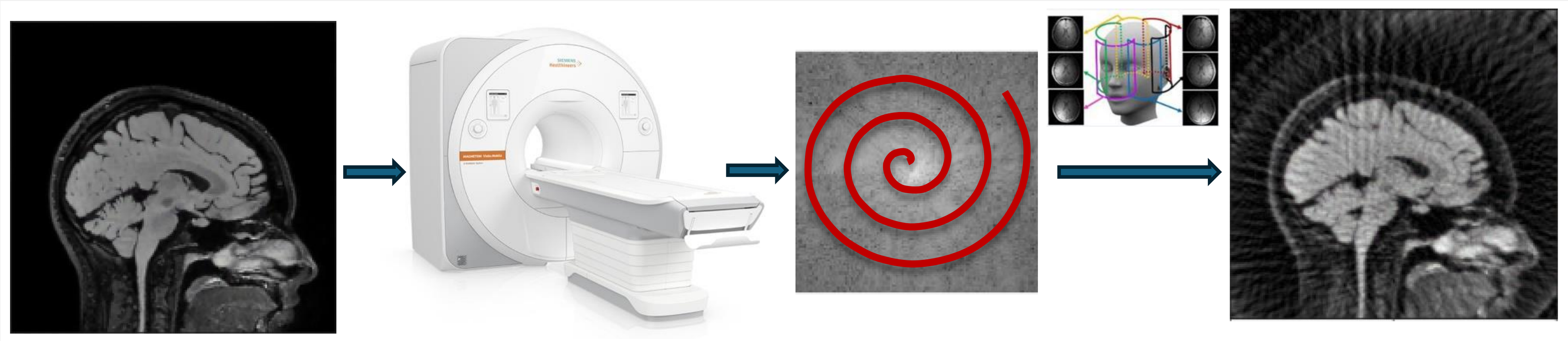
A Convergent Generalized Krylov Subspace Method for Compressed Sensing MRI Reconstruction with Gradient-Driven Denoisers

Tao Hong[1], Umberto Villa[1], and Jeffrey A. Fessler[2]

1. Oden Institute for Computational Engineering & Sciences, UT Austin. 2. Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor.

Background



Model-Based compressive sensing (CS) MRI reconstruction:

$$\mathbf{x}_* = \arg\min_{\mathbf{x}\in\mathbb{C}^N} h(\mathbf{x}) + f(\mathbf{x}),\ \ h(\mathbf{x}) \equiv \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2,\ f(\cdot):\ \text{regularizer} \qquad (1)$$

- $\mathbf{y}$: k-space data, forward model: $\mathbf{A}\in\mathbb{C}^{M\times N} = \{\mathbf{D}\mathbf{F}\mathbf{S}_c\}_c$.
- $\mathbf{D}$: downsampling mask; $\mathbf{F}$: (nonuniform) FFT; $\{\mathbf{S}_c\}_c$: sensitivity maps.

Proximal method for (1) at kth iteration: $\mathbf{x}_{k+1} = \mathrm{prox}_{\alpha f}(\mathbf{x}_k - \alpha\nabla f(\mathbf{x}_k))$

$$\mathrm{prox}_{\alpha f}(\cdot) \equiv \arg\min_{\boldsymbol{x}\in\mathbb{C}^N} \frac{1}{2}\|\mathbf{x}-\cdot\|_2^2 + \alpha f(\mathbf{x}) \Rightarrow \mathbf{D}_{\sigma>0}(\cdot) \equiv \mathrm{prox}_{\alpha f}(\cdot).$$

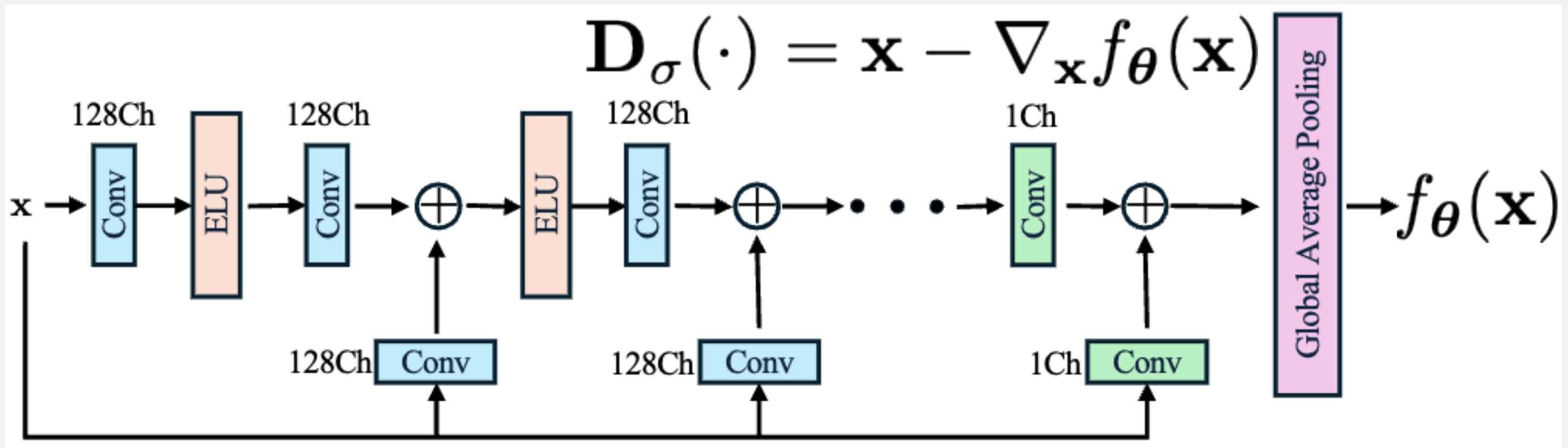
$\mathbf{D}_\sigma(\cdot)$: an abstract denoiser and $\sigma>0$: noise level $\leftrightarrow$ the strength of $f(\cdot)$.

PnP [4]: $\mathbf{x}_{k+1} = \mathbf{D}_\sigma(\mathbf{x}_k - \alpha\nabla f(\mathbf{x}_k))$.

Convergence PnP $\rightarrow$ Assumption: $\mathbf{D}_\sigma(\cdot)$ is nonexpansive ✘

Goal: (1) Bridge the gap between practice and theory—learned priors;
(2) Solve the minimization problem efficiently.
(3) Convergence is guaranteed.

Gradient-Driven Denoisers



$$\min_{\mathbf{x}\in\mathcal{C}} h(\mathbf{x}) + f_{\boldsymbol{\theta}}(\mathbf{x})$$

Observations: 1. h and $f_{\boldsymbol{\theta}}$ are differentiable; 2. $f_{\boldsymbol{\theta}}$ is nonconvex.
Solvers: Gradient based methods, e.g., (accelerated) proximal gradient
Complex quasi-Newton proximal method (CQNPM) [5]:

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{x}\in\mathcal{C}} \langle\nabla f(\mathbf{x}_k), \mathbf{x}-\mathbf{x}_k\rangle + \frac{1}{2\alpha_k}\|\mathbf{x}-\mathbf{x}_k\|_{\mathbf{B}_k}^2 + \underbrace{\frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}_{h(\mathbf{x})},\ \mathbf{B}_k \succ 0 \qquad (2)$$

Messages: 1. CQNPM > first-order methods. 2. Convergence is guaranteed: I) $f_{\boldsymbol{\theta}}$ satisfies Polyak-Łojasiewicz inequality; II) $\nabla_{\mathbf{x}} f_{\boldsymbol{\theta}}$ is Lipschitz continuous.
NB: computing $\mathbf{A}\mathbf{x}$ is quite expensive in multi-coil non-Cartesian sampling MRI.

Q: how to solve (2) efficiently?

Generalized Krylov Subspace Method

Given a subspace basis $\mathbf{V}_k\in\mathbb{C}^{N\times k}$ satisfying $\mathbf{V}_k^{\mathcal{H}}\mathbf{V}_k = \mathbf{I}_k$:

$$\boldsymbol{\beta}_k = \arg\min_{\mathbf{x}=\mathbf{V}_k\boldsymbol{\beta}} \underbrace{\frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \bar{f}(\mathbf{x},\mathbf{x}_k,\mathbf{B}_k,\alpha_k)}_{\bar{F}(\mathbf{x},\mathbf{x}_k)}, \qquad (3)$$

where $\bar{f}(\mathbf{x},\mathbf{x}_k,\mathbf{B}_k,\alpha_k) \equiv \langle\nabla f(\mathbf{x}_k),\mathbf{x}\rangle + \frac{1}{2\alpha_k}\|\mathbf{x}-\mathbf{x}_k\|_{\mathbf{B}_k}^2$.

$$\boldsymbol{\beta}_k = \arg\min_{\boldsymbol{\beta}\in\mathbb{C}^k} \left\| \begin{bmatrix}\mathbf{A}\mathbf{V}_k\boldsymbol{\beta}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{V}_k\boldsymbol{\beta}\end{bmatrix} - \begin{bmatrix}\mathbf{y}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{w}_k\end{bmatrix}\right\|_2^2, \qquad (4)$$

where $\mathbf{w}_k = \mathbf{x}_k - \alpha_k\mathbf{B}_k^{-1}\nabla f(\mathbf{x}_k)$ and $\bar{\mathbf{B}}_k = \bar{\mathbf{B}}_k^{\frac12}\bar{\mathbf{B}}_k^{\frac12}$ with $\bar{\mathbf{B}}_k = \frac{1}{\alpha_k}\mathbf{B}_k$. Then

$$\boldsymbol{\beta}_k = (\mathbf{V}_k^{\mathcal{H}}\mathbf{A}^{\mathcal{H}}\mathbf{A}\mathbf{V}_k + \mathbf{V}_k^{\mathcal{H}}\bar{\mathbf{B}}_k\mathbf{V}_k)^{-1}\mathbf{V}_k^{\mathcal{H}}(\mathbf{A}^{\mathcal{H}}\mathbf{y} + \bar{\mathbf{B}}_k\mathbf{w}_k) \qquad \mathcal{C} = \mathbb{C}^N \qquad (5)$$

$$\boldsymbol{\beta}_k = \arg\min_{(\mathbf{V}_k\boldsymbol{\beta})\in\mathcal{C}} \left\| \begin{bmatrix}\mathbf{A}\mathbf{V}_k\boldsymbol{\beta}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{V}_k\boldsymbol{\beta}\end{bmatrix} - \begin{bmatrix}\mathbf{y}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{w}_k\end{bmatrix}\right\|_2^2 \qquad \mathcal{C} \neq \mathbb{C}^N \qquad (6)$$

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{z}\in\mathcal{C}} \left\| \begin{bmatrix}\mathbf{A}\mathbf{V}_k\mathbf{V}_k^{\mathcal{H}}\mathbf{z}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{V}_k\mathbf{V}_k^{\mathcal{H}}\mathbf{z}\end{bmatrix} - \begin{bmatrix}\mathbf{y}\\ \bar{\mathbf{B}}_k^{\frac12}\mathbf{w}_k\end{bmatrix}\right\|_2^2 \qquad (7)$$

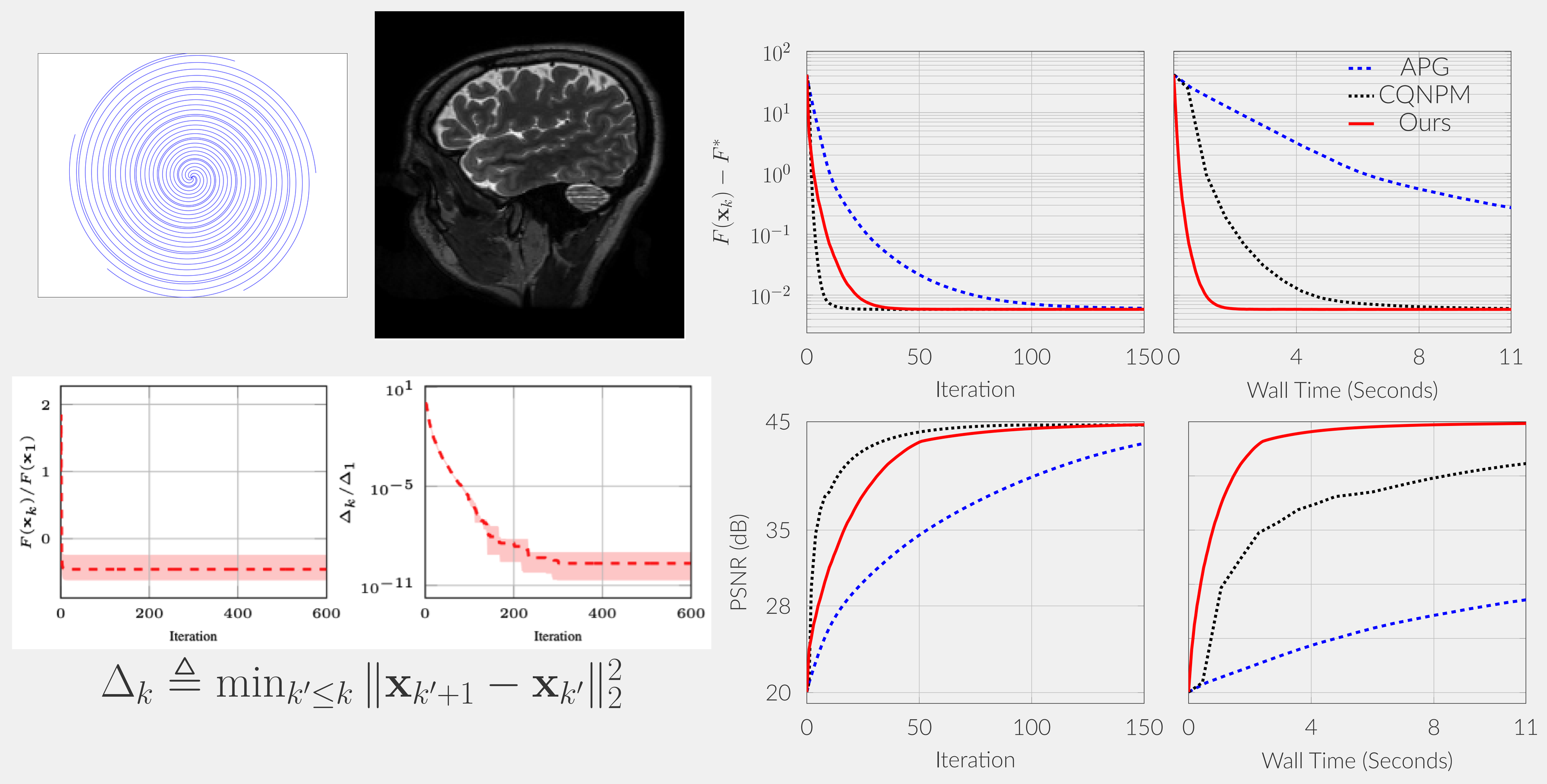
Algorithm 1 Our Method

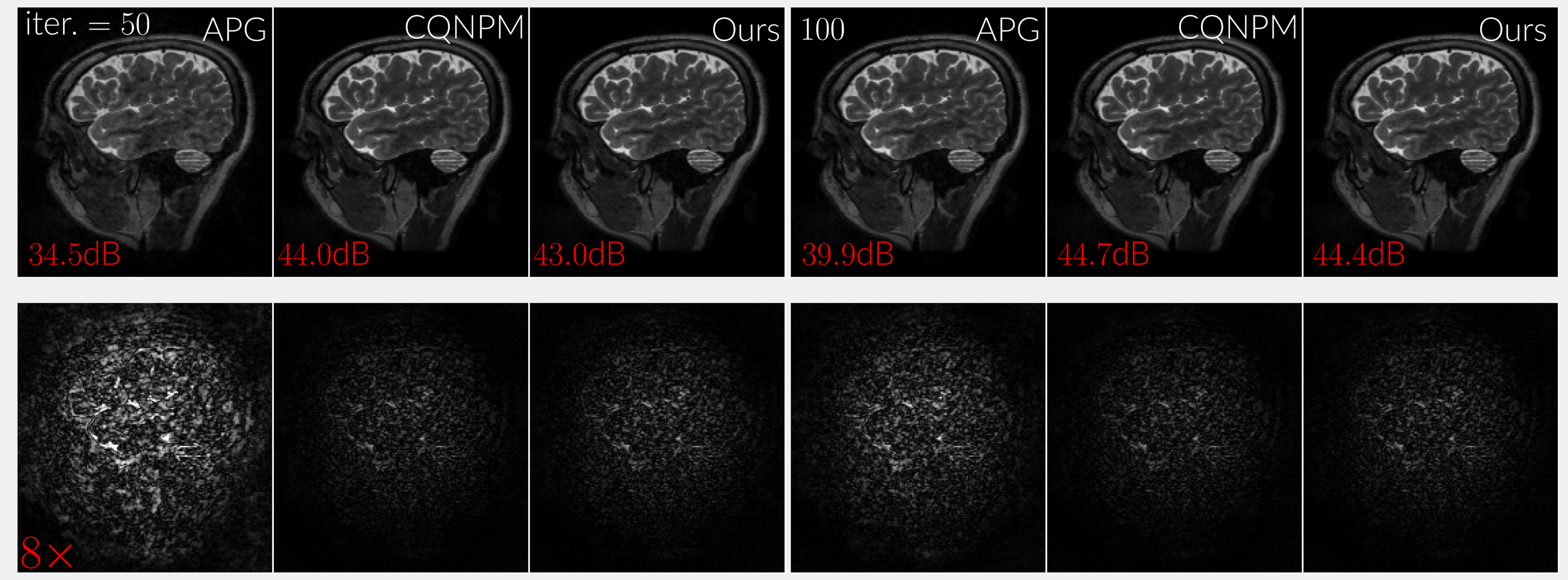
Initialization: $\mathbf{x}_1$, stepsize $\alpha_k > 0$, $\mathbf{V}_1 = \frac{\mathbf{A}^{\mathcal{H}}\mathbf{y}}{\|\mathbf{A}^{\mathcal{H}}\mathbf{y}\|}$, $\mathbf{A}\mathbf{V}_1$, maximal number of subspace iterations K, and maximal number of total iterations `Max_Iter`
Iteration:

```
for k = 1, 2, ..., Max_Iter do
    Compute ∇f(x_k)
    Set B_k ≻ 0
    Compute β_k with (5) or (6)
    Compute x_{k+1} ← V_k β_k
    if k ≤ K then
        Compute r_k ← ∇_x F̄(x_{k+1}, x_k).
        r̃_k ← (I − V_k V_k^H) r_k
        if ‖r̃_k‖ ≠ 0 then
            v_{k+1} ← r̃_k / ‖r̃_k‖
            V_{k+1} ← [V_k v_{k+1}]
            AV_{k+1} ← [AV_k Av_{k+1}]
        else
            V_{k+1} ← V_k
            AV_{k+1} ← AV_k
        end if
    else
        V_{k+1} ← I_N
    end if
end for
```

NB: Convergence is guaranteed under the Kurdyka–Łojasiewicz inequality.

Results





Acknowledgements

This work was funded by NIH grants R01 EB034261, R01 EB035618, and R21 EB034344. The author gratefully acknowledges the support of the Travel Award funded by NSF CBMS Award No. 2430460, which supported the presentation of portions of this work.

References

[1] Shreyas Chaudhari, Srinivasa Pranav, and José MF Moura. Gradient networks. *IEEE Transactions on Signal Processing*, 73:324 – 339, 2024.

[2] Regev Cohen, Yochai Blau, Daniel Freedman, and Ehud Rivlin. It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. *NeurIPS*, 34:18152–18164, 2021.

[3] Tao Hong, Umberto Villa, and Jeffrey A Fessler. A convergent generalized Krylov subspace method for compressed sensing MRI reconstruction with gradient-driven denoisers. *arXiv preprint arXiv:2508.11219*, 2025.

[4] Tao Hong, Xiaojian Xu, Jason Hu, and Jeffrey A Fessler. Provable preconditioned plug-and-play approach for compressed sensing MRI reconstruction. *IEEE Transactions on Computational Imaging*, 10:372 – 384, 2024.

[5] Tao Hong, Zhaoyi Xu, Se Young Chun, Luis Hernandez-Garcia, and Jeffrey A Fessler. Convergent complex quasi-Newton proximal methods for gradient-driven denoisers in compressed sensing MRI reconstruction. *IEEE Transactions on Computational Imaging*, 2025. to appear.

[6] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. *ICLR*, 2022.