Jeffrey A. Fessler

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

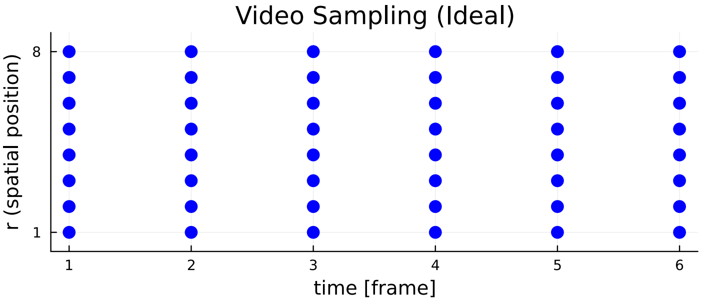
http://web.eecs.umich.edu/~fessler

BIRS-UBC-O: Leveraging Model- and Data-Driven Methods in Medical Imaging
2023-06-29

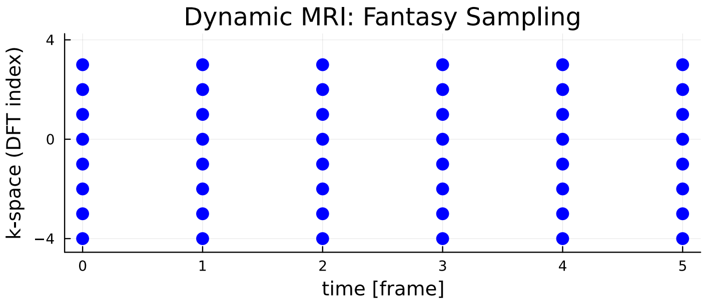
Acknowledgments:
Caroline Crockett, Raj Nadakuditi, Rodrigo A. Lobos, Javier Salazar Cavazos

Outline

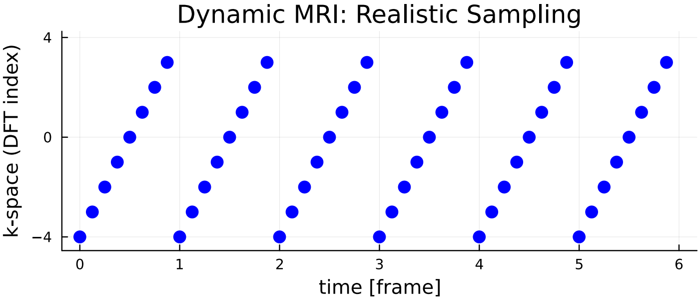
- Video: sampling in real space



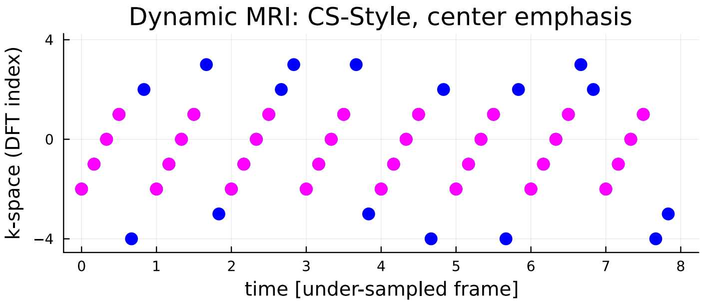
- MRI: sampling in k-space (k-t sampling)



- MRI: sampling in k-space (k-t sampling)

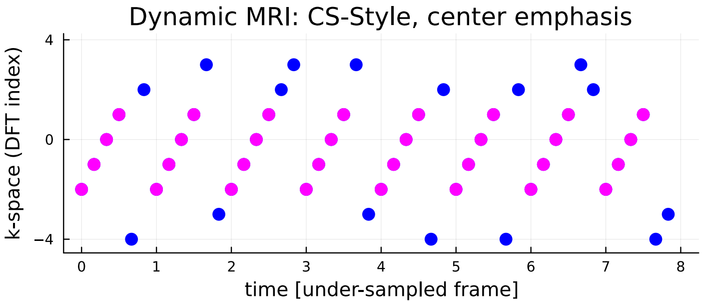


- MRI: sampling in k-space (k-t sampling)



-

- MRI: sampling in k-space (k-t sampling)



- better temporal resolution, but spatially under-sampled

- Measurement model:

$$\boldsymbol{y}_t = \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T$$

$\boldsymbol{y}_t \in \mathbb{C}^{M_t}$: k-space data for tth frame
$\boldsymbol{x}_t \in \mathbb{C}^{N}$: latent image for tth frame (vec of 2D or 3D array)
$\boldsymbol{A}_t \in \mathbb{C}^{M_t \times N}$: forward model for tth frame

- Stack data: $\boldsymbol{y} \triangleq \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_T \end{bmatrix}, \ \boldsymbol{\varepsilon} \triangleq \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{bmatrix} \in \mathbb{C}^M, \quad M \triangleq \sum_{t=1}^{T} M_t$
- Latent space-time matrix $\boldsymbol{X} \triangleq \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_T \end{bmatrix} \in \mathbb{C}^{N \times T}$
- Linear forward model:

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{\varepsilon}, \qquad \mathcal{A} : \mathbb{C}^{N \times T} \mapsto \mathbb{C}^M$$

- Goal: estimate $\boldsymbol{X}$ from $\boldsymbol{y}$ given $\mathcal{A}$
- Under-determined $M < N\,T$, so regularization is essential

Outline

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank:

$$\hat{\boldsymbol{X}} = \operatorname*{arg\,min}_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{X}\|_*$$

$\|\boldsymbol{X}\|_*$: nuclear norm

-

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{X}\|_*$$

  $\|\boldsymbol{X}\|_*$: nuclear norm

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank + temporally sparse:

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{L}} + \hat{\boldsymbol{S}}, \qquad (\hat{\boldsymbol{L}}, \hat{\boldsymbol{S}}) = \arg\min_{\boldsymbol{L},\boldsymbol{S}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{L} + \boldsymbol{S}) - \boldsymbol{y}\|_2^2 + \beta_1 \|\boldsymbol{L}\|_* + \beta_2 \|\boldsymbol{S}\boldsymbol{T}\|_{1,1}$$

  for some temporal sparsifying transform $\boldsymbol{T}$
  $\|\boldsymbol{X}\|_{1,1} = \|\mathrm{vec}(\boldsymbol{X})\|_1$

-

Dynamic MRI recon using global low-rank models

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{X}\|_*$$

  $\|\boldsymbol{X}\|_*$: nuclear norm

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank + temporally sparse:

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{L}} + \hat{\boldsymbol{S}}, \qquad (\hat{\boldsymbol{L}}, \hat{\boldsymbol{S}}) = \arg\min_{\boldsymbol{L},\boldsymbol{S}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{L} + \boldsymbol{S}) - \boldsymbol{y}\|_2^2 + \beta_1 \|\boldsymbol{L}\|_* + \beta_2 \|\boldsymbol{S}\boldsymbol{T}\|_{1,1}$$

  for some temporal sparsifying transform $\boldsymbol{T}$
  $\|\boldsymbol{X}\|_{1,1} = \|\mathrm{vec}(\boldsymbol{X})\|_1$

- Both easily solved by proximal optimized gradient method (POGM)
  (A. Taylor et al., SIAM JO, 2017) [1] with adaptive restart (D. Kim & JF, JOTA 2018) [2]
  C. Lin & JF, IEEE T-CI 2019 [3]

-

Dynamic MRI recon using global low-rank models

- If $\boldsymbol{X}$ is assumed to be (globally) low-rank:

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}}{\arg\min} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{X}\|_*$$

  $\|\boldsymbol{X}\|_*$: nuclear norm
- If $\boldsymbol{X}$ is assumed to be (globally) low-rank + temporally sparse:

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{L}} + \hat{\boldsymbol{S}}, \qquad (\hat{\boldsymbol{L}}, \hat{\boldsymbol{S}}) = \underset{\boldsymbol{L},\boldsymbol{S}}{\arg\min} \frac{1}{2} \|\mathcal{A}(\boldsymbol{L} + \boldsymbol{S}) - \boldsymbol{y}\|_2^2 + \beta_1 \|\boldsymbol{L}\|_* + \beta_2 \|\boldsymbol{S}\boldsymbol{T}\|_{1,1}$$

  for some temporal sparsifying transform $\boldsymbol{T}$
  $\|\boldsymbol{X}\|_{1,1} = \|\mathrm{vec}(\boldsymbol{X})\|_1$
- Both easily solved by proximal optimized gradient method (POGM) (A. Taylor et al., SIAM JO, 2017) [1] with adaptive restart (D. Kim & JF, JOTA 2018) [2]
  C. Lin & JF, IEEE T-CI 2019 [3]
- Both “data driven” because temporal basis from data learned

- Composite cost function:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \Psi(\boldsymbol{X}), \quad \Psi(\boldsymbol{X}) = f(\boldsymbol{X}) + g(\boldsymbol{X}),$$

$$\underbrace{f(\boldsymbol{X}) = \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2\,,}_{\text{smooth}} \quad \underbrace{g(\boldsymbol{X}) = \beta \|\boldsymbol{X}\|_*,}_{\text{prox friendly}} \quad \underbrace{L_{\nabla f} = \|\mathcal{A}^* \mathcal{A}\|_2}_{\text{Lipschitz constant}}$$

- Proximal gradient method (PGM) / Iterative soft thresholding algorithm (ISTA):

$$\boldsymbol{X}_{k+1} = \mathrm{SVST}\left(\boldsymbol{X}_k - \frac{1}{L_{\nabla f}} \nabla f(\boldsymbol{X}_k), \frac{\beta}{L_{\nabla f}}\right), \quad \nabla f(\boldsymbol{X}_k) = \mathcal{A}^* \left(\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\right)$$

- Singular value soft thresholding (SVST):

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{Diag}\{\sigma_k\} \boldsymbol{V}' \Longrightarrow \mathrm{SVST}(\boldsymbol{X}, \gamma) = \operatorname{prox}_{\gamma \|\cdot\|_*}(\boldsymbol{X}) = \boldsymbol{U} \operatorname{Diag}\{[\sigma_k - \gamma]_+\} \boldsymbol{V}'$$

- FISTA and POGM are similar, with momentum terms

OGM extension for composite problems by Taylor et al. [1]:

Proximal optimized gradient method (POGM)

Input: $F^{(1)} \in \mathcal{F}_{0,L}(\mathbb{E})$, $F^{(2)} \in \mathcal{F}_{0,\infty}(\mathbb{E})$, $x_0 \in \mathbb{E}$, $y_0 = x_0$, $\theta_0 = 1$.

For $k = 1 : N$

$$y_k = x_{k-1} - \frac{1}{L} B^{-1} \nabla F^{(1)}(x_{k-1})$$

$$z_k = y_k + \frac{\theta_{k-1} - 1}{\theta_k}(y_k - y_{k-1}) + \frac{\theta_{k-1}}{\theta_k}(y_k - x_{k-1}) + \frac{\theta_{k-1} - 1}{L\gamma_{k-1}\theta_k}(z_{k-1} - x_{k-1})$$

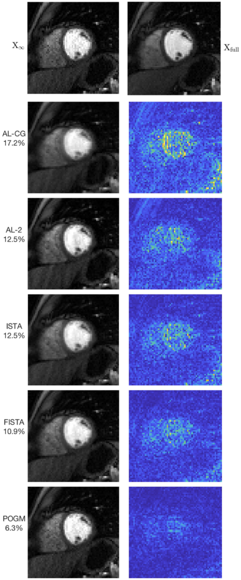
$$x_k = \operatorname{prox}_{\gamma_k F^{(2)}}(z_k)$$

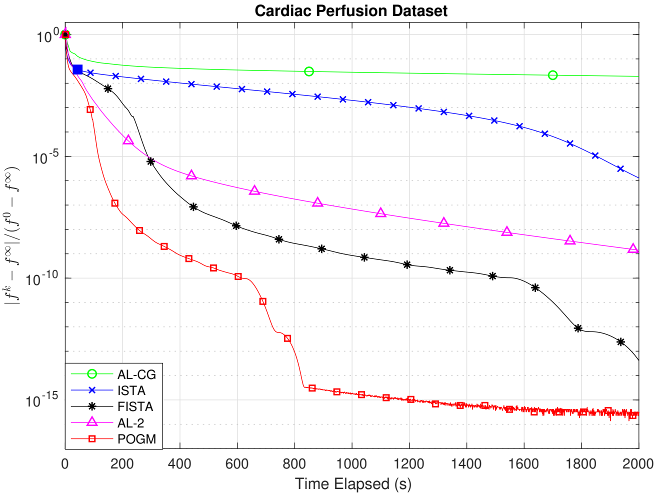
In this algorithm, we use the sequence $\gamma_k = \frac{1}{L}\frac{2\theta_{k-1}+\theta_k-1}{\theta_k}$ and the inertial coefficients proposed in [23]:

$$\theta_k = \begin{cases} \frac{1+\sqrt{4\theta_{k-1}^2+1}}{2}, & i \le N-1, \\ \frac{1+\sqrt{8\theta_{k-1}^2+1}}{2}, & i = N. \end{cases}$$

Simply trying to generalize OGM using the standard proximal step on the primary sequence $\{y_i\}$ (as for FPGM1) does not lead to a converging algorithm. We obtained

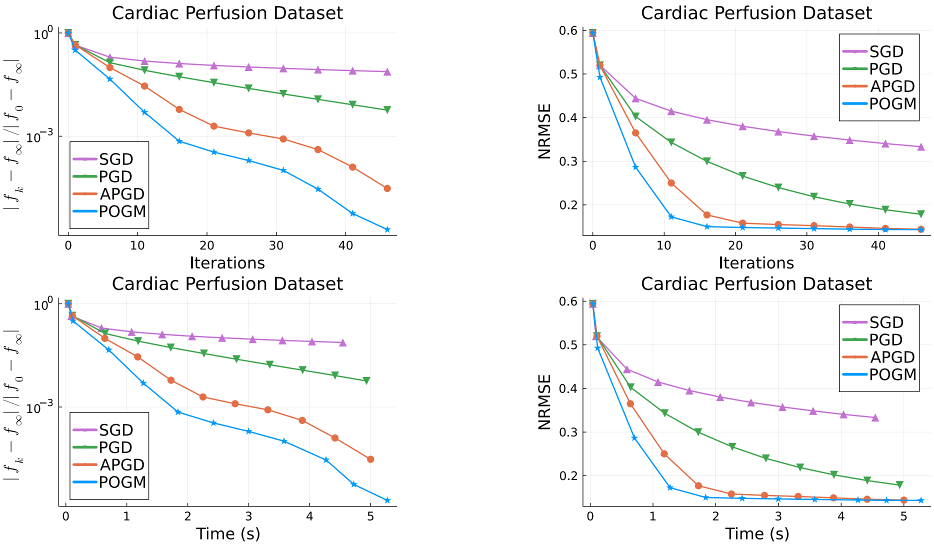
Global low-rank + sparse results





C. Lin & JF, IEEE T-CI 2019 [3]

Global low-rank results



POGM works well in practice, corroborating “worst-case” bounds

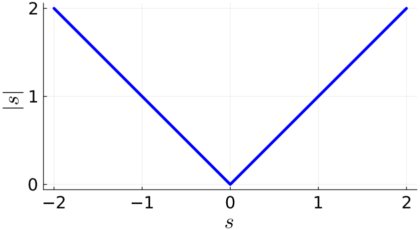
To smooth or not smooth?

- Nuclear norm regularizer is a non-smooth function of $\boldsymbol{X}$:

$$\|\boldsymbol{X}\|_* = \sum_k \sigma_k(\boldsymbol{X})$$

Singular value of 1×1 matrix $[s]$ is $\sigma_1([s]) = |s|$

- Requires "complicated" methods like accelerated first-order proximal gradient methods, ADMM, ...
-

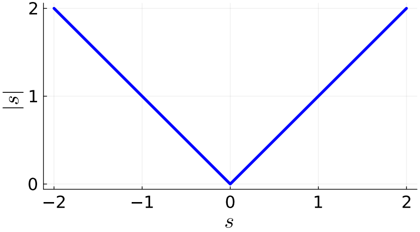


- Nuclear norm regularizer is a non-smooth function of $\boldsymbol{X}$:

$$\|\boldsymbol{X}\|_* = \sum_k \sigma_k(\boldsymbol{X})$$

Singular value of 1×1 matrix $[s]$ is $\sigma_1([s]) = |s|$

- Requires "complicated" methods like accelerated first-order proximal gradient methods, ADMM, ...
- $\|\boldsymbol{X}\|_*$ is a relaxation of rank$\{\boldsymbol{X}\}$

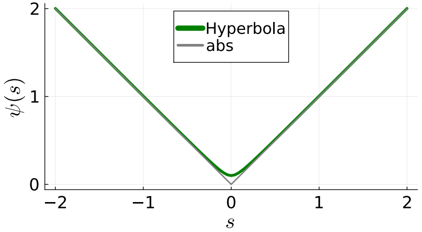


To smooth or not smooth?

- Smooth regularizer:

$$R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- where ψ satisfies Huber's conditions [4]:
  - $\psi(-s) = \psi(s)$
  - ψ differentiable
  - $\omega_\psi(s) \triangleq \dot{\psi}(s) \,/\, s$ is bounded
    $\Longrightarrow \dot{\psi}$ is Lipschitz
-



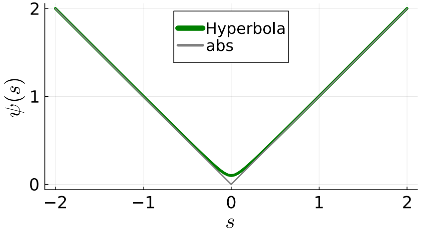
- hyperbola: $\psi(s) = \sqrt{s^2 + \delta^2}$
-

To smooth or not smooth?

- Smooth regularizer:

$$R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- where ψ satisfies Huber's conditions [4]:
  - $\psi(-s) = \psi(s)$
  - ψ differentiable
  - $\omega_\psi(s) \triangleq \dot{\psi}(s) \,/ s$ is bounded
    $\implies \dot{\psi}$ is Lipschitz
- Enables gradient-based optimization algorithms like nonlinear conjugate gradient (CG) and quasi-Newton
- Faster rate? [5]



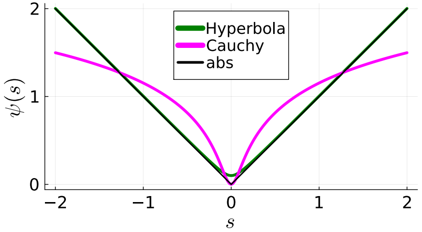
- hyperbola: $\psi(s) = \sqrt{s^2 + \delta^2}$
-

To smooth or not smooth?

- Smooth regularizer:

$$R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- where ψ satisfies Huber's conditions [4]:
  - $\psi(-s) = \psi(s)$
  - ψ differentiable
  - $\omega_\psi(s) \triangleq \dot{\psi}(s)/s$ is bounded
    $\Longrightarrow \dot{\psi}$ is Lipschitz
- Enables gradient-based optimization algorithms like nonlinear conjugate gradient (CG) and quasi-Newton
- Faster rate? [5]



- hyperbola: $\psi(s) = \sqrt{s^2 + \delta^2}$
- Graduated non-convexity?

Smooth regularizer:

$$R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- Convexity: ψ convex $\Longrightarrow$ R convex [6]
- Gradient: (A. Lewis, J. Convex. Analysis, 1995) [7, 8]:

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{Diag}\{\boldsymbol{\sigma}\} \boldsymbol{V}'$$

$$\Longrightarrow \nabla R(\boldsymbol{X}) = \boldsymbol{U} \operatorname{Diag}\left\{\dot{\psi}_{\cdot}(\boldsymbol{\sigma})\right\} \boldsymbol{V}'$$

-

Regularizer properties

Smooth regularizer:

$$R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- Convexity: ψ convex $\Longrightarrow$ R convex [6]
- Gradient: (A. Lewis, J. Convex. Analysis, 1995) [7, 8]:

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{Diag}\{\boldsymbol{\sigma}\} \boldsymbol{V}'$$

$$\Longrightarrow \nabla R(\boldsymbol{X}) = \boldsymbol{U} \operatorname{Diag}\left\{\dot{\psi}_{\cdot}(\boldsymbol{\sigma})\right\} \boldsymbol{V}'$$

- Smoothness Theorem: Lipschitz constant

$$L_{\nabla R} = \omega_{\psi}(0) = \ddot{\psi}(0) .$$

Proof builds on Qi & Yang, SIAM J. MAA 2003 [9]

- Smooth cost function:

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}}{\arg\min}\, \Psi(\boldsymbol{X}), \quad \Psi(\boldsymbol{X}) \triangleq \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{X}), \quad R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- Gradient (A. Lewis, J. Convex. Analysis, 1995) [7]:

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{Diag}\{\sigma_k\} \boldsymbol{V}' \Longrightarrow \nabla \Psi(\boldsymbol{X}) = \mathcal{A}^* (\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}) + \beta \underbrace{\sum_k \dot{\psi}(\sigma_k(\boldsymbol{X}))\, \boldsymbol{u}_k \boldsymbol{v}_k'}_{\nabla R(\boldsymbol{X})}$$

-

- Smooth cost function:

$$\hat{\boldsymbol{X}} = \operatorname*{arg\,min}_{\boldsymbol{X}} \Psi(\boldsymbol{X}), \quad \Psi(\boldsymbol{X}) \triangleq \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{X}), \quad R(\boldsymbol{X}) = \sum_k \psi(\sigma_k(\boldsymbol{X}))$$

- Gradient (A. Lewis, J. Convex. Analysis, 1995) [7]:

$$\boldsymbol{X} = \boldsymbol{U} \operatorname{Diag}\{\sigma_k\} \boldsymbol{V}' \Longrightarrow \nabla \Psi(\boldsymbol{X}) = \mathcal{A}^* (\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}) + \beta \underbrace{\sum_k \dot{\psi}(\sigma_k(\boldsymbol{X})) \, \boldsymbol{u}_k \, \boldsymbol{v}_k'}_{\nabla R(\boldsymbol{X})}$$

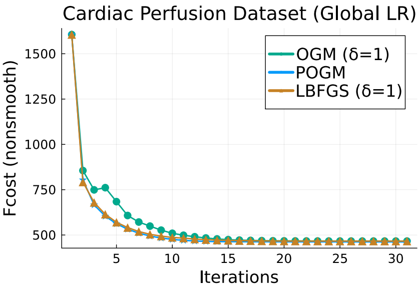
- Line search function:

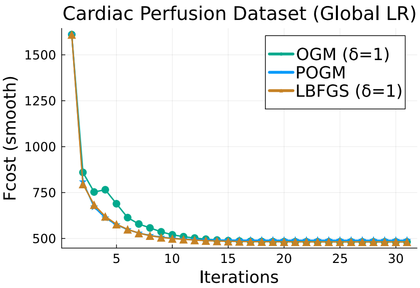
$$h(\alpha) \triangleq \Psi(\boldsymbol{X} + \alpha \boldsymbol{\Delta}), \qquad \dot{h}(\alpha) = \mathcal{A}(\boldsymbol{\Delta})' (\mathcal{A}(\boldsymbol{X} + \alpha \boldsymbol{\Delta}) - \boldsymbol{y}) + \beta \, \langle \nabla R(\boldsymbol{X} + \alpha \boldsymbol{\Delta}), \, \boldsymbol{\Delta} \rangle_{\mathrm{F}}$$

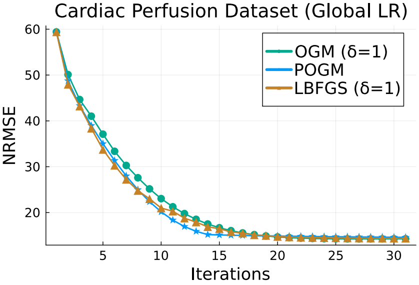
- Theorem: h is smooth; $\dot{h}$ has Lipschitz constant

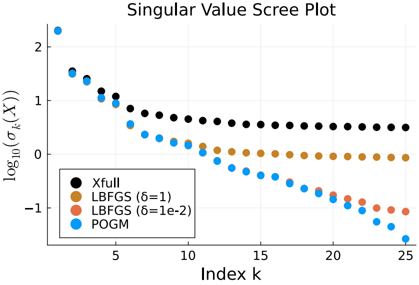
$$L_{\dot{h}} = \|\mathcal{A}(\boldsymbol{\Delta})\|_2^2 + \beta \, \omega_\psi(0) \, \|\boldsymbol{\Delta}\|_{\mathrm{F}}^2$$

Smooth regularizer results



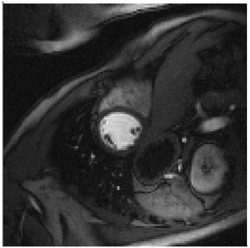




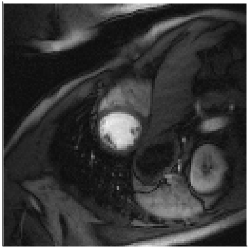


Global low-rank images

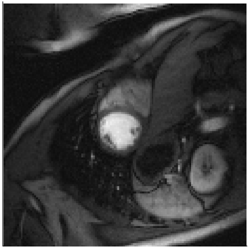
Fully Sampled Ref.



POGM - NRMSE: 14.6 %

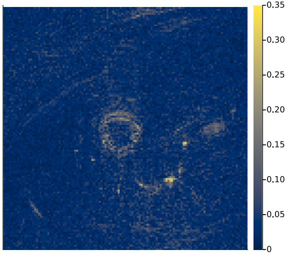


LBFGS - NRMSE: 14.4 %

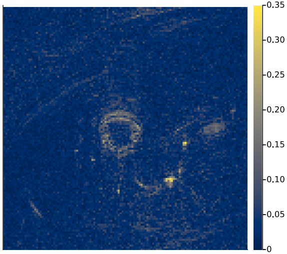


Initial Image

POGM - NRMSE: 14.6 %



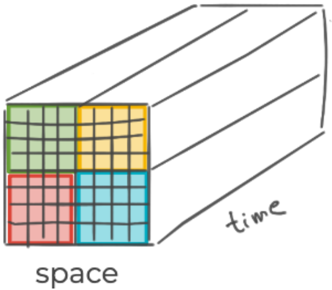
LBFGS - NRMSE: 14.4 %



- If space-time patches of $\boldsymbol{X}$ are assumed to be (locally) low-rank:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_{p=1}^{P} \|\mathcal{P}_p(\boldsymbol{X})\|_*$$

  $\mathcal{P}_p$: picks the pth space-time patch from $\boldsymbol{X}$
- Low-rank modeling may be more reasonable locally
-

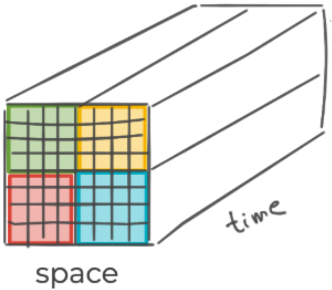


Dynamic MRI recon using local low-rank models

- If space-time patches of $\boldsymbol{X}$ are assumed to be (locally) low-rank:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_{p=1}^{P} \|\mathcal{P}_p(\boldsymbol{X})\|_*$$

  $\mathcal{P}_p$: picks the pth space-time patch from $\boldsymbol{X}$
- Low-rank modeling may be more reasonable locally
- Numerous applications, *e.g.*:
  - matrix approximation [10]
  - dynamic MRI [11, 12]
  - multi-contrast and quantitative MRI [13–15]
  - MR denoising [16]
  - MR motion correction [17]
  - fMRI denoising [18–20]
  - fMRI dynamic image reconstruction [21]



To overlap or not?

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_{p=1}^{P} \|\mathcal{P}_p(\boldsymbol{X})\|_*$$

- Non-overlapping patches
  - prox-friendly: separate SVST for each space-time patch
  - Suitable for proximal gradient methods like POGM *only* in this case
-

To overlap or not?

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_s \sum_{p=1}^{P} \|\mathcal{P}_p(\mathrm{Shift}_s(\boldsymbol{X}))\|_*$$

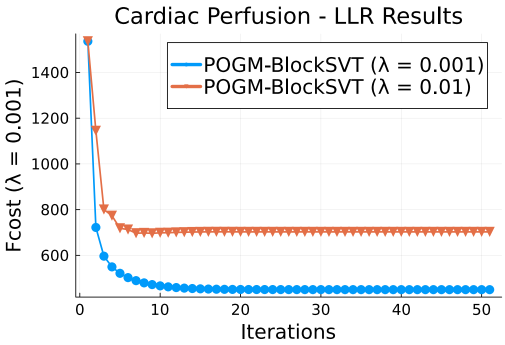
- Non-overlapping patches
  - prox-friendly: separate SVST for each space-time patch
  - Suitable for proximal gradient methods like POGM *only* in this case
- Overlapping patches (stride $<$ patch size)
  - shift invariant (if stride $= 1$) $\Longrightarrow$ no block artifacts
  - no (known) simple proximal operator
  -

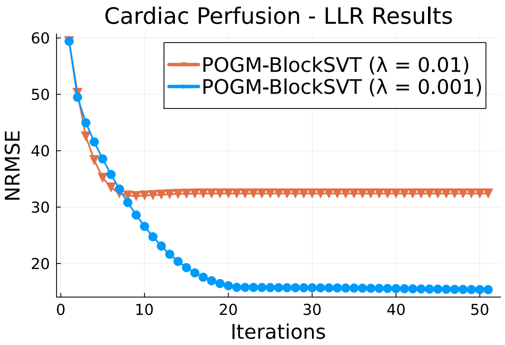
To overlap or not?

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_{s} \sum_{p=1}^{P} \|\mathcal{P}_p(\mathrm{Shift}_s(\boldsymbol{X}))\|_*$$

- Non-overlapping patches
  - prox-friendly: separate SVST for each space-time patch
  - Suitable for proximal gradient methods like POGM *only* in this case
- Overlapping patches (stride $<$ patch size)
  - shift invariant (if stride $= 1$) $\implies$ no block artifacts
  - no (known) simple proximal operator
  - Optimization algorithm options:
    - subgradient descent
    - cycle spinning approximation (akin to stochastic proximal gradient method)
    - proximal averaging [22, 23]: prox of sum $\approx$ sum of prox ?
    - ADMM with *numerous* auxiliary variables

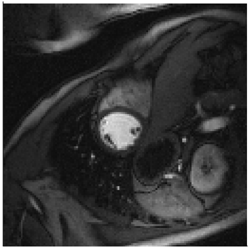
Locally low-rank results: non-overlapping patches



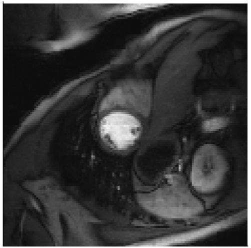


Locally low-rank results: non-overlapping patches

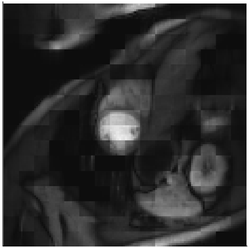
Fully Sampled Ref.



POGM (λ = 0.001)

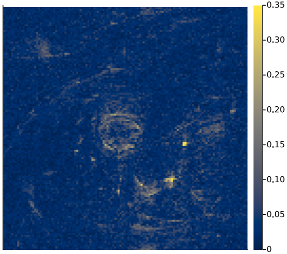


POGM (λ = 0.01)

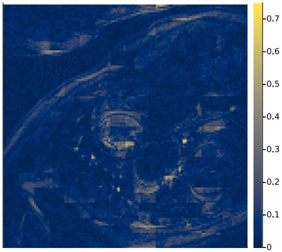


Initial Image

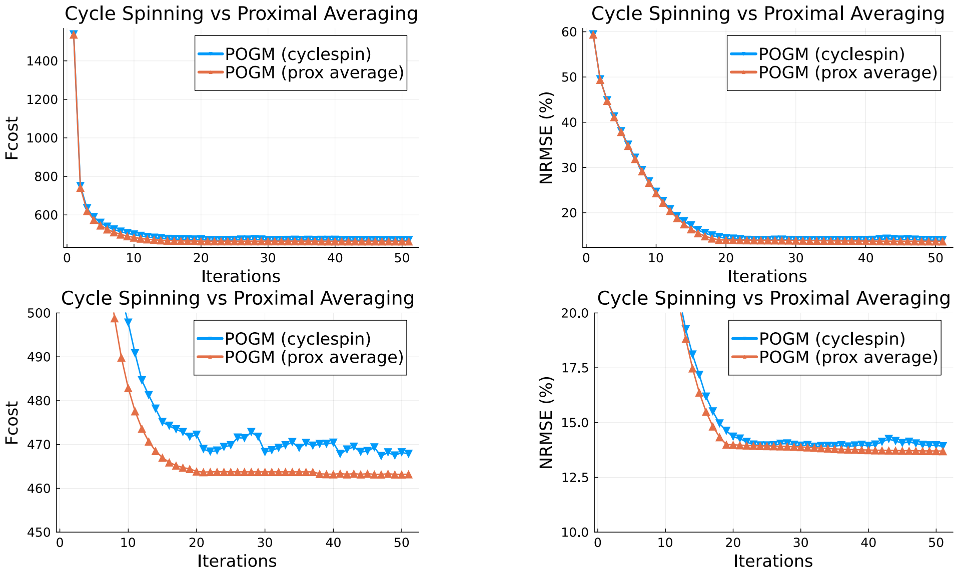
POGM - NRMSE: 15.4 %



POGM - NRMSE: 32.4 %

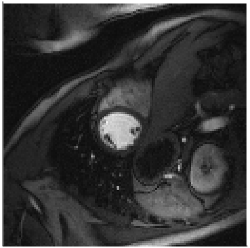


POGM with ad hoc LLR modifications

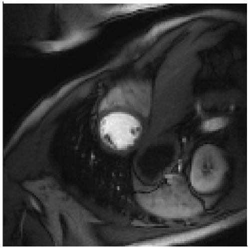


POGM for LLR (overlapping patches) with proximal average

Fully Sampled Ref.

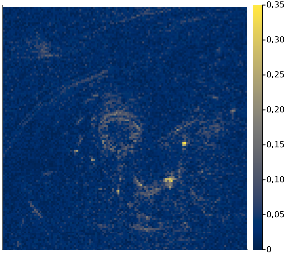


POGM - NRMSE 13.7 %



Initial Image

POGM - NRMSE 13.7 %



$$\hat{\boldsymbol{X}} = \operatorname*{arg\,min}_{\boldsymbol{X}} \Psi(\boldsymbol{X}), \quad \Psi(\boldsymbol{X}) \triangleq \frac{1}{2} \|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta \sum_{p=1}^{P} R_p(\boldsymbol{X}), \quad R_p(\boldsymbol{X}) \triangleq \sum_k \psi(\sigma_k(\mathcal{P}_p(\boldsymbol{X})))$$

Now we can easily apply gradient-based methods like CG & quasi-Newton:

- Gradient (corollary to previous theorem):

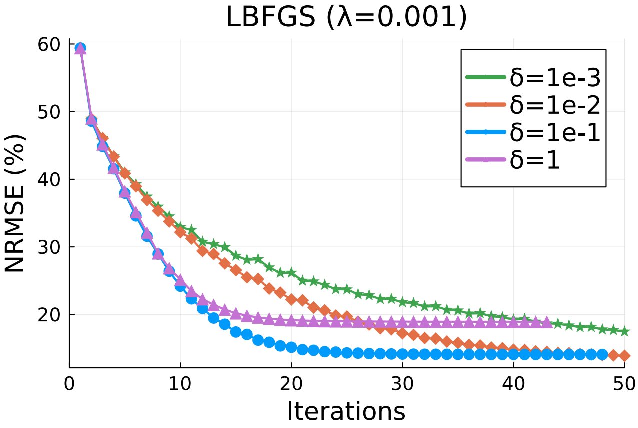
$$\nabla \Psi(\boldsymbol{X}) = \mathcal{A}^*(\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}) + \beta \sum_{p=1}^{P} \boldsymbol{U}_p \operatorname{Diag}\left\{\dot{\psi}.(\boldsymbol{\sigma}_p)\right\} \boldsymbol{V}_p', \quad \mathcal{P}_p(\boldsymbol{X}) = \boldsymbol{U}_p \operatorname{Diag}\{\boldsymbol{\sigma}_p\} \boldsymbol{V}_p'$$

- Lipschitz constant (for line-search step):

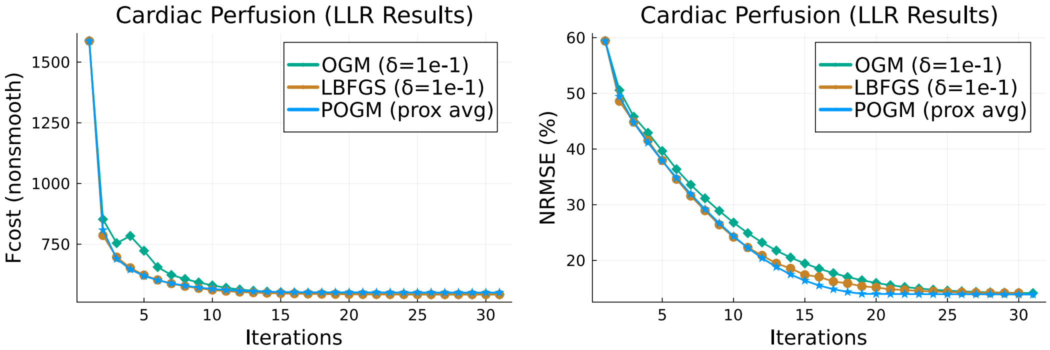
$$L_{\nabla \Psi} = \|\mathcal{A}^* \mathcal{A}\|_2 + \beta P \, \omega_{\psi}(0)$$
$$L_{\dot{h}} = L_{\nabla \Psi} \|\boldsymbol{\Delta}\|_{\mathrm{F}}^2$$

Tuning smoothness parameter

Hyperbola potential function: $\psi(\sigma) = \sqrt{\sigma^2 + \delta^2} \approx |\sigma|$ for $|\sigma| \gg \delta$

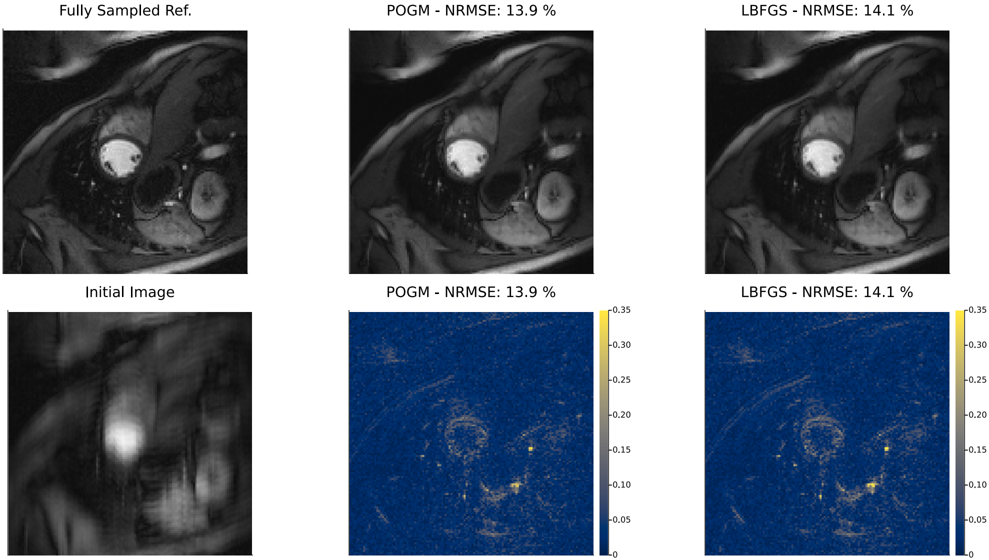


Smooth LLR results



- POGM with proximal averaging descends, but to what?
- Room to optimize LBFGS

LLR images


Fully Sampled Ref.
POGM - NRMSE: 13.9 %
LBFGS - NRMSE: 14.1 %
Initial Image
POGM - NRMSE: 13.9 %
LBFGS - NRMSE: 14.1 %
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0

Outline

Summary / future directions

- Smooth approximation to nuclear norm
  - leads to similar dynamic image reconstruction results
  - Lipschitz gradient enables efficient convex optimization with convergence
  - POGM with cycle-spinning or proximal averaging works unexpectedly well but what convergence theory?
-

Summary / future directions

- Smooth approximation to nuclear norm
  - leads to similar dynamic image reconstruction results
  - Lipschitz gradient enables efficient convex optimization with convergence
  - POGM with cycle-spinning or proximal averaging works unexpectedly well but what convergence theory?
- Future
  - Non-convexity
    - Non-convex ψ
    - Regularize tail singular values [24]: $R(\boldsymbol{X}) = \sum_{k=\hat{r}+1} \psi(\sigma_k(\boldsymbol{X}))$
  - Reduce computation time
    - Quadratic majorizer for better line search?
    - Exploit parallelism in code
    - Inter-twine Newton-like methods with proximal methods? [25]
    - Iteration-dependent regularization parameter? [13]
    - Compare POGM with OptISTA [26]

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

Bibliography I

[1] 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.

[2] 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.

[3] 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.

[4] P. J. Huber. *Robust statistics*. New York: Wiley, 1981.

[5] A. I. Cohen. "Rate of convergence of several conjugate gradient algorithms." In: *SIAM J. Numer. Anal.* 9.2 (1972), 248–59.

[6] C. Davis. "All convex invariant functions of Hermitian matrices." In: *Archiv der Mathematik* 8.4 (1957), 276–8.

[7] A. S. Lewis. "The convex analysis of unitarily invariant matrix functions." In: *J. Convex Analysis* 2.1 (1995), 173–83.

[8] A. S. Lewis. "Derivatives of spectral functions." In: *Mathematics of Operations Research* 21.3 (1996), 576–88.

[9] H. Qi and X. Yang. "Semismoothness of spectral functions." In: *SIAM J. Matrix. Anal. Appl.* 25.3 (2003), 766–83.

[10] J. Lee, S. Kim, G. Lebanon, and Y. Singer. "Local low-rank matrix approximation." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 28. 2013, 82–90.

[11] J. Trzasko and A. Manduca. "Local versus global low-rank promotion in dynamic MRI series reconstruction." In: *Proc. Intl. Soc. Mag. Res. Med.* 2011, p. 4371.

[12] Z. Ke, W. Huang, Z-X. Cui, J. Cheng, S. Jia, H. Wang, X. Liu, H. Zheng, L. Ying, Y. Zhu, and D. Liang. "Learned low-rank priors in dynamic MR imaging." In: *IEEE Trans. Med. Imag.* 40.12 (2021), 3698–710.

[13] T. Zhang, J. M. Pauly, and I. R. Levesque. "Accelerating parameter mapping with a locally low rank constraint." In: *Mag. Res. Med.* 73.2 (2015), 655–61.

[14] B. Yaman, S. Weingartner, N. Kargas, N. D. Sidiropoulos, and M. Akcakaya. “Locally low-rank tensor regularization for high-resolution quantitative dynamic MRI.” In: *Proc. Intl. Wkshp. Comp. Adv. Multi-Sensor Adapt. Proc.* 2017.

[15] J. I. Tamir, M. Uecker, W. Chen, P. Lai, M. T. Alley, S. S. Vasanawala, and M. Lustig. “T2 shuffling: sharp, multicontrast, volumetric fast spin-echo imaging.” In: *Mag. Res. Med.* 77.1 (Jan. 2017), 180–95.

[16] J. Lu, C. Xu, Z. Hu, X. Liu, Q. Jiang, D. Meng, and Z. Lin. “A new nonlocal low-rank regularization method with applications to magnetic resonance image denoising.” In: *Inverse Prob.* 38.6 (May 2022), p. 065012.

[17] X. Chen et al. “Improved structured low-rank reconstruction for 3D multi-shot EPI with joint motion modelling.” In: *ISMRM Workshop on Data Sampling and Image Reconstruction*. 2023.

[18] L. Vizioli, S. Moeller, L. Dowdle, M. Akcakaya, F. D. Martino, E. Yacoub, and K. Ugurbil. “Lowering the thermal noise barrier in functional brain mapping with magnetic resonance imaging.” In: *Nature Comm.* 12 (2021), p. 5181.

[19] P-A. Comby, Z. Amor, A. Vignaud, and P. Ciuciu. *Denoising of fMRI volumes using local low rank methods*. 2023.

[20] N. K. Meyer, D. Kang, D. F. Black, N. G. Campeau, K. M. Welker, E. M. Gray, M-H. In, Y. Shu, J. Huston, M. A. Bernstein, and J. D. Trzasko. “Enhanced clinical task-based fMRI metrics through locally low-rank denoising of complex-valued data.” In: *The Neuroradiology J.* (2022), p. 19714009221122171.

[21] S. Guo, J. A. Fessler, and D. C. Noll. “High-resolution oscillating steady-state fMRI using patch-tensor low-rank reconstruction.” In: *IEEE Trans. Med. Imag.* 39.12 (Dec. 2020), 4357–68.

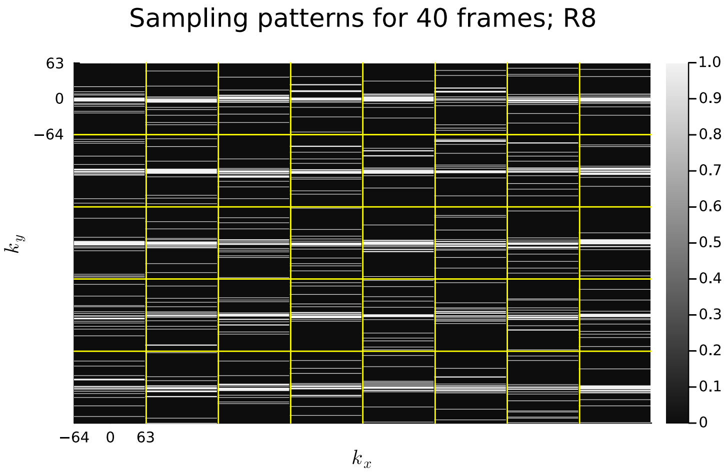
[22] H. Bauschke, R. Goebel, Y. Lucet, and X. Wang. “The proximal average: basic theory.” In: *SIAM J. Optim.* 19.2 (2008), 766–85.

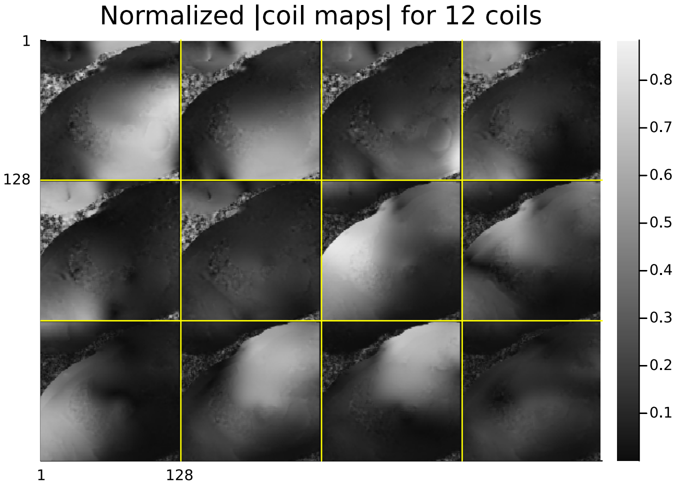
[23] L. W. Zhong and J. T. Kwok. “Accelerated stochastic gradient method for composite regularization.” In: *AISTATS*. 2014, 1086–94.

[24] J. A. S. Cavazos, J. A. Fessler, and L. Balzano. “Sample-wise heteroscedastic PCA with tail singular value regularization.” In: *Proc. IEEE Intl. Conf. on Sampling Theory and Appl. (SampTA)*. To appear. 2023.

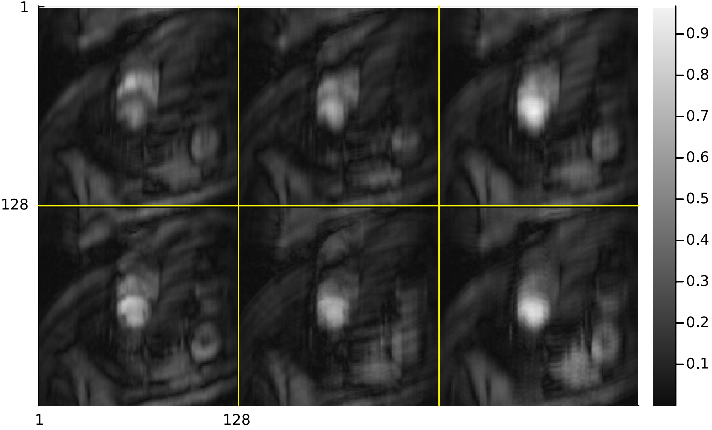
[25] G. Bareilles, F. Iutzeler, and Jerome Malick. "Newton acceleration on manifolds identified by proximal gradient methods." In: *Mathematical Programming* 200 (2023), 37–70.

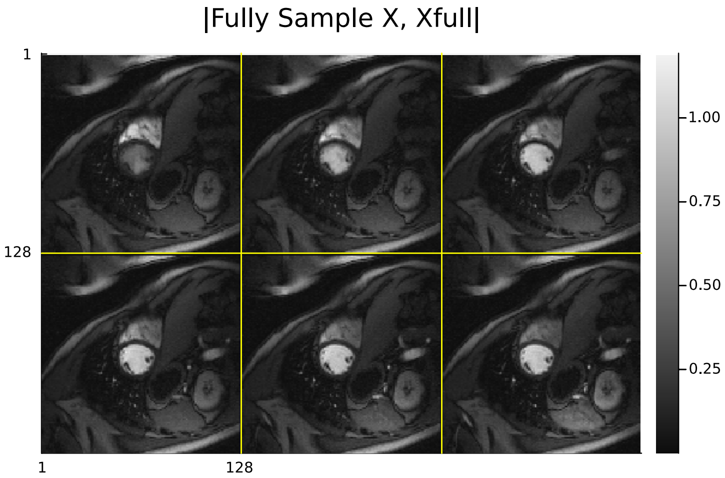
[26] U. Jang, S. D. Gupta, and E. K. Ryu. *Computer-assisted design of accelerated composite optimization methods: optISTA*. 2023.


Sampling patterns for 40 frames; R8
63
0
−64
k_y
−64
0
63
k_x
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

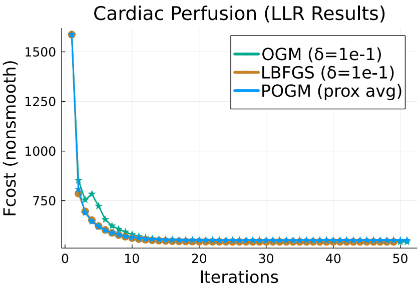

Normalized |coil maps| for 12 coils
1
128
1
128
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

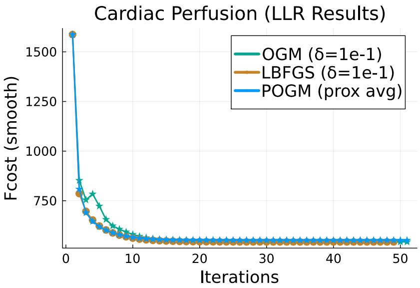
|Initial L|

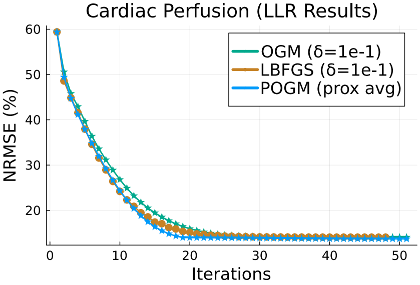


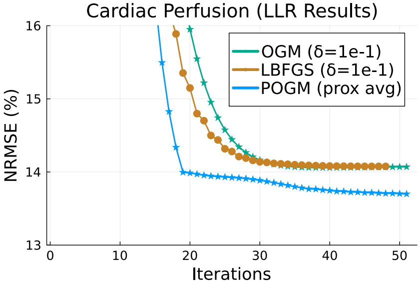

|Fully Sample X, Xfull|
1
128
1
128
1.00
0.75
0.50
0.25

OGM curves


Cardiac Perfusion (LLR Results)
Fcost (nonsmooth)
1500
1250
1000
750
0
10
20
30
40
50
Iterations
OGM (δ=1e-1)
LBFGS (δ=1e-1)
POGM (prox avg)


Cardiac Perfusion (LLR Results)
Fcost (smooth)
1500
1250
1000
750
0
10
20
30
40
50
Iterations
OGM (δ=1e-1)
LBFGS (δ=1e-1)
POGM (prox avg)


Cardiac Perfusion (LLR Results)
NRMSE (%)
60
50
40
30
20
0
10
20
30
40
50
Iterations
OGM (δ=1e-1)
LBFGS (δ=1e-1)
POGM (prox avg)


Cardiac Perfusion (LLR Results)
NRMSE (%)
16
15
14
13
0
10
20
30
40
50
Iterations
OGM (δ=1e-1)
LBFGS (δ=1e-1)
POGM (prox avg)

OGM images