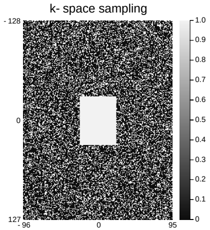
Data driven methods for image reconstruction in MRI (and CT)

Jeffrey A. Fessler

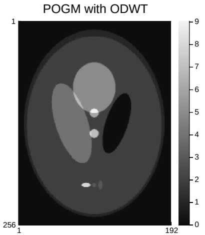
EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

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



SIAM Imaging Science 2020 Tutorial

2020-07-14



Key references

[1] http://doi.org/10.1109/JPROC.2019.2936204
Sai Ravishankar, Jong Chul Ye, J A Fessler. Image reconstruction: from sparsity to data-adaptive methods and machine learning. Proc. IEEE, 108(1):86-109, Jan. 2020.

[2] http://doi.org/10.1109/MSP.2019.2943645
J A Fessler. Optimization methods for MR image reconstruction.
IEEE Sig. Proc. Mag., 37(1):33-40, Jan. 2020.

[3] http://arxiv.org/abs/1907.11818
Il Yong Chun, Zhengyu Huang, Hongki Lim, J A Fessler. Momentum-Net: Fast and convergent iterative neural network for inverse problems. 2019.

Slides:
https://web.eecs.umich.edu/~fessler/papers/files/talk/20/siam.pdf

Overview

Outline

History

- Goal of accelerating MRI scans has a long history
  - gradient waveforms: echo-planar imaging [4], echo-volumar imaging [5], spiral imaging [6, 7], . . .
  - partial Fourier [8, 9, 10]
  - parallel imaging [11, 12, 13]
    sensitivity encoded imaging, initially with regular under-sampling [14, 15, 16], soon after with irregular / non-Cartesian sampling [17, 18]
  - Data-driven choice of phase encodes: Yue Cao & David Levin, 1993 [19]
    recent work optimizes sampling patterns: [20, 21, 22, 23, 24, 25, 26, 27, 28]
- Recent history (since 2006) driven by compressed sensing
  - Conf. papers 2004-2006 [29, 30, 31, 32]
  - Journal papers in 2007 [33, 34, 35, 36]
  - Review papers (!) in 2008 [37] [38]
  - FDA approval for CS in MRI 10 years later [39, 40, 41, 42] "HyperSense" "Compressed SENSE"

Every great idea has its precursors: Y. Bresler's group was solving $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0$ s.t. $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\| \leq \varepsilon$ in 1998 [43]

- static vs dynamic imaging
- single-coil vs multiple-coil data (parallel MRI)
- "SENSE" methods model coil sensitivities in the image domain
- "GRAPPA" methods model the effect of coil sensitivity in k-space
- "calibrationless" methods that use low-rank properties [45, 46, 47]
- clinical anatomical imaging vs quantitative [48, 49, 50, 51]
- smooth vs non-smooth cost functions

Code

- Jupyter notebooks with code in the language Julia [52] illustrate many of the methods shown in this tutorial
- Notebooks use the Michigan Image Reconstruction Toolbox in Julia (MIRT.jl) http://github.com/JeffFessler/MIRT.jl
- For demo notebooks, see: https://github.com/JeffFessler/mirt-demo
- Installation-less test drives using https://mybinder.org/
- Matlab version of MIRT: http://web.eecs.umich.edu/~fessler/code/index.html

Why Julia?

- adopts best ideas of Matlab and Python (among others),
- interactive yet built on a compiler (solves "2 language problem"), ...
- open source, ...
- package manager facilitates exact reproducible research.
- My Julia tutorial: https://www.ima.umn.edu/2019-2020/SW10.14-18.19/28302

From http://github.com/JeffFessler/MIRT.jl

Examples

You can test drive some jupyter notebooks in your browser without installing any local software by using the free service at https://mybinder.org/

- Introduction: launch binder
- MRI compressed sensing demo: launch binder

If binder is too slow for you, you can view static html versions:

- http://web.eecs.umich.edu/~fessler/demo/00-isbi.html
- http://web.eecs.umich.edu/~fessler/demo/01-odwt.html

You can also view the notebook code directly:

- demo/

Measurement model in MRI

RF pulses gradient waveforms tissue properties	$\rightarrow$	Bloch Eqn Nonlinear	$\rightarrow$	transverse magnetization $\boldsymbol{x}$	$\rightarrow$	$\boldsymbol{A}$ or $\boldsymbol{F}$ encoding (coil, phase, ...) Linear	$\rightarrow$	recorded signal $\boldsymbol{s}$

Simplified signal model:

$$\boldsymbol{s} = \boldsymbol{F}\boldsymbol{x}, \quad F_{ij} = \exp(-\imath 2\pi \vec{\nu}_i \cdot \vec{r}_j)$$

- $\boldsymbol{s} \in \mathbb{C}^M$ signal samples recorded by an ideal MR receive coil
- $\boldsymbol{x} \in \mathbb{C}^N$ discretized version of the transverse magnetization
- $\vec{\nu}_i$ k-space sample location of the ith sample (units cycles/cm)
- $\vec{r}_j$ spatial coordinates of the center of the jth pixel (units cm)
- $\boldsymbol{F} \in \mathbb{C}^{M \times N}$ Fourier encoding matrix

Classical MR image reconstruction

For Cartesian sampling $\boldsymbol{s} = \boldsymbol{F}\boldsymbol{x} = \texttt{fft}(\boldsymbol{x})$ and $\boldsymbol{F}^{-1}\,\boldsymbol{s} = \frac{1}{N}\boldsymbol{F}'\,\boldsymbol{s} = \texttt{ifft}(\boldsymbol{s})$

Prevalent in clinical MRI

Reconstruction is "interesting" when considering

- non-Cartesian sampling [56],
- "accelerated" scanning with $M < N$ samples [16, 35],
- non-Fourier effects like magnetic field inhomogeneity [57],
- multiple receive coils.

$$\boldsymbol{y}_l = \boldsymbol{s}_l + \boldsymbol{\varepsilon}_l, \qquad \boldsymbol{s}_l = \boldsymbol{F}\boldsymbol{C}_l\boldsymbol{x} \qquad l = 1, \ldots, L$$

- $\boldsymbol{y}_l \in \mathbb{C}^M$: noisy samples recorded by the lth of of L receive coils
- $\boldsymbol{C}_l$: $N \times N$ diagonal matrix containing the lth coil sensitivity pattern.

Combining yields a standard forward model in MRI:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_L \end{bmatrix} = \boldsymbol{y} = \underbrace{(\boldsymbol{I}_L \otimes \boldsymbol{F})\boldsymbol{C}}_{\boldsymbol{A}} \boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 \\ \vdots \\ \boldsymbol{C}_L \end{bmatrix} \tag{1}$$

- $\boldsymbol{A} \in \mathbb{C}^{ML \times N}$: system matrix
- $\boldsymbol{y} \in \mathbb{C}^{ML}$: measured k-space data
- $\boldsymbol{x} \in \mathbb{C}^N$: latent image
- $\boldsymbol{\varepsilon}$: complex white Gaussian noise [58]

Extensions consider other physics effects like relaxation and field inhomogeneity [55].

Goal: recover image $\boldsymbol{x}$ from data $\boldsymbol{y}$

Outline

When $ML \geq N$, linear model (1) is over-determined so use ordinary least-squares:

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{C}^N} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 = (\boldsymbol{A}'\boldsymbol{A})^{-1} \boldsymbol{A}'\boldsymbol{y}$$
$$= \left(\sum_{l=1}^L \boldsymbol{C}_l' \boldsymbol{F}' \boldsymbol{F} \boldsymbol{C}_l\right)^{-1} \left(\sum_{l=1}^L \boldsymbol{C}_l' \boldsymbol{F}' \boldsymbol{y}\right). \tag{2}$$

- For fully sampled Cartesian k-space data where $\boldsymbol{F}^{-1} = \frac{1}{N}\boldsymbol{F}'$, it simplifies to $\hat{\boldsymbol{x}} = \left(\sum_{l=1}^L \boldsymbol{C}_l' \boldsymbol{C}_l\right)^{-1} \left(\sum_{l=1}^L \boldsymbol{C}_l' \boldsymbol{F}^{-1} \boldsymbol{y}\right)$, optimal coil combination approach [13].
- For regularly under-sampled Cartesian data, $\boldsymbol{F}'\boldsymbol{F}$ has a simple block structure $\Longrightarrow$ SENSE reconstruction [16].

Smooth regularization

Regularization is essential for

- under-sampled problems ($ML < N$)
- poorly conditioned problems, *e.g.*, non-Cartesian sampling

Simplest form is quadratically regularized LS (RLS):

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{C}^N} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \frac{1}{2} \|\boldsymbol{T}\boldsymbol{x}\|_2^2, \tag{3}$$

- $\beta > 0$: regularization parameter
- $\boldsymbol{T}$: $K \times N$ matrix transform such as finite differences (spatial smoothness model).
  - The conjugate gradient (CG) algorithm is well-suited [17, 57]
  - The Hessian matrix $\boldsymbol{A}'\boldsymbol{A} + \beta\, \boldsymbol{T}'\boldsymbol{T}$ often is approximately Toeplitz [59]
    $\Longrightarrow$ CG with circulant preconditioning [60]
  - RLS is now passé, but CG often used as an inner step [61],
    even in recent deep learning methods for MRI reconstruction [62]

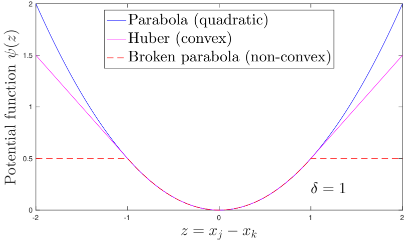
- Quadratic regularizer blurs edges when $\boldsymbol{T}$ is finite differences.
- Edge-preserving regularizer reduces such blur:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{C}^N} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta\, \psi(\boldsymbol{T}\boldsymbol{x}) \,. \tag{4}$$

ψ: nonquadratic potential function, typically convex and smooth

Huber function [63], hyperbola [64, 65],

Fair potential $\psi(z) = \delta^2 \left(|z/\delta| - \log(1 + |z/\delta|)\right)$



- Roots in Bayesian methods for Markov random fields [66, 67]

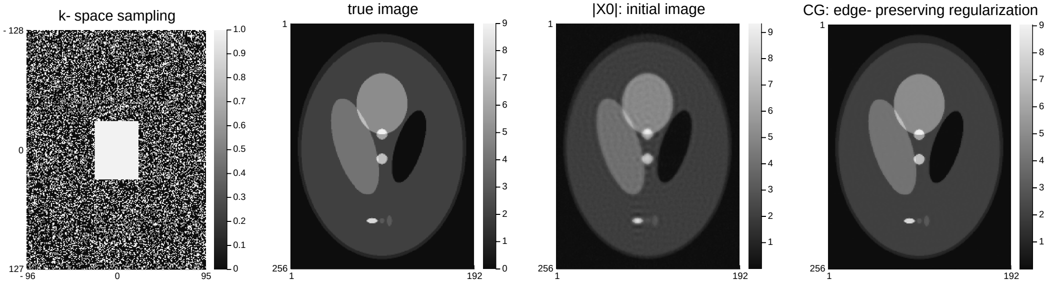
Algorithm options:

- Nonlinear CG algorithm
- 3MG (majorize-minimize memory gradient) algorithm [65]
- optimized gradient method (OGM) [68]

  optimal worst-case first-order method for convex cost functions with Lipschitz continuous gradients [69]

  OGM convergence rate bound $2\times$ better than Nesterov's fast gradient method [70]
- line-search OGM [71].

Edge-preserving regularization: example



ψ: Fair potential, $\delta = 0.1$

$\boldsymbol{T}$: finite differences

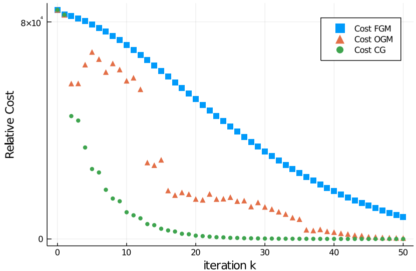
$=$ corner-rounded TV

Demo notebook: `01-recon`

`https://github.com/JeffFessler/mirt-demo`

Final NRMSE: 1.55%

(Cost function is locally strongly convex.)



Introduction

Brief review of classic methods ($>$ 10 years old)

Sparsity regularizers: Basic

- Sparsity models: synthesis form
- Proximal methods: ISTA/FISTA/POGM
- Sparsity models: analysis form

Sparsity regularizers: Advanced

Adaptive regularizers

Denoising-based "regularization"

Deep-learning approaches for image reconstruction

Looking forward

Bibliography

Received: 1 January 2020 | Revised: 28 April 2020 | Accepted: 30 April 2020
DOI: 10.1002/mrm.28338
FULL PAPER
Magnetic Resonance in Medicine

Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge

Florian Knoll[1] | Tullie Murrell[2] | Anuroop Sriram[2] | Nafissa Yakubova[2] | Jure Zbontar[2] | Michael Rabbat[2] | Aaron Defazio[2] | Matthew J. Muckley[1] | Daniel K. Sodickson[1] | C. Lawrence Zitnick[2] | Michael P. Recht[1]

“the winners … chose approaches that used a combination of a learned prior and a data-fidelity term that encodes information about the MR physics of the acquisition, in line with approaches that can be seen as neural network extensions of classic iterative image reconstruction methods” [72]

- Synthesis model:
  Assume $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{z}$
  $\boldsymbol{B}$: $N \times K$ matrix ("basis"), usually wide (over complete)
  $\boldsymbol{z} \in \mathbb{C}^K$ sparse coefficient vector
  $\Longrightarrow$ use $\|\boldsymbol{z}\|_1$

- Analysis model:
  Assumes $\boldsymbol{T}\boldsymbol{x}$ is sparse
  $\boldsymbol{T}$: $K \times N$ transformation matrix, usually tall
  $\Longrightarrow$ use $\|\boldsymbol{T}\boldsymbol{x}\|_1$
  Most likely used in recent FDA-approved CS methods.

- Equivalent if $\boldsymbol{B} = \boldsymbol{T}^{-1}$ (but usually both are non-square)
- optimization trade-off: $\|\boldsymbol{z}\|_1$ is easier than $\|\boldsymbol{T}\boldsymbol{x}\|_1$, but $K > N$

All models are wrong, but some models are useful...

Sparsity regularizers: Synthesis approach

- Typical compressed sensing cost function for a synthesis model:

$$\hat{\boldsymbol{x}} = \boldsymbol{B}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} = \underset{\boldsymbol{z}\in\mathbb{C}^K}{\arg\min} \, \frac{1}{2} \|\boldsymbol{ABz} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 \tag{5}$$

- 1-norm is convex relaxation of the ℓ_0 counting measure
  encourages coefficients $\hat{\boldsymbol{z}}$ to be sparse.
- Also known as the LASSO problem [73, 74]
- Numerous optimization algorithms
  proximal optimized gradient method (POGM) particularly effective
  (Taylor et al., 2017) [75, 76]

Outline

- Classical approach: iterative soft thresholding algorithm (ISTA) [77]
  aka proximal gradient method (PGM) and proximal forward-backward splitting [78],

$$\boldsymbol{z}_{k+1} = \operatorname{soft}\left(\boldsymbol{z}_k - \boldsymbol{D}^{-1} \boldsymbol{B}' \boldsymbol{A}' (\boldsymbol{A}\boldsymbol{B}\boldsymbol{z}_k - \boldsymbol{y}), \beta / \boldsymbol{d}\right), \tag{6}$$

  - soft thresholding : $\operatorname{soft}(z, c) = \operatorname{sign}(z) \max(|z| - c, 0)$
  - $\boldsymbol{D} = \operatorname{Diag}\{\boldsymbol{d}\}$: diagonal matrix satisfying $\boldsymbol{D} \succeq \boldsymbol{B}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{B}$ [79]

    Slight generalization of usual ISTA where $\boldsymbol{D} = L\boldsymbol{I}$ and $L = \|\boldsymbol{B}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{B}\|_2$.
- $O(1/k)$ convergence bound of ISTA is undesirably slow
- fast iterative soft thresholding algorithm (FISTA) [80, 81], has $O(1/k^2)$ bound
  aka fast proximal gradient method (FPGM)

Proximal optimized gradient method (POGM)

Composite cost function:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \underbrace{f(\boldsymbol{x})}_{\text{smooth}} + \underbrace{g(\boldsymbol{x})}_{\text{prox friendly}}$$

- Recent extension: proximal optimized gradient method (POGM)
- worst-case convergence bound about $2\times$ better than FISTA/FPGM [75, 76].
- ISTA / FISTA / POGM nearly equally simple to implement
  equivalent computation time per iteration, dominated by ∇f and prox_g
- POGM converges faster than FISTA empirically [82, 83, 84, 85],
  when combined with adaptive restart [86, 76].
- Active research area [87, 88, 85]

Initialize $\boldsymbol{w}_0 = \boldsymbol{x}_0$, $\theta_0 = 1$. Then `for` $k = 1:N$:

$$\theta_k = \begin{cases} \frac{1}{2}\left(1 + \sqrt{4\theta_{k-1}^2 + 1}\right), & k < N \\ \frac{1}{2}\left(1 + \sqrt{8\theta_{k-1}^2 + 1}\right), & k = N \end{cases} \qquad \gamma_k = \frac{1}{L} \frac{2\theta_{k-1} + \theta_k - 1}{\theta_k}$$

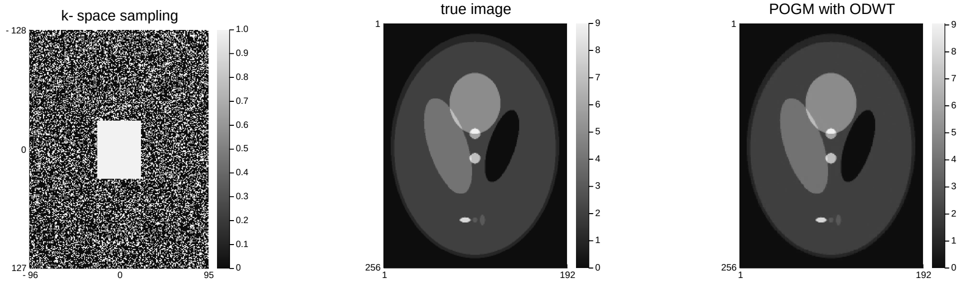
$$\boldsymbol{w}_k = \boldsymbol{x}_{k-1} - \frac{1}{L} \nabla f(\boldsymbol{x}_{k-1})$$

$$\boldsymbol{z}_k = \boldsymbol{w}_k + \frac{\theta_{k-1} - 1}{\theta_k}(\boldsymbol{w}_k - \boldsymbol{w}_{k-1}) + \frac{\theta_{k-1}}{\theta_k}(\boldsymbol{w}_k - \boldsymbol{x}_{k-1}) + \frac{\theta_{k-1} - 1}{L\gamma_{k-1}\theta_k}(\boldsymbol{z}_{k-1} - \boldsymbol{x}_{k-1})$$

$$\boldsymbol{x}_k = \operatorname{prox}_{\gamma_k g}(\boldsymbol{z}_k) = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}_k\|_2^2 + \gamma_k g(\boldsymbol{x})$$

POGM method [75] for minimizing $f(\boldsymbol{x}) + g(\boldsymbol{x})$ where f is convex with L-Lipschitz smooth gradient and g is convex. See [76] for adaptive restart version.

POGM illustrated



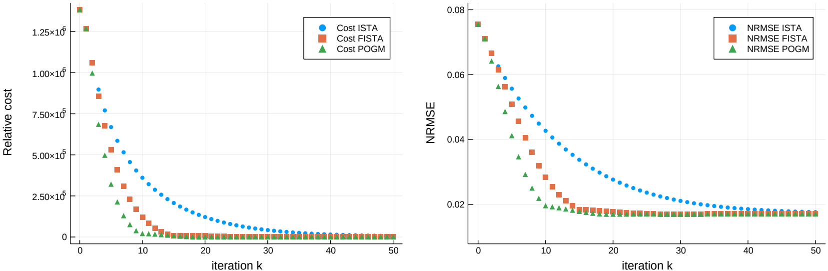
Demo:

`01-recon`

B: ODWT

Final NRMSE:

1.71%



- Typical optimization problem for analysis sparsity model:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 \quad (7)$$

  $\boldsymbol{T}$: sparsifying operator
  - wavelet transform
  - finite differences, aka total variation (TV) [33]
  - both [38]
- FDA-approved methods for compressed sensing MRI presumably related to (7).
- The analysis optimization problem (7) is harder than the synthesis form (5) due to the matrix $\boldsymbol{T}$ within 1-norm.

Proximal methods for analysis regularizer

PGM for analysis regularizer problem (7):

$$\tilde{\boldsymbol{x}}_k \triangleq \boldsymbol{x}_k - \frac{1}{L}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}) \quad \text{(gradient step)}$$

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{L}{2}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}_k\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 = \operatorname{prox}_{\frac{\beta}{L}\|\boldsymbol{T}\cdot\|_1}(\tilde{\boldsymbol{x}}_k) \tag{8}$$

$L = \|\boldsymbol{A}\|_2^2$: Lipschitz constant

- No simple solution for the proximity operator
- Inner iterative methods are required, typically involving dual formulations [89, 81]
- Perhaps main drawback of analysis regularization
- $\Longrightarrow$ PGM / FPGM / POGM less attractive for (7)

Circumventing the challenge of the matrix $\boldsymbol{T}$ in the 1-norm:

- Corner rounding

  Use smooth approximation: $|z| \approx \sqrt{|z|^2 + \epsilon}$
  - Akin to edge-preserving regularization
  - Proximal operators have no thresholding effect that induces sparsity
- Penalty approach

  Replace (7) with the following alternative [90]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R_\alpha(\boldsymbol{x}), \quad R_\alpha(\boldsymbol{x}) = \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \alpha \|\boldsymbol{z}\|_1 .$$

  - Fact: $R_\alpha(\boldsymbol{x}) = \psi(\boldsymbol{T}\boldsymbol{x}, \alpha)$ where ψ is the Huber function $\Longrightarrow$ corner rounding
  - Requires choosing parameter α.
- Iterative reweighted least-squares (FOCUSS) [36], akin to corner rounding

Variable splitting: AL / ADMM I

Replace $\|\boldsymbol{T}\boldsymbol{x}\|_1$ in (7) with exactly equivalent constrained minimization problem involving auxiliary variable(s):

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z} \,:\, \boldsymbol{z}=\boldsymbol{T}\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 . \quad (9)$$

- split Bregman algorithm [91]
- augmented Lagrangian (AL) methods [92, 61]
- alternating direction multiplier method (ADMM) [93, 94]
- Douglas-Rachford splitting method

Corresponding augmented Lagrangian for constrained problem for (9):

$$L(\boldsymbol{x}, \boldsymbol{z}; \gamma, \mu) = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 + \text{real}\{\langle \gamma, \boldsymbol{T}\boldsymbol{x} - \boldsymbol{z} \rangle\} + \frac{\mu}{2} \|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z}\|_2^2,$$

$\gamma \in \mathbb{C}^K$: Lagrange multipliers
$\mu > 0$: AL penalty parameter (affects the convergence rate, not $\hat{\boldsymbol{x}}$)
Scaled dual variable: $\boldsymbol{\eta} \triangleq (1/\mu)\gamma \Longrightarrow$ scaled augmented Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\eta}, \mu) = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 + \frac{\mu}{2} \left(\|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{\eta}\|_2^2 - \|\boldsymbol{\eta}\|_2^2 \right).$$

AL approach alternates between

- descent updates of primal variables $\boldsymbol{x}$, $\boldsymbol{z}$,
- ascent update of scaled dual variable $\boldsymbol{\eta}$.

$$L(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\eta}, \mu) = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 + \frac{\mu}{2} \left(\|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{\eta}\|_2^2 - \|\boldsymbol{\eta}\|_2^2 \right).$$

$\boldsymbol{z}$ update is simply soft thresholding:

$$\boldsymbol{z}_{k+1} = \text{soft}(\boldsymbol{T}\boldsymbol{x}_k + \boldsymbol{\eta}_k, \beta/\mu).$$

$\boldsymbol{x}$ update minimizes a quadratic function (use PCG):

$$\boldsymbol{x}_{k+1} = (\boldsymbol{A}'\boldsymbol{A} + \mu \boldsymbol{T}'\boldsymbol{T})^{-1}(\boldsymbol{A}'\boldsymbol{y} + \mu \boldsymbol{T}'(\boldsymbol{z}_{k+1} - \boldsymbol{\eta}_k)).$$

$\boldsymbol{\eta}$ update is ascent:

$$\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k + (\boldsymbol{T}\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1}).$$

- Adaptive methods for tuning μ [94, 95, 96].
- Parallel ADMM updates are also possible [97, 98].

AL/ADMM for parallel MRI I

Properties of $\boldsymbol{A} = \boldsymbol{F}_L \boldsymbol{C}$, $\boldsymbol{F}_L \triangleq \boldsymbol{I}_L \otimes \boldsymbol{F}$:
$\boldsymbol{F}'\boldsymbol{F}$: circulant (for Cartesian sampling) or Toeplitz (for non-Cartesian sampling)
$\boldsymbol{C}_l$: diagonal coil sensitivity matrix
Alternative variable split with simpler updates [61]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{C}^N} \min_{\boldsymbol{u} \in \mathbb{C}^{NL} : \boldsymbol{u} = \boldsymbol{C}\boldsymbol{x}} \min_{\boldsymbol{z} \in \mathbb{C}^K : \boldsymbol{z} = \boldsymbol{T}\boldsymbol{v}} \min_{\boldsymbol{v} \in \mathbb{C}^N : \boldsymbol{v} = \boldsymbol{x}} \frac{1}{2} \left\| \boldsymbol{F}_L \boldsymbol{u} - \boldsymbol{y} \right\|_2^2 + \beta \left\| \boldsymbol{z} \right\|_1 , \qquad (10)$$

$\boldsymbol{z}$ update : soft thresholding
$\boldsymbol{x}$ update : involves diagonal matrix $\boldsymbol{C}'\boldsymbol{C}$
$\boldsymbol{v}$ update : involves $\boldsymbol{T}'\boldsymbol{T}$ that is circulant or Toeplitz $\Longrightarrow$ use FFT or DCT
$\boldsymbol{u}$ update : involves $\boldsymbol{F}_L'\boldsymbol{F}_L$ that is circulant or Toeplitz.

- Primary drawback: more AL penalty parameters; (condition number criteria can help [61]).
- Many variations, *e.g.*, [98] [99].

A key idea behind duality-based methods:

$$\|\boldsymbol{T}\boldsymbol{x}\|_1 = \max_{\boldsymbol{z}\in\mathcal{Z}} \operatorname{real}\{\langle \boldsymbol{z},\ \boldsymbol{T}\boldsymbol{x}\rangle\}, \qquad \mathcal{Z} \triangleq \left\{\boldsymbol{z} \in \mathbb{C}^K : \|\boldsymbol{z}\|_\infty \leq 1\right\}.$$

Thus the analysis regularized problem (7) is equivalent to constrained problem:

$$\arg\min_{\boldsymbol{x}} \max_{\boldsymbol{z}\in\mathcal{Z}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \operatorname{real}\{\langle \boldsymbol{z},\ \boldsymbol{T}\boldsymbol{x}\rangle\}. \tag{11}$$

Primal-dual methods typically alternate between updating primal variable $\boldsymbol{x}$ and dual variable $\boldsymbol{z}$, using more convenient alternatives to (11).

- Separate multiplication by $\boldsymbol{A}$ and by $\boldsymbol{A}'$ without requiring inner CG iterations.
- Convergence guarantees and $O(1/k^2)$ rates [100, 101, 99, 102, 103, 104, 105, 106, 107].
- Require running a power iteration to find a Lipschitz constant.
- Like AL methods, have algorithm parameters to tune for convergence rate.

Outline

Instead of a single latent image $\boldsymbol{x}$, define a latent image for each coil $\boldsymbol{x}_l \triangleq \boldsymbol{C}_l \boldsymbol{x}$.
Measurement model: $\boldsymbol{y}_l = \boldsymbol{F}\boldsymbol{x}_l + \boldsymbol{\varepsilon}_l$.
Reconstruct L images $\boldsymbol{X} = [\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_L]$ from measurements $\boldsymbol{Y} = [\boldsymbol{y}_1 \ \ldots \ \boldsymbol{y}_L] \in \mathbb{C}^{M \times L}$.
Account for some relationships between those L images.
Multiplication by the smooth sensitivity map $\boldsymbol{C}_l$ in the image domain
$\Longrightarrow$ convolution with a small kernel in the frequency domain
$\Longrightarrow$ any point in k-space $\approx$ a linear combination of its neighbors
$\Longrightarrow$ GRAPPA consistency condition [18] $\mathrm{vec}(\boldsymbol{X}) \approx \boldsymbol{G}\,\mathrm{vec}(\boldsymbol{X})$
$\boldsymbol{G}$ involves small k-space kernels learned from calibration k-space data
"SPIRiT" [108] optimization problems like:

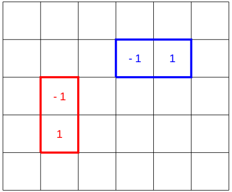
$$\hat{\boldsymbol{X}} = \mathop{\arg\min}_{\boldsymbol{X} \in \mathbb{C}^{N \times L}} \frac{1}{2} \|\boldsymbol{F}\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{Frob}}^2 + \beta_1 \frac{1}{2} \|(\boldsymbol{G} - \boldsymbol{I})\,\mathrm{vec}(\boldsymbol{X})\|_2^2 + \beta_2 R(\boldsymbol{X}),$$

$R(\boldsymbol{X})$: regularizer that encourages joint image sparsity [109].
Needs no sensitivity maps; use CG [108] or ADMM [110, 111].

Subspace and joint sparsity approaches go further by circumventing finding the calibration matrix $\boldsymbol{G}$ [112, 45, 46, 113]

Patch-based regularization

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

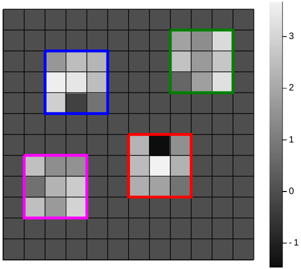


Larger patches provide more context
for distinguishing signal from noise.

cf. CNN approaches

Patch-based regularizers:

- synthesis models
- analysis methods



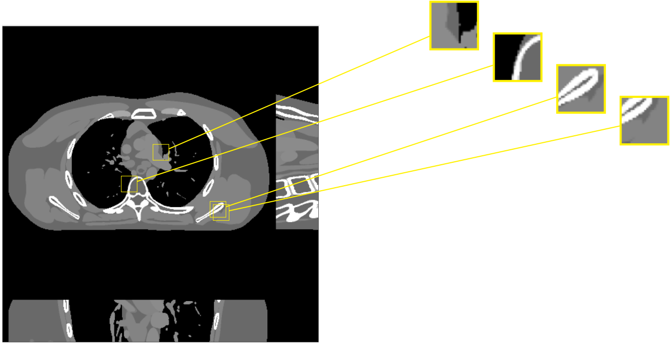
Patch-wise dictionary sparsity model

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



Patch-based regularization: synthesis approach

Patch synthesis model uses sparse linear combination of patch atoms: $\boldsymbol{P}_p \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_p$

$\boldsymbol{P}_p \in \{0,1\}^{d \times N}$: extracts pth of P d-pixel patches from image $\boldsymbol{x}$

$\boldsymbol{D} \in \mathbb{C}^{d \times J}$: dictionary of J patch atoms

$\boldsymbol{z}_p \in \mathbb{C}^{J}$: sparse coefficient vector for pth patch.

Natural regularizer for patch synthesis sparsity model [114]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{D} \boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

Use alternating minimization algorithms for optimization:

- $\boldsymbol{x}$ update is quadratic (PCG),
- $\boldsymbol{z}_p$ update is sparse coding (POGM).
- Basic approach would require storing JN coefficient values $\{\boldsymbol{z}_p\}$.

Patch analysis model assumes: $\boldsymbol{T}\boldsymbol{P}_p\boldsymbol{x}$ tends to be sparse
$\boldsymbol{P}_p \in \{0,1\}^{d \times N}$: extracts pth of P d-pixel patches from image $\boldsymbol{x}$
$\boldsymbol{T} \in \mathbb{C}^{K \times d}$: patch sparsifying transform (*e.g.*, DCT)
Natural regularizer for patch analysis sparsity model [115]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{T}\boldsymbol{P}_p\boldsymbol{x} - \boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

Use alternating minimization algorithms for optimization:

- $\boldsymbol{x}$ update is quadratic (PCG),
  Hessian w.r.t. $\boldsymbol{x}$: $\boldsymbol{A}'\boldsymbol{A} + \beta \sum_p \boldsymbol{P}_p' \boldsymbol{T}' \boldsymbol{T}\boldsymbol{P}_p$, simplifies when $\boldsymbol{T}$ is unitary
- $\boldsymbol{z}_p$ update is simply soft thresholding (easier than sparse coding!)
- Efficient implementation uses accumulator: $\tilde{\boldsymbol{x}}_t \triangleq \sum_{p=1}^{P} \boldsymbol{P}_p' \boldsymbol{T}' \operatorname{soft}(\boldsymbol{T}\boldsymbol{P}_p\boldsymbol{x}_t, \alpha)$,
  avoids storing all KN coefficient values $\{\boldsymbol{z}_p\}$.

Preview of unrolled loop NN methods

Patch analysis sparsity formulation reiterated:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{T}\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

Alternating minimization, aka two-block block coordinate descent (BCD):

- $\boldsymbol{x}$ update : quadratic
- $\boldsymbol{z}_p$ update : soft thresholding of transform coefficients

Combining:

$$\tilde{\boldsymbol{x}}_t = \sum_{p=1}^{P} \boldsymbol{P}_p' \boldsymbol{T}' \operatorname{soft}(\boldsymbol{T}\boldsymbol{P}_p \boldsymbol{x}_t, \alpha) \qquad \text{denoising, makes “prior”}$$

$$\boldsymbol{x}_{t+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \frac{1}{2} \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_t\|_2^2 \qquad \text{physics / k-space data.}$$

Cost function decreases monotonically here

May replace denoising step with NLM, BM3D, CNN, ...

Outline

Convolutional sparsity: synthesis model

Close relative of patch models is convolutional sparsity models [116, 117, 118].
Convolutional synthesis model assumes $\boldsymbol{x} \approx \sum_{k=1}^{K} \boldsymbol{h}_k * \boldsymbol{z}_k$

- $\boldsymbol{h}_k$: filters (typically learned from training data) [119]
- $\boldsymbol{z}_k$: sparse coefficient images (requires NK storage, challenging in 3D or dynamic)

Corresponding regularizer :

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_k\}} \frac{1}{2} \left\| \boldsymbol{x} - \sum_{k=1}^{K} \boldsymbol{h}_k * \boldsymbol{z}_k \right\|_2^2 + \alpha \|\boldsymbol{z}_k\|_1 .$$

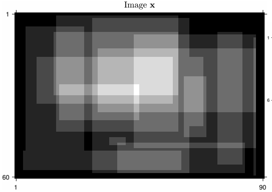
Use alternating minimization:

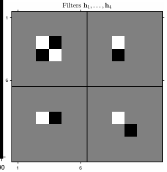
- $\boldsymbol{x}$ update : quadratic (PCG)
- $\boldsymbol{z}_k$ update : sparse coding (POGM)

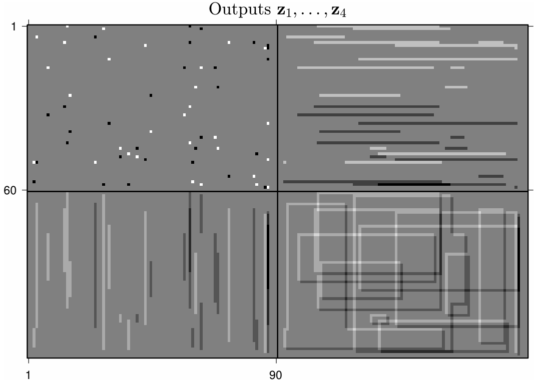
Assumption: For a plausible image $\boldsymbol{x}$, the filter outputs $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ are sparse, for some filters $\{\boldsymbol{h}_k\}_{k=1}^K$ [117]

- 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):







Convolutional sparsity: analysis model

Convolution analysis model assumes $\boldsymbol{h}_k * \boldsymbol{x}$ is sparse. Corresponding regularizer:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_k\}} \sum_{k=1}^{K} \frac{1}{2} \|\boldsymbol{h}_k * \boldsymbol{x} - \boldsymbol{z}_k\|_2^2 + \alpha \|\boldsymbol{z}_k\|_1 .$$

Use alternating minimization algorithm [116]:

- $\boldsymbol{x}$ update : quadratic (GD or PCG),
- $\boldsymbol{z}_k$ update : soft thresholding.
- Implement efficiently without storing NK sparse coefficients:

$$\tilde{\boldsymbol{x}}_t = \sum_{k=1}^{K} \boldsymbol{h}_k^{(-)} * \mathrm{soft}(\boldsymbol{h}_k * \boldsymbol{x}_{t-1}) \qquad \text{(denoising, accumulator)}$$

$$\boldsymbol{x}_{t+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \frac{1}{2} \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_t\|_2^2 \qquad \text{(physics)}$$

Summary of sparsity

Image models / cost functions:

- Image sparse synthesis (*e.g.*, wavelet basis)
- Image transform/analysis sparsity
  (*e.g.*, finite differences for total variation, wavelets, ...)
- Patch dictionary sparsity
- Patch transform sparsity
- Convolutional sparsity

Algorithm components:

- gradient-based (CG)
- proximal operations (soft/d thresholding)
- alternating minimization
- duality (AL, ADMM, primal-dual, ...)

Outline

Adaptive regularization

- Key required component for sparsity models:
  - $\boldsymbol{D}$ dictionary for patch-based synthesis model
  - $\boldsymbol{T}$ transform for patch-based analysis model
  - $\{\boldsymbol{h}_k\}$ filters for convolutional models
- Options:
  - Hand-crafted mathematical models, *e.g.*, discrete cosine transform (DCT)
  - Population adaptive approach:
    learn model from "good quality" (*e.g.*, fully sampled) representative training data
  - Patient adaptive ("blind") approach:
    adapt model for each patient by jointly optimizing image $\boldsymbol{x}$,
    sparse coefficients $\boldsymbol{Z}$ and model component $(\boldsymbol{D}$ or $\boldsymbol{T}$ or $\{\boldsymbol{h}_k\})$ [114, 115]
  - Hybrids of population and patient adaptive approaches [120]
- For adaptive schemes, constraints on the learned components are essential.

June 2018 special issue of IEEE Trans. on Medical Imaging [121]:

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*

Properties of a convex regularizer

When learning the regularizer, convex functions allow most theory and avoid local minima!

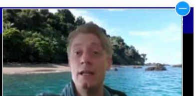
In the interpretation $R(u) = -\log(p(u))$, a convex regularizer makes no sense for images!

0.5  $+0.5$  $=$ 

$$\frac{1}{2}R(u_1) + \frac{1}{2}R(u_2) \geq R\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right)$$

Needs to have lower R(u) than one of the other images!

July 10 2020 Daniel Cremers, Michael Möller



Adaptive regularization options

- Data
  - Population adaptive methods
  - 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

X-ray CT with learned sparsifying transforms

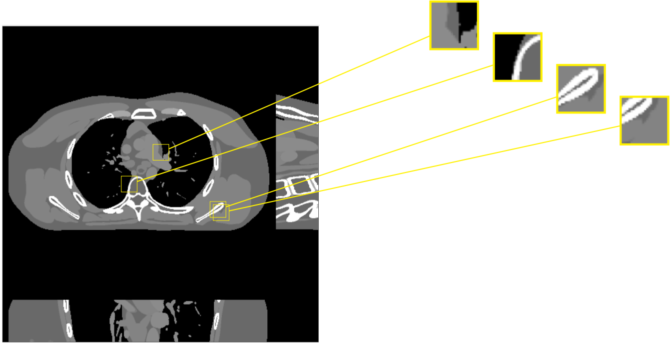
- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

Patch-wise transform sparsity model

Assumption: if $\boldsymbol{x}$ is a plausible image, then each patch transform $\boldsymbol{T}\boldsymbol{P}_m\boldsymbol{x}$ 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)

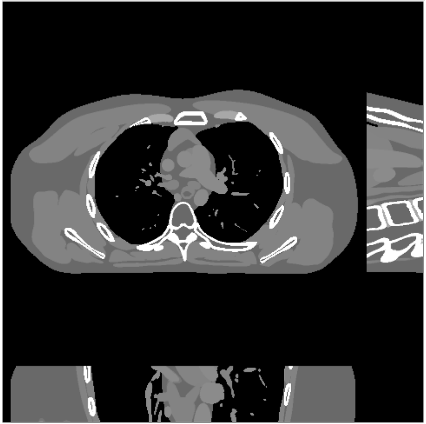
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find transform $\boldsymbol{T}_*$ that best sparsifies their patches:

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

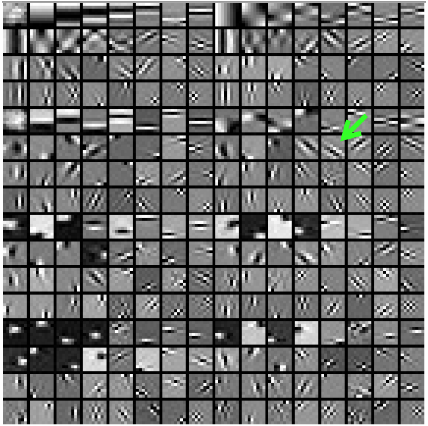
- Encourage aggregate sparsity, not patch-wise sparsity (cf K-SVD [122])
- Non-convex due to unitary constraint and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [123]
  - $\boldsymbol{z}$ update : simple hard thresholding
  - $\boldsymbol{T}$ update : orthogonal Procrustes problem (SVD)
  - Subsequence convergence guarantees [123]

Example of learned sparsifying transform

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 [124]:

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

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

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

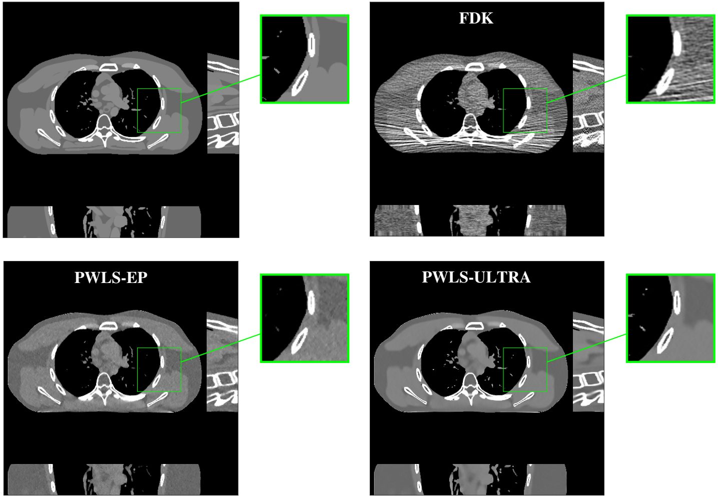
Alternating minimization optimizer:

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

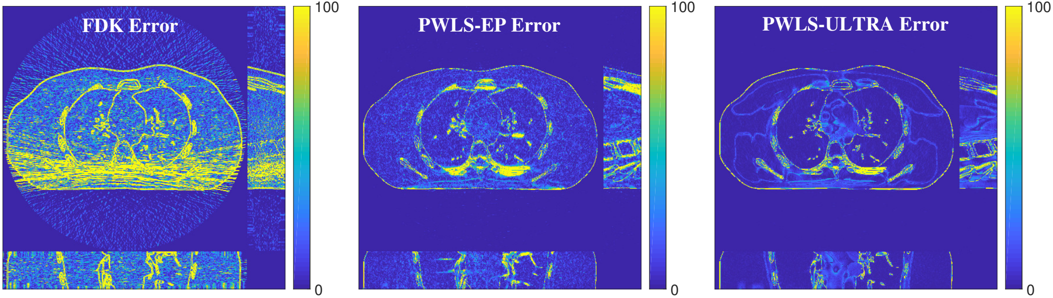
  Linearized augmented Lagrangian method (LALM) [125]

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

X. Zheng, S. Ravishankar, Y. Long, JF:
IEEE T-MI, June 2018 [124].



3D X-ray CT simulation Error maps



	X-ray Intensity	FDK	EP	ST $\boldsymbol{T}_*$	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**

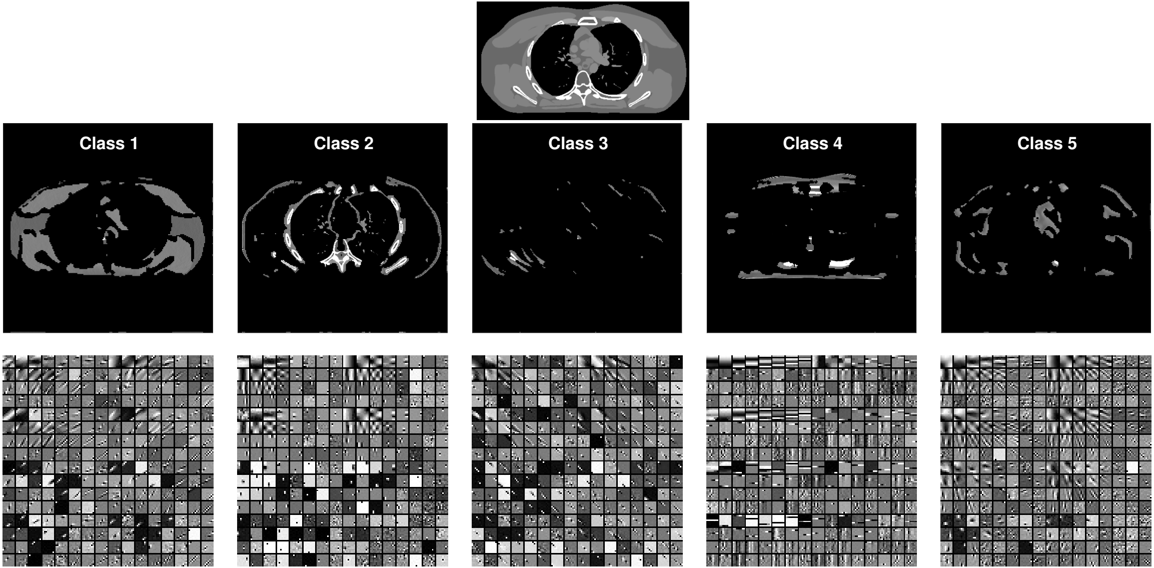
- Physics / statistics provides dramatic improvement
- Data adaptive regularization further reduces RMSE

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:

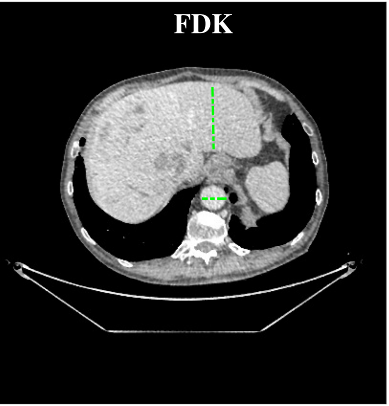
$$\left\{\hat{\boldsymbol{T}}_k\right\} = \mathop{\arg\min}_{\{\boldsymbol{T}_k \text{ unitary}\}} \min_{\{\boldsymbol{z}_{l,m}\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left(\min_{k \in \{1,\ldots,K\}} \left\| \boldsymbol{T}_k \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m} \right\|_2^2 + \alpha \left\|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

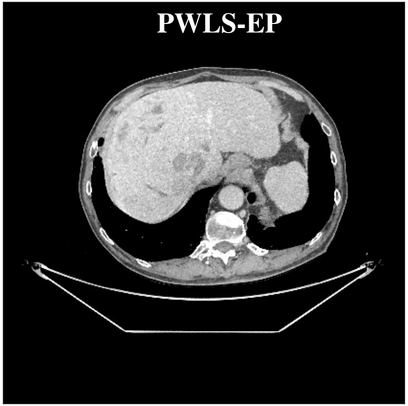
- Joint unsupervised clustering / sparsification
- Further nonconvexity due to clustering
- Efficient alternating minimization algorithm [126]

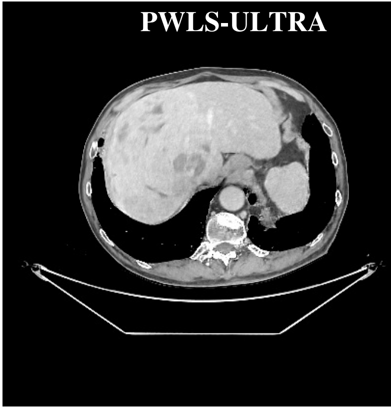
Example: 3D X-ray CT learned set of transforms



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







Zheng et al., IEEE T-MI, June 2018 [124] (Special issue on machine learning for image reconstruction)

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

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

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

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.

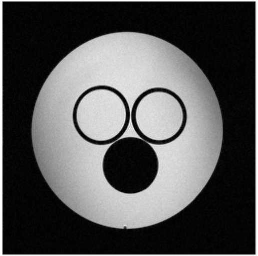
In words: of the many images...

Alternating (nested) minimization:

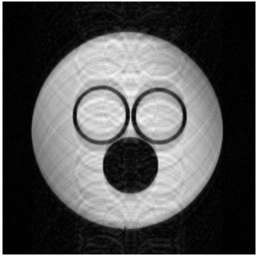
- Fixing $\boldsymbol{x}$ and $\boldsymbol{D}$, update each row of $\boldsymbol{Z} = [\boldsymbol{z}_1 \ \ldots \ \boldsymbol{z}_M]$ sequentially via hard-thresholding.
- Fixing $\boldsymbol{x}$ and $\boldsymbol{Z}$, update $\boldsymbol{D}$ using SOUP-DIL [127].
- Fixing $\boldsymbol{Z}$ and $\boldsymbol{D}$, 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 due to $\mathcal{D}$, $\boldsymbol{D}\boldsymbol{z}_m$, 0-norm, but monotone decreasing and some convergence theory [127].

2D CS MRI results I

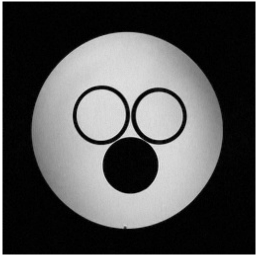
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI



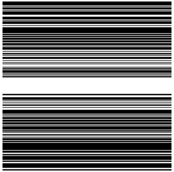
6×6 patches

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

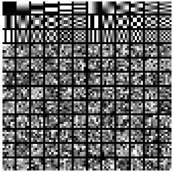
$\boldsymbol{D}_0$: [DCT | random]

[127]

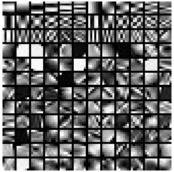
Sampling ($2.5\times$)



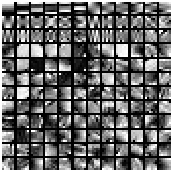
Initial $\boldsymbol{D}$



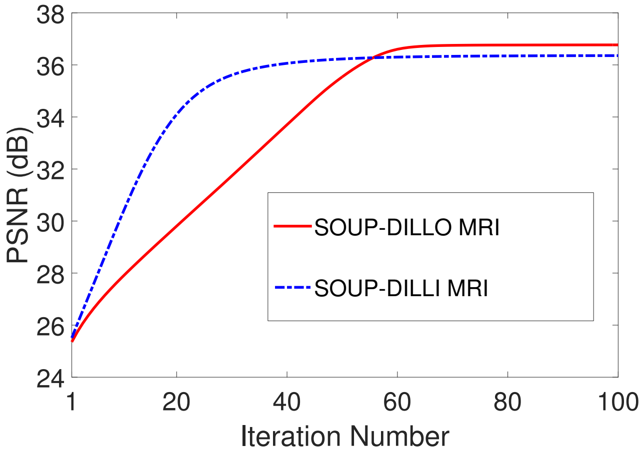
Learned real$\{\boldsymbol{D}\}$



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



2D CS MRI results II

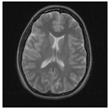


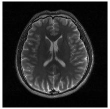
(SNR vs fully sampled image.)
Using $\|\boldsymbol{z}_m\|_0$ leads to higher SNR than $\|\boldsymbol{z}_m\|_1$.
Adaptive case is non-convex anyway...

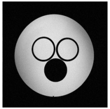
Matlab code: http://web.eecs.umich.edu/~fessler/irt/reproduce/

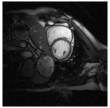
https://gitlab.eecs.umich.edu/fessler/soupdil_dinokat

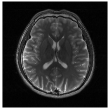
2D CS MRI results III


(a)

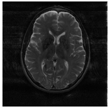

(b)


(c)


(d)


(e)


(f)

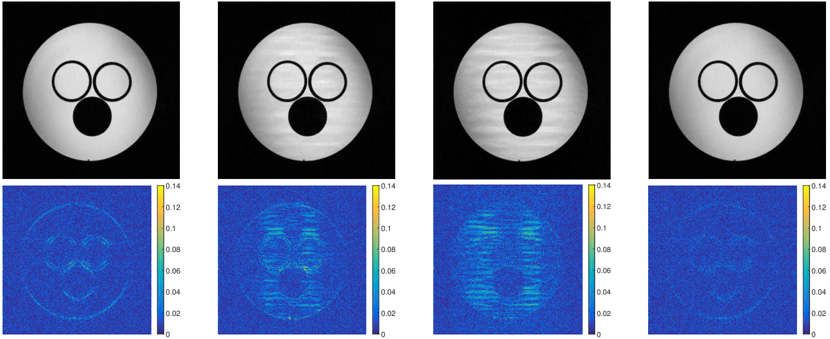

(g)

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.				[108]	[128]	[114]	[127]	[127]

2D CS MRI results IV

DLMRI PANO FDLCP SOUP-DILLO



[114] [128] [129] [127]

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

Outline

Denoising based "regularization"

Patch-based and convolutional sparsity models lead to a denoising step for the current image estimate $\boldsymbol{x}_t$ at iteration t

Many alternative denoising methods:

- nonlocal means (NLM) [130]
- block-matching 3D (BM3D) [131]
- ...

To adapt most such denoising methods for image reconstruction:

- plug-and-play ADMM [132, 133]
- Regularization by denoising (RED) [134, 135, 136]

Outline

- Learn sparsifying transform or dictionary for patches from training data
  - interpretable (?) optimization formulations
  - local prior information only (patch size)
  - perhaps slower computation due to optimization iterations
- Train neural network (aka deep learning)
  - less interpretable
  - possibly more global prior information
  - slow training, but perhaps faster computation after trained

Overview:

- image-domain learning [137, 138]
- k-space or data-domain learning
  e.g., RAKI [139], [140]
- transform learning (direct from k-space to image)
  e.g., AUTOMAP [141]
- hybrid-domain learning (unrolled loop, *e.g.*, variational network)
  alternate between denoising/dealiasing and reconstruction from k-space
  e.g., [142, 143, 144, 145, 146, 140] ...

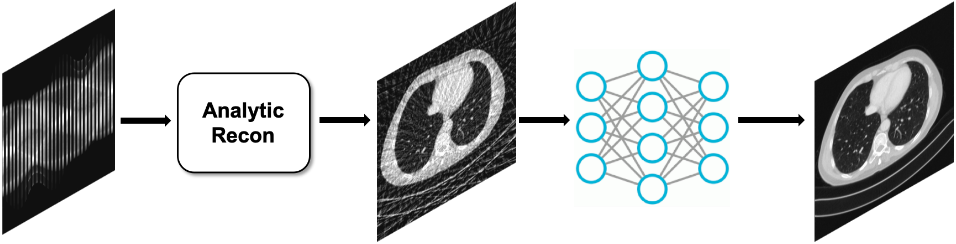


Figure courtesy of Jong Chul Ye, KAIST University.

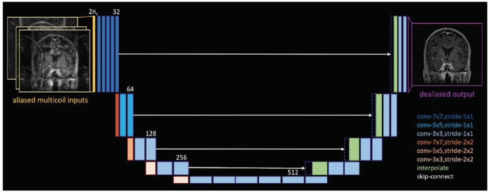
+ simple and fast
– aliasing is spatially widespread, requires deep network

Dangers of image-domain learning I

Investigating Robustness to Unseen Pathologies in Model-Free Deep Multicoil Reconstruction

Gopal Nataraj[1] and Ricardo Otazo[1,2]
[1]Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center
[2]Dept. of Radiology, Memorial Sloan Kettering Cancer Center

Introduction
Speed is often claimed as a key advantage of deep learning (DL) for undersampled parallel MRI reconstruction [1]. However, the only DL approach that to our knowledge has studied generalizability to pathologies unseen in training [2] requires repeated application of the MR acquisition model and its adjoint, just as in iterative methods. In contrast, model-free DL reconstruction has the potential to be much faster. Prior model-free DL work [3] proposes to learn a mapping directly from k-space, but with



[147] ISMRM 2020 Workshop on Data Sampling & Image Reconstruction

Dangers of image-domain learning II

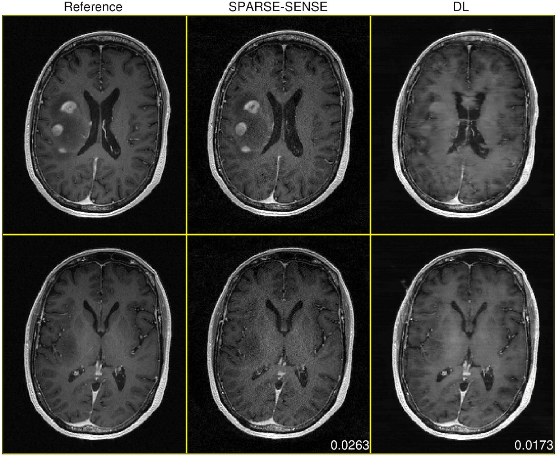


Figure 3: Reconstructions in a case of anaplastic astrocytoma, a rare malignant brain tumor. SPARSE-SENSE and DL reconstructions are from the same 4x-accelerated retrospectively undersampled acquisition. DL achieves lower whole-volume MAE than SPARSE-SENSE, but fails to properly reconstruct regions near the tumor.

Image-domain learning variations

- Use NN output as a "prior" for iterative reconstruction [137, 148]:

$$\hat{\boldsymbol{x}}_\beta = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{x} - \boldsymbol{x}_{\text{NN}}\|_2^2 = (\boldsymbol{A}'\boldsymbol{A} + \beta\boldsymbol{I})^{-1}(\boldsymbol{A}'\boldsymbol{y} + \beta\boldsymbol{x}_{\text{NN}})$$

  - For single-coil Cartesian case:
    - no iterations are needed (solve with FFTs)
    - $\lim_{\beta \to 0} \hat{\boldsymbol{x}}_\beta$ replaces missing k-space data with FFT of $\boldsymbol{x}_{\text{NN}}$
  - Iterations needed for parallel MRI and/or non-Cartesian sampling (PCG)

- Learn residual (aliasing artifacts), then subtract [149, 150]

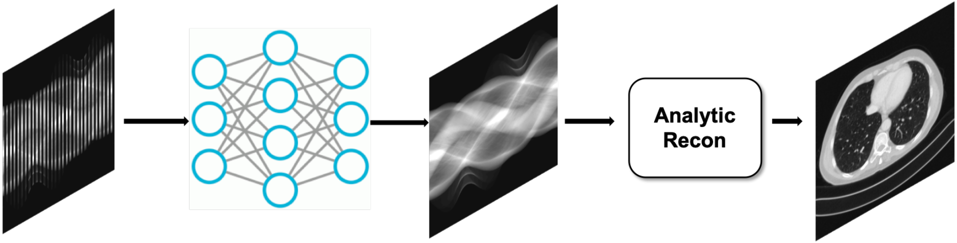


Figure courtesy of Jong Chul Ye, KAIST University.

+ simple and fast ("nonlinear GRAPPA")
+ "database-free" : learn from auto-calibration data
− perhaps harder to represent local image features?

DL for IR: transform learning

Figure courtesy of Jong Chul Ye, KAIST University.

+ in principle, purely data driven; potential to avoid model mismatch
– high memory requirement for fully connected layers

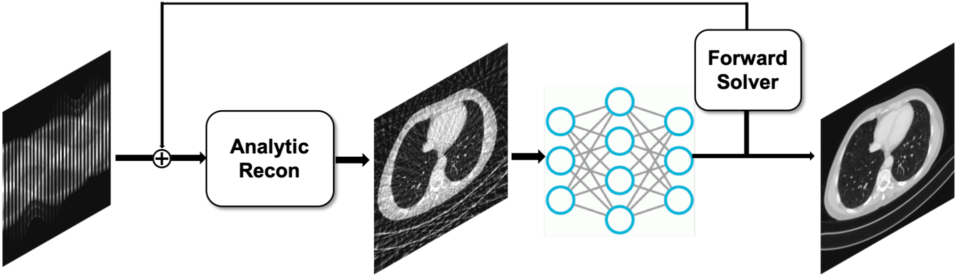


Figure courtesy of Jong Chul Ye, KAIST University.

+ physics-based use of k-space data & image-domain priors
+ interpretable connections to optimization approaches
– more computation to due to “iterations” (layers) and repeated FFT/NUFFT

Unrolled / unfolded loops

- learned ISTA (LISTA) [151]
  aka proximal gradient method / forward-backward splitting [152]
- half-quadratic [153]
- reaction-diffusion (GD) [154, 155]
- gradient descent / Landweber [156, 143]
- ADMM [142, 157]
- iterative hard thresholding (IHT) [158]
- approximate message passing (AMP) [159]
- accelerated gradient method [160]
- primal dual [161]
- primal dual with line search [162]
- alternating minimization [163]
- block coordinate descent (BCD-Net) [164, 165, 166]
- block proximal gradient with momentum (BPGM: Momentum-Net) [167, 3]
- And more [168, 62, 169, 170, 171, 172]

Surveys: [173, 174]

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [175, 176]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$
  - Synthesis form [177]:

$$\hat{\boldsymbol{x}} = G(\hat{\boldsymbol{z}}), \qquad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{A} G(\boldsymbol{z}) - \boldsymbol{y}\|_2^2$$

Caveat: $\hat{\boldsymbol{x}} \in \operatorname{Range}(G)$, non-convex minimization

Nonlinear encoder methods for ML-based IR

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [175, 176]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$
  - Synthesis form [177]:

$$\hat{\boldsymbol{x}} = G(\hat{\boldsymbol{z}}), \qquad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{A} G(\boldsymbol{z}) - \boldsymbol{y}\|_2^2$$

    Caveat: $\hat{\boldsymbol{x}} \in \text{Range}(G)$, non-convex minimization

  - Regularizer form:

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

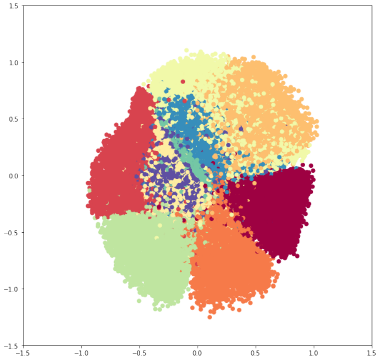
$$R_{\text{encoder}}(\boldsymbol{x}) = \min_{\boldsymbol{z}} \|\boldsymbol{x} - G(\boldsymbol{z})\|_p^p$$

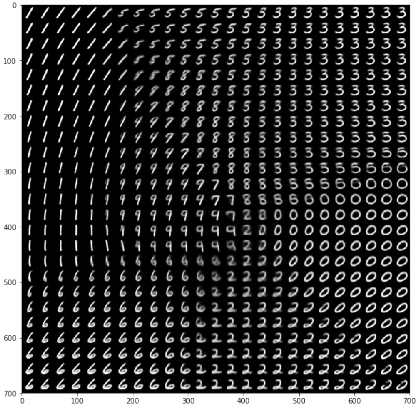
    Caveat: expensive non-convex double minimization, but more robust to encoder

From `jupyter` notebook for [178] (13 layer CNN with ≈ 300K learned parameters) at

https://github.com/skolouri/swae/blob/master/MNIST_SlicedWassersteinAutoEncoder_Circle.ipynb

$z \in \mathbb{R}^2 \quad \mapsto \quad x = G(z) \in \mathbb{R}^{28\times 28}$





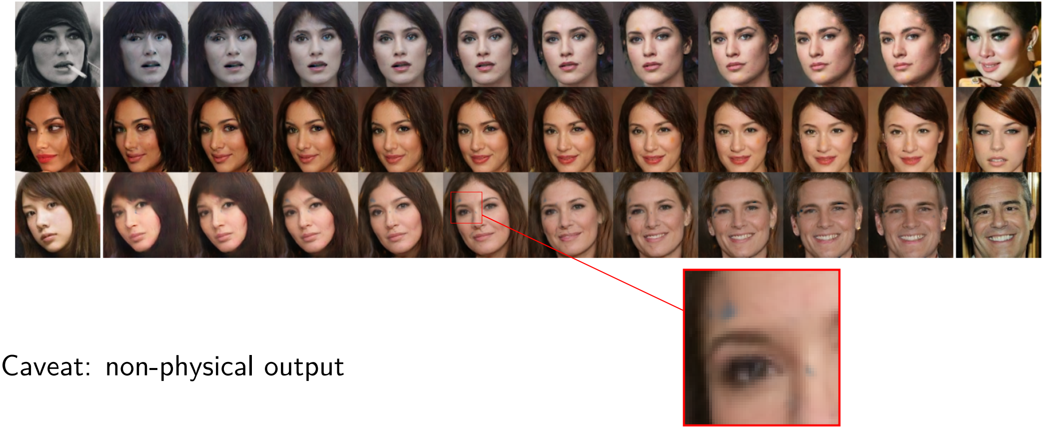
Caveat: Where is 4?

Generative Adversarial Networks (GAN) example

From Google's [179]:

Much more realistic than linear interpolation (averaging)
"setting a new milestone in visual quality" [179]

From Google's [179]:



Caveat: non-physical output

Model based image reconstruction using deep learned priors (MODL) [169, 62]

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \|\mathrm{CNN}(\boldsymbol{x})\|_2^2$$

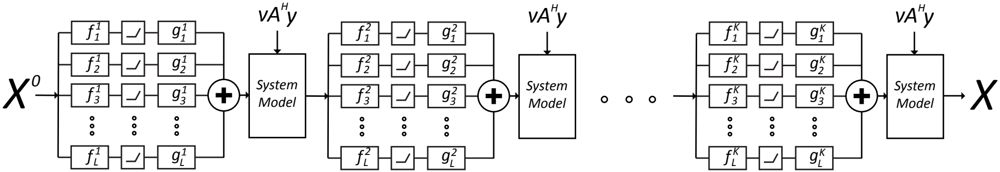
- $\mathrm{CNN}(\boldsymbol{x}) = \boldsymbol{x} - \mathrm{denoise}(\boldsymbol{x})$ predicts noise and aliasing patterns (*cf.* ResNet principle [149])
- Demonstrated robustness to changes in acceleration factors

Caveats to NN methods

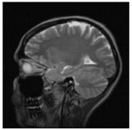
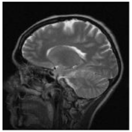
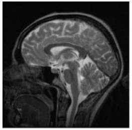
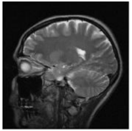
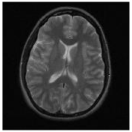
- Training data size (but self supervision [180])
- Local minimizers of training loss functions
- Sensitivity to adversarial examples (for classification problems)
- Enormous design space (architectures, parameters)
- Training loss functions, evaluation metrics vs clinical tasks
- Generalizability
  - noise level
  - coil sensitivity
  - k-space sampling
- Stability [181]
- Memory (especially 3D and dynamic)
- ...

Caveat: careful comparisons needed I

Unrolled loop method with 20 layers trained with $1.3 \cdot 10^6$ MR image 8×8 patches [163]



Tested with 5 different images:

Caveat: careful comparisons needed II

Results:

UF	Image	Zero-filled	Sparse MRI	UTMRI	Proposed
3.3×	1	25.6	26.7	28.3	28.2
	2	25.2	26.6	27.9	27.8
	3	26.0	27.3	29.3	28.9
	4	25.4	26.7	28.2	28.1
	5	27.2	28.9	30.6	30.3
Avg. PSNR change	-	-	**1.36**	**2.98**	**2.78**
5×	1	24.7	25.9	27.6	27.5
	2	24.2	25.5	27.2	27.0
	3	24.9	26.3	28.5	28.0
	4	24.4	25.7	27.6	27.4
	5	26.2	27.9	29.8	29.5
Avg. PSNR change	-	-	**1.38**	**3.26**	**3.0**
Approx recon time	-	-	**100s**	**240s**	**50s**

Sparse MRI [35] total variation and wavelets
UTMRI [126] (union of learned sparsifying transforms): adaptive, not "deep"

Outline

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:

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

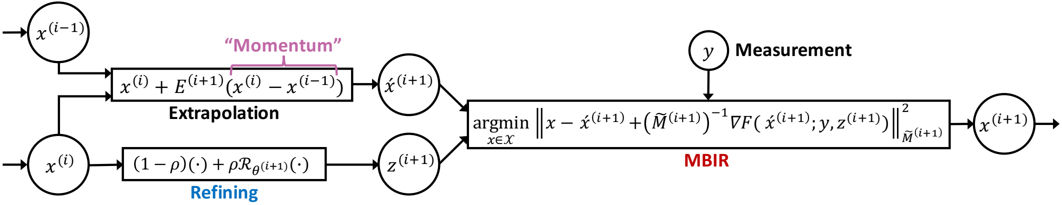
$$\text{Image:}\ \boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x}} F(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{z}^{(i)})$$

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

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

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

Unrolled loop network with momentum and quadratic majorizer [3]:



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

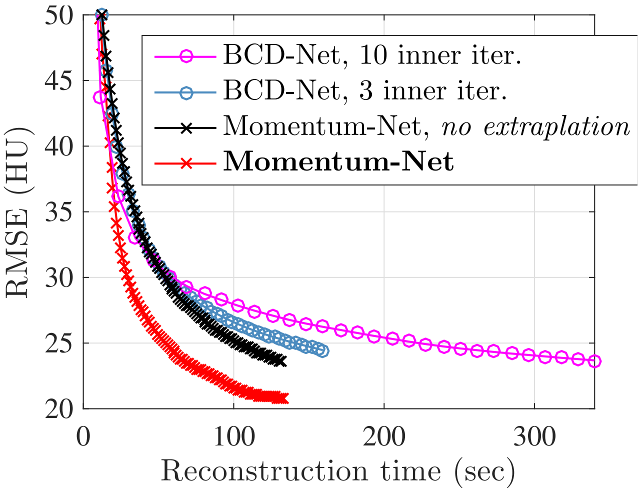
Momentum-Net benefits

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

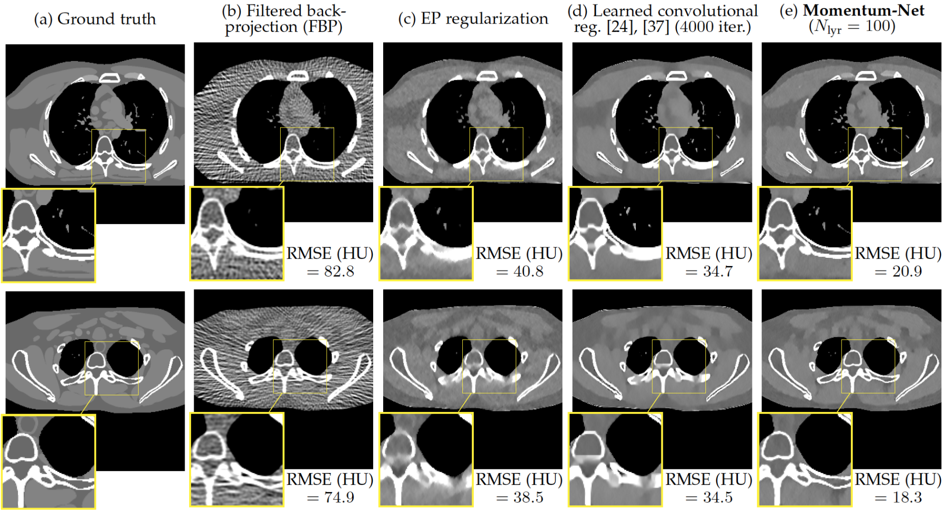
[3] Il Yong Chun, Zhengyu Huang, Hongki Lim, J A Fessler
Momentum-Net: Fast and convergent iterative neural network for inverse problems

http://arxiv.org/abs/1907.11818

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.

Introduction

Brief review of classic methods ($>$ 10 years old)

Sparsity regularizers: Basic

Sparsity regularizers: Advanced

Adaptive regularizers

Denoising-based “regularization”

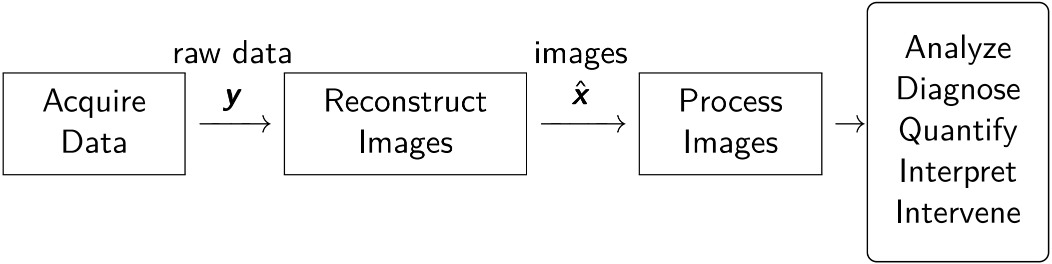
Deep-learning approaches for image reconstruction

Looking forward

Bibliography

- All presented sparsity models are applicable
- and many more: low rank, tensors...
- Challenge for supervised ML methods: all dynamic scans are under-sampled
- DL methods are appearing, *e.g.*, [144]

Overview of medical imaging:



A more speculative opportunity for machine learning:

raw data $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{ML-based “magic”}}$ $\rightarrow$ interpretation

...

A more speculative opportunity for machine learning:

$$\begin{array}{c} \text{raw data} \\ \boldsymbol{y} \end{array} \rightarrow \boxed{\begin{array}{c} \text{ML-based} \\ \text{“magic”} \end{array}} \rightarrow \text{interpretation}$$

- CT sinogram to vessel diameter [182, 183]
- k-space to ???

Caveat: seeing is believing...

One more opportunity for ML in medical imaging:

$$\boxed{\text{Acquire Data}} \xrightarrow[\text{}]{\text{raw data } \boldsymbol{y}} \boxed{\text{Reconstruct Images}} \xrightarrow{\text{images } \hat{\boldsymbol{x}}} \boxed{\begin{array}{c}\text{Analyze}\\ \text{Diagnose}\\ \text{Quantify}\\ \text{Interpret}\\ \text{Intervene}\end{array}}$$

...

One more opportunity for ML in medical imaging:

$$\boxed{\text{Acquire Data}} \xrightarrow[\ ]{\text{raw data } \boldsymbol{y}} \boxed{\text{Reconstruct Images}} \xrightarrow[\ ]{\text{images } \hat{\boldsymbol{x}}} \boxed{\begin{array}{c}\text{Analyze}\\ \text{Diagnose}\\ \text{Quantify}\\ \text{Interpret}\\ \text{Intervene}\end{array}}$$

k-space sampling design using ML methods:
"Learning-based compressive MRI" [25, 184]
(Volkan Cevher group, June 2018 IEEE T-MI)
Caveat: single coil only so far; hard to generalize to parallel MRI?

Acknowledgements

Thanks to numerous graduate students, postdocs, collaborators.

Especially for the results shown in this talk:
Prof. Sai Ravishankar, Prof. Il Yong Chun, Prof. Donghwan Kim, Prof. Yong Long, Xuehang Zheng,
Prof. Raj Nadakuditi, Prof. Doug Noll

Resources

Talk: http://web.eecs.umich.edu/~fessler/papers/files/talk/20/siam.pdf
code: https://github.com/JeffFessler/MIRT.jl

[1] S. Ravishankar, J. C. Ye, and J. A. Fessler. "Image reconstruction: from sparsity to data-adaptive methods and machine learning." In: *Proc. IEEE* 108.1 (Jan. 2020), 86–109. DOI: 10.1109/JPROC.2019.2936204 (cit. on p. 2).

[2] J. A. Fessler. "Optimization methods for MR image reconstruction." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 33–40. DOI: 10.1109/MSP.2019.2943645 (cit. on p. 2).

[3] I. Y. Chun et al. *Momentum-Net: Fast and convergent iterative neural network for inverse problems*. 2019. URL: http://arxiv.org/abs/1907.11818 (cit. on pp. 2, 85, 97, 98).

[4] M. K. Stehling, R. Turner, and P. Mansfield. "Echo-planar imaging: magnetic resonance imaging in a fraction of a second." In: *Science* 254.5028 (Oct. 1991), 43–50. DOI: 10.1126/science.1925560 (cit. on p. 5).

[5] P. Mansfield, R. Coxon, and J. Hykin. "Echo-volumar imaging (EVI) of the brain at 3.0 T: first normal volunteer and functional imaging results." In: *J. Comp. Assisted Tomo.* 19.6 (Nov. 1995), 847–52 (cit. on p. 5).

[6] A. Macovski. "Volumetric NMR imaging with time-varying gradients." In: *Mag. Res. Med.* 2.1 (Feb. 1985), 29–40. DOI: 10.1002/mrm.1910020105 (cit. on p. 5).

[7] C. H. Meyer et al. "Fast spiral coronary artery imaging." In: *Mag. Res. Med.* 28.2 (Dec. 1992), 202–13. DOI: 10.1002/mrm.1910280204 (cit. on p. 5).

[8] D. A. Feinberg et al. "Halving MR imaging time by conjugation: demonstration at 3.5 kG." In: *Radiology* 161.2 (Nov. 1986), 527–31. DOI: 10.1148/radiology.161.2.3763926 (cit. on p. 5).

[9] P. Margosian, F. Schmitt, and D. E. Purdy. "Faster MR imaging: Imaging with half the data." In: *Health Care Instrum.* 1.6 (1986), 195–7 (cit. on p. 5).

[10] D. C. Noll, D. G. Nishimura, and A. Macovski. "Homodyne detection in magnetic resonance imaging." In: *IEEE Trans. Med. Imag.* 10.2 (June 1991), 154–63. DOI: 10.1109/42.79473 (cit. on p. 5).

[11] J. W. Carlson. "An algorithm for NMR imaging reconstruction based on multiple RF receiver coils." In: *J. Mag. Res.* 74.2 (Sept. 1987), 376–80. DOI: '10.1016/0022-2364(87)90348-9' (cit. on p. 5).

[12] M. Hutchinson and U. Raff. "Fast MRI data acquisition using multiple detectors." In: *Mag. Res. Med.* 6.1 (Jan. 1988), 87–91. DOI: 10.1002/mrm.1910060110 (cit. on p. 5).

[13] P. B. Roemer et al. "The NMR phased array." In: *Mag. Res. Med.* 16.2 (Nov. 1990), 192–225. DOI: 10.1002/mrm.1910160203 (cit. on pp. 5, 13).

[14] J. B. Ra and C. Y. Rim. "Fast imaging using subencoding data sets from multiple detectors." In: *Mag. Res. Med.* 30.1 (July 1993), 142–5. DOI: 10.1002/mrm.1910300123 (cit. on p. 5).

[15] D. K. Sodickson and W. J. Manning. "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays." In: *Mag. Res. Med.* 38.4 (Oct. 1997), 591–603. DOI: 10.1002/mrm.1910380414 (cit. on p. 5).

[16] K. P. Pruessmann et al. "SENSE: sensitivity encoding for fast MRI." In: *Mag. Res. Med.* 42.5 (Nov. 1999), 952–62. DOI: '10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S' (cit. on pp. 5, 10, 13).

[17] K. P. Pruessmann et al. "Advances in sensitivity encoding with arbitrary k-space trajectories." In: *Mag. Res. Med.* 46.4 (Oct. 2001), 638–51. DOI: 10.1002/mrm.1241 (cit. on pp. 5, 14).

[18] M. A. Griswold et al. "Generalized autocalibrating partially parallel acquisitions (GRAPPA)." In: *Mag. Res. Med.* 47.6 (June 2002), 1202–10. DOI: 10.1002/mrm.10171 (cit. on pp. 5, 37).

[19] Y. Cao and D. N. Levin. "Feature-recognizing MRI." In: *Mag. Res. Med.* 30.3 (Sept. 1993), 305–17. DOI: 10.1002/mrm.1910300306 (cit. on p. 5).

[20] N. Chauffert et al. "Variable density sampling with continuous trajectories." In: *SIAM J. Imaging Sci.* 7.4 (2014), 1962–92. DOI: 10.1137/130946642 (cit. on p. 5).

[21] C. Boyer et al. "On the generation of sampling schemes for magnetic resonance imaging." In: *SIAM J. Imaging Sci.* 9.4 (2016), 2039–72. DOI: 10.1137/16m1059205 (cit. on p. 5).

Bibliography III

[22] N. Chauffert et al. "A projection algorithm for gradient waveforms design in magnetic resonance imaging." In: *IEEE Trans. Med. Imag.* 35.9 (Sept. 2016), 2026–39. DOI: 10.1109/tmi.2016.2544251 (cit. on p. 5).

[23] C. Lazarus et al. "SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI." In: *Mag. Res. Med.* 81.6 (June 2019), 3643–61. DOI: 10.1002/mrm.27678 (cit. on p. 5).

[24] S. Ravishankar and Y. Bresler. "Adaptive sampling design for compressed sensing MRI." In: *Proc. Int'l. Conf. IEEE Engr. in Med. and Biol. Soc.* 2011, 3751–5. DOI: 10.1109/IEMBS.2011.6090639 (cit. on p. 5).

[25] L. Baldassarre et al. "Learning-based compressive subsampling." In: *IEEE J. Sel. Top. Sig. Proc.* 10.4 (June 2016), 809–22. DOI: 10.1109/jstsp.2016.2548442 (cit. on pp. 5, 107).

[26] G. Godaliyadda et al. "A framework for dynamic image sampling based on supervised learning." In: *IEEE Trans. Computational Imaging* 4.1 (Mar. 2018), 1–16. DOI: 10.1109/tci.2017.2777482 (cit. on p. 5).

[27] C. D. Bahadir, A. V. Dalca, and M. R. Sabuncu. "Learning-based optimization of the under-sampling pattern in MRI." In: *Information Processing in Medical Im.* 2019, 780–92. DOI: 10.1007/978-3-030-20351-1_61 (cit. on p. 5).

[28] H. K. Aggarwal and M. Jacob. "Joint optimization of sampling pattern and priors in model based deep learning." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2020, 926–9. DOI: 10.1109/ISBI45749.2020.9098639 (cit. on p. 5).

[29] E. Candes et al. "Image reconstruction from highly undersampled data using total variation minimization." In: *MRA Workshop, Oct. 8, London Ontario.* 2004, p. 147 (cit. on p. 5).

[30] M. Lustig et al. "Faster imaging with randomly perturbed, undersampled spirals and $|L|_1$ reconstruction." In: *Proc. Intl. Soc. Mag. Res. Med.* 2005, p. 685. URL: http://cds.ismrm.org/ismrm-2005/Files/00685.pdf (cit. on p. 5).

[31] S. J. LaRoque, E. Y. Sidky, and X. Pan. "Image Reconstruction from Sparse Data in Echo-Planar Imaging." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* Vol. 5. 2006, 3166–9. DOI: 10.1109/NSSMIC.2006.356547 (cit. on p. 5).

[32] M. Lustig, D. L. Donoho, and J. M. Pauly. "Rapid MR imaging with "compressed sensing" and randomly under-sampled 3DFT trajectories." In: *Proc. Intl. Soc. Mag. Res. Med.* 2006, p. 695. URL: http://dev.ismrm.org//2006/0695.html (cit. on p. 5).

[33] K. T. Block, M. Uecker, and J. Frahm. "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint." In: *Mag. Res. Med.* 57.6 (June 2007), 1086–98. DOI: 10.1002/mrm.21236 (cit. on pp. 5, 27).

[34] E. Candes and J. Romberg. "Sparsity and incoherence in compressive sampling." In: *Inverse Prob.* 23.3 (June 2007), 969–86. DOI: 10.1088/0266-5611/23/3/008 (cit. on p. 5).

[35] M. Lustig, D. Donoho, and J. M. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging." In: *Mag. Res. Med.* 58.6 (Dec. 2007), 1182–95. DOI: 10.1002/mrm.21391 (cit. on pp. 5, 10, 94).

[36] J. C. Ye et al. "Projection reconstruction MR imaging using FOCUSS." In: *Mag. Res. Med.* 57.4 (Apr. 2007), 764–75. DOI: 10.1002/mrm.21202 (cit. on pp. 5, 29).

[37] E. J. Candes and M. B. Wakin. "An introduction to compressive sampling." In: *IEEE Sig. Proc. Mag.* 25.2 (Mar. 2008), 21–30. DOI: 10.1109/MSP.2007.914731 (cit. on p. 5).

[38] M. Lustig et al. "Compressed sensing MRI." In: *IEEE Sig. Proc. Mag.* 25.2 (Mar. 2008), 72–82. DOI: 10.1109/MSP.2007.914728 (cit. on pp. 5, 27).

[39] FDA. *510k premarket notification of HyperSense (GE Medical Systems)*. 2017. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K162722 (cit. on p. 5).

[40] FDA. *510k premarket notification of Compressed Sensing Cardiac Cine (Siemens)*. 2017. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K163312 (cit. on p. 5).

[41] FDA. *510k premarket notification of Compressed SENSE*. 2018. URL: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173079.pdf (cit. on p. 5).

[42] L. Geerts-Ossevoort et al. *Compressed SENSE*. Philips white paper 4522 991 31821 Nov. 2018. 2018. URL: https://philipsproductcontent.blob.core.windows.net/assets/20180109/619119731f2a42c4acd4a863008a46c7.pdf (cit. on p. 5).

[43] G. Harikumar, C. Couvreur, and Y. Bresler. "Fast optimal and suboptimal algorithms for sparse solutions to linear inverse problems." In: *Proc. IEEE Conf. Acoust. Speech Sig. Proc.* Vol. 3. 1998, 1877–80. DOI: 10.1109/ICASSP.1998.681830 (cit. on p. 5).

[44] J. Hamilton, D. Franson, and N. Seiberlich. "Recent advances in parallel imaging for MRI." In: *Prog. in Nuclear Magnetic Resonance Spectroscopy* 101 (Aug. 2017), 71–95. DOI: 10.1016/j.pnmrs.2017.04.002.

[45] P. J. Shin et al. "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion." In: *Mag. Res. Med.* 72.4 (Oct. 2014), 959–70. DOI: 10.1002/mrm.24997 (cit. on pp. 6, 38).

[46] A. Balachandrasekaran, M. Mani, and M. Jacob. *Calibration-free B0 correction of EPI data using structured low rank matrix recovery*. 2018. URL: http://arxiv.org/abs/1804.07436 (cit. on pp. 6, 38).

[47] J. P. Haldar and K. Setsompop. "Linear predictability in MRI reconstruction: Leveraging shift-invariant Fourier structure for faster and better imaging." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 69–82. DOI: 10.1109/MSP.2019.2949570 (cit. on p. 6).

[48] H-L. M. Cheng et al. "Practical medical applications of quantitative MR relaxometry." In: *J. Mag. Res. Im.* 36.4 (Oct. 2012), 805–24. DOI: 10.1002/jmri.23718 (cit. on p. 6).

[49] B. Zhao, F. Lam, and Z-P. Liang. "Model-based MR parameter mapping with sparsity constraints: parameter estimation and performance bounds." In: *IEEE Trans. Med. Imag.* 33.9 (Sept. 2014), 1832–44. DOI: 10.1109/TMI.2014.2322815 (cit. on p. 6).

[50] G. Nataraj et al. "Dictionary-free MRI PERK: Parameter estimation via regression with kernels." In: *IEEE Trans. Med. Imag.* 37.9 (Sept. 2018), 2103–14. DOI: 10.1109/TMI.2018.2817547 (cit. on p. 6).

[51] B. B. Mehta et al. "Magnetic resonance fingerprinting: a technical review." In: *Mag. Res. Med.* 81.1 (Jan. 2019), 25–46. DOI: 10.1002/mrm.27403 (cit. on p. 6).

[52] J. Bezanson et al. "Julia: A fresh approach to numerical computing." In: *SIAM Review* 59.1 (2017), 65–98. DOI: 10.1137/141000671 (cit. on p. 7).

[53] G. A. Wright. "Magnetic resonance imaging." In: *IEEE Sig. Proc. Mag.* 14.1 (Jan. 1997), 56–66. DOI: 10.1109/79.560324.

[54] M. Doneva. "Mathematical models for magnetic resonance imaging reconstruction: an overview of the approaches, problems, and future research areas." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 24–32. DOI: 10.1109/MSP.2019.2936964.

[55] J. A. Fessler. "Model-based image reconstruction for MRI." In: *IEEE Sig. Proc. Mag.* 27.4 (July 2010). Invited submission to special issue on medical imaging, 81–9. DOI: 10.1109/MSP.2010.936726 (cit. on p. 11).

[56] J. A. Fessler and B. P. Sutton. "Nonuniform fast Fourier transforms using min-max interpolation." In: *IEEE Trans. Sig. Proc.* 51.2 (Feb. 2003), 560–74. DOI: 10.1109/TSP.2002.807005 (cit. on p. 10).

[57] B. P. Sutton, D. C. Noll, and J. A. Fessler. "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities." In: *IEEE Trans. Med. Imag.* 22.2 (Feb. 2003), 178–88. DOI: 10.1109/TMI.2002.808360 (cit. on pp. 10, 14).

[58] A. Macovski. "Noise in MRI." In: *Mag. Res. Med.* 36.3 (Sept. 1996), 494–7. DOI: 10.1002/mrm.1910360327 (cit. on p. 11).

[59] J. A. Fessler et al. "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity." In: *IEEE Trans. Sig. Proc.* 53.9 (Sept. 2005), 3393–402. DOI: 10.1109/TSP.2005.853152 (cit. on p. 14).

[60] R. H. Chan and M. K. Ng. "Conjugate gradient methods for Toeplitz systems." In: *SIAM Review* 38.3 (Sept. 1996), 427–82. DOI: 10.1137/S0036144594276474 (cit. on p. 14).

[61] S. Ramani and J. A. Fessler. "Parallel MR image reconstruction using augmented Lagrangian methods." In: *IEEE Trans. Med. Imag.* 30.3 (Mar. 2011), 694–706. DOI: 10.1109/TMI.2010.2093536 (cit. on pp. 14, 30, 33).

[62] H. K. Aggarwal, M. P. Mani, and M. Jacob. "MoDL: model-based deep learning architecture for inverse problems." In: *IEEE Trans. Med. Imag.* 38.2 (Feb. 2019), 394–405. DOI: 10.1109/tmi.2018.2865356 (cit. on pp. 14, 85, 91).

[63] R. Boubertakh et al. "Non-quadratic convex regularized reconstruction of MR images from spiral acquisitions." In: *Signal Processing* 86.9 (Sept. 2006), 2479–94. DOI: 10.1016/j.sigpro.2005.11.011 (cit. on p. 15).

[64] S. Husse, Y. Goussard, and M. Idiert. "Extended forms of Geman & Yang algorithm: application to MRI reconstruction." In: *Proc. IEEE Conf. Acoust. Speech Sig. Proc.* Vol. 3. 2004, 513–16. DOI: 10.1109/ICASSP.2004.1326594 (cit. on p. 15).

[65] A. Florescu et al. "A majorize-minimize memory gradient method for complex-valued inverse problems." In: *Signal Processing* 103 (Oct. 2014), 285–95. DOI: 10.1016/j.sigpro.2013.09.026 (cit. on pp. 15, 16).

[66] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images." In: *IEEE Trans. Patt. Anal. Mach. Int.* 6.6 (Nov. 1984), 721–41. DOI: 10.1109/TPAMI.1984.4767596 (cit. on p. 15).

[67] J. Besag. "On the statistical analysis of dirty pictures." In: *J. Royal Stat. Soc. Ser. B* 48.3 (1986), 259–302. URL: http://www.jstor.org/stable/2345426 (cit. on p. 15).

[68] D. Kim and J. A. Fessler. "Optimized first-order methods for smooth convex minimization." In: *Mathematical Programming* 159.1 (Sept. 2016), 81–107. DOI: 10.1007/s10107-015-0949-3 (cit. on p. 16).

[69] Y. Drori. "The exact information-based complexity of smooth convex minimization." In: *J. Complexity* 39 (Apr. 2017), 1–16. DOI: 10.1016/j.jco.2016.11.001 (cit. on p. 16).

[70] Y. Nesterov. "A method of solving a convex programming problem with convergence rate $O(1/k^2)$." In: *Soviet Math. Dokl.* 27.2 (1983), 372–76. URL: http://www.core.ucl.ac.be/~nesterov/Research/Papers/DAN83.pdf (cit. on p. 16).

[71] Y. Drori and A. B. Taylor. "Efficient first-order methods for convex minimization: a constructive approach." In: *Mathematical Programming* (2020). DOI: 10.1007/s10107-019-01410-2 (cit. on p. 16).

[72] F. Knoll et al. "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge." In: *Mag. Res. Med.* (2020). DOI: 10.1002/mrm.28338 (cit. on p. 19).

[73] R. Tibshirani. "Regression shrinkage and selection via the LASSO." In: *J. Royal Stat. Soc. Ser. B* 58.1 (1996), 267–88. URL: http://www.jstor.org/stable/2346178 (cit. on p. 21).

[74] E. J. Candes and Y. Plan. "Near-ideal model selection by ℓ_1 minimization." In: *Ann. Stat.* 37.5a (2009), 2145–77. DOI: 10.1214/08-AOS653 (cit. on p. 21).

[75] A. B. Taylor, J. M. Hendrickx, and Francois Glineur. "Exact worst-case performance of first-order methods for composite convex optimization." In: *SIAM J. Optim.* 27.3 (Jan. 2017), 1283–313. DOI: 10.1137/16m108104x (cit. on pp. 21, 24, 25).

[76] D. Kim and J. A. Fessler. "Adaptive restart of the optimized gradient method for convex optimization." In: *J. Optim. Theory Appl.* 178.1 (July 2018), 240–63. DOI: 10.1007/s10957-018-1287-4 (cit. on pp. 21, 24, 25).

[77] I. Daubechies, M. Defrise, and C. De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." In: *Comm. Pure Appl. Math.* 57.11 (Nov. 2004), 1413–57. DOI: 10.1002/cpa.20042 (cit. on p. 23).

[78] P. Combettes and V. Wajs. "Signal recovery by proximal forward-backward splitting." In: *SIAM J. Multi. Mod. Sim.* 4.4 (2005), 1168–200. DOI: 10.1137/050626090 (cit. on p. 23).

[79] M. J. Muckley, D. C. Noll, and J. A. Fessler. "Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA)." In: *IEEE Trans. Med. Imag.* 34.2 (Feb. 2015), 578–88. DOI: 10.1109/TMI.2014.2363034 (cit. on p. 23).

[80] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." In: *SIAM J. Imaging Sci.* 2.1 (2009), 183–202. DOI: 10.1137/080716542 (cit. on p. 23).

[81] A. Beck and M. Teboulle. "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems." In: *IEEE Trans. Im. Proc.* 18.11 (Nov. 2009), 2419–34. DOI: 10.1109/TIP.2009.2028250 (cit. on pp. 23, 28).

[82] L. El Gueddari et al. "Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI." In: *Proc. IEEE SAM.* 2018, 415–9. DOI: 10.1109/SAM.2018.8448776 (cit. on p. 24).

[83] C. Y. Lin and J. A. Fessler. "Accelerated methods for low-rank plus sparse image reconstruction." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 48–51. DOI: 10.1109/ISBI.2018.8363520 (cit. on p. 24).

[84] C. Y. Lin and J. A. Fessler. "Efficient dynamic parallel MRI reconstruction for the low-rank plus sparse model." In: *IEEE Trans. Computational Imaging* 5.1 (Mar. 2019), 17–26. DOI: 10.1109/TCI.2018.2882089 (cit. on p. 24).

[85] Z. Ramzi, P. Ciuciu, and J-L. Starck. "Benchmarking proximal methods acceleration enhancements for CS-acquired MR image analysis reconstruction." In: *Sig. Proc. with Adapt. Sparse Struct. Rep. SPARS*. 2019. URL: https://hal.inria.fr/hal-02298569 (cit. on p. 24).

[86] B. O'Donoghue and E. Candes. "Adaptive restart for accelerated gradient schemes." In: *Found. Comp. Math.* 15.3 (June 2015), 715–32. DOI: 10.1007/s10208-013-9150-3 (cit. on p. 24).

[87] J. Liang and C-B. Schonlieb. *Faster FISTA*. 2018. URL: http://arxiv.org/abs/1807.04005 (cit. on p. 24).

[88] J. Liang and C-B. Schonlieb. *Improving FISTA: Faster, smarter and greedier*. 2018. URL: http://arxiv.org/abs/1811.01430 (cit. on p. 24).

[89] A. Chambolle. "An algorithm for total variation minimization and applications." In: *J. Math. Im. Vision* 20.1-2 (Jan. 2004), 89–97. DOI: 10.1023/B:JMIV.0000011325.36760.1e (cit. on p. 28).

[90] Y. Wang et al. "A new alternating minimization algorithm for total variation image reconstruction." In: *SIAM J. Imaging Sci.* 1.3 (2008), 248–72. DOI: 10.1137/080724265 (cit. on p. 29).

[91] T. Goldstein and S. Osher. "The split Bregman method for L1-regularized problems." In: *SIAM J. Imaging Sci.* 2.2 (2009), 323–43. DOI: 10.1137/080725891 (cit. on p. 30).

[92] J. Aelterman et al. "Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data." In: *Signal Processing* 91.12 (Jan. 2011), 2731–42. DOI: 10.1016/j.sigpro.2011.04.033 (cit. on p. 30).

[93] J. Eckstein and D. P. Bertsekas. "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators." In: *Mathematical Programming* 55.1-3 (Apr. 1992), 293–318. DOI: 10.1007/BF01581204 (cit. on p. 30).

[94] S. Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers." In: *Found. & Trends in Machine Learning* 3.1 (2010), 1–122. DOI: 10.1561/2200000016 (cit. on pp. 30, 32).

[95] Z. Xu, M. A. T. Figueiredo, and T. Goldstein. "Adaptive ADMM with spectral penalty parameter selection." In: *aistats*. 2017, 718–27. URL: http://proceedings.mlr.press/v54/xu17a.html (cit. on p. 32).

[96] B. Wohlberg. *ADMM penalty parameter selection by residual balancing*. 2017. URL: http://arxiv.org/abs/1704.06209 (cit. on p. 32).

[97] J. Eckstein. "Parallel alternating direction multiplier decomposition of convex programs." In: *J. Optim. Theory Appl.* 80.1 (Jan. 1994), 39–62. DOI: 10.1007/BF02196592 (cit. on p. 32).

[98] M. Le and J. A. Fessler. "Efficient, convergent SENSE MRI reconstruction for non-periodic boundary conditions via tridiagonal solvers." In: *IEEE Trans. Computational Imaging* 3.1 (Mar. 2017), 11–21. DOI: 10.1109/TCI.2016.2626999 (cit. on pp. 32, 33).

[99] Y. Chen et al. "Fast algorithms for image reconstruction with application to partially parallel MR imaging." In: *SIAM J. Imaging Sci.* 5.1 (2012), 90–118. DOI: 10.1137/100792688 (cit. on pp. 33, 34).

[100] A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging." In: *J. Math. Im. Vision* 40.1 (2011), 120–145. DOI: 10.1007/s10851-010-0251-1 (cit. on p. 34).

[101] T. Pock and A. Chambolle. "Diagonal preconditioning for first order primal-dual algorithms in convex optimization." In: *Proc. Intl. Conf. Comp. Vision*. 2011, 1762–9. DOI: 10.1109/ICCV.2011.6126441 (cit. on p. 34).

[102] P. L. Combettes and J. C. Pesquet. "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators." In: *Set-Valued Var Anal* 20.2 (June 2012), 307–30. DOI: 10.1007/s11228-011-0191-y (cit. on p. 34).

[103] L. Condat. "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms." In: *J. Optim. Theory Appl.* 158.2 (2013), 460–79. DOI: 10.1007/s10957-012-0245-9 (cit. on p. 34).

[104] E. Y. Sidky, J. H. Jorgensen, and X. Pan. "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm." In: *Phys. Med. Biol.* 57.10 (May 2012), 3065–92. DOI: 10.1088/0031-9155/57/10/3065 (cit. on p. 34).

[105] B. Công Vu. "A splitting algorithm for dual monotone inclusions involving cocoercive operators." In: *Adv. in Comp. Math.* 38.3 (Apr. 2013), 667–81. DOI: 10.1007/s10444-011-9254-8 (cit. on p. 34).

[106] T. Valkonen. "A primal-dual hybrid gradient method for nonlinear operators with applications to MRI." In: *Inverse Prob.* 30.5 (May 2014), p. 055012. DOI: 10.1088/0266-5611/30/5/055012 (cit. on p. 34).

[107] F. Ong, M. Uecker, and M. Lustig. *Accelerating non-Cartesian MRI reconstruction convergence using k-space preconditioning*. 2019. URL: http://arxiv.org/abs/1902.09657 (cit. on p. 34).

[108] 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. DOI: 10.1002/mrm.22428 (cit. on pp. 37, 71).

[109] M. Murphy et al. "Fast ℓ_1-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime." In: *IEEE Trans. Med. Imag.* 31.6 (June 2012), 1250–62. DOI: 10.1109/TMI.2012.2188039 (cit. on p. 37).

[110] D. Weller, S. Ramani, and J. A. Fessler. "Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction." In: *IEEE Trans. Med. Imag.* 33.2 (Feb. 2014), 351–61. DOI: 10.1109/TMI.2013.2285046 (cit. on p. 37).

[111] J. Duan, Y. Liu, and P. Jing. "Efficient operator splitting algorithm for joint sparsity-regularized SPIRiT-based parallel MR imaging reconstruction." In: *Mag. Res. Im.* 46 (Feb. 2018), 81–9. DOI: 10.1016/j.mri.2017.10.013 (cit. on p. 37).

[112] M. Uecker et al. "ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA." In: *Mag. Res. Med.* 71.3 (Mar. 2014), 990–1001. DOI: 10.1002/mrm.24751 (cit. on p. 38).

[113] L. El Gueddari et al. "Calibrationless OSCAR-based image reconstruction in compressed sensing parallel MRI." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2019, 1532–6 (cit. on p. 38).

[114] 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. DOI: 10.1109/TMI.2010.2090538 (cit. on pp. 42, 51, 71, 72).

[115] 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. DOI: 10.1137/141002293 (cit. on pp. 43, 51).

[116] I. Y. Chun and J. A. Fessler. "Convolutional dictionary learning: acceleration and convergence." In: *IEEE Trans. Im. Proc.* 27.4 (Apr. 2018), 1697–712. DOI: 10.1109/TIP.2017.2761545 (cit. on pp. 46, 48).

[117] I. Y. Chun and J. A. Fessler. *Convolutional analysis operator learning: acceleration and convergence*. 2018. URL: http://arxiv.org/abs/1802.05584 (cit. on pp. 46, 47).

[118] T. Nguyen-Duc and W-K. Jeong. "Compressed sensing dynamic MRI reconstruction using multi-scale 3D convolutional sparse coding with elastic net regularization." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 332–5. DOI: 10.1109/ISBI.2018.8363586 (cit. on p. 46).

[119] B. Wohlberg. "Efficient algorithms for convolutional sparse representations." In: *IEEE Trans. Im. Proc.* 25.1 (Jan. 2016), 301–15. DOI: 10.1109/TIP.2015.2495260 (cit. on p. 46).

[120] A. Lahiri, S. Ravishankar, and J. A. Fessler. "Combining supervised and semi-blind dictionary (Super-BReD) learning for MRI reconstruction." In: *Proc. Intl. Soc. Mag. Res. Med.* To appear. 2020, p. 3456 (cit. on p. 51).

[121] G. Wang et al. "Image reconstruction is a new frontier of machine learning." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1289–96. DOI: 10.1109/TMI.2018.2833635 (cit. on p. 52).

[122] 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. DOI: 10.1109/TSP.2006.881199 (cit. on p. 58).

[123] 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. DOI: 10.1109/TSP.2015.2405503 (cit. on p. 58).

[124] X. Zheng et al. "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. DOI: 10.1109/TMI.2018.2832007 (cit. on pp. 60, 61, 64, 65).

[125] 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. DOI: 10.1109/TMI.2015.2508780 (cit. on p. 60).

[126] 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. DOI: 10.1109/TCI.2016.2567299 (cit. on pp. 63, 94).

[127] 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. DOI: 10.1109/TCI.2017.2697206 (cit. on pp. 68, 69, 71, 72).

[128] X. Qu et al. "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator." In: *Med. Im. Anal.* 18.6 (Aug. 2014), 843–56. DOI: 10.1016/j.media.2013.09.007 (cit. on pp. 71, 72).

[129] Z. Zhan et al. "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction." In: *IEEE Trans. Biomed. Engin.* 63.9 (Sept. 2016), 1850–61. DOI: 10.1109/tbme.2015.2503756 (cit. on p. 72).

[130] A. Buades, B. Coll, and J-M. Morel. "The staircasing effect in neighborhood filters and its solution." In: *IEEE Trans. Im. Proc.* 15.6 (June 2006), 1499–505. DOI: 10.1109/TIP.2006.871137 (cit. on p. 74).

[131] K. Dabov et al. "Image denoising by sparse 3-D transform-domain collaborative filtering." In: *IEEE Trans. Im. Proc.* 16.8 (Aug. 2007), 2080–95. DOI: 10.1109/TIP.2007.901238 (cit. on p. 74).

[132] 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. DOI: 10.1109/tci.2016.2629286 (cit. on p. 74).

[133] G. T. Buzzard et al. "Plug-and-play unplugged: optimization-free reconstruction using consensus equilibrium." In: *SIAM J. Imaging Sci.* 11.3 (Jan. 2018), 2001–20. DOI: 10.1137/17m1122451 (cit. on p. 74).

[134] Y. Romano, M. Elad, and P. Milanfar. "The little engine that could: Regularization by denoising (RED)." In: *SIAM J. Imaging Sci.* 10.4 (2017), 1804–44. DOI: 10.1137/16M1102884 (cit. on p. 74).

[135] E. T. Reehorst and P. Schniter. *Regularization by denoising: clarifications and new interpretations*. 2018. URL: http://arxiv.org/abs/1806.02296 (cit. on p. 74).

[136] R. Ahmad et al. "Plug and play methods for magnetic resonance imaging." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 105–16. DOI: 10.1109/MSP.2019.2949470 (cit. on p. 74).

[137] S. Wang et al. "Exploiting deep convolutional neural network for fast magnetic resonance imaging." In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1778. URL: http://archive.ismrm.org/2016/1778.html (cit. on pp. 77, 81).

[138] D. Lee, J. Yoo, and J. C. Ye. *Deep artifact learning for compressed sensing and parallel MRI*. 2017. URL: http://arxiv.org/abs/1703.01120 (cit. on p. 77).

[139] M. Akcakaya et al. "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging." In: *Mag. Res. Med.* 81.1 (Jan. 2019), 439–53. DOI: 10.1002/mrm.27420 (cit. on p. 77).

[140] Y. Han and J. C. Ye. "K-space deep learning for accelerated MRI." In: *IEEE Trans. Med. Imag.* 39.2 (Feb. 2020), 377–86. DOI: 10.1109/TMI.2019.2927101 (cit. on p. 77).

[141] B. Zhu et al. "Image reconstruction by domain-transform manifold learning." In: *Nature* 555 (Mar. 2018), 487–92. DOI: 10.1038/nature25988 (cit. on p. 77).

[142] Y. Yang et al. "Deep ADMM-net for compressive sensing MRI." In: *Neural Info. Proc. Sys.* 2016, 10–18. URL: https://papers.nips.cc/paper/6406-deep-admm-net-for-compressive-sensing-mri (cit. on pp. 77, 85).

Bibliography XV

[143] K. Hammernik et al. "Learning a variational network for reconstruction of accelerated MRI data." In: *Mag. Res. Med.* 79.6 (June 2018), 3055–71. DOI: 10.1002/mrm.26977 (cit. on pp. 77, 85).

[144] J. Schlemper et al. "A deep cascade of convolutional neural networks for dynamic MR image reconstruction." In: *IEEE Trans. Med. Imag.* 37.2 (Feb. 2018), 491–503. DOI: 10.1109/tmi.2017.2760978 (cit. on pp. 77, 102).

[145] T. M. Quan, T. Nguyen-Duc, and W-K. Jeong. "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1488–97. DOI: 10.1109/TMI.2018.2820120 (cit. on p. 77).

[146] D. Lee et al. "Deep residual learning for accelerated MRI using magnitude and phase networks." In: *IEEE Trans. Biomed. Engin.* 65.9 (Sept. 2018), 1985–95. DOI: 10.1109/TBME.2018.2821699 (cit. on p. 77).

[147] G. Nataraj and R. Otazo. "Investigating robustness to unseen pathologies in model-free deep multicoil reconstruction." In: *ISMRM Workshop on Data Sampling and Image Reconstruction*. 2020 (cit. on p. 79).

[148] G. Yang et al. "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1310–21. DOI: 10.1109/TMI.2017.2785879 (cit. on p. 81).

[149] K. He et al. "Deep residual learning for image recognition." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2016, 770–8. DOI: 10.1109/CVPR.2016.90 (cit. on pp. 81, 91).

[150] K. Zhang et al. "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising." In: *IEEE Trans. Im. Proc.* 26.7 (July 2017), 3142–55. DOI: 10.1109/tip.2017.2662206 (cit. on p. 81).

[151] K. Gregor and Y. LeCun. "Learning fast approximations of sparse coding." In: *Proc. Intl. Conf. Mach. Learn.* 2010. URL: http://yann.lecun.com/exdb/publis/pdf/gregor-icml-10.pdf (cit. on p. 85).

[152] T. Meinhardt et al. "Learning proximal operators: using denoising networks for regularizing inverse imaging problems." In: *Proc. Intl. Conf. Comp. Vision*. 2017, 1799–808. DOI: 10.1109/ICCV.2017.198 (cit. on p. 85).

[153] U. Schmidt and S. Roth. "Shrinkage fields for effective image restoration." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2014, 2774–81. DOI: 10.1109/CVPR.2014.349 (cit. on p. 85).

[154] Y. Chen, W. Yu, and T. Pock. "On learning optimized reaction diffusion processes for effective image restoration." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2015, 5261–9. DOI: 10.1109/CVPR.2015.7299163 (cit. on p. 85).

[155] Y. Chen and T. Pock. "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration." In: *IEEE Trans. Patt. Anal. Mach. Int.* 39.6 (June 2017), 1256–72. DOI: 10.1109/TPAMI.2016.2596743 (cit. on p. 85).

[156] K. Hammernik et al. "Learning a variational model for compressed sensing MRI reconstruction." In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1088. URL: http://archive.ismrm.org/2016/1088.html (cit. on p. 85).

[157] Y. Yang et al. *ADMM-net: A deep learning approach for compressive sensing MRI*. 2017. URL: http://arxiv.org/abs/1705.06869 (cit. on p. 85).

[158] B. Xin et al. *Maximal sparsity with deep networks?* 2016. URL: http://arxiv.org/abs/1605.01636 (cit. on p. 85).

[159] P. Schniter. "Recent advances in approximate message passing." In: *spars-17*. 2017, plenary (cit. on p. 85).

[160] H-Y. Liu et al. "Compressive imaging with iterative forward models." In: *Proc. IEEE Conf. Acoust. Speech Sig. Proc.* 2017 (cit. on p. 85).

[161] J. Adler and O. Oktem. "Learned primal-dual reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1322–32. DOI: 10.1109/tmi.2018.2799231 (cit. on p. 85).

[162] Y. Malitsky and T. Pock. "A first-order primal-dual algorithm with linesearch." In: *SIAM J. Optim.* 28.1 (Jan. 2018), 411–32. DOI: 10.1137/16m1092015 (cit. on p. 85).

[163] S. Ravishankar et al. "Deep dictionary-transform learning for image reconstruction." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 1208–12. DOI: 10.1109/ISBI.2018.8363788 (cit. on pp. 85, 93).

[164] I. Y. Chun and J. A. Fessler. "Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery." In: *Proc. IEEE Wkshp. on Image, Video, Multidim. Signal Proc.* 2018, 1–5. DOI: 10.1109/IVMSPW.2018.8448694 (cit. on p. 85).

[165] I. Y. Chun and J. A. Fessler. *Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery*. 2018. URL: http://arxiv.org/abs/1802.07129 (cit. on p. 85).

[166] H. Lim et al. "Application of trained deep BCD-net to iterative low-count PET image reconstruction." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* 2018, 1–4. DOI: 10.1109/NSSMIC.2018.8824563 (cit. on p. 85).

[167] I. Y. Chun et al. "Fast and convergent iterative image recovery using trained convolutional neural networks." In: *Allerton Conf. on Comm., Control, and Computing*. Invited. 2018, 155–9. DOI: 10.1109/ALLERTON.2018.8635932 (cit. on p. 85).

[168] J. R. Hershey, J. L. Roux, and F. Weninger. *Deep unfolding: Model-based inspiration of novel deep architectures*. 2014. URL: http://arxiv.org/abs/1409.2574 (cit. on p. 85).

[169] H. K. Aggarwal, M. P. Mani, and M. Jacob. "Model based image reconstruction using deep learned priors (MODL)." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 671–4. DOI: 10.1109/ISBI.2018.8363663 (cit. on pp. 85, 91).

[170] K. H. Jin, M. McCann, and M. Unser. "BPConvNet for compressed sensing recovery in bioimaging." In: *spars-17*. 2017 (cit. on p. 85).

[171] H. Chen et al. "LEARN: Learned experts: assessment-based reconstruction network for sparse-data CT." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1333–47. DOI: 10.1109/TMI.2018.2805692 (cit. on p. 85).

[172] D. Wu et al. *End-to-end abnormality detection in medical imaging*. 2018. URL: http://arxiv.org/abs/1711.02074 (cit. on p. 85).

[173] D. Liang et al. "Deep MRI reconstruction: Unrolled optimization algorithms meet neural networks." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 141–51. DOI: 10.1109/MSP.2019.2950557 (cit. on p. 85).

[174] V. Monga, Y. Li, and Y. C. Eldar. *Algorithm unrolling: interpretable, efficient deep learning for signal and image processing*. 2020. URL: http://arxiv.org/abs/1912.10557 (cit. on p. 85).

[175] I. J. Goodfellow et al. *Generative adversarial networks*. 2014. URL: http://arxiv.org/abs/1406.2661 (cit. on pp. 86, 87).

[176] X. Chen et al. “InfoGAN: interpretable representation learning by information maximizing generative adversarial nets.” In: *Neural Info. Proc. Sys.* 2016, 2172–80. URL: https://papers.nips.cc/paper/6399-infogan-interpretable-representation-learning-by-information-maximizing-generative-adversarial-nets (cit. on pp. 86, 87).

[177] A. Bora et al. “Compressed sensing using generative models.” In: *Proc. Intl. Conf. Mach. Learn.* Vol. 70. 2017, 537–46. URL: http://proceedings.mlr.press/v70/bora17a.html (cit. on pp. 86, 87).

[178] S. Kolouri et al. *Sliced-Wasserstein autoencoder: an embarrassingly simple generative model*. 2018. URL: http://arxiv.org/abs/1804.01947 (cit. on p. 88).

[179] D. Berthelot, T. Schumm, and L. Metz. *BEGAN: boundary equilibrium generative adversarial networks*. 2017. URL: http://arxiv.org/abs/1703.10717 (cit. on pp. 89, 90).

[180] B. Yaman et al. “Self-supervised learning of physics-based reconstruction neural networks without fully-sampled reference data.” In: *Mag. Res. Med.* (2020). DOI: 10.1002/mrm.28378 (cit. on p. 92).

[181] V. Antun et al. *On instabilities of deep learning in image reconstruction - Does AI come at a cost?* 2019. URL: http://arxiv.org/abs/1902.05300 (cit. on p. 92).

[182] E. Haneda et al. “CT sinogram analysis using deep learning.” In: *Proc. 5th Intl. Mtg. on Image Formation in X-ray CT*. 2018, 419–22 (cit. on p. 105).

[183] Q. De Man et al. “A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms.” In: *Med. Phys.* 46.12 (Dec. 2019), e790–800. DOI: 10.1002/mp.13640 (cit. on p. 105).

[184] B. Gozcu et al. “Learning-based compressive MRI.” In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1394–406. DOI: 10.1109/TMI.2018.2832540 (cit. on p. 107).