Medical imaging inverse problems using optimization and machine learning

Jeffrey A. Fessler
William L. Root Professor of EECS

EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

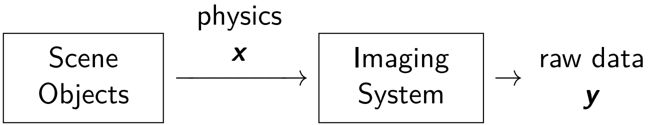
http://web.eecs.umich.edu/~fessler

Work with Sai Ravishankar, Il Yong Chun, Raj Nadakuditi, Yong Long, Xuehang Zheng, ...

CHEPS Seminar

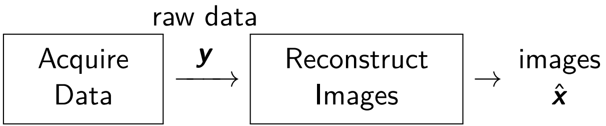
2019-10-08

- Forward problem (data acquisition):



SPECT, PET, X-ray CT, MRI, optical

- Inverse problem (image formation):



- Image reconstruction topics: physics models, measurement statistical models, regularization / object priors, optimization.

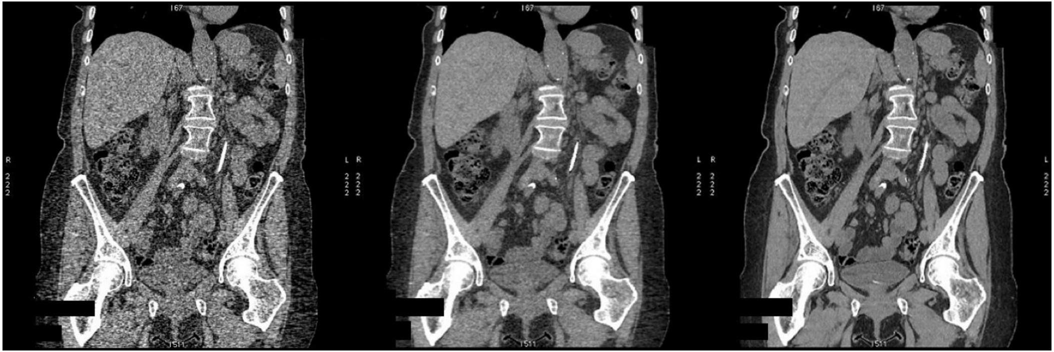
Generations of medical image reconstruction methods

1. 70's "Analytical" methods (integral equations)
   FBP for SPECT / PET / X-ray CT, IFFT for MRI, ...
2. 80's Algebraic methods (as in "linear algebra")
   Solve $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
3. 90's Statistical methods
   - LS / ML methods
   - regularized / Bayesian methods
4. 00's Compressed sensing methods
   (mathematical sparsity models)
5. 10's Adaptive / data-driven methods
   machine learning, deep learning, ...

Improving X-ray CT image reconstruction

J. Fessler

- A picture is worth 1000 words
- (and perhaps several 1000 seconds of computation?)



Thin-slice FBP	ASIR (denoise)	Statistical
Seconds	A bit longer	Much longer

Today's talk: less about computation, more about image quality
Right image used edge-preserving regularization

Safety / health relevance: X-ray dose and diagnostic accuracy

History: Milestones in iterative image reconstruction

Commercial availability of iterative methods for human scanners per FDA 510(k) dates:

- PET/SPECT
  Unregularized OS-EM ≈ 1997
- X-ray CT
  Regularized MBIR [2011-11-09 for GE Veo]
  (Installed at UM in Jan. 2012)
- PET
  Regularized EM variant (Q.Clear) 2014-03-21
- MRI
  Compressed sensing! (Sparsity-based regularization)
  [2017-01-27 for Siemens Cardiac Cine]
  [2017-04-20 for GE HyperSense]
- Ultrasound?

Accelerating MR imaging using adaptive regularization

(a) 4× under-sampled MR k-space

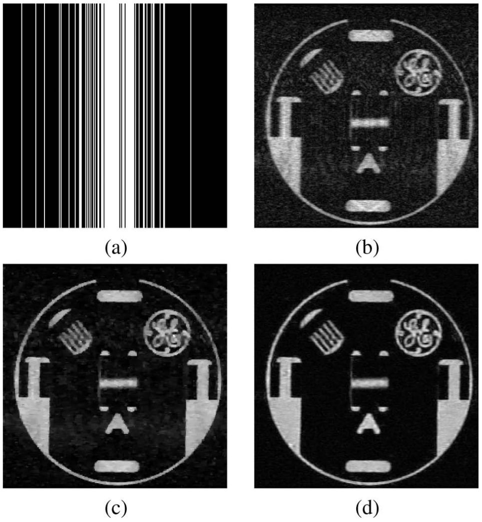
(b) zero-filled reconstruction

(c) "compressed sensing" reconstruction with TV regularization

(d) adaptive dictionary learning regularization [1, Fig. 10]

Safety / health relevance:

- scan time
- motion
- image quality



Outline

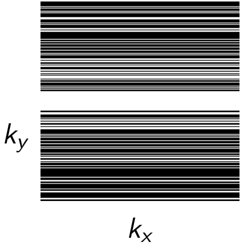
Ill-posed inverse problems

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

$\boldsymbol{y}$: measurements $\boldsymbol{\varepsilon}$: noise

$\boldsymbol{x}$: unknown image $\boldsymbol{A}$: system matrix (typically wide)

- compressed sensing (*e.g.*, MRI) ($\boldsymbol{A}$ "random" rows of DFT)



- deblurring (restoration) ($\boldsymbol{A}$ Toeplitz)
- in-painting ($\boldsymbol{A}$ subset of rows of $\boldsymbol{I}$)
- denoising (not ill posed) ($\boldsymbol{A} = \boldsymbol{I}$)

Why under-sample?

Why under-sample in MRI?

- Reduce scan time (?)
  - Patient comfort
  - Scan cost / throughput
  - Motion artifacts (Philips at ISMRM 2017)
- Improve spatial resolution (collect higher k-space lines)
- Improve scan diversity for quantitative MRI
- Improve temporal resolution trade-off in dynamic MRI

Why under-sample or reduce intensity in CT?

- Reduce X-ray dose

(But under-sampling leads to ill-posed inverse problems...)

Inverse problems via MAP estimation

Unknown image $\boldsymbol{x}$ $\rightarrow$ [System model $\mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x})$] $\rightarrow$ Data $\boldsymbol{y}$ $\rightarrow$ [Estimator] $\rightarrow$ Recon. image $\hat{\boldsymbol{x}}$

If we have a prior $\mathrm{p}(\boldsymbol{x})$, then the MAP estimate is:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \,.$$

For gaussian measurement errors and a linear forward model:

$$-\log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) \equiv \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2$$

where $\|\boldsymbol{y}\|_{\boldsymbol{W}}^2 = \boldsymbol{y}' \boldsymbol{W} \boldsymbol{y}$ and $\boldsymbol{W}^{-1} = \mathrm{Cov}\{\boldsymbol{y} \mid \boldsymbol{x}\}$ is known
($\boldsymbol{A}$ from physics, $\boldsymbol{W}$ from statistics)

Priors for MAP estimation

J. Fessler

UNIVERSITY OF MICHIGAN

- If all images $\boldsymbol{x}$ are "plausible" (have non-zero probability) then

$$\mathrm{p}(\boldsymbol{x}) \propto \mathrm{e}^{-\mathsf{R}(\boldsymbol{x})} \Longrightarrow -\log \mathrm{p}(\boldsymbol{x}) \equiv \mathsf{R}(\boldsymbol{x})$$

  (from fantasy / imagination / wishful thinking / data)

- MAP $\equiv$ regularized weighted least-squares (WLS) estimation:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\arg\max} \; \log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x})$$
$$= \underset{\boldsymbol{x}}{\arg\min} \; \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{x})$$

- A regularizer $\mathsf{R}(\boldsymbol{x})$, aka log prior, is essential for high-quality solutions to ill-conditioned / ill-posed inverse problems.
- Why ill-posed? Often high ambitions...

Classical regularizers ("hand crafted")

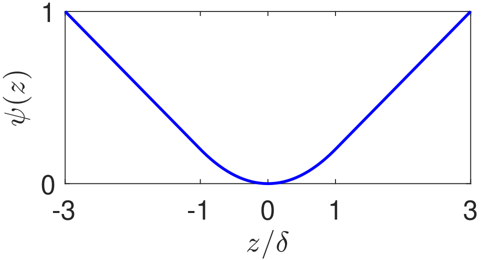
- Tikhonov regularization (IID gaussian prior)
- Roughness penalty (Basic MRF prior)
- Sparsity in ambient space
- Edge-preserving regularization
- Total-variation (TV) regularization
- Black-box denoiser like NLM

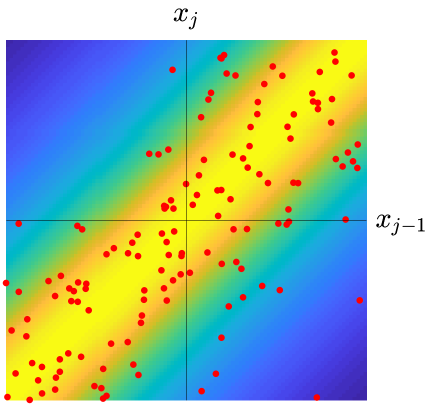
Edge-preserving regularization

Neighboring pixels tend to have similar values except near edges:

$$\mathsf{R}(\boldsymbol{x}) = \beta \sum_j \psi(x_j - x_{j-1})$$

Potential function ψ:



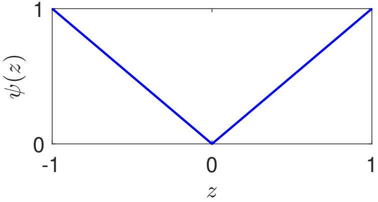


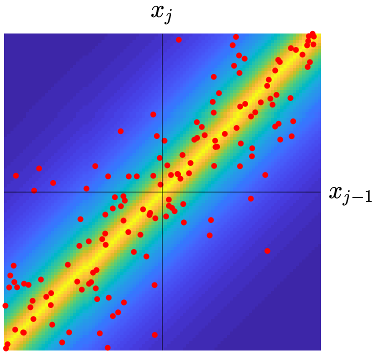
- Equivalent to improper prior (agnostic to DC value)
- Accounts for spatial correlations, but only very locally
- Used clinically now for low-dose X-ray CT image reconstruction

Neighboring pixels tend to have similar values except near edges ("gradient sparsity"):

$$\mathsf{R}(\boldsymbol{x}) = \beta \, \mathrm{TV}(\boldsymbol{x}) = \beta \left\| \boldsymbol{\Delta x} \right\|_1$$
$$= \beta \sum_j |x_j - x_{j-1}|$$

Potential function ψ:





- Equivalent to improper prior (agnostic to DC value)
- Accounts for correlations, but only very locally
- Well-suited to piece-wise constant Shepp-Logan phantom!
- Used in many academic publications...

Many more "conventional" regularizers / priors

J. Fessler

- Transforms: wavelets, curvelets, . . .
- Markov random field models
- Graphical models
- . . .

All "hand crafted" ...

Outline

- Data
  - Population adaptive methods (*e.g.*, X-ray CT)
  - Patient adaptive methods (*e.g.*, dynamic MRI?)
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transforms) approach

Many options...

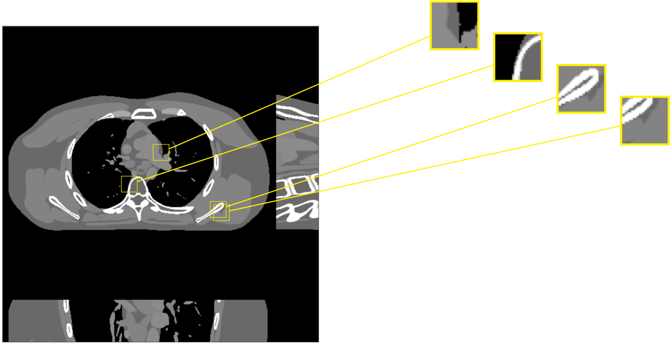
Outline

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

Patch-wise transform sparsity model

Assumption: if $\boldsymbol{x}$ is a plausible image, then each $\boldsymbol{\Omega P}_m \boldsymbol{x}$ is sparse.

- $\boldsymbol{P}_m \boldsymbol{x}$ extracts the mth of M patches from $\boldsymbol{x}$
- $\boldsymbol{\Omega}$ is a square sparsifying transform matrix



Sparsifying transform learning (population adaptive)

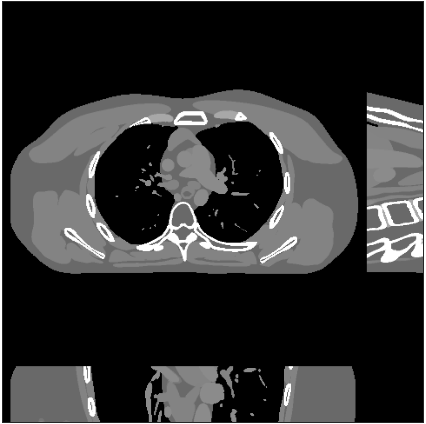
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find transform $\boldsymbol{\Omega}_*$ that best sparsifies their patches:

$$\boldsymbol{\Omega}_* = \underset{\boldsymbol{\Omega} \text{ unitary}}{\arg\min} \; \min_{\{\boldsymbol{z}_{l,m}\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left\| \boldsymbol{\Omega} \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,m} \right\|_0$$

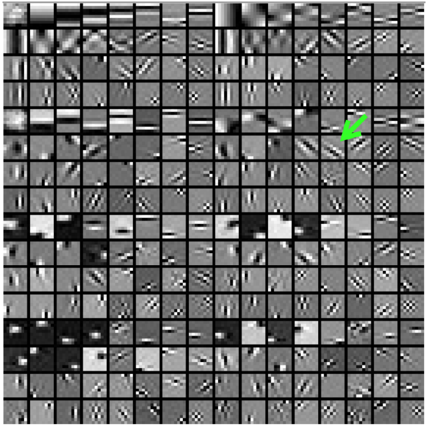
- Encourage aggregate sparsity, not patch-wise sparsity (cf K-SVD [2])
- Non-convex due to unitary constraint and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [3]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - $\boldsymbol{\Omega}$ update is an orthogonal Procrustes problem (SVD)
  - Subsequence convergence guarantees [3]

Example of learned sparsifying transform

3D X-ray training data



Parts of learned sparsifier $\boldsymbol{\Omega}_*$



(2D slices in x-y, x-z, y-z, from 3D image volume)

$8 \times 8 \times 8$ patches $\Longrightarrow$ $\boldsymbol{\Omega}_*$ is $8^3 \times 8^3 = 512 \times 512$

top 8×8 slice of 256 of the 512 rows of $\boldsymbol{\Omega}_*$ $\uparrow$

Regularizer based on learned sparsifying transform

Regularized inverse problem [4]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta\, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{\Omega}_* \boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{z}_m\|_2^2 + \alpha \|\boldsymbol{z}_m\|_0 .$$

$\boldsymbol{\Omega}_*$ adapted to population training data

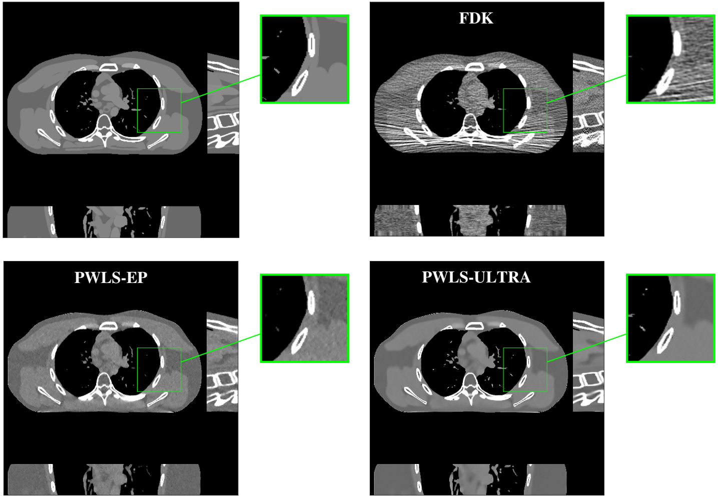
Alternating minimization optimizer:

- $\boldsymbol{z}_m$ update is simple hard thresholding
- $\boldsymbol{x}$ update is a quadratic problem: many options

  Linearized augmented Lagrangian method (LALM) [5]

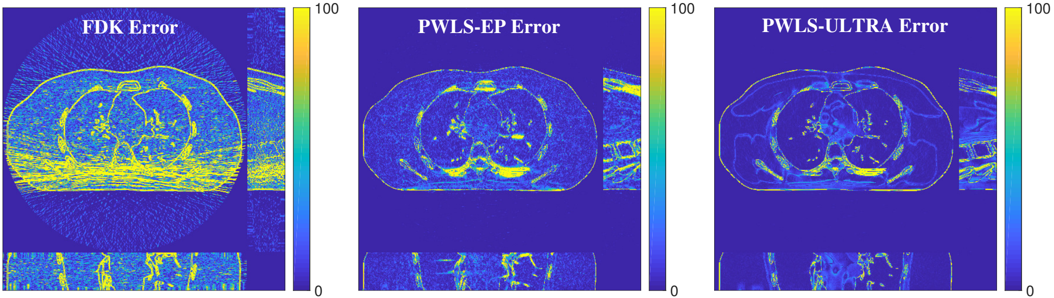
Example: low-dose 3D X-ray CT simulation

X. Zheng, S. Ravishankar, Y. Long, JF: IEEE T-MI, June 2018 [4]



3D X-ray CT simulation Error maps

J. Fessler



	X-ray Intensity	FDK	EP	ST Ω_*	ULTRA	ULTRA-$\{\tau_j\}$
RMSE in HU	1×10^4	67.8	34.6	32.1	30.7	**29.2**
	5×10^3	89.0	41.1	37.3	35.7	**34.2**

- Physics / statistics provides dramatic improvement
- Data adaptive regularization further reduces RMSE

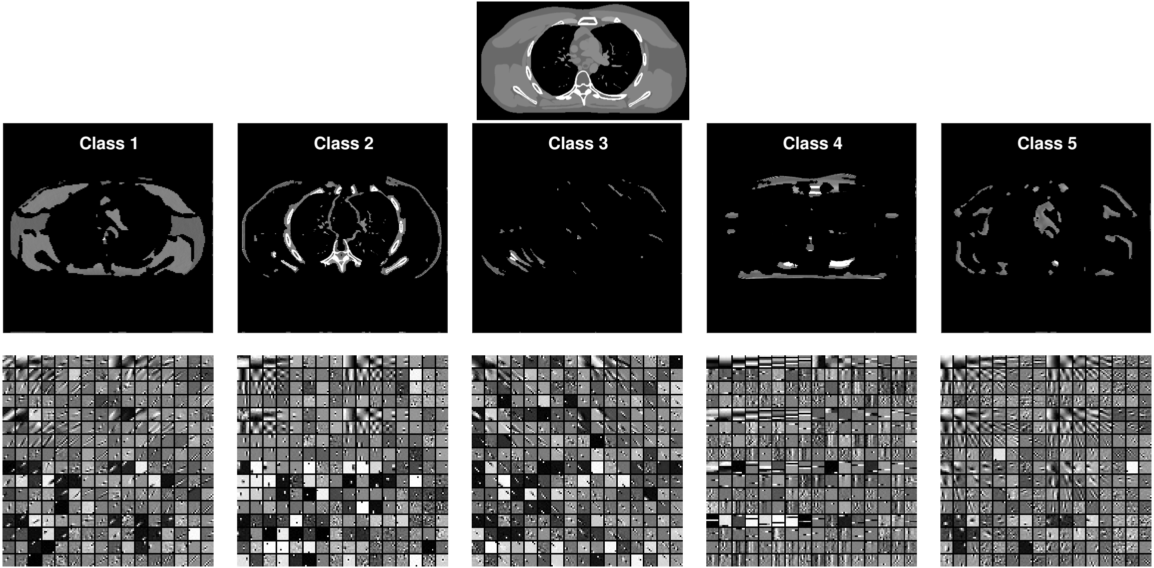
Union of Learned TRAnsforms (ULTRA)

Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find a set of transforms $\left\{\hat{\boldsymbol{\Omega}}_k\right\}_{k=1}^{K}$ that best sparsify image patches:

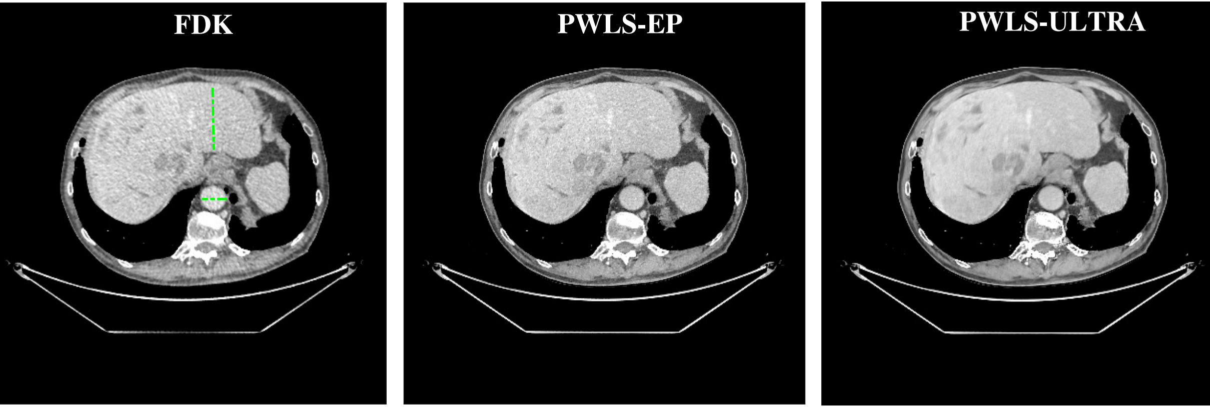
$$\left\{\hat{\boldsymbol{\Omega}}_k\right\} = \underset{\{\boldsymbol{\Omega}_k \text{ unitary}\}}{\arg\min} \; \min_{\left\{k_{l,m} \in \{1,\ldots,K\}\right\}} \; \min_{\left\{\boldsymbol{z}_{l,m}\right\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left\| \boldsymbol{\Omega}_{k_{l,m}} \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,m} \right\|_0$$

- Joint unsupervised clustering / sparsification
- Further nonconvexity due to clustering
- Efficient alternating minimization algorithm [6]

Example: 3D X-ray CT learned set of transforms



Example: 3D X-ray CT ULTRA for chest scan



Zheng et al., IEEE T-MI, June 2018 [4]

X-ray CT with learned convolutional filters

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

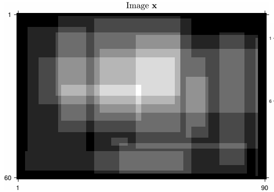
Drawback of basic patch-based methods:
$512 \times 512 \times 512$ 3D X-ray CT image volume
$8 \times 8 \times 8$ patches
$\Longrightarrow 512^3 \cdot 8^3 \cdot 4 = 256$ Gbyte of patch data for stride=1

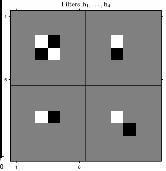
Convolutional sparsity model

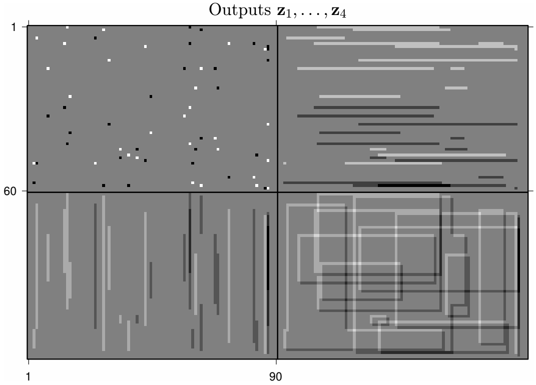
Assumption: There is a set of filters $\{\boldsymbol{h}_k\}_{k=1}^K$
such that the images $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ are sparse for a plausible image $\boldsymbol{x}$.

- For more plausible images, $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ is more sparse.
- $*$ denotes convolution
- Inherently shift invariant and no patches

Example (hand crafted filters):







Sparsifying filter learning (population adaptive)

J. Fessler

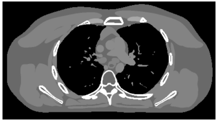
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find filters $\left\{\hat{\boldsymbol{h}}_k\right\}_{k=1}^K$ that best sparsify them:

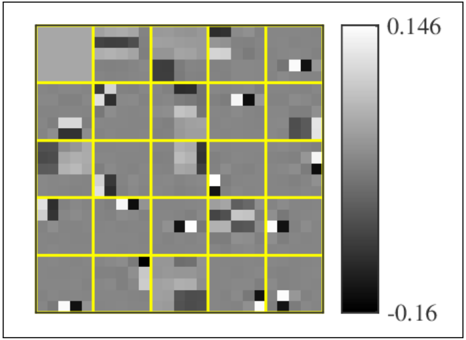
$$\left\{\hat{\boldsymbol{h}}_k\right\} = \underset{\{\boldsymbol{h}_k\}\in\mathcal{H}}{\arg\min} \min_{\{\boldsymbol{z}_{l,k}\}} \sum_{l=1}^{L} \sum_{k=1}^{K} \|\boldsymbol{h}_k * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}\|_2^2 + \alpha \|\boldsymbol{z}_{l,k}\|_0$$

- To encourage filter diversity:
  - $\mathcal{H} = \{\boldsymbol{H} : \boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}\}, \ \boldsymbol{H} = [\boldsymbol{h}_1 \ \ldots \ \boldsymbol{h}_K]$
  - *cf.* tight-frame condition $\sum_{k=1}^K \|\boldsymbol{h}_k * \boldsymbol{x}\|_2^2 \propto \|\boldsymbol{x}\|_2^2$
- Encourage aggregate sparsity, period
- Non-convex due to constraint $\mathcal{H}$ and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [7]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - Filter update uses diagonal majorizer, proximal map (SVD)
  - Subsequence convergence guarantees [7]

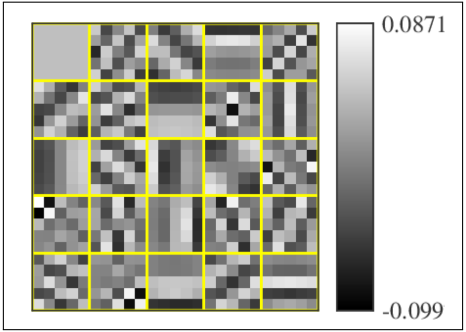
Examples of learned sparsifying filters

2D X-ray CT training data and learned 5×5 sparsifying filters $\left\{\hat{\boldsymbol{h}}_k\right\}$ [7]:





$\alpha = 10^{-4}$



$\alpha = 2 \times 10^{-3}$

Regularizer based on learned sparsifying filters

Regularized inverse problem [7]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \succeq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \arg\min_{\{\boldsymbol{z}_k\}} \sum_{k=1}^{K} \left\| \hat{\boldsymbol{h}}_k * \boldsymbol{x} - \boldsymbol{z}_k \right\|_2^2 + \alpha \, \|\boldsymbol{z}_k\|_0 .$$

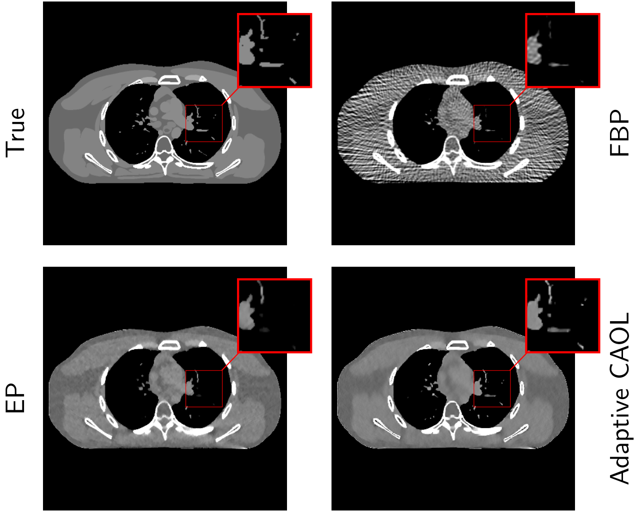
$\left\{ \hat{\boldsymbol{h}}_k \right\}$ adapted to population training data

Block proximal gradient with majorizer (BPG-M) optimizer:

- $\boldsymbol{z}_k$ update is simple hard thresholding
- $\boldsymbol{x}$ update is a quadratic problem: diagonal majorizer

I. Y. Chun, JF, 2018, arXiv 1802.05584 [7]

Example: sparse-view 2D X-ray CT simulation



Quantitative results

J. Fessler

123 views (out of usual 984) $\Longrightarrow$ $8\times$ dose reduction

RMSE (in HU):

FBP	82.8
EP	40.8
Adaptive filters	**35.2**

- Physics / statistics provides dramatic improvement
- Data-adaptive regularization further reduces RMSE

Extension to multiple layers (cf CNN) I

Convolutional sparsity model: $\boldsymbol{h}_k * \boldsymbol{x}$ is sparse for $k = 1, \ldots, K_1$

Learning 1 "layer" of filters:

$$\{\hat{\boldsymbol{h}}_k^{[1]}\} = \underset{\{\boldsymbol{h}_k^{[1]}\} \in \mathcal{H}}{\arg\min} \min_{\{\boldsymbol{z}_{l,k}^{[1]}\}} \sum_{l=1}^{L} \sum_{k=1}^{K_1} \left\| \boldsymbol{h}_k^{[1]} * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}^{[1]} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,k}^{[1]} \right\|_0$$

Learning 2 layers of filters [7]:

$$\left(\{\hat{\boldsymbol{h}}_k^{[1]}\}, \{\hat{\boldsymbol{h}}_k^{[2]}\}\right) = \mathop{\arg\min}_{\{\boldsymbol{h}_k^{[1]}\},\{\boldsymbol{h}_k^{[2]}\}\in\mathcal{H}} \min_{\{\boldsymbol{z}_{l,k}^{[1]}\}} \min_{\{\boldsymbol{z}_{l,k}^{[2]}\}} \sum_{l=1}^{L}\sum_{k=1}^{K_1} \left\|\boldsymbol{h}_k^{[1]} * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}^{[1]}\right\|_2^2 + \alpha \left\|\boldsymbol{z}_{l,k}^{[1]}\right\|_0 + \sum_{l=1}^{L}\sum_{k=1}^{K_2} \left\|\boldsymbol{h}_k^{[2]} * \left(\boldsymbol{P}_k \boldsymbol{z}_l^{[1]}\right) - \boldsymbol{z}_{l,k}^{[2]}\right\|_2^2 + \alpha \left\|\boldsymbol{z}_{l,k}^{[2]}\right\|_0$$

Here $\boldsymbol{P}_k$ is a pooling operator for the output of first layer

Block proximal gradient with majorizer (BPG-M) optimizer

I. Y. Chun, JF, 2018, arXiv 1802.05584 [7]

Use multi-level learned filters as (interpretable?) regularizer for CT.

Outline

MR with adapted dictionary

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

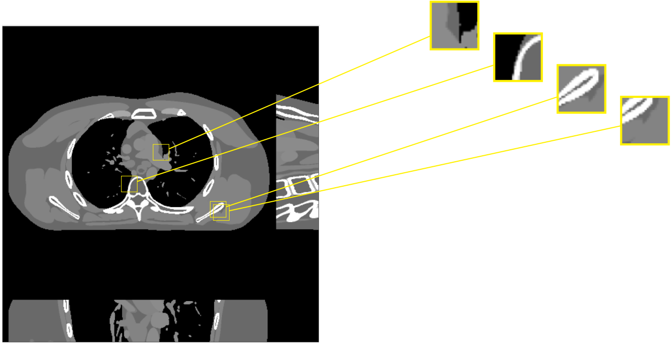
Patch-wise dictionary sparsity model

Assumption: if $\boldsymbol{x}$ is a plausible image, then each patch has

$$\boldsymbol{P}_m \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_m,$$

for a sparse coefficient vector $\boldsymbol{z}_m$. (Synthesis approach.)

- $\boldsymbol{P}_m \boldsymbol{x}$ extracts the mth of M patches from $\boldsymbol{x}$
- $\boldsymbol{D}$ is a (typically overcomplete) dictionary for patches



Dictionary-blind MR image reconstruction:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \beta \, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathcal{D}} \min_{\boldsymbol{Z}' \in \mathcal{C}} \sum_{m=1}^{M} \left(\|\boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_m\|_2^2 + \lambda^2 \|\boldsymbol{z}_m\|_0 \right)$$

where $\boldsymbol{P}_m$ extracts mth of M image patches.

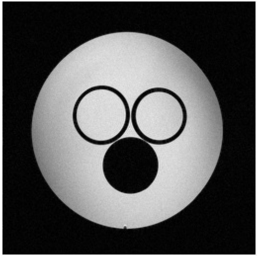
In words: of the many images...

Alternating (nested) minimization:

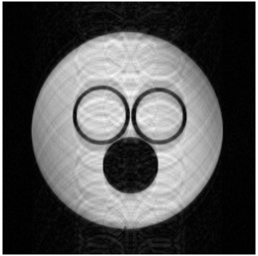
- Fixing $\boldsymbol{x}$ and $\boldsymbol{D}$, update each row of $\boldsymbol{Z} = [\boldsymbol{z}_1 \; \ldots \; \boldsymbol{z}_M]$ sequentially via hard-thresholding.
- Fixing $\boldsymbol{x}$ and $\boldsymbol{Z}$, update $\boldsymbol{D}$ using SOUP-DIL [8].
- Fixing $\boldsymbol{Z}$ and $\boldsymbol{D}$, updating $\boldsymbol{x}$ is a quadratic problem.
  - Efficient FFT solution for single-coil Cartesian MRI.
  - Use CG for non-Cartesian and/or parallel MRI.
- Non-convex, but monotone decreasing and some convergence theory [8].

2D CS MRI results I

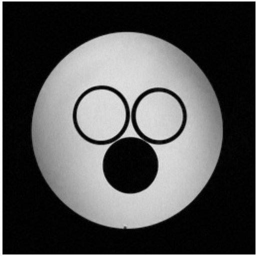
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI

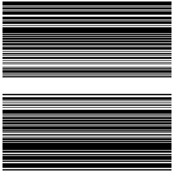


6×6 patches

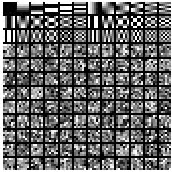
$\boldsymbol{D} \in \mathbb{C}^{6^2 \times 144}$

[8]

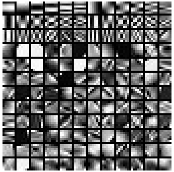
Sampling ($2.5\times$)



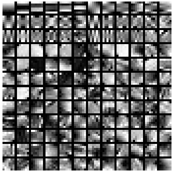
Initial $\boldsymbol{D}$



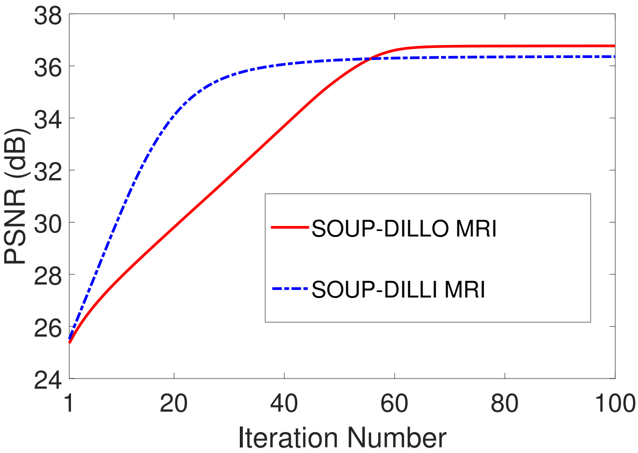
Learned real$\{\boldsymbol{D}\}$



imag$\{\boldsymbol{D}\}$



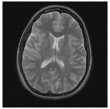
2D CS MRI results II

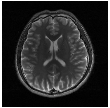


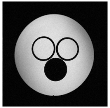
(SNR compared to fully sampled image.)
Using $\|z_m\|_0$ leads to higher SNR than $\|z_m\|_1$.
Adaptive case is non-convex anyway...

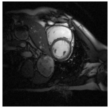
2D CS MRI results III

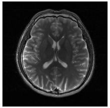
J. Fessler


(a)

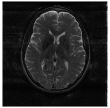

(b)


(c)


(d)

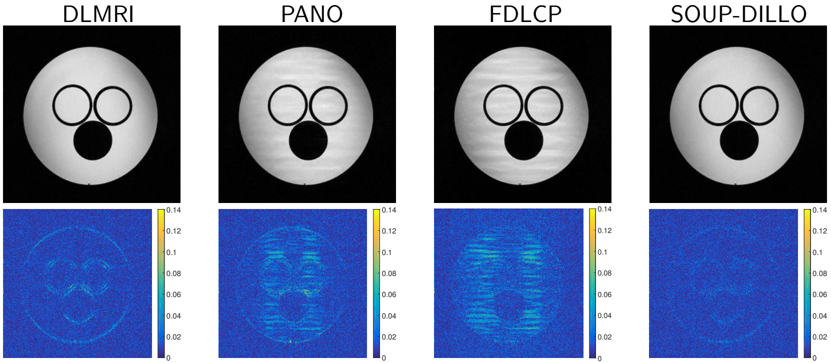

(e)


(f)


(g)

Im.	Samp.	Acc.	0-fill	Sparse MRI	PANO	DLMRI	SOUP-DILLI	SOUP-DILLO
a	Cart.	7x	27.9	28.6	**31.1**	**31.1**	30.8	**31.1**
b	Cart.	2.5x	27.7	31.6	41.3	40.2	38.5	**42.3**
c	Cart.	2.5x	24.9	29.9	34.8	36.7	36.6	**37.3**
c	Cart.	4x	25.9	28.8	**32.3**	32.1	32.2	**32.3**
d	Cart.	2.5x	29.5	32.1	36.9	38.1	36.7	**38.4**
e	Cart.	2.5x	28.1	31.7	40.0	38.0	37.9	**41.5**
f	2D rand.	5x	26.3	27.4	30.4	30.5	30.3	**30.6**
g	Cart.	2.5x	32.8	39.1	41.6	41.7	42.2	**43.2**
Ref.				[9]	[10]	[1]	[8]	[8]

2D CS MRI results IV



[1] [10] [11] [8]

Summary: 2D static MR reconstruction from under-sampled data with adaptive dictionary learning and convergent algorithm, faster than K-SVD approach of DLMRI.

Summary

- Data-driven / adaptive regularization
  - Beneficial for low-dose CT and under-sampled MRI reconstruction
  - Dictionary atom structure (*e.g.*, low rank) further helpful for dynamic MRI
  - Block proximal methods provide reasonably computational efficiency
  - Convergence theory (unlike KSVD)

- Future work:
  - Synthesis (*e.g.*, dictionary) vs analysis (*e.g.*, transform learning) formulations
    Begs for some principled model comparison...
  - Online methods for reduced memory, better adaptation [12–15]
  - Adaptive methods versus "deep" methods?
  - Prospective use

June 2018 special issue of IEEE Trans. on Medical Imaging [16]:

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 37, NO. 6, JUNE 2018 1289

Image Reconstruction Is a New Frontier of Machine Learning

Ge Wang, *Fellow, IEEE*, Jong Chu Ye, *Senior Member, IEEE*, Klaus Mueller, *Senior Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

Bibliography I

[1] S. Ravishankar and Y. Bresler. "MR image reconstruction from highly undersampled k-space data by dictionary learning." In: *IEEE Trans. Med. Imag.* 30.5 (May 2011), 1028–41.

[2] M. Aharon, M. Elad, and A. Bruckstein. "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation." In: *IEEE Trans. Sig. Proc.* 54.11 (Nov. 2006), 4311–22.

[3] S. Ravishankar and Y. Bresler. "l_0 sparsifying transform learning with efficient optimal updates and convergence guarantees." In: *IEEE Trans. Sig. Proc.* 63.9 (May 2015), 2389–404.

[4] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler. "PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1498–510.

[5] H. Nien and J. A. Fessler. "Relaxed linearized algorithms for faster X-ray CT image reconstruction." In: *IEEE Trans. Med. Imag.* 35.4 (Apr. 2016), 1090–8.

[6] S. Ravishankar and Y. Bresler. "Data-driven learning of a union of sparsifying transforms model for blind compressed sensing." In: *IEEE Trans. Computational Imaging* 2.3 (Sept. 2016), 294–309.

[7] I. Y. Chun and J. A. Fessler. "Convolutional analysis operator learning: acceleration, convergence, application, and neural networks." In: *IEEE Trans. Im. Proc.* (2018). Submitted.

[8] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler. "Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems." In: *IEEE Trans. Computational Imaging* 3.4 (Dec. 2017), 694–709.

[9] M. Lustig and J. M. Pauly. "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space." In: *Mag. Res. Med.* 64.2 (Aug. 2010), 457–71.

[10] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen. "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator." In: *Med. Im. Anal.* 18.6 (Aug. 2014), 843–56.

[11] Z. Zhan, J-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction." In: *IEEE Trans. Biomed. Engin.* 63.9 (Sept. 2016), 1850–61.

[12] S. Ravishankar, B. Wen, and Y. Bresler. "Online sparsifying transform learning - Part I: algorithms." In: *IEEE J. Sel. Top. Sig. Proc.* 9.4 (June 2015), 625–36.

[13] S. Ravishankar and Y. Bresler. "Online sparsifying transform learning - Part II: convergence analysis." In: *IEEE J. Sel. Top. Sig. Proc.* 9.4 (June 2015), 637–46.

[14] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler. "Efficient online dictionary adaptation and image reconstruction for dynamic MRI." In: *Proc., IEEE Asilomar Conf. on Signals, Systems, and Comp.* 2017, 835–9.

[15] B. E. Moore and S. Ravishankar. "Online data-driven dynamic image restoration using DINO-KAT models." In: *Proc. IEEE Intl. Conf. on Image Processing*. 2017, 3590–4.

[16] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler. "Image reconstruction is a new frontier of machine learning." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1289–96.