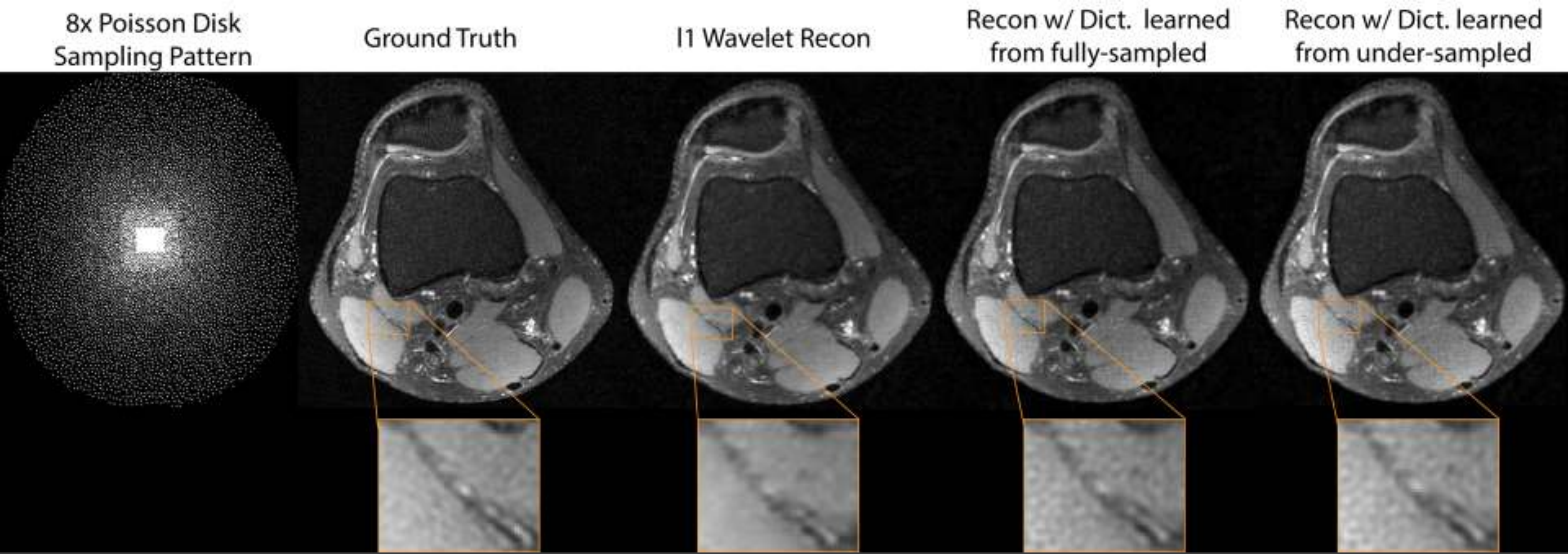


From Frank Ong
F Ong et al., ISMRM 18



From Frank Ong
F Ong et al., ISMRM 18

CSC MRI



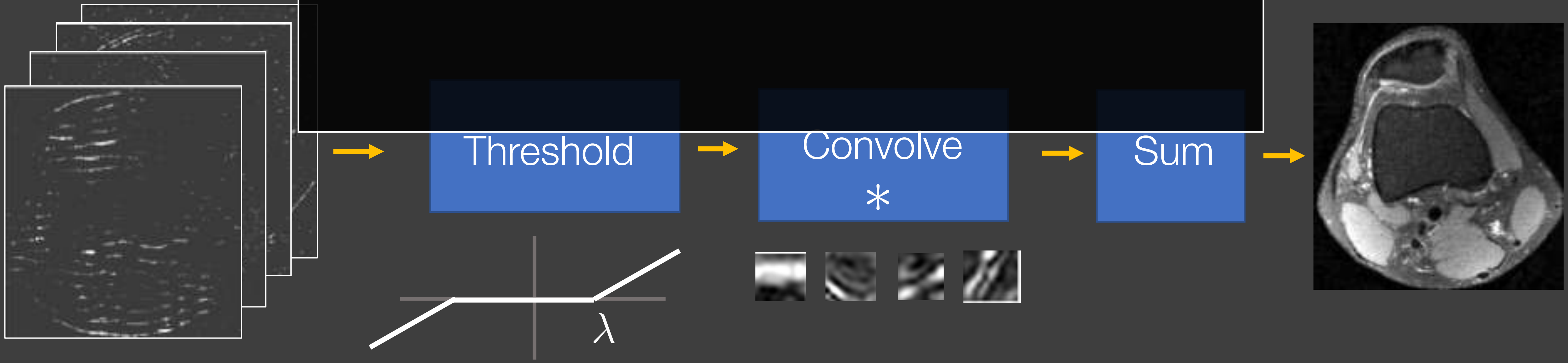
From Frank Ong
F Ong et al., ISMRM 18

From CSC to deep learning

CSC recovery algorithm:

- Thres[redacted]
- Conv[redacted]
- Sum [redacted]
- Repea[redacted]

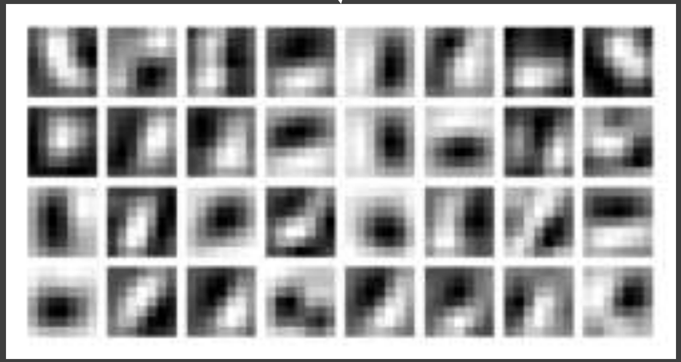
One layer of a convolutional neural network?



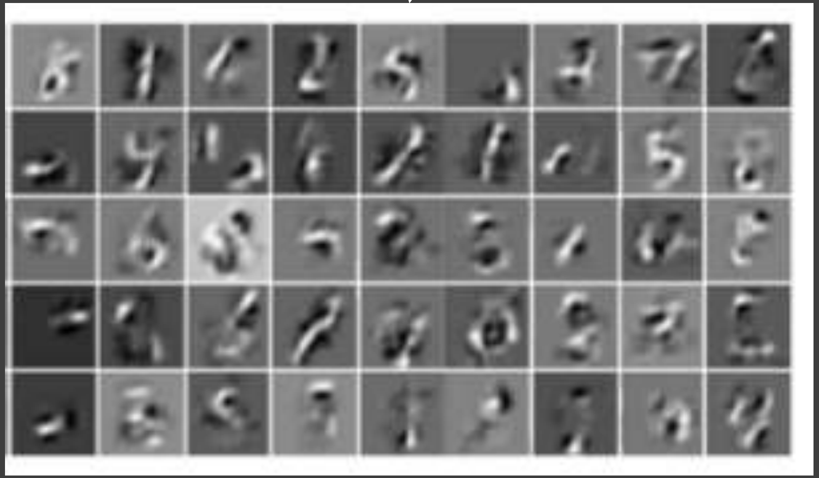
Multi-layer convolutional sparse coding

- Apply CSC multiple times, learn filters at each stage
- Like deep learning, but with performance guarantees!!

CSC Block → CSC Block → … → CSC Block →



dictionary filters



Sulam et al., IEEE TSP, 2018. DOI: 10.1109/TSP.2018.2846226

Summary

- Dictionary learning extends compressed sensing to learned sparsifying transforms
- Application to MRI is straightforward
- Choice of hyper-parameters greatly impacts performance
- Connections between Dictionary Learning (DL) and Deep Learning (DL) in more than just initials

www.jtsense.com

Slides/images
Miki Lustig, Joseph Cheng,
Josh Trzasko, Frank Ong