Jeffrey A. Fessler

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

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

ISMRM Data Sampling & Image Reconstruction Workshop
Sedona, AZ
2026-01-12

Declaration: No relevant financial interests or relationships to disclose

Disclosure

Declaration of Relevant Financial Interests or Relationships
Speaker Name: Jeff Fessler
I have no conflicts of interest with regards to MR topics.

Outline

(with ML connections...)

- Spin preparation RF pulse(s)
  $m(\vec{r}, 0)$ magnetization encoding relaxation, diffusion, velocity, ...
- Bloch equation (or Bloch-McConnell or Bloch-Torrey)
  for transverse magnetization free precession after ith pulse

$$m(\vec{r}, t) = m(\vec{r}, 0)\, \mathrm{e}^{-R_2(\vec{r})t}\, \mathrm{e}^{-\imath \phi_i(\vec{r}, t)}$$

  phase ϕ depends on gradient fields, off-resonance, ...
- Signal model for ith readout:

$$s_i(t) = \int m(\vec{r}, t)\, \mathrm{d}\vec{r} = \int m(\vec{r}, 0)\, \mathrm{e}^{-R_2(\vec{r})t}\, \mathrm{e}^{-\imath \phi_i(\vec{r}, t)}\, \mathrm{d}\vec{r}$$

- Discretization of integral:

$$\boldsymbol{s} \approx \boldsymbol{A}\boldsymbol{x}, \quad x_j \triangleq m(\vec{r}_j, 0)$$

  $\boldsymbol{A}$ depends on physics: gradients, coil maps, ...

- Noisy measurements (thermal noise): $\boldsymbol{y} = \boldsymbol{s} + \varepsilon = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- probability model: $\varepsilon \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

$$\mathsf{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \right)$$

(after pre-whitening of correlation between multiple receive coils)

- negative log-likelihood:

$$-\log \mathsf{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) = \frac{1}{2\sigma^2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

apt for any k-space sampling pattern, including non-Cartesian and non-Fourier encoding

- **Gauss-Markov theorem** provides some theoretical basis for this data-fit term
-

- Noisy measurements (thermal noise): $\boldsymbol{y} = \boldsymbol{s} + \varepsilon = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- probability model: $\varepsilon \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

$$\mathsf{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \right)$$

(after pre-whitening of correlation between multiple receive coils)

- negative log-likelihood:

$$-\log \mathsf{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) = \frac{1}{2\sigma^2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

apt for any k-space sampling pattern, including non-Cartesian and non-Fourier encoding

- **Gauss-Markov theorem** provides some theoretical basis for this data-fit term
- Other important statistical models in MRI
  - physiological "noise" in fMRI
  - particle diffusion distributions
  - ...

Image reconstruction and gradient

- Goal of (basic) image reconstruction is to recover magnetization image $\boldsymbol{x}$ from k-space data $\boldsymbol{y}$ using negative log-likelihood:

$$f(\boldsymbol{x}) = -\log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

  (Drop constant σ^2: unimportant for regularized methods, but more important for Bayesian methods.)

- Iterative methods are used when $\boldsymbol{A}$ is non-square (and hence non-invertible)
- Most iterative methods use the *gradient* of the data-fit term:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})$$

  where $\boldsymbol{A}'$ denotes the *adjoint* of matrix/operator $\boldsymbol{A}$

-

Image reconstruction and gradient

- Goal of (basic) image reconstruction is to recover magnetization image $\boldsymbol{x}$ from k-space data $\boldsymbol{y}$ using negative log-likelihood:

$$f(\boldsymbol{x}) = -\log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

(Drop constant σ^2: unimportant for regularized methods, but more important for Bayesian methods.)

- Iterative methods are used when $\boldsymbol{A}$ is non-square (and hence non-invertible)
- Most iterative methods use the *gradient* of the data-fit term:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})$$

where $\boldsymbol{A}'$ denotes the *adjoint* of matrix/operator $\boldsymbol{A}$

- The "$\boldsymbol{A}$" written on paper is rarely a (dense) matrix in MRI recon software.

Adjoints

- Dynamic MRI data-fit term:

$$f(\boldsymbol{X}) = \frac{1}{2} \|\mathcal{A}\boldsymbol{X} - \boldsymbol{y}\|_2^2$$

  - Casorati matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_T \end{bmatrix} \in \mathbb{C}^{N \times T}$ (# of voxels $\times$ # of time frames)
  - Linear operator $\mathcal{A}$ maps array $\boldsymbol{X}$ into k-space signal data vector $\boldsymbol{s} \in \mathbb{C}^M$:

$$\mathcal{A}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A}_1 \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{A}_T \boldsymbol{x}_T \end{bmatrix} \qquad (\boldsymbol{A}_t \text{ is encoding for } t\text{th frame})$$

-

Adjoint in dynamic MRI

- Dynamic MRI data-fit term:

$$f(\boldsymbol{X}) = \frac{1}{2} \|\mathcal{A}\boldsymbol{X} - \boldsymbol{y}\|_2^2$$

  - Casorati matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_T \end{bmatrix} \in \mathbb{C}^{N \times T}$ (# of voxels $\times$ # of time frames)
  - Linear operator $\mathcal{A}$ maps array $\boldsymbol{X}$ into k-space signal data vector $\boldsymbol{s} \in \mathbb{C}^M$:

$$\mathcal{A}\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A}_1 \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{A}_T \boldsymbol{x}_T \end{bmatrix} \qquad (\boldsymbol{A}_t \text{ is encoding for } t\text{th frame})$$

- Gradient of dynamic MRI data-fit term:

$$\nabla f(\boldsymbol{X}) = \mathcal{A}'(\mathcal{A}\boldsymbol{X} - \boldsymbol{y})$$

- $\mathcal{A}$ is not a matrix, but it has a well-defined adjoint needed for the gradient:

$$\mathcal{A}'\boldsymbol{y} = \begin{bmatrix} \boldsymbol{A}_1' \boldsymbol{y}_1 & \dots & \boldsymbol{A}_T' \boldsymbol{y}_T \end{bmatrix} \qquad \text{(maps vector to matrix)}$$

- Linear operator $\mathcal{A} : \mathcal{X} \mapsto \mathcal{Y}$
  $\mathcal{X}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$
  $\mathcal{Y}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$
- The **(Hermitian) adjoint** of $\mathcal{A}$ is the *unique* operator $\mathcal{A}'$ satisfying

$$\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, \mathcal{A}'\boldsymbol{y} \rangle_{\mathcal{X}}, \quad \forall \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{y} \in \mathcal{Y}$$

-

Adjoints

- Linear operator $\mathcal{A} : \mathcal{X} \mapsto \mathcal{Y}$
  $\mathcal{X}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$
  $\mathcal{Y}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$
- The **(Hermitian) adjoint** of $\mathcal{A}$ is the *unique* operator $\mathcal{A}'$ satisfying

$$\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, \mathcal{A}'\boldsymbol{y} \rangle_{\mathcal{X}}, \quad \forall \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{y} \in \mathcal{Y}$$

- An adjoint usually is not inverse nor pseudoinverse: $\mathcal{A}' \neq \mathcal{A}^{-1}$ and $\mathcal{A}' \neq \mathcal{A}^{+}$
- Notable exception: single-coil MRI, full Cartesian k-space sampling, unitary FFT
-

Adjoints

- Linear operator $\mathcal{A} : \mathcal{X} \mapsto \mathcal{Y}$
  $\mathcal{X}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$
  $\mathcal{Y}$ vector space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$
- The **(Hermitian) adjoint** of $\mathcal{A}$ is the *unique* operator $\mathcal{A}'$ satisfying

$$\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, \mathcal{A}'\boldsymbol{y} \rangle_{\mathcal{X}}, \quad \forall \boldsymbol{x} \in \mathcal{X}, \ \boldsymbol{y} \in \mathcal{Y}$$

- An adjoint usually is not inverse nor pseudoinverse: $\mathcal{A}' \neq \mathcal{A}^{-1}$ and $\mathcal{A}' \neq \mathcal{A}^{+}$
- Notable exception: single-coil MRI, full Cartesian k-space sampling, unitary FFT
- Be wary of code that provides untested "adjoint" functions:
  - "Adjoint" code with density compensation
  - "Adjoint" code for `fft` based on `ifft`
    FFTW library provides `bfft` for adjoint of `fft`

Adjoint testing

Testing adjoints of complicated linear operators

- Random vector test (large scale) using definition:

$$\langle \mathcal{A}\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} \approx \langle \boldsymbol{x}, \mathcal{A}'\boldsymbol{y} \rangle_{\mathcal{X}}, \qquad \text{for random } \boldsymbol{x} \in \mathcal{X} \text{ and } \boldsymbol{y} \in \mathcal{Y}$$

- Exhaustive test (small scale):

$$\langle \mathcal{A}\boldsymbol{e}_j, \boldsymbol{e}_i \rangle_{\mathcal{Y}} \approx \langle \boldsymbol{e}_j, \mathcal{A}'\boldsymbol{e}_i \rangle_{\mathcal{X}}, \qquad \forall i, j$$

Recon Package	Random Tests	Exhaustive Tests
BART (C)	Y	
MatMRI (Matlab)	Y	
MIRTorch (pytorch)	Y	
MIRT.jl (Julia)	Y	Y
MIRT classic (Matlab)	Y	Y
PySap-MRI	Y	
SigPy	Y	

- Another option: let `autodiff` do the work [1]
- "FAST-DIPS: Adjoint-free analytic steps and hard-constrained likelihood correction for diffusion-prior inverse problems"

  ICLR 2026 Submission `https://openreview.net/forum?id=voMeZVAkKL`

Preconditioning

Preconditioning

- A **preconditioner** helps find the *same solution* but *faster*.
- Solving system of equations
  - $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  - $\boldsymbol{A}$ is square and invertible
  - Left preconditioner: $\boldsymbol{P}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$. Same solution if $\boldsymbol{P}$ is invertible.
-

Preconditioning

- A **preconditioner** helps find the *same solution* but *faster*.
- Solving system of equations
  - $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  - $\boldsymbol{A}$ is square and invertible
  - Left preconditioner: $\boldsymbol{P}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$. Same solution if $\boldsymbol{P}$ is invertible.
- In iterative MRI reconstruction, $\boldsymbol{A}$ is rarely square or invertible!
- Solving an optimization problem based on log-likelihood
  $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}), \qquad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x})$
-

Preconditioning

- A **preconditioner** helps find the *same solution* but *faster*.
- Solving system of equations
  - $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  - $\boldsymbol{A}$ is square and invertible
  - Left preconditioner: $\boldsymbol{P}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$. Same solution if $\boldsymbol{P}$ is invertible.
- In iterative MRI reconstruction, $\boldsymbol{A}$ is rarely square or invertible!
- Solving an optimization problem based on log-likelihood
  $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}), \qquad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x})$
- Left-multiplying $\boldsymbol{A}$ *changes* the solution [2, 3] and increases noise [4]:
  $\boldsymbol{x}_? = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{W}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{W}\boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x})$ ("weighting" not "preconditioning")
-

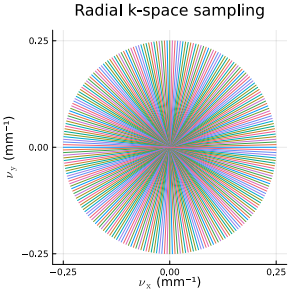
Preconditioning

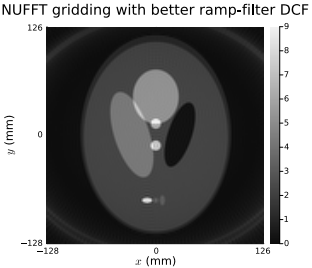
- A **preconditioner** helps find the *same solution* but *faster*.
- Solving system of equations
  - $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  - $\boldsymbol{A}$ is square and invertible
  - Left preconditioner: $\boldsymbol{P}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{b}$. Same solution if $\boldsymbol{P}$ is invertible.
- In iterative MRI reconstruction, $\boldsymbol{A}$ is rarely square or invertible!
- Solving an optimization problem based on log-likelihood
  $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}), \qquad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x})$
- Left-multiplying $\boldsymbol{A}$ *changes* the solution [2, 3] and increases noise [4]:
  $\boldsymbol{x}_? = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{W}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{W}\boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x})$ ("weighting" not "preconditioning")
- One should precondition in a way that *preserves* the solution [5], *e.g.*:
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{P}\nabla\Psi(\boldsymbol{x}_k)$. Ideally $\boldsymbol{P} \approx (\nabla^2\Psi)^{-1}$ (not any DCF)

Preconditioning simulation

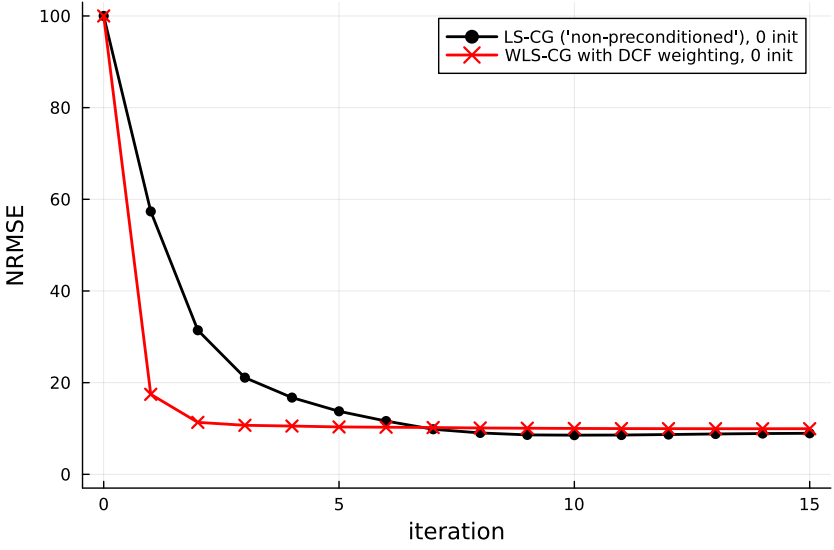
`https://juliaimagerecon.github.io/Examples/generated/mri/6-precon`

- analytical k-space from ellipse Fourier transform (no inverse crime) [6, 7]
- 128×128 image reconstruction
- 126 radial lines with 129 samples each
- Examine NRMSE $= 100\% \cdot \|\hat{\boldsymbol{x}}_k - \boldsymbol{x}_{\text{true}}\|_2 \,/\, \|\boldsymbol{x}_{\text{true}}\|_2$ vs iteration $k = 0, 1, \ldots$

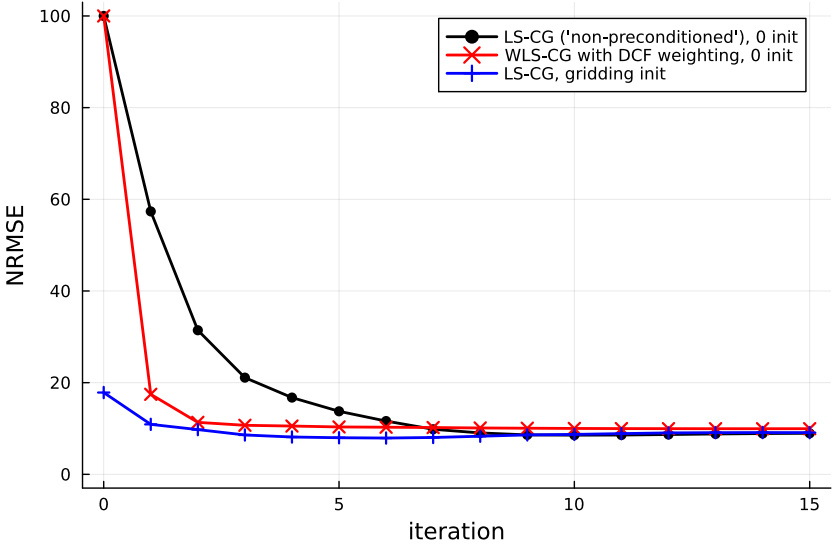




Density compensation and convergence rate for CG



Density compensation and convergence rate for CG



Preconditioning: Further reading

- Pruessmann et al., MRM 2001 [2]
- Gabr et al., MRM 2006 [4]
- Baron et al., MRM 2018 [3]
- Ong et al., IEEE T-MI 2020 [5]
-

Preconditioning: Further reading

- Pruessmann et al., MRM 2001 [2]
- Gabr et al., MRM 2006 [4]
- Baron et al., MRM 2018 [3]
- Ong et al., IEEE T-MI 2020 [5]
- Hong et al., IEEE T-CI 2024 [8]

1476 IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 10, 2024

Provable Preconditioned Plug-and-Play Approach for Compressed Sensing MRI Reconstruction

Tao Hong, *Member, IEEE*, Xiaojian Xu, *Member, IEEE*, Jason Hu, *Graduate Student Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{x} - (\boldsymbol{x}_k - \alpha \boldsymbol{P} \nabla f(\boldsymbol{x}_k))\|_2^2 + \frac{\beta}{\|\boldsymbol{P}^{-1}\|_2} R(\boldsymbol{x}).$$

- Hong et al., IEEE T-CI 2025 [9]

1630 IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 11, 2025

Using Randomized Nyström Preconditioners to Accelerate Variational Image Reconstruction

Tao Hong, *Member, IEEE*, Zhaoyi Xu, Jason Hu, *Graduate Student Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

- 2025-12-05 arXiv submission: (MRM submit) [10]

Fast and Robust Diffusion Posterior Sampling for MR Image Reconstruction Using the Preconditioned Unadjusted Langevin Algorithm

Moritz Blumenthal*[1], Tina Holliber*[1], Jonathan I. Tamir[2,3], and Martin Uecker[1,4]

-

- 2025-12-05 arXiv submission: (MRM submit) [10]

Fast and Robust Diffusion Posterior Sampling for MR Image Reconstruction Using the Preconditioned Unadjusted Langevin Algorithm

Moritz Blumenthal*[1], Tina Holliber*[1], Jonathan I. Tamir[2,3], and Martin Uecker[1,4]

-

$$\mathbf{x}_t^{k+1} = \mathbf{x}_t^k + \gamma M_t \left[A^H(\mathbf{y} - A\mathbf{x}_t^k) + \nabla_{\bar{\mathbf{x}}} \log p_t(\mathbf{x}_t^k) \right] + \sqrt{2\gamma}\mathbf{z}^k \quad \mathbf{z}^k \sim \mathbb{C}\mathcal{N}(\mathbf{0}, M_t^{-1}) \tag{8}$$

Regularization/Priors

- Priors or regularizers are essential for under-sampled data

$$\hat{\boldsymbol{x}}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x})$$

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

$$\mathrm{p}(\boldsymbol{x}) \propto \mathrm{e}^{-\beta R(\boldsymbol{x})}$$

- Modeling choices: $\mathrm{p}(\boldsymbol{x})$ and/or $\beta R(\boldsymbol{x})$
- Iterative optimization algorithms: choices/challenges

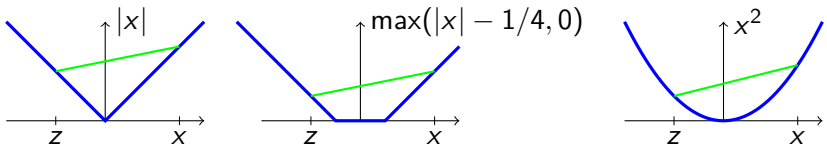
Convexity

Convex or non-convex regularizers?

- Convex function definition:

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{z}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{z}), \ \forall \boldsymbol{x}, \boldsymbol{z}, \ \forall 0 \leq \alpha \leq 1$$

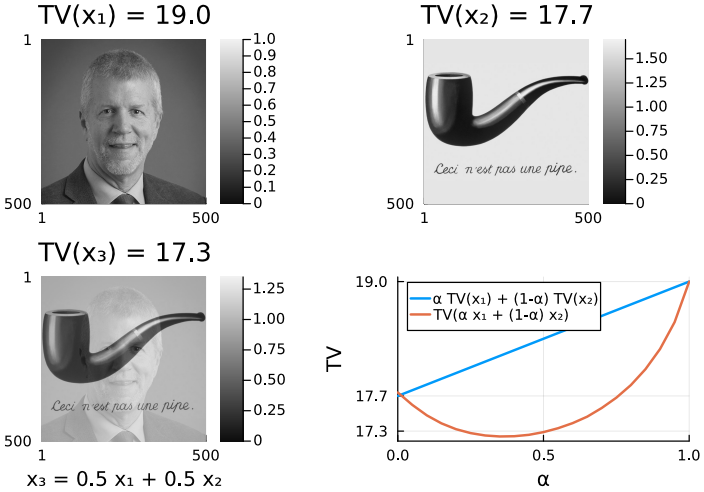
- 1D examples:



- All local minimizers are global minimizers.
- Popular convex regularizers in MRI:
  - Total variation (TV); in 1D: $R(\boldsymbol{x}) = \sum_n |x_n - x_{n-1}| = \|\boldsymbol{D}\boldsymbol{x}\|_1$, $\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & \dots \\ & \ddots & & \\ \dots & 0 & -1 & 1 \end{bmatrix}$
  - 1-norm of wavelet transform: $R(\boldsymbol{x}) = \|\boldsymbol{W}\boldsymbol{x}\|_1$, $\|\boldsymbol{v}\|_1 = \sum_i |v_i|$

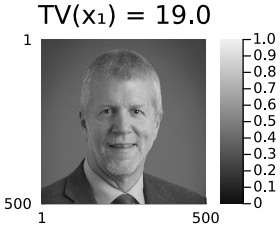
https://radiology.wisc.edu/profile/james-pipe-2758

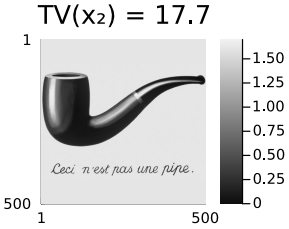
The Treachery of Images: René Magritte, 1929

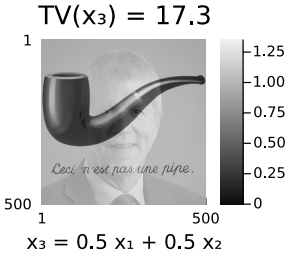


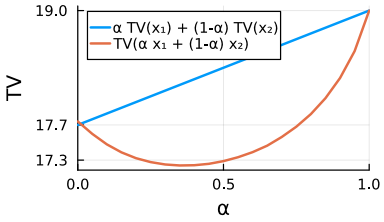
https://radiology.wisc.edu/profile/james-pipe-2758

The Treachery of Images: René Magritte, 1929





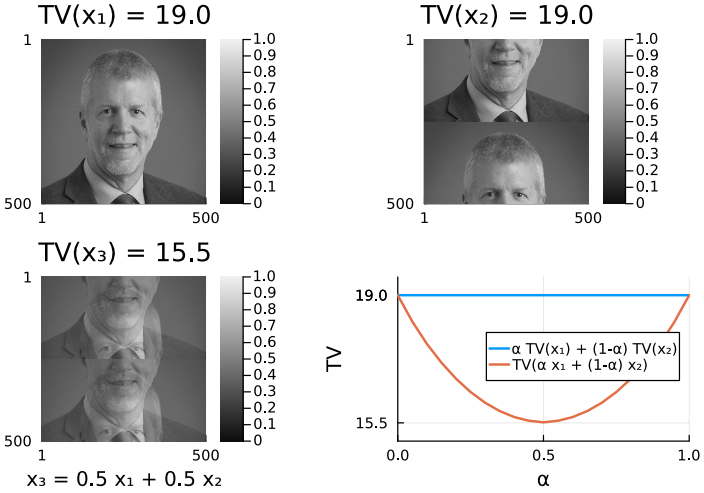




- TV vs "sparsity"?
- Inherent to convex regularizers
- Bayesian view: $p(\boldsymbol{x}) \propto e^{-\beta R(\boldsymbol{x})}$

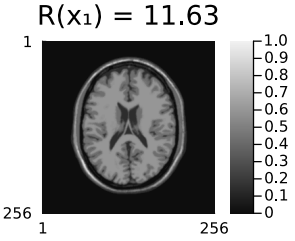
https://radiology.wisc.edu/profile/james-pipe-2758

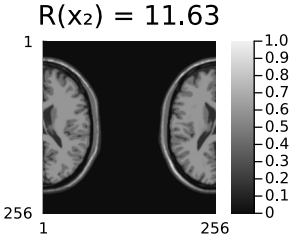
The Treachery of Images: René Magritte, 1929

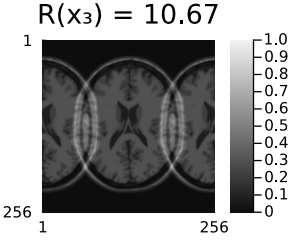


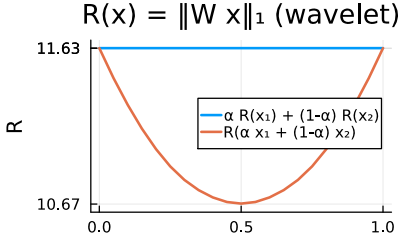
- TV vs "sparsity"?
- Inherent to convex regularizers
- Bayesian view: $p(\boldsymbol{x}) \propto e^{-\beta R(\boldsymbol{x})}$

Wavelet regularizer with 1-norm (of detail coefficients)



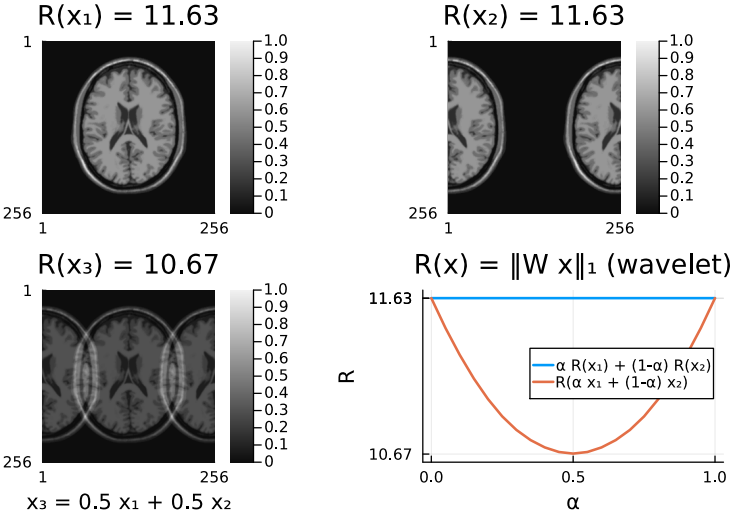






- BrainWeb phantom [11]
-

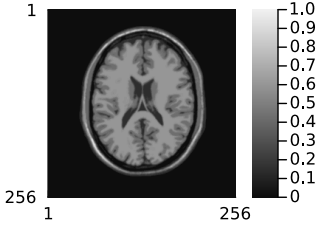
Wavelet regularizer with 1-norm (of detail coefficients)



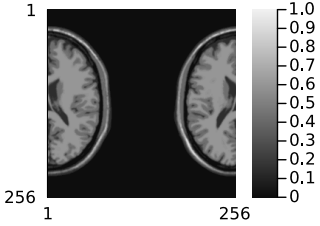
- BrainWeb phantom [11]
- Similar plots in Lobos et al. IEEE T-MI 2018 [12] for structured low-rank regularizer

Non-convex wavelet regularizer

R(x₁) = 90.01

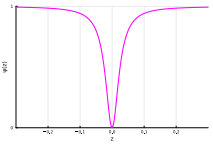


R(x₂) = 90.01

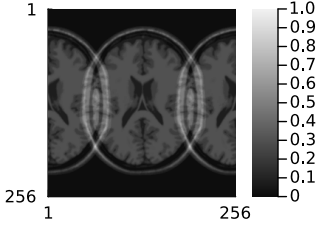


Geman/McClure potential [13]:

$$\psi(z) = \frac{|z/\delta|^2}{1+|z/\delta|^2}$$

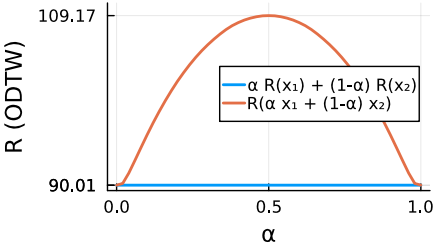


R(x₃) = 109.17



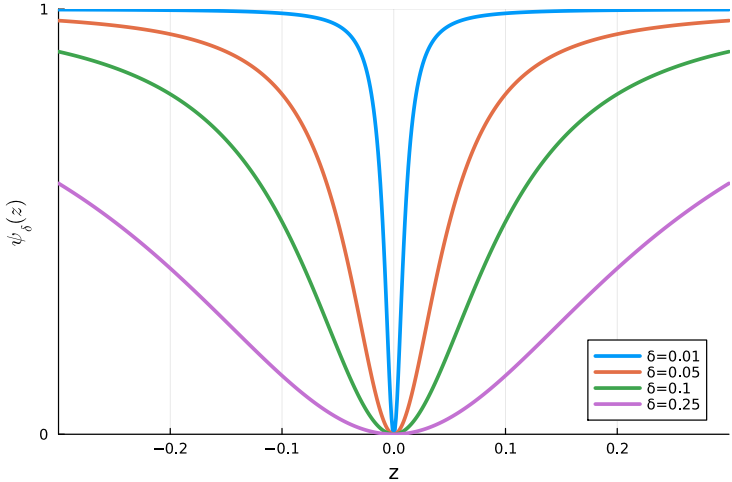
x₃ = 0.5 x₁ + 0.5 x₂

R(x): Geman/McClure



$$R(\boldsymbol{x}) = \sum_i \psi([\boldsymbol{W}\boldsymbol{x}]_i) \approx \|\boldsymbol{W}\boldsymbol{x}\|_0$$

Graduated non-convexity can aid optimization



- One convex run to get a decent initial $\boldsymbol{x}_0$
  followed by non-convex optimization to get $\hat{\boldsymbol{x}}$ often works fine [14]

Convexity take-away

- Convex regularizers cannot really measure "sparsity"
- Most neural network "regularizers" are non-convex
- Some notable exceptions:
  - Goujon et al., SIAM IS 2024 [15]
  - Chand & Jacob, ISMRM 2025 #5153 [16]
- Comparing a fancy deep network to TV/Wavelet with 1-norm is unfair?

Proximal operators

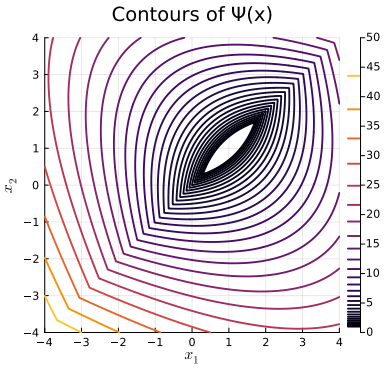
Proximal operators: Why

- Classical compressed sensing uses 1-norm regularizers:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{C}^N} \Psi(\boldsymbol{x}), \quad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1$$

  for some choices of matrix $\boldsymbol{T}$ (*e.g.*, finite differences or wavelets...)

- The 1-norm causes non-differentiable Ψ:
-



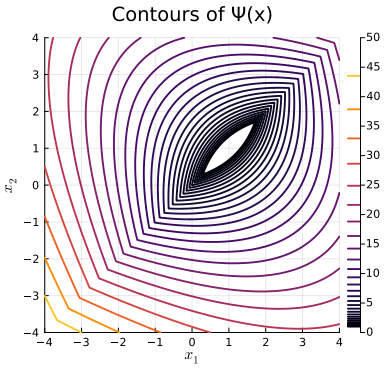
Proximal operators: Why

- Classical compressed sensing uses 1-norm regularizers:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{C}^N} \Psi(\boldsymbol{x}), \quad \Psi(\boldsymbol{x}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1$$

  for some choices of matrix $\boldsymbol{T}$ (*e.g.*, finite differences or wavelets...)

- The 1-norm causes non-differentiable Ψ:
- Options:
  - Use subgradient methods
    slow convergence
  - Round corner of absolute value
    - make Ψ differentiable
    - lose "sparsity" but...
  - Use proximal methods
    - convergence guarantees
    - closely related to many ML methods



Proximal operators

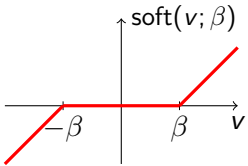
Given a function $f : \mathcal{X} \mapsto \mathbb{R}$ the **proximal operator** is

$$\text{prox}_f(\boldsymbol{v}) \triangleq \arg\min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{v}\|_{\mathcal{X}}^2 + f(\boldsymbol{x}) \right\}.$$

cf Fourier transform

Important and easy case:

$$f(\boldsymbol{x}) = \beta \|\boldsymbol{x}\|_1 \implies \text{prox}_f(\boldsymbol{v}) = \text{soft}_\cdot(\boldsymbol{v}, \beta)$$



Proximal operators

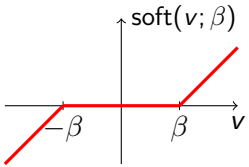
Given a function $f : \mathcal{X} \mapsto \mathbb{R}$ the **proximal operator** is

$$\text{prox}_f(\boldsymbol{v}) \triangleq \arg\min_{\boldsymbol{x} \in \mathcal{X}} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{v}\|_{\mathcal{X}}^2 + f(\boldsymbol{x}) \right\}.$$

cf Fourier transform

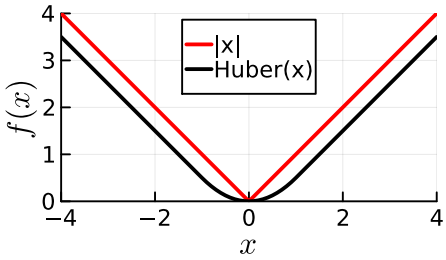
Important and easy case:

$$f(\boldsymbol{x}) = \beta \|\boldsymbol{x}\|_1 \implies \text{prox}_f(\boldsymbol{v}) = \text{soft}_\cdot(\boldsymbol{v}, \beta)$$

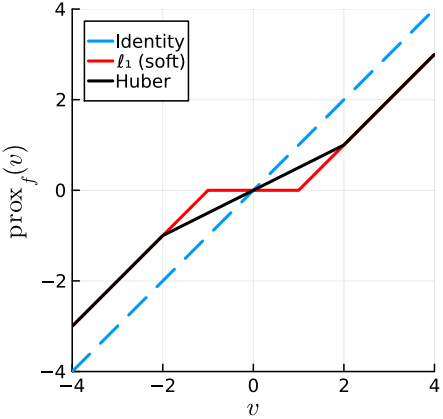


- flat region $\implies$ sparsity
- closely related to ReLU in NN models

Corner rounding: Huber function



Corner rounding $\Longrightarrow$ less sparsity



Proximal operator subtleties

- Easy case: for a **unitary matrix** $\boldsymbol{W}$ (*e.g.*, orthogonal wavelet transform):

$$f(\boldsymbol{x}) = \beta \|\boldsymbol{W}\boldsymbol{x}\|_1 \implies \text{prox}_f(\boldsymbol{v}) = \boldsymbol{W}' \,\text{soft}\,.(\boldsymbol{W}\boldsymbol{v}, \beta)$$

-

Proximal operator subtleties

- Easy case: for a **unitary matrix** $\boldsymbol{W}$ (*e.g.*, orthogonal wavelet transform):

$$f(\boldsymbol{x}) = \beta \left\| \boldsymbol{W}\boldsymbol{x} \right\|_1 \implies \text{prox}_f(\boldsymbol{v}) = \boldsymbol{W}' \, \text{soft}\,.(\boldsymbol{W}\boldsymbol{v}, \beta)$$

- The finite-difference matrix (*cf.* TV) $\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & \dots \\ & \ddots & \ddots & \\ \dots & 0 & -1 & 1 \end{bmatrix}$ is *not* unitary, so

$$f(\boldsymbol{x}) = \beta \left\| \boldsymbol{D}\boldsymbol{x} \right\|_1 \not\Rightarrow \text{prox}_f(\boldsymbol{v}) = \boldsymbol{D}' \, \text{soft}_{\cdot}(\boldsymbol{D}\boldsymbol{x}, \beta)\,.$$

(not hypothetical...)

- Iterative algorithms (*e.g.*, ADMM, dual methods) are needed for TV proximal mapping [17, 18].

Proximal gradient optimization methods

- Regularized image reconstruction problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left(\tfrac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}) \right)$$

- For prox-friendly regularizers, use (convergent!) proximal gradient method (PGM):

$$\boldsymbol{x}_{k+1} = \text{prox}_{\beta R}(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}))$$

- $R(\boldsymbol{x}) = \beta \|\boldsymbol{W}\boldsymbol{x}\|_1$ leads to Iterative Shrinkage-Thresholding Algorithm (ISTA) [19]:

$$\boldsymbol{x}_{k+1} = \mathcal{D}_\beta(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y})), \qquad \mathcal{D}_\beta(\boldsymbol{x}) = \text{prox}_{\beta\|\boldsymbol{W}\boldsymbol{x}\|_1}(\boldsymbol{x}) = \boldsymbol{W}' \,\text{soft}\,.(\boldsymbol{W}\boldsymbol{x}, \beta)$$

(Use FISTA [20] or OPTISTA [21] or POGM [22] with restart [23] instead, not ADMM! [24, 25])

-

Proximal gradient optimization methods

- Regularized image reconstruction problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left(\tfrac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}) \right)$$

- For prox-friendly regularizers, use (convergent!) proximal gradient method (PGM):

$$\boldsymbol{x}_{k+1} = \text{prox}_{\beta R}(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}))$$

- $R(\boldsymbol{x}) = \beta \|\boldsymbol{W}\boldsymbol{x}\|_1$ leads to Iterative Shrinkage-Thresholding Algorithm (ISTA) [19]:

$$\boldsymbol{x}_{k+1} = \mathcal{D}_\beta(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y})), \qquad \mathcal{D}_\beta(\boldsymbol{x}) = \text{prox}_{\beta\|\boldsymbol{W}\boldsymbol{x}\|_1}(\boldsymbol{x}) = \boldsymbol{W}' \, \text{soft}.(\boldsymbol{W}\boldsymbol{x}, \beta)$$

(Use FISTA [20] or OPTISTA [21] or POGM [22] with restart [23] instead, not ADMM! [24, 25])

- Denoising operation *cf.* plug-and-play methods, unrolled deep networks, etc.
  Replace $\mathcal{D}_\beta$ with ML denoiser (PNP survey [26])

Proximal gradient optimization methods

- Regularized image reconstruction problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left(\tfrac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}) \right)$$

- For prox-friendly regularizers, use (convergent!) proximal gradient method (PGM):

$$\boldsymbol{x}_{k+1} = \text{prox}_{\beta R}(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}))$$

- $R(\boldsymbol{x}) = \beta \|\boldsymbol{W}\boldsymbol{x}\|_1$ leads to Iterative Shrinkage-Thresholding Algorithm (ISTA) [19]:

$$\boldsymbol{x}_{k+1} = \mathcal{D}_\beta(\boldsymbol{x}_k - \alpha \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y})), \qquad \mathcal{D}_\beta(\boldsymbol{x}) = \text{prox}_{\beta\|\boldsymbol{W}\boldsymbol{x}\|_1}(\boldsymbol{x}) = \boldsymbol{W}' \,\text{soft}\,.(\boldsymbol{W}\boldsymbol{x}, \beta)$$

(Use FISTA [20] or OPTISTA [21] or POGM [22] with restart [23] instead, not ADMM! [24, 25])

- Denoising operation *cf.* plug-and-play methods, unrolled deep networks, etc.
  Replace $\mathcal{D}_\beta$ with ML denoiser (PNP survey [26])
- Convergence? [27–35]

1534 IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 11, 2025

Convergent Complex Quasi-Newton Proximal Methods for Gradient-Driven Denoisers in Compressed Sensing MRI Reconstruction

Tao Hong, *Member, IEEE*, Zhaoyi Xu, Se Young Chun, *Member, IEEE*, Luis Hernandez-Garcia, and Jeffrey A. Fessler, *Fellow, IEEE*

$$\boldsymbol{x}_{k+1} = \operatorname{prox}^{\boldsymbol{B}_k}_{\alpha_k f}(\boldsymbol{x}_k - \alpha_k \boldsymbol{P}_k \nabla R_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

1534 IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 11, 2025

Convergent Complex Quasi-Newton Proximal Methods for Gradient-Driven Denoisers in Compressed Sensing MRI Reconstruction

Tao Hong, *Member, IEEE*, Zhaoyi Xu, Se Young Chun, *Member, IEEE*, Luis Hernandez-Garcia, and Jeffrey A. Fessler, *Fellow, IEEE*

$$\boldsymbol{x}_{k+1} = \text{prox}_{\alpha_k f}^{\boldsymbol{B}_k}(\boldsymbol{x}_k - \alpha_k \boldsymbol{P}_k \nabla R_{\boldsymbol{\theta}}(\boldsymbol{x}))$$

- Preconditioning (iteration-dependent) $\boldsymbol{B}_k$, $\boldsymbol{P}_k$ via quasi-Newton rank-1 updates
- Non-convex learned regularizer $R_{\boldsymbol{\theta}}$
- (Weighted) proximal operator (uses adjoint of $\boldsymbol{A}$)
- Convergence

Low-rank models

Smooth optimization using global and local low-rank regularizers

Rodrigo A. Lobos[†*], Javier Salazar Cavazos[†], Raj Rao Nadakuditi[†], and Jeffrey A. Fessler[†]

- Differentiable tail singular value regularizer:

$$R(\boldsymbol{X}) = \sum_{k=K+1}^{r} \psi(\sigma_k(\boldsymbol{X}))$$

- Non-convex if $K > 0$ or if ψ is non-convex.
- Differentiable if ψ is.
- Nonlinear CG or quasi-Newton optimizers
- Reduced NRMSE LLR reconstructions for dynamic MRI
  Lobos et al., SIAM IS 2026 [36, 37]

Summary

- Many opportunities to "abuse math" in iterative image reconstruction field
  - Mismatched adjoint operators
  - DCF-based weighting that does not match the log-likelihood
  - Incorrect proximal mappings
  - Replacing proximal operators with deep networks with unknown properties
  - ...
-

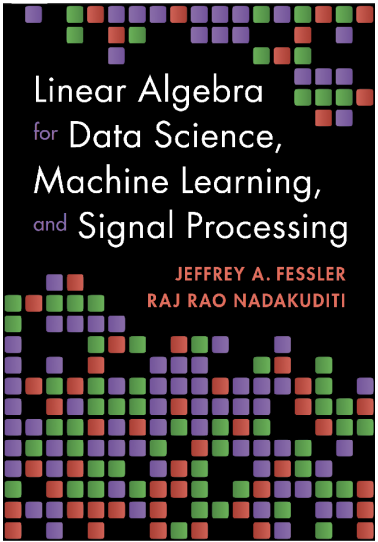
Summary

- Many opportunities to "abuse math" in iterative image reconstruction field
  - Mismatched adjoint operators
  - DCF-based weighting that does not match the log-likelihood
  - Incorrect proximal mappings
  - Replacing proximal operators with deep networks with unknown properties
  - ...
- Other topics:
  - Early stopping of CG as a means of regularization
  - Hidden cost function modifications like corner rounding
  - Using NRMSE and SSIM instead of task-relevant metrics
  - Inverse crimes [6, 38]
-

Summary

- Many opportunities to "abuse math" in iterative image reconstruction field
  - Mismatched adjoint operators
  - DCF-based weighting that does not match the log-likelihood
  - Incorrect proximal mappings
  - Replacing proximal operators with deep networks with unknown properties
  - ...
- Other topics:
  - Early stopping of CG as a means of regularization
  - Hidden cost function modifications like corner rounding
  - Using NRMSE and SSIM instead of task-relevant metrics
  - Inverse crimes [6, 38]
- Take-away
  - Mathematical foundations can also help develop ML-based IR methods with theoretical underpinnings
  - **May we live in interesting times**...

Book



- Online demos:
  `https://github.com/JeffFessler/book-la-demo`
- Topics include: low-rank matrix approximation, robust PCA, photometric stereo, video foreground/background separation, spectral clustering, matrix completion, ...
- Cambridge Univ. Press, 2024

Resources

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

Bibliography I

[1] Anonymous. "FAST-DIPS: Adjoint-free analytic steps and hard-constrained likelihood correction for diffusion-prior inverse problems." In: *Proc. Intl. Conf. on Learning Representations*. 2026.

[2] K. P. Pruessmann, M. Weiger, P. Boernert, and P. Boesiger. "Advances in sensitivity encoding with arbitrary k-space trajectories." In: *Mag. Res. Med.* 46.4 (Oct. 2001), 638–51.

[3] C. A. Baron, N. Dwork, J. M. Pauly, and D. G. Nishimura. "Rapid compressed sensing reconstruction of 3D non-Cartesian MRI." In: *Mag. Res. Med.* 79.5 (May 2018), 2685–92.

[4] R. E. Gabr, P. Aksit, P. A. Bottomley, A-B. M. Youssef, and Y. M. Kadah. "Deconvolution-interpolation gridding (DING): Accurate reconstruction for arbitrary k-space trajectories." In: *Mag. Res. Med.* 56.6 (Dec. 2006), 1182–91.

[5] F. Ong, M. Uecker, and M. Lustig. "Accelerating non-Cartesian MRI reconstruction convergence using k-space preconditioning." In: *IEEE Trans. Med. Imag.* 39.5 (2020), 1646–54.

[6] J. Kaipioa and E. Somersalo. "Statistical inverse problems: Discretization, model reduction and inverse crimes." In: *J. Comp. Appl. Math.* 198.2 (Jan. 2007), 493–504.

[7] M. Guerquin-Kern, D. V. D. Ville, K. Pruessmann, and M. Unser. "Analytical form of Shepp-Logan phantom for parallel MRI." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2010, 261–4.

[8] T. Hong, X. Xu, J. Hu, and J. A. Fessler. "Provable preconditioned plug-and-play approach for compressed sensing MRI reconstruction." In: *IEEE Trans. Computational Imaging* 10 (2024), 1476–88.

[9] T. Hong, Z. Xu, J. Hu, and J. A. Fessler. "Using randomized Nyström preconditioners to accelerate variational image reconstruction." In: *IEEE Trans. Computational Imaging* 11 (2025), 1630–43.

[10] M. Blumenthal, T. Holliber, J. I. Tamir, and M. Uecker. *Fast and robust diffusion posterior sampling for MR image reconstruction using the preconditioned unadjusted Langevin algorithm*. 2025.

[11] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans. "Design and construction of a realistic digital brain phantom." In: *IEEE Trans. Med. Imag.* 17.3 (June 1998), 463–8.

[12] R. A. Lobos, T. H. Kim, W. S. Hoge, and J. P. Haldar. "Navigator-free EPI ghost correction with structured low-rank matrix models: new theory and methods." In: *IEEE Trans. Med. Imag.* 37.11 (Nov. 2018), 2390–402.

[13] S. Geman and D. E. McClure. "Bayesian image analysis: an application to single photon emission tomography." In: *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.* 1985, 12–8.

[14] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and I. G. Zubal. "Bayesian reconstruction for emission tomography via deterministic annealing." In: *Information Processing in Medical Im.* Ed. by H H Barrett and A F Gmitro. Vol. 687. Lecture Notes in Computer Science. Berlin: Springer Verlag, 1993, pp. 322–38.

[15] A. Goujon, S. Neumayer, and M. Unser. "Learning weakly convex regularizers for convergent image-reconstruction algorithms." In: *SIAM J. Imaging Sci.* 17.1 (2024), 91–115.

[16] J. R. Chand and M. Jacob. "Convex implicit multi-scale (i-MuSE) energy framework: bridging compressed sensing and diffusion models." In: *Proc. Intl. Soc. Mag. Res. Med.* 2025, p. 5153.

[17] 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.

[18] L. Calatroni and A. Chambolle. "Backtracking strategies for accelerated descent methods with smooth composite objectives." In: *SIAM J. Optim.* 29.3 (Jan. 2019), 1772–98.

[19] 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.

[20] 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.

[21] U. Jang, S. D. Gupta, and E. K. Ryu. *Computer-assisted design of accelerated composite optimization methods: optISTA*. 2023.

Bibliography III

[22] 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.

[23] 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.

[24] 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.

[25] J. A. Fessler. "Optimization methods for MR image reconstruction." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 33–40.

[26] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg. "Plug-and-play methods for integrating physical and learned models in computational imaging: theory, algorithms, and applications." In: *IEEE Sig. Proc. Mag.* 40.1 (2023), 85–97.

[27] 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.

[28] Y. Sun, B. Wohlberg, and U. S. Kamilov. "An online plug-and-play algorithm for regularized image reconstruction." In: *IEEE Trans. Computational Imaging* 5.3 (Sept. 2019), 395–408.

[29] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin. "Plug-and-play methods provably converge with properly trained denoisers." In: *Proc. Intl. Conf. Mach. Learn.* 2019.

[30] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov. "Provable convergence of plug-and-play priors with MMSE denoisers." In: *IEEE Signal Proc. Letters* 27 (2020), 1280–4.

[31] P. Nair, R. Gavaskar, and K. Chaudhury. "Fixed-point and objective convergence of plug-and-play algorithms." In: *IEEE Trans. Computational Imaging* 7 (2021), 337–48.

[32] Y. Sun, Z. Wu, X. Xu, B. Wohlberg, and U. S. Kamilov. "Scalable plug-and-play ADMM With convergence guarantees." In: *IEEE Trans. Computational Imaging* 7 (2021), 849–63.

Bibliography IV

[33] H. Y. Tan, S. Mukherjee, J. Tang, and C-B. Schonlieb. "Provably convergent plug-and-play quasi-Newton methods." In: *SIAM J. Imaging Sci.* 17.2 (2024), 785–819.

[34] S. Hurault, A. Chambolle, A. Leclaire, and N. Papadakis. "Convergent plug-and-play with proximal denoiser and unconstrained regularization parameter." In: *J. Math. Im. Vision* 66.4 (June 2024), 616–38.

[35] T. Hong, Z. Xu, S. Y. Chun, L. Hernandez-Garcia, and J. A. Fessler. "Convergent complex quasi-Newton proximal method for gradient-driven denoisers in compressed sensing MRI reconstruction." In: *IEEE Trans. Computational Imaging* 11 (2025), 1534–47.

[36] R. A. Lobos, J-S. Cavazos, R. R. Nadakuditi, and J. A. Fessler. "Smooth optimization using global and local low-rank regularizers." In: *SIAM J. Imaging Sci.* (2026). To appear.

[37] R. A. Lobos, J. S. Cavazos, R. R. Nadakuditi, and J. A. Fessler. *Smooth optimization algorithms for global and locally low-rank regularizers*. 2025.

[38] E. Shimron, J. I. Tamir, K. Wang, and M. Lustig. "Implicit data crimes: Machine learning bias arising from misuse of public data." In: *Proc. Natl. Acad. Sci.* 119.13 (Mar. 2022), e2117203119.