CT image reconstruction using adaptive signal models

Jeffrey A. Fessler

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

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

Tufts Conference on Modern Challenges in Imaging
2019-08-07

Acknowledgments: Doug Noll, Yong Long, Sai Ravishankar, Raj Nadakuditi,
Jon Nielsen, Gopal Nataraj, Il Yong Chun, Xuehang Zheng, ...

Declaration: No relevant financial interests or relationships to disclose

Outline

Introduction

Adaptive regularization
- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

Summary

Bibliography

Outline

Introduction

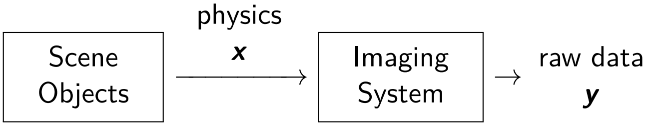
Adaptive regularization

Summary

Bibliography

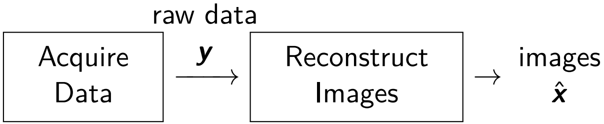
Image reconstruction background

- 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...

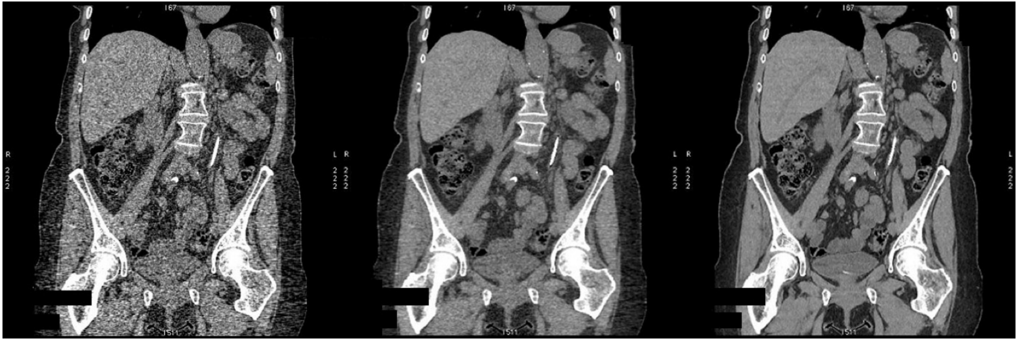
Generations of medical image reconstruction methods

J. Fessler

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
   - Bayesian methods (Markov random fields, ...)
   - regularized methods
4. 00's Compressed sensing methods
   (mathematical sparsity models)
5. 10's Adaptive / data-driven methods
   machine learning, deep learning, ...

Two important milestones for clinical CT

- Model-based image reconstruction (MBIR) FDA approved circa 2012 [1]



- Deep-learning image reconstruction FDA approved 2019 [2, 3]

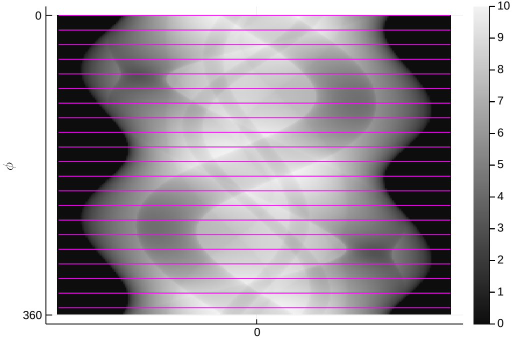
Sparse-view CT: Under-determined inverse problem

J. Fessler

Data model:

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

- $\boldsymbol{y}$: measurements (sinogram)
- $\boldsymbol{\varepsilon}$: noise
- $\boldsymbol{x}$: unknown image
- $\boldsymbol{A}$: system matrix (often wide)



Inverse problems via MAP estimation

J. Fessler

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

- 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:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} \; \log \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}} \; \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{x})\end{aligned}$$

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

Non-adaptive regularizers

- Tikhonov regularization (IID gaussian prior)
- Markov random field (MRF) models
- Roughness penalty (cf MRF prior)
- Edge-preserving regularization (used in clinical CT scanners)
- Total-variation (TV) regularization (*not* used in clinical CT scanners)
- Black-box denoiser like NLM, *e.g.*, plug-and-play ADMM [4]
- Sparsity in ambient space
- Sparsifying transforms: wavelets, curvelets, . . .
- Graphical models
- . . .

All “hand crafted” from statistical / mathematical models ...

Introduction

Adaptive regularization

- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

Summary

Bibliography

Adaptive regularization methods for inverse problems

- 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

Introduction

Adaptive regularization

- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

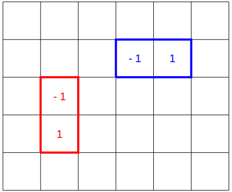
Summary

Bibliography

Patch-based regularization and TV

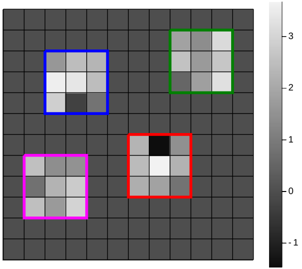
J. Fessler

Anisotropic discrete TV regularizer:
$R(\boldsymbol{x}) = \|\boldsymbol{T}\boldsymbol{x}\|_1$
where $\boldsymbol{T}$ is finite-differences
$\equiv$ patches of size 2×1.



Larger patches provide more context for distinguishing signal from noise.

cf. CNN approaches



Patch-based regularizers:

- synthesis models
- analysis methods

X-ray CT with learned sparsifying transforms

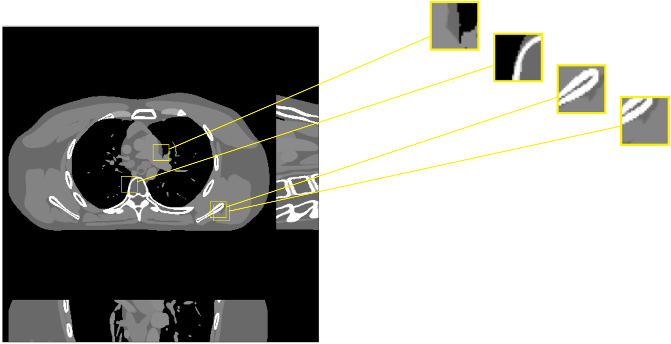
- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

Patch-wise transform sparsity model

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

- $\boldsymbol{P}_m\boldsymbol{x}$ extracts the mth of M patches from $\boldsymbol{x}$
- $\boldsymbol{T}$ is a (often square) sparsifying transform matrix.

What $\boldsymbol{T}$?



Sparsifying transform learning (population adaptive)

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

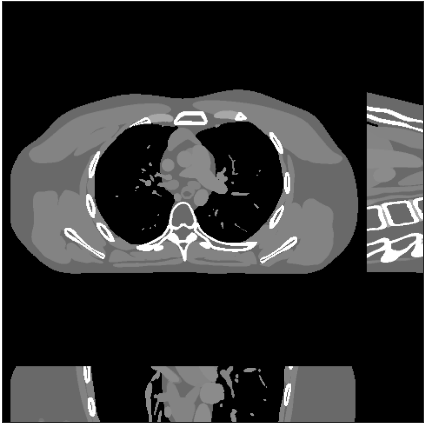
$$\boldsymbol{T}_* = \underset{\boldsymbol{T} \text{ unitary}}{\arg\min} \min_{\{\boldsymbol{z}_{l,m}\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \|\boldsymbol{T}\boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m}\|_2^2 + \alpha \|\boldsymbol{z}_{l,m}\|_0$$

- Encourage aggregate sparsity, not patch-wise sparsity (cf K-SVD [5])
- Non-convex due to unitary constraint and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [6]
  - $\boldsymbol{z}$ update : simple hard thresholding
  - $\boldsymbol{T}$ update : orthogonal Procrustes problem (SVD)
  - Subsequence convergence guarantees [6]

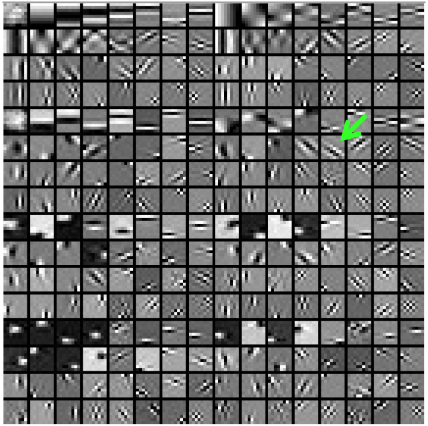
Example of learned sparsifying transform

J. Fessler

3D X-ray training data



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



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

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

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

Regularizer based on learned sparsifying transform

Regularized inverse problem [7]:

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

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

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

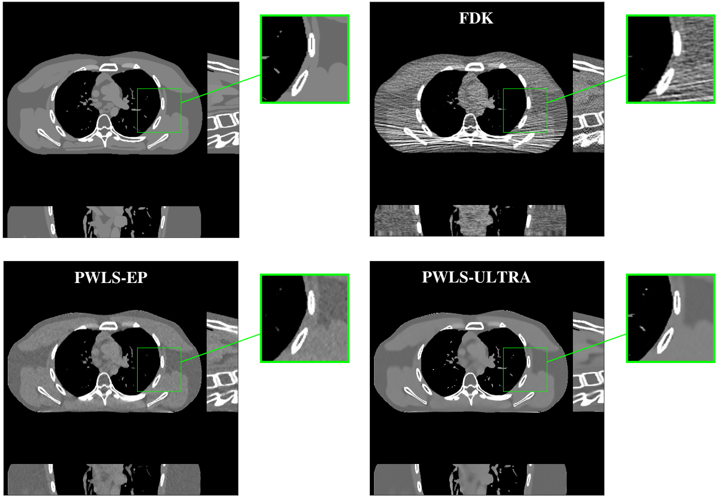
Alternating minimization optimizer:

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

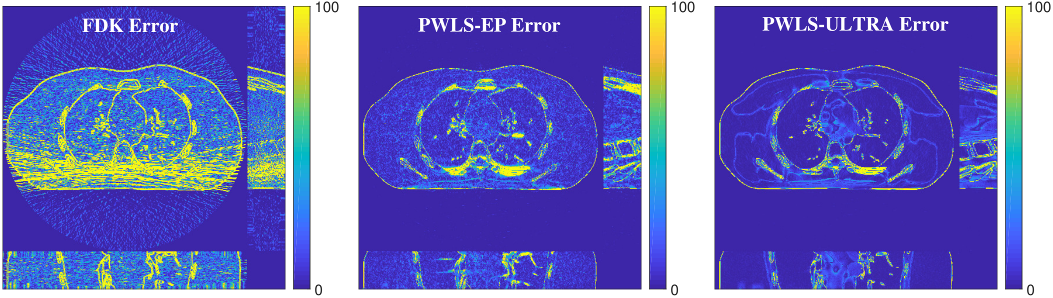
  Linearized augmented Lagrangian method (LALM) [8]

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

X. Zheng, S. Ravishankar,
Y. Long, JF:
IEEE T-MI, June 2018 [7]



3D X-ray CT simulation Error maps



	X-ray Intensity	FDK	EP	ST T_*	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

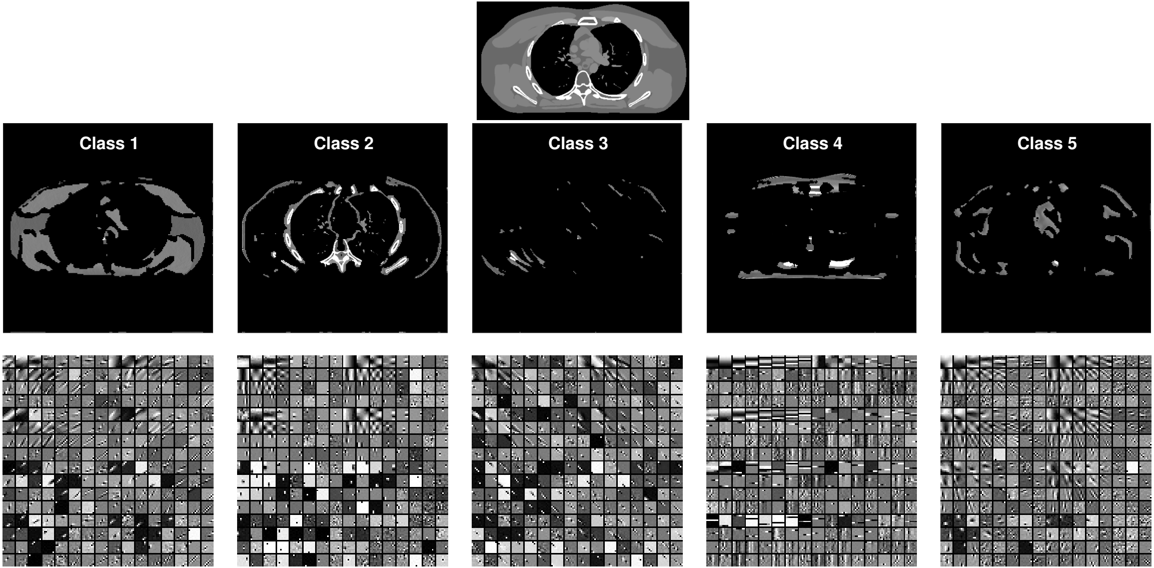
Union of Learned TRAnsforms (ULTRA)

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

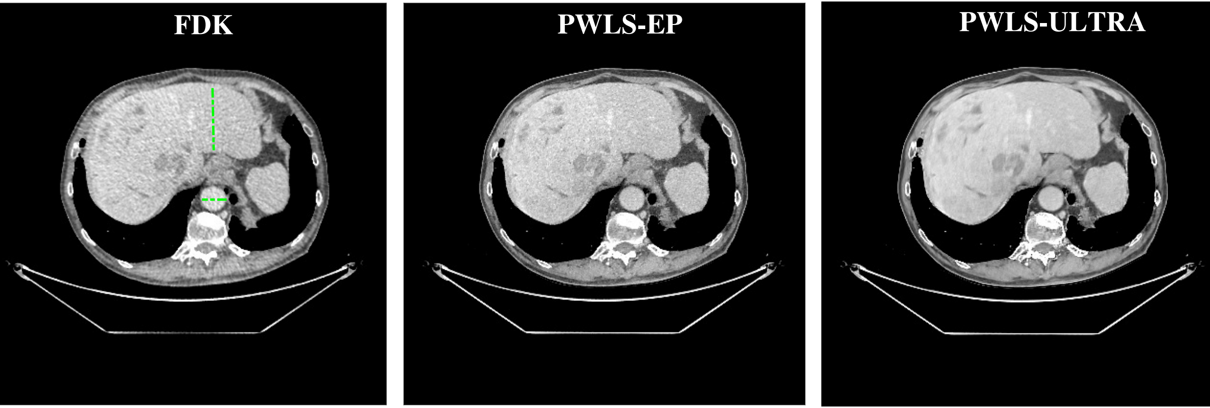
$$\left\{\hat{\boldsymbol{T}}_k\right\} = \underset{\{\boldsymbol{T}_k \text{ unitary}\}}{\arg\min} \, \min_{\{\boldsymbol{z}_{l,m}\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left(\min_{k \in \{1,\ldots,K\}} \left\| \boldsymbol{T}_k \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m} \right\|_2^2 + \alpha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

- Joint unsupervised clustering / sparsification
- Further nonconvexity due to clustering
- Efficient alternating minimization algorithm [9]

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 [7]

Matlab code: `http://web.eecs.umich.edu/~fessler/irt/reproduce/`

`https://github.com/xuehangzheng/PWLS-ULTRA-for-Low-Dose-3D-CT-Image-Reconstruction`

Outline

Introduction

Adaptive regularization

- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

Summary

Bibliography

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

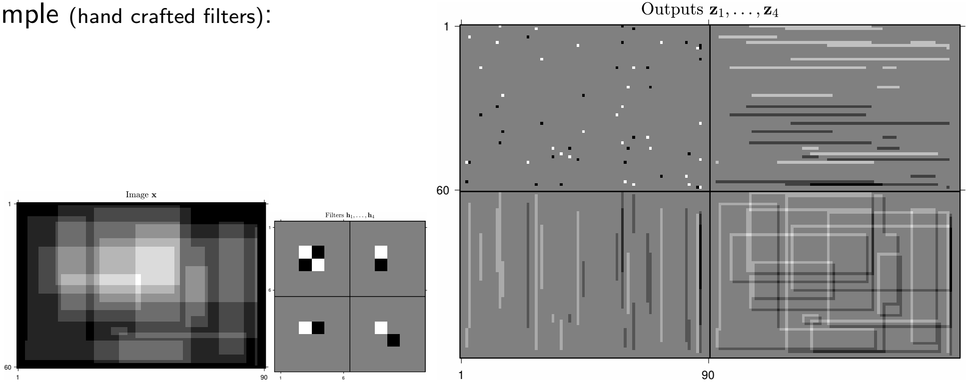
Drawback of basic patch-based methods:
$512 \times 512 \times 512$ 3D X-ray CT image volume
$8 \times 8 \times 8$ patches
$\Longrightarrow 512^3 \cdot 8^3 \cdot 4 = 256$ Gbyte of patch data for stride=1

Convolutional sparsity: analysis model

Assumption: For a plausible image $\boldsymbol{x}$, the filter outputs $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ are sparse, for some filters $\{\boldsymbol{h}_k\}_{k=1}^{K}$ [10]

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

Example (hand crafted filters):



Sparsifying filter learning (population adaptive)

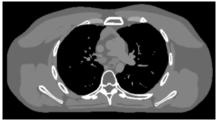
Given training images $\boldsymbol{x}_1, \dots, \boldsymbol{x}_L$ from a representative population, find filters $\left\{\hat{\boldsymbol{h}}_k\right\}_{k=1}^{K}$ that best sparsify them:

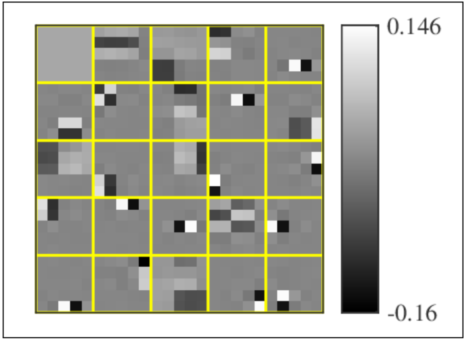
$$\left\{\hat{\boldsymbol{h}}_k\right\} = \mathop{\arg\min}_{\{\boldsymbol{h}_k\}\in\mathcal{H}} \min_{\{\boldsymbol{z}_{l,k}\}} \sum_{l=1}^{L} \sum_{k=1}^{K} \|\boldsymbol{h}_k * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}\|_2^2 + \alpha \|\boldsymbol{z}_{l,k}\|_0$$

- To encourage filter diversity:
  - $\mathcal{H} = \{\boldsymbol{H} : \boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}\}, \ \boldsymbol{H} = [\boldsymbol{h}_1 \ \dots \ \boldsymbol{h}_K]$
  - *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 [11]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - Filter update uses diagonal majorizer, proximal map (SVD)
  - Subsequence convergence guarantees [11]

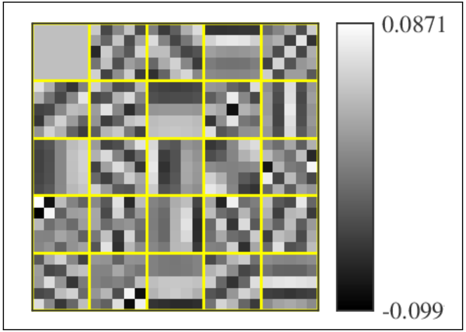
Examples of learned sparsifying filters

2D X-ray CT training data and learned 5 × 5 sparsifying filters $\left\{\hat{\boldsymbol{h}}_k\right\}$ [11]:





$\alpha = 10^{-4}$



$\alpha = 2\times10^{-3}$

Regularizer based on learned sparsifying filters

Regularized inverse problem [11]:

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

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

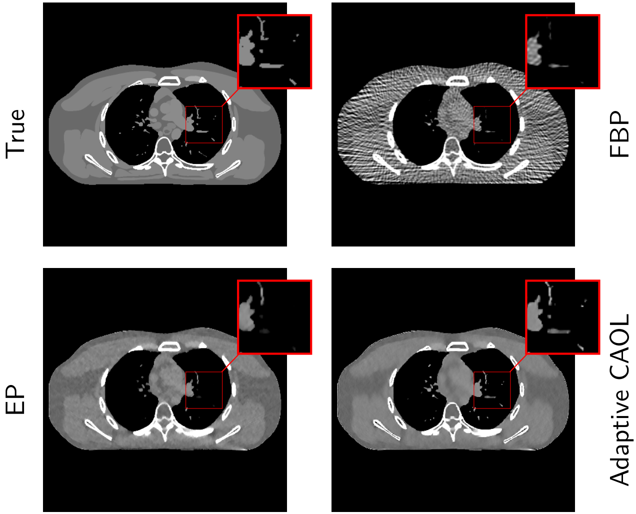
$\left\{ \hat{\boldsymbol{h}}_k \right\}$ adapted to population training data

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 [11]

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



123 views
(out of usual 984)
$\Longrightarrow 8\times$ dose reduction
25 filters 5×5

RMSE (in HU):

FBP	82.8
EP	40.8
Adaptive filters	**35.2**

- Physics / statistics provides dramatic improvement
- Data-adaptive regularization further reduces RMSE, improves fine details

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:

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

Extension to multiple layers (cf CNN) II

Learning 2 layers of filters [11]:

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

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 [11]

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

Outline

Introduction

Adaptive regularization

- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

Summary

Bibliography

MR with adapted patch dictionary

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

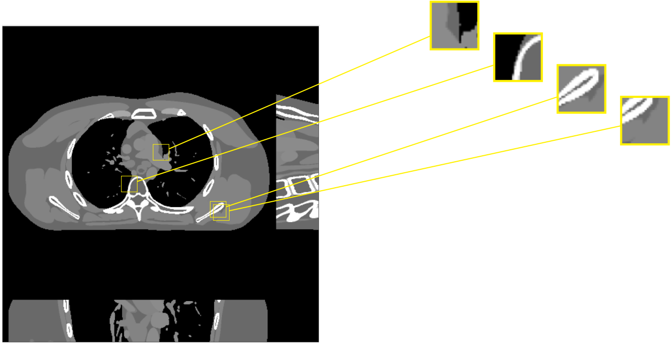
Patch-wise dictionary sparsity model

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

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

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

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



MR reconstruction using adaptive dictionary regularizer

Dictionary-blind MR image reconstruction:

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

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

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

In words: of the many images...

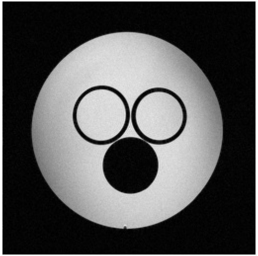
Alternating (nested) minimization:

- 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 [12].
- 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 [12].

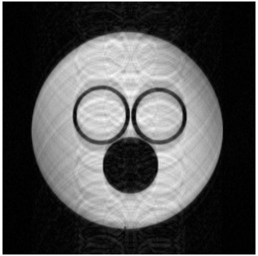
2D CS MRI results I

J. Fessler

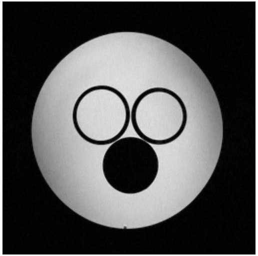
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI

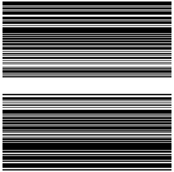


6×6 patches

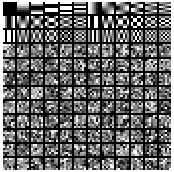
$\boldsymbol{D} \in \mathbb{C}^{6^2 \times 144}$

$\boldsymbol{D}_0$: [DCT | random] [12]

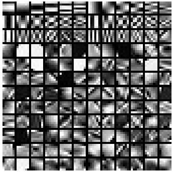
Sampling ($2.5\times$)



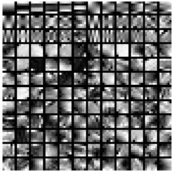
Initial $\boldsymbol{D}$



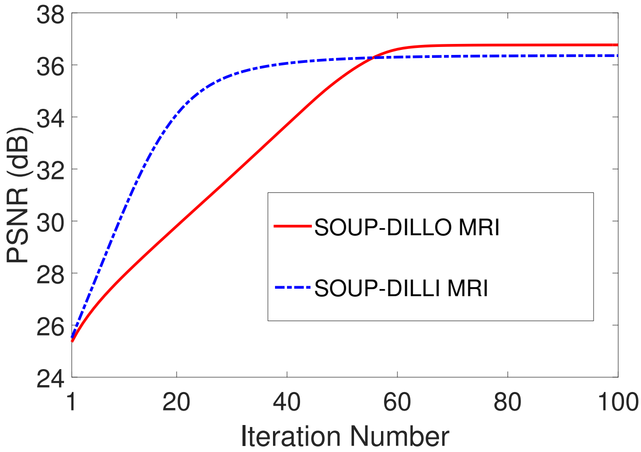
Learned real$\{\boldsymbol{D}\}$



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



2D CS MRI results II

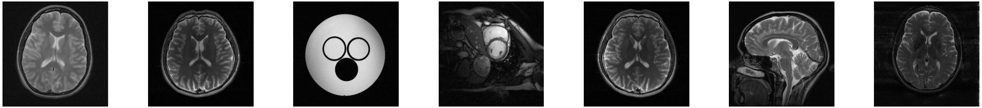


(SNR vs fully sampled image.)
Using $\|\boldsymbol{z}_m\|_0$ leads to higher SNR than $\|\boldsymbol{z}_m\|_1$.
Adaptive case is non-convex anyway...

Matlab code: `http://web.eecs.umich.edu/~fessler/irt/reproduce/`

`https://gitlab.eecs.umich.edu/fessler/soupdil_dinokat`

2D CS MRI results III

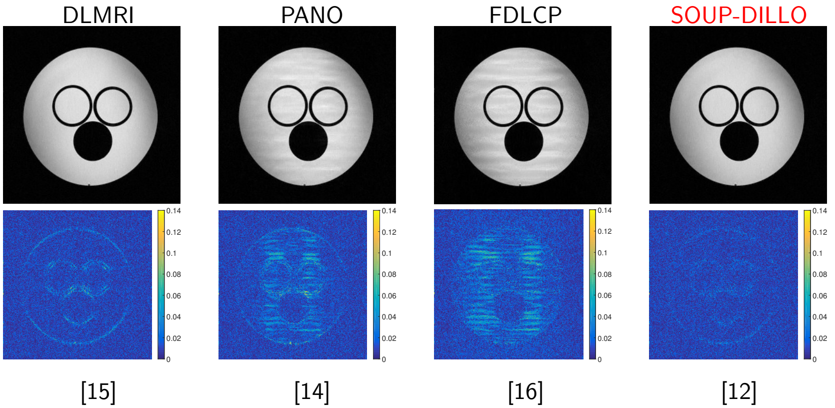


(a) (b) (c) (d) (e) (f) (g)

PSNR:

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.				[13]	[14]	[15]	[12]	[12]

2D CS MRI results IV



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

Summary of patch-based, data-driven adaptive regularizers

J. Fessler

UNIVERSITY OF MICHIGAN

Use training data to learn:

- dictionary $\boldsymbol{D}$ (for patches)
- sparsifying transform(s) $\boldsymbol{T}$ (for patches)
- or convolutional versions thereof [10, 17]

ML-based regularized optimization problem using M image patches:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R_{\mathrm{ML}}(\boldsymbol{x})$$

$$R_{\mathrm{ML-DL}}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{P}_m\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_m\|_2^2 + \alpha \|\boldsymbol{z}_m\|_0$$

$$R_{\mathrm{ML-ST}}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{T}\boldsymbol{P}_m\boldsymbol{x} - \boldsymbol{z}_m\|_2^2 + \alpha \|\boldsymbol{z}_m\|_0$$

Alternative: blind adaptive learned dictionary [15] or learned sparsifying transform [18].
Double minimization (so very "deep?") More interpretable than CNNs?

Outline

Introduction

Adaptive regularization

- Patch-based adaptive regularizers
- Convolutional adaptive regularizers
- Blind dictionary learning
- Supervised adaptive regularization

Summary

Bibliography

Convolutional sparsity revisted

J. Fessler

Cost function for convolutional sparsity regularization:

$$\arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \left(\min_{\zeta} \sum_{k=1}^K \frac{1}{2} \|\boldsymbol{h}_k * \boldsymbol{x} - \zeta_k\|_2^2 + \alpha \|\zeta_k\|_1 \right)$$

Alternating minimization updates:

$$\text{Sparse code:} \qquad \zeta_k^{(n+1)} = \mathrm{soft}\{\boldsymbol{h}_k * \boldsymbol{x}^{(n)}, \alpha\}$$

$$\text{Image: } \boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} F(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{z}^{(n)})$$

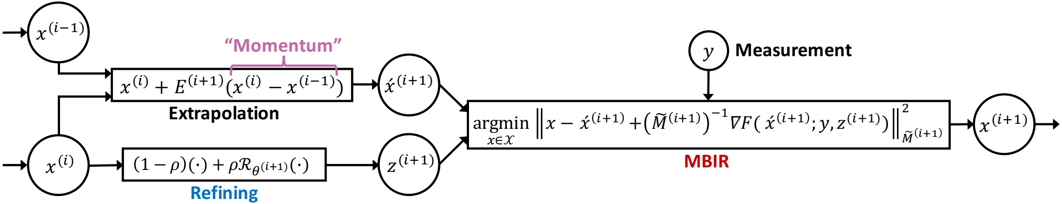
$$F(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{z}^{(n)}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \left(\sum_{k=1}^K \frac{1}{2} \left\|\boldsymbol{h}_k * \boldsymbol{x} - \zeta_k^{(n+1)}\right\|_2^2 + \alpha \left\|\zeta_k^{(n+1)}\right\|_1 \right)$$

$$= \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}^{(n)}\|_2^2 \quad \text{(quadratic but } \textit{large} \Longrightarrow \text{majorize)}$$

$$\boldsymbol{z}^{(n)} = \mathcal{R}(\boldsymbol{z}^{(n)}) = \sum_{k=1}^K \mathrm{flip}(\boldsymbol{h}_k) * \mathrm{soft}\{\boldsymbol{h}_k * \boldsymbol{x}^{(n)}\} \quad \text{(denoise} \Longrightarrow \text{learn)}$$

Momentum-Net overview

Unrolled loop network with momentum and quadratic majorizer:



- Diagonal majorizer: $\boldsymbol{M} = \operatorname{diag}\{\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A} \boldsymbol{1}\} + \beta \boldsymbol{I} \succeq \boldsymbol{A}' \boldsymbol{W} \boldsymbol{A} + \beta \boldsymbol{I}$
- Learn image mapper ("refiner") $\mathcal{R}$ from training data (supervised).
  cf CNN: filter $\rightarrow$ threshold $\rightarrow$ filter

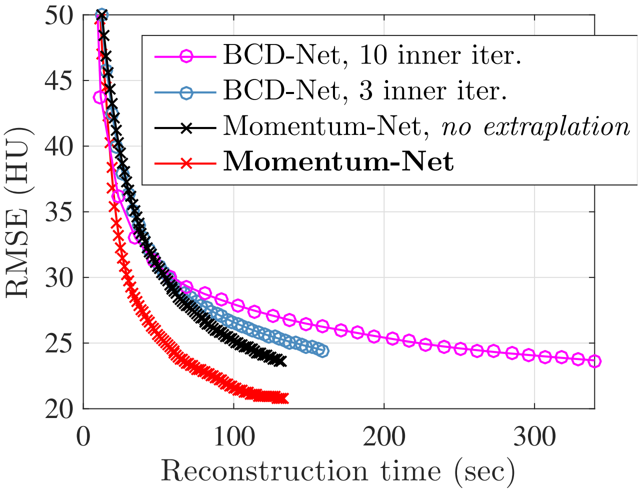
Momentum-Net benefits

- Image mapper $\mathcal{R}$ is shallow
  $\Longrightarrow$ less risk of over-fitting / hallucination
- Momentum accelerates convergence (fewer layers)
- First unrolled loop approach to have convergence theory
  (under suitable assumptions on $\mathcal{R}$)
- Image update uses original CT sinogram $\boldsymbol{y}$ and imaging physics $\boldsymbol{A}$

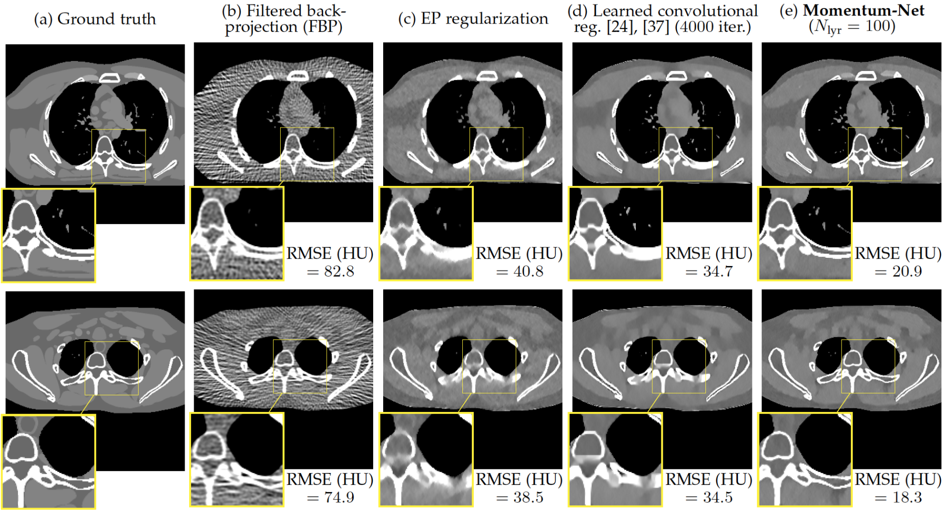
[19]
Il Yong Chun, Zhengyu Huang, Hongki Lim, J A Fessler
Momentum-Net: Fast and convergent iterative neural network for inverse problems
http://arxiv.org/abs/1907.11818

Momentum-Net preliminary results

Illustration of benefits of momentum:



Momentum-Net preliminary image results



Sparse-view CT with 123/984 views, $I_0 = 10^5$, 800-1200 mod. HU display.

J. Fessler

Introduction
Adaptive regularization
Summary
Bibliography

Summary

- CT image reconstruction has evolved greatly in the 50+ years since Allan Cormack's seminal papers [20, 21]
  - physics
  - statistics
  - regularization and optimization
  - data adaptive methods inspired by machine learning
- Machine learning has great potential for medical imaging
- Much excitement but many challenges
- Image reconstruction seems especially suitable for ML ideas
- Data-driven, adaptive regularizers beneficial for low-dose CT
- More comparisons between model-based methods with adaptive regularizers and CNN-based methods needed

Resources

Talk and code available online at
`http://web.eecs.umich.edu/~fessler`

[1] P. J. Pickhardt, M. G. Lubner, D. H. Kim, J. Tang, J. A. Ruma, A. Muñoz del Rio, and G-H. Chen. "Abdominal CT with model-based iterative reconstruction (MBIR): Initial results of a prospective Trial comparing ultralow-dose with standard-dose imaging." In: *Am. J. Roentgenol.* 199.6 (Dec. 2012), 1266–74.

[2] FDA. *510k premarket notification of AiCE Deep Learning Reconstruction (Canon)*. 2019.

[3] FDA. *510k premarket notification of Deep Learning Image Reconstruction (GE Medical Systems)*. 2019.

[4] S. H. Chan, X. Wang, and O. A. Elgendy. "Plug-and-play ADMM for image restoration: fixed-point convergence and applications." In: *IEEE Trans. Computational Imaging* 3.1 (Mar. 2017), 84–98.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] I. Y. Chun and J. A. Fessler. *Convolutional analysis operator learning: acceleration and convergence*. 2018.

[11] I. Y. Chun and J. A. Fessler. "Convolutional analysis operator learning: acceleration and convergence." In: *IEEE Trans. Im. Proc.* (2019). Submitted.

Bibliography II

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] I. Y. Chun and J. A. Fessler. “Convolutional dictionary learning: acceleration and convergence.” In: *IEEE Trans. Im. Proc.* 27.4 (Apr. 2018), 1697–712.

[18] S. Ravishankar and Y. Bresler. “Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to MRI.” In: *SIAM J. Imaging Sci.* 8.4 (2015), 2519–57.

[19] I. Y. Chun, Z. Huang, H. Lim, and J. A. Fessler. *Momentum-Net: Fast and convergent iterative neural network for inverse problems*. 2019.

[20] A. M. Cormack. “Representation of a function by its line integrals, with some radiological applications.” In: *J. Appl. Phys.* 34.9 (Sept. 1963), 2722–7.

[21] A. M. Cormack. “Representation of a function by its line integrals, with some radiological applications II.” In: *japplphy* 35.10 (Oct. 1964), 2908–13.